

Mark Podolskij
Robert Stelzer
Steen Thorbjørnsen
Almut E.D. Veraart *Editors*

The Fascination of Probability, Statistics and their Applications

In Honour of Ole E. Barndorff-Nielsen

The Fascination of Probability, Statistics and their Applications

Mark Podolskij · Robert Stelzer
Steen Thorbjørnsen · Almut E.D. Veraart
Editors

The Fascination of Probability, Statistics and their Applications

In Honour of Ole E. Barndorff-Nielsen

Editors

Mark Podolskij
Department of Mathematics
University of Aarhus
Aarhus C
Denmark

Steen Thorbjørnsen
Department of Mathematics
University of Aarhus
Aarhus C
Denmark

Robert Stelzer
Institute of Mathematical Finance
Ulm University
Ulm
Germany

Almut E.D. Veraart
Department of Mathematics
Imperial College London
London
UK

ISBN 978-3-319-25824-9

ISBN 978-3-319-25826-3 (eBook)

DOI 10.1007/978-3-319-25826-3

Library of Congress Control Number: 2015954604

Mathematics Subject Classification (2010): 60J70, 60G51, 86A60, 60K40, 76F55, 75M35, 60H15, 35Q30, 76F55, 60G99, 62E10, 60G51, 30E10, 62F25, 62G15, 62G20, 62M09, 60E07, 60G09, 60G51, 60G10, 60H10, 60H07, 60J70, 35R05, 47D07, 86A05, 60H35, 65C30, 60H20, 60H15, 60G57, 60G51, 60B20, 60G51, 60G57, 60E07, 60H05, 60H15, 62M20, 60G55, 60F05, 60G15, 60G17, 60G22, 62M10, 62M1, 60G105, 60F05, 60F17, 60G51, 60G07, 60J10, 62H12, 62M05, 91G70, 91B30, 65C50, 60G25, 62M20, 86A05, 97M30, 93E20, 60H10, 60H20, 60G51, 91G80, 91B70, 60G40, 93E20, 91G80, 60G60, 60G48, 68M15, 60K20, 60F99, 60J27

JEL classification code: C10, C22, C80, G11, C70

Springer Cham Heidelberg New York Dordrecht London

© Springer International Publishing Switzerland 2016

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made.

Printed on acid-free paper

Springer International Publishing AG Switzerland is part of Springer Science+Business Media
(www.springer.com)

Preface

This book is our little present for Ole Eiler Barndorff-Nielsen on the occasion of his 80th birthday. The four of us are all still rather young and thus only met Ole for the first time during the last 15 years. For all of us meeting, working and discussing research with Ole has been a wonderful experience. His kindness, his wide scientific interests, his great insights (both mathematically and in actual applications), his eagerness to cooperate with young scientists, and his many other qualities, have been a great gift and inspiration. Looking at his oeuvre one immediately sees that he has worked in many different areas, all usually connected to probability and statistics, and made a deep impact in many of them.

In this Preface we do not attempt in any way to summarise his achievements, neither can this book cover all aspects of Ole's work. But we are very glad that many colleagues with whom Ole collaborated in some way or another agreed to contribute to this book. We hope that the contributions collected here create a fascinating journey through probability, statistics and their applications in relation to Ole and his work. Of course, we could have asked many more colleagues to contribute (and many more would have certainly said yes), but like for many things in life there is a limit for the reasonable size of a book. Likewise, given Ole's current industriousness and continuing interest to investigate new scientific questions, this book will certainly fail to include some areas which Ole will look into in the future. So instead of expounding more inadequate words, we want to let the contributions in this book speak for themselves and in the meantime please join us in wishing:

Tillykke med fødselsdagen, Ole, og vi glæder os til mange flere år i dit inspirerende selskab! Happy Birthday, Ole, and we are looking forward to many more years in your inspiring company!

Let us now give you a little tour guide for your journey through probability, statistics and their applications when reading this book. At first we have a piece on Ole's (scientific) life written by Eva Vedel-Jensen, Mark Podolskij and further collaborators as a "pleasant" starter.

Then we go medias in res in an area which Ole has been excited about for decades—sand and turbulence. There are three papers on this area: Michael Sørensen develops a model for the grain size distribution in a natural sand deposit based on the normal inverse Gaussian distribution, a special case of the generalised hyperbolic distribution; Björn Birnir's article focuses on the closely related question of whether or not one can prove that the turbulence in the wind which shapes natural sand deposits causes a mass distribution which is described well by the generalised hyperbolic distribution; and finally, José Ulises Márquez and Jürgen Schmiegel introduce a new modelling framework for turbulent velocity fields based on the class of so-called Brownian semistationary processes, where again the normal inverse Gaussian distribution features as an important ingredient.

Next, a paper by Gérard Letac reminds us of perhaps the first big topic in Ole's works: Exponential Families. This is followed by another (theoretical) statistics paper written by Per Mykland and Jianming Ye which focuses on cumulants and Bartlett identities in Cox regressions, which is related to Ole's work on accuracy in statistical theory. Two papers, one by Martin Drapatz and Alexander Lindner and one by Peter Tankov, look at the (probability) theory of infinitely divisible distributions—another area to which Ole has contributed throughout the years. While the former paper investigates the concept of exchangeability in the context of infinitely divisible distribution, the latter article gives a comprehensive review of recent results in the area of Lévy copulas.

Thereafter, there are four papers reflecting Ole's various interests in (the theory of) stochastic processes. First, Fred Espen Benth and Asma Khedher consider generalised Ornstein-Uhlenbeck processes, which exhibit stochastic mean-reversion, and investigate weak stationarity of such processes. Next, Jorge Ramirez, Enrique Thomann and Ed Waymire study the continuity of local times with applications in physics and biology. Bohan Chen, Carsten Chong and Claudia Klüppelberg then present numerical methods for simulating stochastic Volterra equations when the driving space-time Lévy noise is of pure jump type. Finally, Victor Pérez-Abreu and Alfonso Rocha-Arteaga study the eigenvalue-process associated to a matrix-valued Lévy process.

In empirical applications, stochastic processes have to be estimated, so Neil Shephard and Justin Yang consider likelihood inference in trawl processes—a model that has previously been introduced in a paper with Ole as one of the authors. This is followed by two papers on time series, one by Thibault Jaisson and Mathieu Rosenbaum, who consider different asymptotic regimes of nearly unstable autoregressive processes, and another one by Alessandra Luati and Tommaso Proietti, who introduce the concept of generalised partial autocorrelations in the context of stationary processes.

Aarhus Institute of Advanced Studies | Aarhus University | Denmark

**AARHUS CONFERENCE ON PROBABILITY, STATISTICS
AND THEIR APPLICATIONS**

15-19 June 2015

**Celebrating the scientific achievements of
Ole E. Barndorff-Nielsen**

Invited Speakers

Søren Asmussen (Aarhus University)
 Andreas Basse-O'Connor (Aarhus University)
 Fred Espen Benth (University of Oslo)
 Jean Bertoin (University of Zürich)
 Bjørn Brinr (University of California Santa Barbara)
 José Manuel Corcuera (University of Barcelona)
 David Cox (University of Oxford)
 Mark Davis (Imperial College London)
 Giulia di Nunno (University of Oslo)
 Peter Reinhard Hansen (European University Institute)
 Friedrich Hubalek (Vienna University of Technology)
 Jean Jacod (UMC Paris VI)
 Eva B. Vedel Jensen (Aarhus University)
 Søren Johansen (University of Copenhagen)
 Andreas Kyprianou (University of Bath)
 Gérard Letac (Université Paul Sabatier, Toulouse)
 Alexander Lindner (Ulm University)
 Alessandra Lualaba (University of Bologna)
 Asger Lunde (Aarhus University)
 Makoto Maejima (Keio University)
 Per Mykland (University of Chicago)
 Mikko Pakkanen (Imperial College London)
 Giovanni Peccati (Luxembourg University)
 Victor Pérez-Abreu (CIMAT, Mexico)
 Victor Rivera (CIMAT, Mexico)
 Mathieu Rosenbaum (UPMC, Paris VI)
 Jan Rosinski (University of Basel)
 Jürgen Schmiegel (Aarhus University)
 Neil Shephard (Harvard University)
 Albert Shiryaev (Steklov Mathematical Institute)
 Michael Sørensen (University of Copenhagen)
 Peter Tankov (Université Paul Sabatier, Paris VII)
 Ed Waymire (Oregon State University)
 Nakahiro Yoshida (University of Tokyo)
 Bernt Øksendal (University of Oslo)

Scope of the Conference

The topics covered at the conference include, but are not limited to, limit theory, stochastic analysis, statistical inference for stochastic processes, mathematical finance and turbulence.

In particular, the conference honors the numerous scientific achievements of Ole E. Barndorff-Nielsen on the occasion of his 80th birthday and some social events will be dedicated to celebrate him.

Deadline for registration: 15 April 2015

Organizers

Mark Podolski (Aarhus University)
 Robert Stelzer (Ulm University)
 Steen Thorbjørnsen (Aarhus University)
 Almut Veraart (Imperial College London)

Please find further information
 at the homepage:
thiele.au.dk/events/conferences/2015/aarhus/
 or contact Oddbjørn Wethelund:
oddbjorg@math.au.dk

THIELE CENTRE
 FOR APPLIED MATHEMATICS IN NATURAL SCIENCE

<http://thiele.au.dk/events/conferences/2015/aarhus/>

Poster announcing the “Aarhus Conference on Probability, Statistics and Their Applications: Celebrating the scientific achievements of Ole E. Barndorff-Nielsen” held in June 2015 © posterdesign and concept by Daniela Mayer, DesignZone, daniela@designzone.info

Especially over the last two decades Ole has enjoyed working on the econometric analysis of high-frequency financial data with a special focus on volatility estimation and inference. So we are pleased that three teams of researchers have decided to contribute their recent research findings in this topic area to this Festschrift. First, Jean Jacod and Victor Todorov tackle the problem of efficiently estimating integrated volatility in the presence of jump processes of infinite variation. Next, Masayuki Uchida and Nakahiro Yoshida develop a model selection procedure in the context of a diffusion model which aims for a good prediction of the corresponding spot volatility component. Finally, Peter Reinhard Hansen, Guillaume Horel, Asger Lunde and Ilya Archakov introduce a new estimator for multivariate stochastic volatility within a Markov chain framework.

Volatility reflects risk and the last years have reminded us (the hard way) of the importance of understanding our risks. Therefore, we have Paul Embrechts and Edgars Jakobsons investigating the effects of uncertainty on the dependence structure when quantifying risk, whereas Mark Davis introduces a Beaufort Scale of Predictability (a nice reference back to our first topic—wind—by the way). Next comes a paper by Bernt Øksendal, Agnès Sulem and Tusheng Zhang dealing with a stochastic optimal control problem, before Ole's deep interest in modelling financial markets with Lévy processes is reflected in two papers—one by José Manuel Corcuera, José Fajardo, Wim Schoutens and Arturo Valdivia, who derive pricing formulas for contingent convertible bonds, and one by Giulia Di Nunno and Erik Hove Karlsen, who consider the hedging problem under a worst case scenario when the financial market is driven by time-changed Lévy noises.

Finally, a paper in applied probability by Søren Asmussen (one of Ole's current colleagues in Aarhus), Lester Lipsky and Stephen Thompson, which focuses on restart problems in complex systems and develops suitable Markov renewal methods, rounds off our journey through probability, statistics and their applications.

Given all these contributions it was very hard for us to find an appropriate title for the book, even more as we wanted it to be special, to stick out from the crowd of books and to be related to Ole. In the end it was (once again) Ole that inspired us: This time by the title of his contribution together with Preben Blæsild, Jens Ledet Jensen and Michael Sørensen to the book "A Celebration of Statistics: The ISI Centenary Volume A Volume to Celebrate the Founding of the International Statistical Institute in 1885" edited in 1985 by Anthony C. Atkinson and Stephen E. Fienberg. The title of it was simply "The Fascination of Sand".

We would like to sincerely thank all contributors to this book for their excellent contributions and for their efforts to make this book possible within a short time period. It has been a great pleasure to work with all of you! Likewise, we owe a big thank you to all our anonymous reviewers for their dedication and work which were essential for this book. Last but not least, we wholeheartedly thank Eva Nacca—the good spirit of the Institute of Mathematical Finance in Ulm—for taking care of many details for the book and the Springer staff, in particular Marina Reizakis and Catriona Byrne, for being extremely helpful and making the publication possible.

From 15 to 19 June 2015, the “Aarhus Conference on Probability, Statistics and Their Applications: Celebrating the scientific achievements of Ole E. Barndorff-Nielsen” took place at the University of Aarhus and we presented the first preliminary version of this book to Ole at that conference. It was a pleasure to organise this event. As part of this, we would like to take this opportunity to wholeheartedly thank:

- the (other) local organisers: Søren Asmussen, Andreas Basse-O’Connor, Eva B. Vedel-Jensen, Jens Ledet Jensen, Asger Lunde, Jan Pedersen and Jürgen Schmiegel,
- Oddbjørg Wethelund for her tremendous efforts,
- CREATES, CSGB, the Department of Mathematics and the Faculty of Science and Technology at Aarhus University for their great financial support,
- and the invited speakers for their talks—many of them based on contributions to this volume.

To conclude we hope that when you read the contributions now you enjoy them as much as we did when we edited the book.

Aarhus
Ulm
Aarhus
London
June 2015

Mark Podolskij
Robert Stelzer
Steen Thorbjørnsen
Almut E.D. Veraart

Ole E. Barndorff-Nielsen's Scientific Achievements

This year Ole E. Barndorff-Nielsen has celebrated his 80th birthday. He has contributed to the development of mathematical statistics and probability in more than half a century at a multitude of levels and he is still contributing very actively.

Ole Barndorff-Nielsen has played a crucial role in our field both via scientific leadership and scientific impact. Throughout his entire scientific career, Ole E. Barndorff-Nielsen has shown his impressive talent for leadership and extraordinary ability to convey enthusiasm to his surroundings. He has had a leading role in Danish statistics and, in particular, he has created the stochastic group at Aarhus University and thereby influenced the professional scientific life of the large number of researchers affiliated with Aarhus. A main achievement has been the establishment of the Centre for Mathematical Physics and Stochastics (MaPhySto) in 1998. The Centre was funded by a grant from the Danish National Research Foundation and was based at the Department of Mathematical Sciences at Aarhus University with Ole E. Barndorff-Nielsen as Director, but involved principal investigators also from the Universities of Copenhagen, Odense and Aalborg. Under the competent leadership of Ole E. Barndorff-Nielsen, MaPhySto became a very important focus point in the international stochastics community, by its extended programme of workshops and visitors and, not least, via the series of Concentrated Advanced Courses, arranged by MaPhySto that was attended by junior researchers from all over the world. Also at the individual level Ole E. Barndorff-Nielsen has given generous support to many younger statisticians.

Internationally, Ole E. Barndorff-Nielsen has been a leading scientist, in particular via his role in the formation and life of the Bernoulli Society for Mathematical Statistics and Probability. He was the President of the Society in 1993–1995 and the first Editor-in-Chief of the Society's journal *Bernoulli* from 1994 to 2000. Previously he had been Editor of the *International Statistical Review* 1980–1987, and he was Chairman of European Research Centres on Mathematics (ERCOM) 1997–2002. The *Seminaires Européens de Statistiques* were initiated already in 1989 with high-level expository talks on new developments in statistics. Ole E. Barndorff-Nielsen is member of the Royal Danish Academy of Sciences and Letters, member of *Academia Europaea*, Honorary Fellow of the Royal Statistical

Society, the International Statistical Institute and the Danish Society for Theoretical Statistics. He holds Honorary Doctorates from Université Paul Sabatier, Toulouse (June 1993) and Katholieke Universiteit Leuven (April 1999). And in 2001 he was awarded a Humboldt Forschungspreis.

Ole E. Barndorff-Nielsen has a very large scientific production, represented by six research monographs and about 280 research papers. He has influenced the methodology in a wide range of research areas and is a distinguished representative of our profession where detailed analysis of data is put on equal footing with modelling and inference, and advanced mathematical tools are used in the development of statistical methodology. The list of research areas of Ole E. Barndorff-Nielsen includes

- Likelihood and evidence
- Differential geometry in statistical inference
- Quantum stochastics
- Laws, processes and Lévy theory
- The physics of blown sands
- The statistical theory of turbulence
- Financial econometrics
- Ambit stochastics

Till around 1995, the main research focus of Ole E. Barndorff-Nielsen was on the mathematical theory of inference, centred on concepts of *likelihood*, *sufficiency*, *ancillarity*, *conditioning* and *exponential families*. His interest in inference was triggered already around 1956 when he was affiliated to the Biostatistics Department of the Danish State Serum Institute in Copenhagen. The first discussions with Sir David R. Cox concerning inferential questions were initiated by Ole E. Barndorff-Nielsen in 1963. These discussions developed into a long-term collaboration, resulting in a series of joint papers and books, and a close personal relationship—including joint attendances to operas, including Richard Wagner's 'Ring'.

The research on *likelihood and evidence* is presented in five books and a long series of seminal papers written in a time span of about 30 years. A very important contribution came as early as in 1978 with the publication of the Wiley book entitled *Information and Exponential Families in Statistical Theory*, based to a large degree on the Sc.D. thesis *Exponential Families and Conditioning* that was published in 1973. The Wiley book has very recently been re-issued with a new Preface. The roots of the book lie in the writings of R.A. Fisher both as concerns results and the general attitude to statistical science, and this attitude has been the determining factor in the selection of topics. It is also important to draw attention to the two books entitled *Asymptotic Techniques for Use in Statistics and Inference and Asymptotics*, written jointly with Sir David R. Cox, that appeared in 1989 and 1994, respectively. First- and higher-order theory were discussed here in relation to various notions of likelihood.

The series of papers on *likelihood and evidence* appeared in top statistical journals such as *Biometrika*, *Annals of Statistics* and *Journal of the Royal Statistical*

Society B. Different concepts of likelihood were investigated, e.g. profile, modified and directed likelihood, along with factorisations of likelihood, Edgeworth and saddlepoint approximations of likelihood, Bartlett adjustments to the likelihood ratio test statistic and approximate confidence limits. Different definitions of sufficiency and ancillarity (for instance, S, L and M) were studied in detail together with the resulting data reduction principle or conditioning principle. Throughout this period, the exponential families played a central role because of their richness and tractability. Certain general classes of exponential models were studied in further detail, such as exponential transformation models.

One of the most important findings during this *likelihood and evidence* period is the formula that approximates the conditional distribution of the maximum likelihood estimator, given an ancillary statistic. This formula, published in 1983 in *Biometrika*, enables conditional inference in practice. It was found that the formula is exact for many important models. By extending the basic theory of transformation models to the cases where the group action is not free, it was shown that the formula is exact for arbitrary transformation models.

Ole E. Barndorff-Nielsen has also played a leading role in the investigation of *differential geometry in statistical inference* where parametric statistical models are regarded as differentiable manifolds equipped with tensors and connections, derived from statistical considerations. The Fisher information may be interpreted as a metric tensor and the model is now a Riemannian manifold. Exponential families constitute the simplest statistical manifolds in this sense. Ole E. Barndorff-Nielsen found that differential geometry has important applications in the asymptotic theory of statistical inference. The path-breaking work resulted in a simplified form for the asymptotic expansions associated with the likelihood ratio statistic if geometrical invariants were used. This deep relationship between statistics and differential geometry involved a form of higher-order differential geometry which is not readily accessible to a statistical audience. In particular, the theory of strings and phyla played an important role. Much of this research arose from the interest in curved exponential families. Transformation models, in particular exponential transformation models, were also studied from the viewpoint of differential geometry.

The research of Ole E. Barndorff-Nielsen on *quantum stochasticity* focuses on the roles the concepts of sufficiency, ancillarity and exponential families, from non-quantum statistical theory, have in a quantum setting. The motivation came from his deep fascination of quantum physics and the associated aspects of probability. In fact, in view of his interests in the principles of statistical inference, it seems inevitable that he would at some stage be involved with statistical and probabilistic aspects of quantum physics. A cue for this was the 1997 Nobel Prize awarded to three physicists for their development of methods to cool and trap atoms with laser light. But the intricacies of the inherent physical concepts, the differences to ordinary realism, and the advanced mathematics required—which is unfamiliar to most traditionally educated statisticians—made the field fairly difficult to access. Still, Ole Barndorff-Nielsen and co-authors published a series of papers on quantum stochasticity in the beginning of the 2000s, including a discussion paper in *Journal*

of the *Royal Statistical Society B* on quantum statistical inference, a paper on Fisher information in quantum statistics and one on laser cooling and stochastics.

In relation to *laws, processes and Lévy theory*, the hyperbolic distributions and their generalisations as well as models based on Lévy processes are at the centre of Ole E. Barndorff-Nielsen's work. This research has resulted in a marked expansion of the catalogue of available flexible families of distributions, an important prerequisite for a realistic data analysis. For modelling by stochastic processes, Ole E. Barndorff-Nielsen found that deeper probabilistic properties of the family are of considerable interest. In particular, the property of infinite divisibility—or even self-decomposability—is of key importance for stochastic process modelling, as illustrated by the case of the generalised inverse and the normal inverse Gaussian distributions. This research on *laws, processes and Lévy theory* has been published in a series of papers from early 1990s to today (a selection consist of about 45 papers).

Some of these papers deal with self-decomposability and Lévy processes in free probability. Already in the mid 1990s Ole E. Barndorff-Nielsen was intrigued by the new types of convolution for probability measures that had recently arisen in the context of non-commutative probability theory. In particular he became interested in studying the notion of self-decomposability associated with free additive convolution. This lead to a series of co-authored papers in which freely self-decomposable probability measures were introduced and their representation in terms of stochastic integrals with respect to free Lévy processes was established. The developed theory further led to the free analogue of the Lévy-Itô representation for free Lévy processes. Somewhat surprisingly the studies of free infinite divisibility also led to the introduction of so-called Upsilon transformations, which form a rich class of mappings between classes of *classically* infinitely divisible probability measures strongly related to stochastic integration with respect to (classical) Lévy processes.

The project on the *physics of blown sand* was initiated in the 1970s. This colourful project was carried out under Ole E. Barndorff-Nielsen's energetic leadership by a group of statisticians, earth scientists and physicists, colloquially known as the 'Sand Gang'. The project had a practical side where data were collected on the size distribution in sand deposits and on drifting sand, but it also had a strong theoretical side with development of new distributional laws and stochastic models of grain motion and turbulence. The key step that started the whole project was the derivation of the hyperbolic distribution. Brigadier Ralph Alger Bagnold, who had explored the Libyan deserts between the two world wars, had in 1941 published the influential book *The Physics of Blown Sands and Desert Dunes*, where (among many other things) he noted that the log-density of the size distribution of sand samples resembles a hyperbola. Ole E. Barndorff-Nielsen formalised Bagnold's heuristic idea, and in a way characteristic of his research he presented the distribution as a member of the class of generalised hyperbolic distributions, defined as normal variance-mean mixtures, and derived a rich probabilistic theory for this class of distributions. Via David Kendall, Ole E. Barndorff-Nielsen came in contact with Bagnold, who was very pleased that

after 30 years someone had taken up his old suggestion. Ole E. Barndorff-Nielsen became close to Bagnold both as a scientific collaborator and as a friend. Bagnold was awarded a Doctor Philosophiae Honoris Causa degree at Aarhus University. In 1980, they published a joint paper entitled *The pattern of natural sizedistributions in Sedimentology*.

The main aim of the Sand Gang was to pursue some of the interesting and intriguing problems arising out of Bagnold's pioneering work and out of the work of Professor Paul Robert Owen, another early contributor to the physics of wind-blown sand whom Ole E. Barndorff-Nielsen had contacted and invited to Aarhus. A wind tunnel was constructed in the basement of the Physics Department at Aarhus University, which over the years developed into the best facility in the world for empirical studies of sand transport by the wind. The wind tunnel experiments were supplemented by several field experiments in Northern Jutland, in Scotland and in the deserts of North America by members of the Sand Gang—often including Ole E. Barndorff-Nielsen.

A selection of 14 papers on the physics of blown sand covers the period 1977–1991. Much of the work carried out was reported in a two-volume supplement to *Acta Mechanica* on 'Aeolian Grain Transport'. An account of some of the work done by the members of the Sand Gang may be found in the paper entitled *The fascination of sand*, published in 1985 in the ISI Centenary Volume 'A Celebration of Statistics'.

The sand project led naturally to the deep interest of Ole E. Barndorff-Nielsen in the statistical theory of *turbulence* and *financial econometrics*, two areas which, as he has pointed out, have some striking similarities. His contributions to financial econometrics have revolutionised the non-parametric methodology for high-frequency financial data. In the period 2002–2010, Ole E. Barndorff-Nielsen and co-authors have been developing the statistical theory of a class of realised measures named *multipower variations*. Multipower variations possess a fully non-parametric asymptotic theory, which makes it possible to assess the volatility and/or jumps of an underlying semimartingale possibly contaminated by microstructure noise. These new classes of statistics have triggered many new developments in the field of financial econometrics. In another direction, Ole E. Barndorff-Nielsen has made a great contribution to modelling financial data. The *Barndorff-Nielsen-Shephard model*, which is a stochastic volatility model for the price of a financial asset based on Lévy-driven Ornstein-Uhlenbeck processes, is one of the most cited models in the literature.

From early in this century, Ole E. Barndorff-Nielsen has been heavily involved in the mathematical investigation of *ambit stochastics*, which has been originally motivated by turbulence modelling, but later found manifold applications in finance and biology. Ambit processes are spatiotemporal models, which describe the dynamics in a stochastically developing field, for instance a turbulent wind field, along curves embedded in such a field. Beyond the basic noise component, which is typically modelled by a Lévy basis, the processes specifically incorporate stochastic volatility/intermittency and time change. Ole E. Barndorff-Nielsen has been studying limit theorems, concepts of stochastic integration and statistical inference

for ambit fields during the last decade. The mathematical theory has led to new insights in the understanding of turbulence. In particular, realistic stochastic models for the turbulent energy dissipation and velocity time series have been formulated that capture the main stylised features of turbulence. New statistical features discovered by Ole E. Barndorff-Nielsen and collaborators include the representation of empirical densities of velocity increments by normal inverse Gaussian distributions and a new type of universality, termed increment similarity that goes far beyond the traditional scaling analysis. The generalisation of the turbulence models to cover the full spatio-temporal dynamics is one of the research areas where Ole E. Barndorff-Nielsen is currently very active and, as usual, very productive.

There is thus ample reason to honour Ole E. Barndorff-Nielsen by a conference and a proceedings volume on themes that are at the centre of his interests.

Aarhus
Aarhus
Copenhagen
Aarhus
June 2015

Eva B. Vedel Jensen
Mark Podolskij
Michael Sørensen
Steen Thorbjørnsen

Contents

On the Size Distribution of Sand	1
Michael Sørensen	
From Wind-Blown Sand to Turbulence and Back.	15
Björn Birnir	
Modelling Turbulent Time Series by <i>BSS</i>-Processes	29
José Ulises Márquez and Jürgen Schmiegel	
Associated Natural Exponential Families and Elliptic Functions	53
Gérard Letac	
Cumulants and Bartlett Identities in Cox Regression	85
Per Aslak Mykland and Jianming Ye	
Exchangeability and Infinite Divisibility	99
Martin Drapatz and Alexander Lindner	
Lévy Copulas: Review of Recent Results	127
Peter Tankov	
Weak Stationarity of Ornstein-Uhlenbeck Processes with Stochastic Speed of Mean Reversion	153
Fred Espen Benth and Asma Khedher	
Continuity of Local Time: An Applied Perspective	191
Jorge M. Ramirez, Enrique A. Thomann and Edward C. Waymire	
Simulation of Stochastic Volterra Equations Driven by Space-Time Lévy Noise	209
Bohan Chen, Carsten Chong and Claudia Klüppelberg	
On the Process of the Eigenvalues of a Hermitian Lévy process	231
Victor Pérez-Abreu and Alfonso Rocha-Arteaga	

Likelihood Inference for Exponential-Trawl Processes	251
Neil Shephard and Justin J. Yang	
The Different Asymptotic Regimes of Nearly Unstable Autoregressive Processes	283
Thibault Jaisson and Mathieu Rosenbaum	
Generalised Partial Autocorrelations and the Mutual Information Between Past and Future.	303
Alessandra Luati and Tommaso Proietti	
Efficient Estimation of Integrated Volatility in Presence of Infinite Variation Jumps with Multiple Activity Indices	317
Jean Jacod and Viktor Todorov	
Model Selection for Volatility Prediction	343
Masayuki Uchida and Nakahiro Yoshida	
A Markov Chain Estimator of Multivariate Volatility from High Frequency Data	361
Peter Reinhard Hansen, Guillaume Horel, Asger Lunde and Ilya Archakov	
Dependence Uncertainty for Aggregate Risk: Examples and Simple Bounds	395
Paul Embrechts and Edgars Jakobsons	
A Beaufort Scale of Predictability	419
Mark H.A. Davis	
A Stochastic HJB Equation for Optimal Control of Forward-Backward SDEs	435
Bernt Øksendal, Agnès Sulem and Tusheng Zhang	
CoCos with Extension Risk. A Structural Approach	447
José Manuel Corcuera, José Fajardo, Wim Schoutens and Arturo Valdivia	
Hedging Under Worst-Case-Scenario in a Market Driven by Time-Changed Lévy Noises.	465
Giulia Di Nunno and Erik Hove Karlsen	
Markov Renewal Methods in Restart Problems in Complex Systems	501
Søren Asmussen, Lester Lipsky and Stephen Thompson	

On the Size Distribution of Sand

Michael Sørensen

Abstract A model is presented of the development of the size distribution of sand while it is transported from a source to a deposit. The model provides a possible explanation of the log-hyperbolic shape that is frequently found in unimodal grain size distributions in natural sand deposits, as pointed out by Bagnold and confirmed in numerous empirical studies. The model implies that the size distribution of a sand deposit is a logarithmic normal-inverse Gaussian (NIG) distribution, which is one of the generalized hyperbolic distributions. The model modifies a previous model, which implied a log-normal size-distribution, by taking into account that individual grains do not have the same travel time from the source to the deposit. The travel time is assumed to be random so that the wear on the individual grains vary randomly. The model provides an interpretation of the parameters of the NIG-distribution, and relates the mean, variance and skewness of the log-size distribution to the physical parameters of the model. The results might be useful when comparing empirical size-distributions from different deposits. It is argued that size-distributions with the same general shape as the NIG-distributions can be obtained also when some of the model assumptions are relaxed.

Keywords First hitting time · Grain attrition process · Normal-inverse gaussian distribution · Normal variance-mean mixture

1 Introduction

In his influential book on the physics of wind blown sand, Bagnold [1] pointed out that the logarithm of the grain size in natural sand deposits is not normal distributed. Specifically, he noted that the logarithm of the density function of the log-size

M. Sørensen (✉)
Department of Mathematical Sciences, University of Copenhagen,
Universitetsparken 5, 2100 Copenhagen Ø, Denmark
e-mail: michael@math.ku.dk

distribution looks more like a hyperbola than like a parabola, indicating a distribution with (approximately) exponentially decreasing tails. This motivated Barndorff-Nielsen [4] to introduce the hyperbolic distributions, for which the logarithm of the density function is a hyperbola. In a series of papers the hyperbolic distributions have been shown empirically to give a very good description of the log-size distribution of natural sand deposits, aeolian as well as alluvial, and to be a powerful tool for analyzing the spatial and temporal variation of the size distribution of sediments, see e.g. [12, 15, 20], and the review paper by Bagnold and Barndorff-Nielsen [2]. Barndorff-Nielsen [4] introduced a more general class of probability distributions, the generalized hyperbolic distributions, that are similar to the hyperbolic distributions. These were introduced as the normal variance-mean mixtures with the generalized inverse Gaussian distribution as mixing distribution. A particularly important subclass of these distributions is the class of normal-inverse Gaussian (NIG) distributions that gives an even better fit to the size distributions of sand than the hyperbolic distributions. This is because a NIG-distribution has slightly heavier tails than the original hyperbolic distributions, which is quite often the case for natural sand deposits. The NIG-distributions have also turned out to fit the distribution of many types of financial returns. An advantage of the NIG-distributions over the original hyperbolic distributions is that they have much nicer mathematical properties. For instance, simple expressions for the mean, the variance and all other moments can be calculated explicitly. A review of the properties of the generalized hyperbolic distributions and related dynamical models, with a view to applications in finance, can be found in [10].

In this paper, we present a model that provides a possible explanation why sand deposits are log-NIG distributed and gives an interpretation of the parameters of the distribution in terms of physical quantities. This might be useful when comparing the size distributions of sand from different deposits. The model presented here extends a model by Kolmogoroff [14], which implied a log-normal size-distribution, by taking into account the fact that the grains found in a particular deposit may not have taken the same time to arrive at the deposit from their source and therefore have not been subject to the same amount of attrition. Thus the wear on the individual grains vary randomly.

The paper is organized as follows. In Sect. 2 models of the development of the size-distribution and of the grain motion are presented, and the resulting size distribution in the deposit is shown to be a variance-mean mixture, which is identified as a NIG-distribution. The model of grain motion also implies a first order model of the shape of the deposit. In Sect. 3 some basic properties of the normal-inverse Gaussian distribution are reviewed, and the mean value, variance and other parameters of the log-size distribution are related to physical parameters of the model. Various further topics are discussed in Sect. 4. In particular, it is shown that the size-distribution will also be a normal variance-mean mixture if the assumptions about grain motion and the attrition process are relaxed very considerably, and it is pointed out that size-distributions with the same general shape as the NIG-distributions can in some cases be obtained also when these model assumptions have been changed.

An introduction to aeolian sand transport for statisticians can be found in [3].

2 The Model and the Size-Distribution

The basic assumption of the model is that a grain in transit from its source to its present position has experienced a random number of breakage events, for instance forceful collisions or extreme weather conditions, that cause a random fraction of the volume of the grain to break off. By grain size we will mean the grain volume. In geomorphology grain size is usually measured by some, more or less well-defined, grain diameter. If v is the grain volume, then any reasonable measure of the grain diameter must be proportional to $v^{1/3}$, so if the distribution of $\log v$ is a NIG-distribution, then so is the distribution of the log-diameter $\frac{1}{3} \log v$. Therefore it is no restriction that we consider the grain volume. This only implies a change of the location and scale parameters.

Denote the number of such breakage events by N_t , where t denotes the time since the grain started from its source. If we denote the original grain size by S_0 and the fraction of the grain that is broken off in the i th collision by D_i , then the grain size at time t is

$$S_t = S_0 \prod_{i=1}^{N_t} (1 - D_i),$$

and the logarithm of the grain size is

$$\log S_t = \mu + \sum_{i=1}^{N_t} B_i,$$

where $\mu = \log(S_0)$ and $B_i = \log(1 - D_i)$. We assume that the random variables B_i are independent and identically distributed with mean b_1 and second moment b_2 . Let further N_t be a Poisson process with parameter λ that is independent of the random variables $B_1, B_2, \dots$. This means that the number of breakage events in a time interval of length s is Poisson distributed with mean value λs , and that the numbers of such events in disjoint time intervals are independent random variables.

Under these assumptions $\sum_{i=1}^{N_t} B_i$ is a compound Poisson process (finite activity Lévy process) with expectation $\lambda b_1 t$ and variance $\lambda b_2 t$, and by the central limit theorem for Lévy processes,

$$U_\lambda(t) = \frac{1}{\sqrt{\lambda}} \left(\sum_{i=1}^{N_t} B_i - \lambda t b_1 \right) \quad (1)$$

is approximately normal distributed with mean zero and variance $t b_2$ when λ (or λt) is large, i.e.

$$\log S_t \sim N(\mu + \lambda t b_1, \lambda t b_2), \quad (2)$$

for large λ , where $\sim$ denotes approximate distribution. The value of λ depends on properties of the grains, the climate, and the terrain.

All grains do not arrive at the sand deposit from the source at the same time. Some may have been moved back and forth for a long while, whereas other grains may have been moved directly and therefore more quickly to the sand deposit. Obviously the wear on a grain depends on how long it has been transported around, as is clear from (2). Therefore we need a model for the motion of a grain from its source. In experiments carried out in the Aarhus wind tunnel and reported in [9, 19], the motion of individual sand grains was studied. The grains were marked with radioactive gold so that they could be followed by mean of a portable scintillation detector. The experiments showed that the grain is buried rather often while a ripple passes over it. However, when the grain is not buried, its motion is very accurately described by a compound Poisson process. At a time scale of a few minutes, a good model of the grain motion is thus a compound Poisson process with an intensity that in random and quite frequent time intervals is equal to zero and otherwise is non-random and constant. However, the time scale of the motion of the grains from their source to the deposit is of the order of several years. It is therefore quite justified, as a good approximation, to model the motion of a grain by a two-dimensional Brownian motion with drift $X_t = (X_{1t}, X_{2t})$.

We can approximate the locations of the source and the deposit by two points. This is obviously not exactly correct, but relative to the typically long distance between the two, this is a reasonable approximation. We use a coordinate system with the origin at the source and the first axis pointing in the direction of the deposit. Then

$$dX_t = \begin{pmatrix} \nu \\ 0 \end{pmatrix} dt + \eta dW_t, \quad (3)$$

where W is a 2-dimensional standard Wiener process and η is a 2×2 -matrix. The speed of the grain is given by $\nu > 0$, and the infinitesimal variance of X is

$$\Sigma = \eta \eta^T = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}.$$

Thus the motion of the grain in a time interval of length s is two-dimensional normal distributed with mean $(\nu s, 0)$ and covariance matrix Σs , and the displacement of the grain in disjoint time intervals are independent random variables. It is reasonable to assume that there is very little motion in the direction of the second axis, i.e. that $\sigma_2 \ll \sigma_1$. Otherwise, the sand would be scattered over a large area, and there would not be a deposit. Finally, we assume that the Brownian motion of a grain is independent of the breakage process, i.e. independent of the Poisson process N and the random variables $B_1, B_2, \dots$. Motion in the lateral direction will cause extra wear. This effect can be build into the model by assuming that λ is proportional to σ_2 .

Let a denote the distance between the source and the deposit. Then the time it takes for a grain to move from the source to the deposit equals the first hitting time, τ , of the one-dimensional Brownian motion X_{1t} to the level a . It is well known (see e.g. [13],

p. 363) that the distribution of τ is the inverse Gaussian distribution $IG(a/\sigma_1, v/\sigma_1)$. The probability density of the inverse Gaussian distribution $IG(\delta, \gamma)$ is

$$\frac{\delta}{\sqrt{2\pi x^3}} \exp\left(-\frac{1}{2}\gamma^2(x - \delta/\gamma)^2/x\right), \quad x > 0.$$

This result goes back to Schrödinger [18].

The logarithm of the size of a grain that has arrived at the sand deposit at time τ is $\log S_\tau$. Since the Brownian motion is independent of the breakage process, so is τ . Therefore the conditional distribution of the log-size given $\zeta = \lambda b_2 \tau$ is, by (2), approximately

$$\log S_\tau | \zeta \sim N(\mu + \beta\zeta, \zeta), \quad (4)$$

where $\beta = b_1/b_2$. The distribution of ζ is

$$\zeta \sim IG\left(a\sqrt{\lambda b_2}/\sigma_1, v/(\sigma_1\sqrt{\lambda b_2})\right). \quad (5)$$

Suppose the grains in the deposit have only one source, and that once a grain has arrived at the deposit, it is trapped there (the deposit could for instance be a big dune). Then the distribution of the logarithm of the grain size in the whole population of grains in the deposit is the unconditional distribution of $\log S_\tau$. Heuristically, this is approximately equal to the normal variance-mean mixture given by (4) and (5), which is the normal-inverse Gaussian distribution $NIG(\alpha, \beta, \delta, \mu)$ with

$$\alpha = \sqrt{\beta^2 + v^2/(\sigma_1^2 \lambda b_2)}, \quad \delta = \frac{a}{\sigma_1} \sqrt{\lambda b_2},$$

and with β and μ as above.

The probability density function of the $NIG(\alpha, \beta, \delta, \mu)$ -distribution is

$$\frac{\alpha\delta}{\pi} e^{\delta\gamma} \cdot \frac{K_1\left(\alpha\sqrt{\delta^2 + (x - \mu)^2}\right)}{\sqrt{\delta^2 + (x - \mu)^2}} \cdot e^{\beta(x - \mu)}, \quad x \in \mathbb{R},$$

where K_1 is a modified Bessel function of the third kind. The possible values of the parameters are $\alpha > 0$, $\delta > 0$, $|\beta| < \alpha$, while μ can be any real number. The shape of the density is similar to that of the original hyperbolic distribution, except that the tails are slightly heavier. Examples of the shape of the logarithm of the density function are given in Fig. 1. While the expression for the density is relatively complicated, the characteristic function of the $NIG(\alpha, \beta, \delta, \mu)$ -distribution has the following simple form:

$$\exp\left(\delta\sqrt{\alpha^2 - \beta^2}\left[1 - \sqrt{1 + (s^2 - is2\beta)/(\alpha^2 - \beta^2)}\right] + is\mu\right).$$

A mathematically rigorous derivation of the log-size distribution in the deposit is a consequence of the following theorem.

Theorem 1 *The joint distribution of $(U_\lambda(\tau), \tau)$, with $U_\lambda(t)$ given by (1), converges weakly as $\lambda \rightarrow \infty$:*

$$E(\exp(is_1 U_\lambda(\tau) + is_2 \tau)) \rightarrow \psi\left(-\frac{1}{2}b_2 s_1^2 + is_2\right), \quad (6)$$

where ψ is the moment generating function of τ given by

$$\psi(z) = E(e^{z\tau}) = \exp\left(\frac{av}{\sigma_1^2} \left[1 - \sqrt{1 - 2z\sigma_1^2/v^2}\right]\right), \quad (7)$$

which is defined for all complex numbers z with $\Re(z) \leq v^2/(2\sigma_1^2)$. The limit distribution of $U_\lambda(\tau)$ is the NIG($v/(\sigma_1\sqrt{b_2})$, 0 , $a\sqrt{b_2}/\sigma_1$, 0)-distribution.

Proof Let $\varphi(u) = E(e^{iuB_i})$ denote the characteristic function of the distribution of B_i . Then the characteristic function of $U_\lambda(t)$ is

$$E(e^{isU_\lambda(t)}) = \exp\left(\lambda t \left[\varphi(s/\sqrt{\lambda}) - 1 - ib_1 s/\sqrt{\lambda}\right]\right).$$

Note that (2) follows easily from this expression by Taylor expanding φ .

The moment generating function, ψ , of the stopping time τ is given by (7), so the joint characteristic function of $(U_\lambda(\tau), \tau)$ is

$$\begin{aligned} E(\exp(is_1 U_\lambda(\tau) + is_2 \tau)) &= E\left[E(\exp(is_1 U_\lambda(\tau)) \mid \tau) \exp(is_2 \tau)\right] \\ &= E\left(\exp\left(\tau \left(\lambda \left[\varphi(s_1/\sqrt{\lambda}) - 1 - ib_1 s_1/\sqrt{\lambda}\right] + is_2\right)\right)\right) \\ &= \psi\left(\lambda \left[\varphi(s_1/\sqrt{\lambda}) - 1 - ib_1 s_1/\sqrt{\lambda}\right] + is_2\right) \\ &= \psi\left(-\frac{1}{2}b_2 s_1^2 + is_2 + O(\lambda^{-1/2})\right) \\ &\rightarrow \psi\left(-\frac{1}{2}b_2 s_1^2 + is_2\right) \end{aligned}$$

as $\lambda \rightarrow \infty$. For $s_2 = 0$ we recognise the characteristic function of the NIG($v/(\sigma_1\sqrt{b_2})$, 0 , $a\sqrt{b_2}/\sigma_1$, 0)-distribution. $\square$

For λ sufficiently large, the joint distribution of $(U_\lambda(\tau), \tau)$ is approximated by the distribution with characteristic function $\psi(-\frac{1}{2}b_2 s_1^2 + is_2)$. Therefore the distribution of

$$\log(S_\tau) = \sqrt{\lambda}U_\lambda(\tau) + \lambda b_1 \tau + \mu$$

is approximated by the distribution with characteristic function

$$\begin{aligned} \psi & \left(-\frac{1}{2}\lambda b_2 s^2 + i\lambda b_1 s \right) \exp(is\mu) \\ & = \exp \left(\frac{av}{\sigma_1^2} \left[1 - \sqrt{1 + (b_2 s^2 - is2b_1)\lambda\sigma_1^2/v^2} \right] + is\mu \right). \end{aligned}$$

We recognise this as the characteristic function of the NIG($\alpha, \beta, \delta, \mu$)-distribution with $\alpha = \sqrt{\beta^2 + v^2}/(\sigma_1^2\lambda b_2)$, $\beta = b_1/b_2$, and $\delta = a\sqrt{\lambda b_2}/\sigma_1$.

The model of grain motion also implies a result on the *shape of the deposit*. The position of the grain on the second axis when the grain arrives at the deposit is $X_{2\tau}$. The density function of this random variable gives a first approximation to the profile of the deposit in the direction of the y -axis. It is a well-known property of the two-dimensional Brownian motion that X_{1t} and $Z_t = X_{2t} - \omega X_{1t}$ with $\omega = \sigma_{12}/\sigma_1^2$ are independent Brownian motions. It is indeed easy to see that the quadratic variation of (X_{1t}, Z_t) is a diagonal matrix. Since Z_t is $N(-\omega vt, (\sigma_2^2 - \omega\sigma_{12})t)$ -distributed, and because Z_t is independent of τ , it follows that the conditional distribution of $X_{2\tau} = Z_\tau + \omega a$ given $\xi = \tau(\sigma_2^2 - \omega\sigma_{12})$ is

$$X_{2\tau} | \xi \sim N(\omega a - \sigma_{12}v\kappa^{-2}\xi, \xi),$$

where $\kappa^2 = \sigma_1^2\sigma_2^2 - \sigma_{12}^2$. The distribution of ξ is

$$\xi \sim IG(a\kappa/\sigma_1^2, v/\kappa),$$

so it follows in a way similar to the derivation for the size-distribution that the unconditional distribution of $X_{2\tau}$ is the NIG($\alpha, \beta, \delta, \mu$)-distribution with $\mu = \omega a$, $\delta = a\kappa/\sigma_1^2$, $\beta = -\sigma_{12}v\kappa^{-2}$, and $\alpha = \sqrt{\beta^2 + v^2}/\kappa^2$. This result is mainly of theoretical interest. This is partly because the model of grain motion is probably not sufficiently accurate to be used to predict the shape of the deposit, but mainly because the shape of the deposit is determined to a large extent by the interaction between the shape of the deposit and the wind blowing over the deposit, see e.g. [17].

3 Interpretation of the Size Distribution Parameters

If a random variable X is NIG($\alpha, \beta, \delta, \mu$)-distributed, then the mean and variance are

$$E(X) = \mu + \frac{\delta\beta}{\gamma} \quad \text{and} \quad \text{Var}(X) = \frac{\delta\alpha^2}{\gamma^3},$$

where $\gamma = \sqrt{\alpha^2 - \beta^2}$. The skewness is $3\beta/(\alpha\sqrt{\delta\gamma})$, and the kurtosis is $3(1 + 4\beta^2/\alpha^2)/(\delta\gamma)$. When $\beta < 0$, the NIG distribution is negatively skewed.

For the original hyperbolic distribution the shape triangle, introduced in [3], has turned out to be a very useful tool for detecting trends in the distributional shape

when studying samples from several sand deposits or from several positions on the same deposit. For NIG-distributions a completely analogous shape triangle can be plotted, see e.g. [16]. The shape triangle is defined by

$$\{(\chi, \xi) \mid 0 \leq |\chi| < \xi < 1\}, \quad (8)$$

where

$$\chi = \frac{\beta/\alpha}{\sqrt{1+\delta\gamma}} \quad \text{and} \quad \xi = \frac{1}{\sqrt{1+\delta\gamma}}.$$

These quantities are simple natural measures of asymmetry and heavy-tailedness (“kurtosis”) of the normal-inverse Gaussian distributions. This follows by arguments similar to those given in [3] for the hyperbolic distributions. The quantities χ and ξ are invariant under location and scale transformations, and are clearly closely related to the classical skewness and kurtosis given above. When ξ goes to zero, the NIG-distribution approaches the normal distribution. In Fig. 1 NIG log-density functions are plotted for different values of χ and ξ in the shape triangle. Note the parabola of the normal distribution for $\xi = 0$.

For fixed α and β the NIG distributions are closed under convolution. Specifically, suppose that X_1 and X_2 are independent with $X_i \sim \text{NIG}(\alpha, \beta, \delta_i, \mu_i)$, $i = 1, 2$, then

$$X_1 + X_2 \sim \text{NIG}(\alpha, \beta, \delta_1 + \delta_2, \mu_1 + \mu_2).$$

The sum of two independent hyperbolic distributed random variables is not hyperbolic distributed.

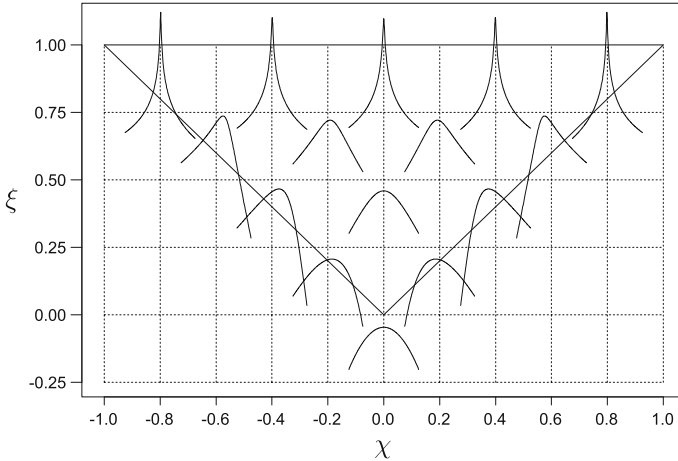


Fig. 1 Normal-inverse Gaussian log densities with mean 0 and variance 1 for different values of the shape parameters χ and ξ . The log-densities are located in the *shape triangle* according to their values of χ and ξ

Because there are simple expressions for the first four moments, it is relatively easy to obtain initial moment estimators of the parameters. However, better estimates are obtained by maximum likelihood estimation. If an empirical size-distribution has been determined by sieving a sand sample, estimators can be obtained by the multinomial pseudo-likelihood procedure outlined in [4]. When comparing the estimated parameters of NIG-distributions fitted to size-distributions from different sand deposits, the model presented in Sect. 2 might be helpful when interpreting the empirical results. Note however, that the model is probably not useful when comparing size-distributions from different locations within the same deposit. This is because the differences between such size-distributions are most likely due to local sorting and are not caused by the random breakage process that causes the NIG-distribution in the model. Local sorting was considered by [5, 7]. If an empirical size distribution is poly-modal, it does not make sense to use the model. Sand with a poly-modal size distribution is probably a mixture of two or more populations of grains, each of which may (or may not) be described by the model. It could for instance be a mixture of sand from two or more sources.

The mean log-size is

$$\mu + a\lambda b_1/\nu = \mu + \lambda\bar{t}b_1 = \mu + \theta,$$

where $\bar{t} = a/\nu$ is the *average travel time* of a grain from the source to the deposit (the expected value of τ), and $\lambda\bar{t}$ is the mean number of breakage events that cause a part of the grain to break off. The quantity $\theta = \lambda\bar{t}b_1$ is the *average effect of the breakage process* during transport from the source to the deposit. Since b_1 is the mean value of $B_i = \log(1 - D_i)$, it is negative. Therefore the mean log-size decreases as the distance between the source and the deposit, a , increases and as the number of collisions the cause breakage increases, while it increases when the mean speed of a grain ν increases.

The variance of the log-size is

$$\frac{a\lambda}{\nu} \left(b_2 + \frac{\lambda\sigma_1^2 b_1^2}{\nu^2} \right) = \theta b_1 (c_0^2 + \lambda\bar{t}\omega^2),$$

where $\omega = \sqrt{\bar{t}\sigma_1^2}/(\bar{t}\nu)$ is the coefficient of variation of the grain position relative to the source at time $\bar{t}$. This ratio measures the random variability of the grain motion relative to the mean distance traveled. The quantity $c_0 = \sqrt{b_2}/b_1$ is closely related to the coefficient of variation of the distribution of $B_i = \log(1 - D_i)$, which equals $\sqrt{c_0^2 - 1}$. Thus c_0 is a measure of the variability of a breakage event relative to a typical value of B_i . The variance of the log-size is an increasing function of a , λ , σ_1^2 , b_1^2 and b_2 , while it is a decreasing function of ν .

The log-size distribution is always negatively skewed because b_1 is negative. Thus positively skewed log-size distributions cannot be explained by the present model. In such cases alternative or additional processes that influence the size distribution

must be sought for. The model predicts that the asymmetry parameter χ is

$$\chi = -\frac{1}{\sqrt{\left(1 + \frac{v^2 b_2}{\sigma_1^2 \lambda b_1^2}\right) (1 + va/\sigma^2)}} = -\frac{1}{\sqrt{[1 + c_0^2 \omega^{-2} (\lambda \bar{t})^{-1}] [1 + \omega^{-2}]}}$$

and that the parameter ξ is

$$\xi = \frac{1}{\sqrt{1 + \omega^{-2}}}.$$

We see that the log-size distribution becomes less normal as the coefficient of variation of the grain motion increases. For a fixed value of the parameter ξ , the asymmetry is determined by the quantity $c_0^2 (\lambda \bar{t})^{-1}$. Thus if the mean number of breakage events increases, then the log-size distribution becomes more skewed, while the asymmetry decreases if the variability of the breakage process increases.

It is not easy to make informed guesses concerning the values of b_1 and b_2 . A very simple possible assumption about the breakage process is that the probability density function of the fraction of the grain that breaks off, D_i , is

$$p(x) = \rho(1-x)^{\rho-1}, \quad 0 < x < 1 \quad \rho > 0. \quad (9)$$

For this model, the distribution of $-B_i = -\log(1 - D_i)$ is an exponential distribution with expectation ρ^{-1} , so $b_1 = -\rho^{-1}$ and $b_2 = 2\rho^{-2}$. In this case, $c_0 = -\sqrt{2}$ irrespective of the value of ρ . However, the breakage process could potentially be much more complicated. To illustrate this, consider another tractable probability density function for D_i :

$$p(x) = [\log(1/(1-x))]^{\kappa-1} (1-x)^{\rho-1} \rho^\kappa / \Gamma(\kappa), \quad 0 < x < 1, \quad \kappa > 0, \rho > 0,$$

where Γ denotes the gamma function. If $\rho > 1$ and $\kappa > 1$, this density function goes to zero for both small and large fractions, and the most likely fractions are somewhere in the interior of the interval between zero and one. If $\rho > 1$ and $\kappa \leq 1$, small fractions are most likely. For $\kappa = 1$ we recover the model (9). In this case $-B_i$ is gamma distributed with shape parameter κ and scale parameter ρ^{-1} . Hence $b_1 = -\kappa/\rho$ and $b_2 = (\kappa + \kappa^2)/\rho^2$, so $c_0^2 = 1 + \kappa^{-1}$. Thus for this breakage model any real number smaller than -1 is a possible value of c_0 .

4 Discussion

The probability density function of $-\sum_{i=1}^{N_t} B_i$ is explicitly known in the particular case where the probability density of D_i is given by (9), so that $-B_i$ is exponentially distributed with expectation ρ^{-1} . Apart from an atom at $N_t = 0$ with the small probability $e^{-\lambda t}$, the density function is

$$f(x) = \sqrt{\rho\lambda t/x} e^{-(\lambda t + \rho x)} I_1(2\sqrt{\rho\lambda tx}), \quad x > 0,$$

where I_1 is a modified Bessel function of the first kind. This result might be used as a starting point in a study of how well a NIG-distribution fits the log-size distribution. Incidentally, this probability density was derived in [11] as the density function of the position of an individual particle in alluvial sediment transport, where it was assumed that the grains move according to a compound Poisson process with exponentially distributed jumps.

The assumption that the grains move according to a Brownian motion is probably a good approximation under homogeneous conditions. The effect of using a different model for the grain motion is that the distribution of the random arrival time τ is changed. In this more general case, Theorem 1 holds with the modification that the moment generating function of the inverse Gaussian distribution ψ is replaced by the moment generating function ψ^* of the new distribution of τ . Therefore, we find in analogy with Sect. 2 that when λ is sufficiently large, then the distribution of $\log(S_\tau)$ is approximated by the distribution with characteristic function

$$\psi^* \left(-\frac{1}{2} \lambda b_2 s^2 + i \lambda b_1 s \right) \exp(is\mu). \quad (10)$$

This is the characteristic function of a normal variance-mean mixture. Suppose the conditional distribution of a random variable X given the random variable v is a normal distribution with mean $\mu + \lambda b_1 v$ and variance $\lambda b_2 v$, and that the distribution of v has the moment generating function ψ^* . Then the unconditional distribution of X has characteristic function (10), see e.g. [8]. The NIG-distributions are variance-mean mixtures, where the mixing distribution is an inverse Gaussian distribution. For many other distributions of τ than the inverse Gaussian distribution, the general shape of the density function of the logarithm of the grain size is similar to that of a NIG-distribution and a hyperbolic distribution. Conditions on the mixing distribution of a normal variance-mean mixture that ensure a tail behaviour similar to that of these distributions were given in [8].

Let us briefly consider a few examples of alternative models for the grain motion. For simplicity we consider one-dimensional models and assume that there is no motion in the direction of the second axis. If, for instance, the motion of a grain is modelled by a Brownian motion without drift, and if there is a reflecting barrier (e.g. a steep cliff) near the source (in the direction opposite to the deposit), then the approximate log-size distribution is a hyperbolic cosine distribution, which has exponentially decreasing tails and is very similar to a hyperbolic distribution. It is, however, not one of the generalized hyperbolic distributions, but belongs to the class of generalized logistic distributions (a.k.a. the z -distributions). The logistic distribution itself can also be obtained in a model involving an absorbing barrier (e.g. a ravine). For details of these two models, see [8]. Diffusion models for which certain first hitting times are generalized inverse Gaussian distributions can be found in [6]. Mixing distributions of this type give rise to generalized hyperbolic distributions.

Another possible modification of the model is to assume that N_t is an inhomogeneous Poisson process with intensity $c\lambda_t$, so that the mean number of breakage events per time unit changes with time. In this case λt should be replaced by $c\Lambda(t)$, with $\Lambda(t) = \int_0^t \lambda_s ds$, in (2). By arguments similar to those in the proof of Theorem 1, it follows that for $c \rightarrow \infty$, the log-size distribution is also in this case approximated by a normal variance-mean mixture, where the mixing distribution is the distribution of $\Lambda(\tau)$. This is true even if λ_t is random such that N_t is a Cox-process.

5 Conclusion

A model was presented that gives a possible explanation of the hyperbolic shape of the log-size distribution found in many natural sand deposits. The distribution was shown to be well approximated by a variance-mean mixture of normal distributions because it is a mixture of the size distributions of grains that have spent different amounts of time in transit between their source and the deposit, and therefore have been subject to a randomly varying amount of attrition. Under the assumptions of the model, the log-size distribution was shown to be approximately NIG-distributed, and an interpretation of differences between the estimated parameters of log-size distributions from different deposits was provided. For instance, the mean, variance and an asymmetry parameters were related to the physical parameters of the model. The model can only be used to interpret unimodal and negatively skewed log-size distributions. It was argued that the general result that the log-size distribution is well approximated by a normal variance-mean mixture of the same general shape as a NIG-distribution is robust to the model assumptions and holds under weaker assumptions about the grain motion and the attrition process.

Acknowledgments I am grateful to Bo Martin Bibby for preparing Fig. 1. The research was supported by the Dynamical Systems Interdisciplinary Network which is financed by a grant from the University of Copenhagen Programme of Excellence.

References

1. Bagnold, R.A.: The Physics of Blown Sand and Desert Dunes. Methuen, London (1941)
2. Bagnold, R.A., Barndorff-Nielsen, O.E.: The pattern of natural size distributions. *Sedimentol.* **27**, 199–207 (1980)
3. Barndorff-Nielsen, O.E., Blæsild, P., Jensen, J.L., Sørensen, M.: The fascination of sand. In: Atkinson, A.C., Fienberg, S.E. (eds.) *A Celebration of Statistics*, pp. 57–87. Springer, New York (1985)
4. Barndorff-Nielsen, O. E.: Exponentially decreasing distributions for the logarithm of particle size. *Proc. R. Soc. Lond. A* **353**, 401–419 (1977)
5. Barndorff-Nielsen, O.E., Christiansen, C.: Erosion, deposition and size distributions of sand. *Proc. R. Soc. Lond. A* **417**, 335–352 (1988)

6. Barndorff-Nielsen, O.E., Halgreen, C.: Infinite divisibility of hyperbolic and generalized inverse gaussian distributions. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **38**, 309–312 (1977)
7. Barndorff-Nielsen, O.E., Sørensen, M.: On the temporal-spatial variation of sediment size distributions. *Acta Mech. [Suppl]* **2**, 23–35 (1991)
8. Barndorff-Nielsen, O.E., Kent, J., Sørensen, M.: Normal variance-mean mixtures and z-distributions. *Int. Stat. Rev.* **50**, 145–159 (1982)
9. Barndorff-Nielsen, O.E., Jensen, J.L., Nielsen, H.L., Rasmussen, K.R., Sørensen, M.: Wind tunnel tracer studies of grain progress. In: Barndorff-Nielsen, O.E., Møller, J.T., Rasmussen, K.R., Willetts, B.B. (eds.) *Proceedings of International Workshop on the Physics of Blown Sand*, vol. 2, pp. 243–251. Dept. Theo. Stat, University of Aarhus, Aarhus (1985)
10. Bibby, B.M., Sørensen, M.: Hyperbolic processes in finance. In: Rachev, S. (ed.) *Handbook of Heavy Tailed Distributions in Finance*, pp. 211–248. Elsevier Science, Amsterdam (2003)
11. Einstein, H.A.: *Der Geschiebetransport als Wahrscheinlichkeitsproblem*. Verlag Rascher and Co., Zürich (1937)
12. Hartmann, D.: Cross-shore selective sorting processes and grain size distributional shape. *Acta Mech. [Suppl]*, **2**, 49–63 (1991)
13. Karlin, S., Taylor, H.M.: *A First Course in Stochastic Processes*. Academic Press, New York (1975)
14. Kolmogoroff, A. N.: Über das logarithmisch normale verteilungsgesetz der dimensionen der teilchen bei zerstückelung. *C. R. Acad. Sci. U.R.S.S.*, **31**, 99–101 (1941)
15. McArthur, D.S.: Distinctions between grain-size distributions of accretion and encroachment deposits in an inland dune. *Sediment. Geol.* **54**, 147–163 (1987)
16. Rydberg, T.H.: The normal inverse gaussian lévy process: simulation and approximation. *Commun. Statist.: Stochast. Models* **13**, 887–910 (1997)
17. Sauermann, G., Kroy, K., Herrmann, H. J.: Continuum saltation model for sand dunes. *Physical Review E*, **64** (2001)
18. Schrödinger, E.: Zur theorie der fall- und steigeversuche an teilchen mit brownscher bewegung. *Physikalische Zeitschrift*, **16** (1915)
19. Sørensen, M.: Statistical analysis of data from a wind tunnel experiment using radioactive grains. Research report, Department of Theoretical Statistics, University of Aarhus (1986)
20. Vincent, P.: Differentiation of modern beach and coastal dune sands—a logistic regression approach using the parameters of the hyperbolic function. *Sediment. Geol.* **49**, 167–176 (1986)

From Wind-Blown Sand to Turbulence and Back

Björn Birnir

In honor of Ole Barndorff-Nielsen on the occasion of his 80th birthday.

Abstract We describe the recently developed Kolmogorov-Obukhov statistical theories of homogeneous turbulence and its extension to boundary layer turbulence. The theories can be used to describe the size distribution of wind-blown sand but the statistical theory of Lagrangian turbulence is still missing, so this task cannot be completed yet. That this can be done was suggested by Ole-Barndorff Nielsen and we show how his Generalized Hyperbolic Distribution gives the continuous part of the probability distribution functions of the turbulent velocity differences.

Keywords Turbulence · Intermittency · Invariant measure · Kolmogorov-Obukhov scaling · Inertial cascade · Navier-Stokes equation · Large deviations · Poisson processes · Central limit theorem · Structure functions · She-Leveque intermittency corrections · Boundary value turbulence · Lagrangian turbulence · Wind-blown sand

1 Introduction

In his book “The physics of blown-sand and desert dunes”, published in 1954, the British military engineer R.A. Bagnold, described the size and mass distribution of wind blown sand [1]. His research was based on years spent in the Saharan Desert studying how the wind moves and forms the desert. He had done careful experiments after his retirement from the military, where he fed sand to the mouth of a wind-tunnel

B. Birnir (✉)

Center for Complex and Nonlinear Science and Department of Mathematics,
University of California, Santa Barbara, CA 93106, USA
e-mail: birnir@math.ucsb.edu, birnirb@gmail.com

B. Birnir

University of Iceland, Reykjavík, Iceland

© Springer International Publishing Switzerland 2016

M. Podolskij et al. (eds.), *The Fascination of Probability,*

Statistics and their Applications, DOI 10.1007/978-3-319-25826-3_2

and observed the mass distribution as a function of the distance down the tunnel. He discovered that if he plotted this distribution on a log-log plot or he plotted the log of the mass distribution as a function of the log of the distance down the tunnel, then he got the lower-half of a hyperbola. This caught the eye of a brilliant young Danish statistician named Ole Barndorff-Nielsen and he found the distribution [3] that he called the Generalized Hyperbolic Distribution (GHD). What he also noticed [2] was that his distributions looked very similar to the distribution of the velocity differences in turbulence, that people had been able to compute in the sixties and were becoming common in turbulence research in the seventies. It was a remarkable observation and insight and it lead to a conjecture; that it was the turbulence in the wind that was sorting the sand and giving its mass distribution the form of the GHD.

The question I want to ask here is: can we prove this relationship? As you will see the answer is no, not yet, but we have made progress and have probably completed 2/3 of the journey towards this goal.

In 1941 Kolmogorov [16] and Obukhov [18] proposed that there exists a statistical theory of turbulence that should allow the computation of all the statistical quantities that can be computed and measured in turbulent systems. These are quantities such as the moments, the structure functions and the probability density functions (PDFs) of the turbulent velocity field. The Kolmogorov-Obukhov'41 theory predicted that the structure functions of turbulence, that are the moments of the velocity differences at distances separated by a lag variable l , should scale with the lag variable to a power $p/3$ for the p th structure function, multiplied by a universal constant. This was found to be inconsistent with observations and in 1962 Kolmogorov [17] and Obukhov [19] presented a refined scaling hypothesis, where the multiplicative constants are not universal and the scaling exponents are modified to $\zeta_p = p/3 + \tau_p$, by the intermittency correction τ_p that are due to intermittency in the turbulent velocity. It was still not clear what the values of τ_p should be, because the log-normal exponents suggested by Kolmogorov turned out again to be inconsistent with observations. Then in 1994 She and Leveque [22] found the correct (log-Poissonian) formulas for τ_p that are consistent with modern simulations and experiments.

We will outline below how the statistical theory of Kolmogorov and Obukhov is derived from the Navier-Stokes equation without getting into any of the technical details. We start with the classical Reynolds decomposition of the velocity into the mean (large scale) flow and the fluctuations or small scale flow. Then we develop a stochastic Navier-Stokes equation [8], for the small scale flow. If we assume that dissipation takes place on all scales in the inertial range (defined below) then it turns out that the noise in this stochastic Navier-Stokes equation is determined by well-known theorems in probability. The additive noise in the stochastic Navier-Stokes equation is generic noise given by the central limit theorem and the large deviation principle. The multiplicative noise consists of jumps multiplying the velocity, modeling jumps in the velocity gradient. We will explain how this form of the noise follows from a very general hypothesis.

Once the form of the noise in the stochastic Navier-Stokes equation for the small scales is determined, we can estimate the structure functions of turbulence and establish the Kolmogorov-Obukhov'62 scaling hypothesis with the She-Leveque inter-

mittency corrections [7]. Then one can compute the invariant measure of turbulence writing the stochastic Navier-Stokes equation as an infinite-dimensional Ito process and solving the linear Kolmogorov-Hopf [14] functional differential equation for the invariant measure. Finally the invariant measure can be projected onto the PDF. The PDFs turn out to be the GHD of Barndorff-Nielsen [2, 4], and compare well with PDFs from simulations and experiments, as was shown by Barndorff-Nielsen et al. [4] in 2004. The details of the proofs can be found in [7] and the background material can be found in [8].

The problem with sand is that it is always blown along a boundary, but not in homogeneous turbulence as described above. So although we have been able to show that the distributions of the moments of the velocity differences in homogeneous turbulence are indeed given by the GHD, as predicted by Ole's remarkable insight, we have not dealt with turbulent boundary layers. The sand is carried by the turbulent boundary layer, so we have only come 1/3 of the way.

The famous log-law of Prandtl (1925) and von Kármán (1930) is the most distinguished characteristic of turbulent boundary flows. It says that the mean-velocity profile in the inertial region satisfies the formula:

$$\langle u \rangle / u_\tau = \kappa^{-1} \ln(z u_\tau / \nu) + B, \quad (1)$$

where $u_\tau = \sqrt{\tau_w / \rho}$ is the friction velocity based on the wall stress τ_w , ρ is the fluid density, ν is the kinematic viscosity, κ is the von Kármán constant and B is also a constant. $\langle \cdot \rangle$ denotes an ensemble average, this can be an average over many experiments or many computer simulations. The log-law is well established both experimentally and numerically, see reviews by Smits, McKeon and Marusic (2011) and Jimenez (2012). It has proven to be an invaluable tool in the theory of turbulent boundary layers.

It was suggested by Meneveau and Marusic (2013) that the log-law of the fluctuations could be generalized to any moment $p \geq 2$ of the fluctuations ($p = 2$ is due to Townsend) by the law

$$\langle (u^+)^{2p} \rangle^{1/p} = B_p - A_p \ln(z/\delta) = D_p(Re_\tau) - A_p \ln(z u_\tau / \nu) \quad (2)$$

where $z u_\tau / \nu$ are the viscous units and $D_p = B_p + A_p \ln Re_\tau$ is a Reynolds number Re_τ dependent offset.

Recently Birnir et al. [11], extended the theory of Birnir [7, 8] developed for homogeneous turbulence to boundary layers to prove (2) with a more physically-based normalization in the inertial range. They showed that the universal constants satisfy the relationship

$$A_p = \left(\frac{1}{l^*} \right)^{\zeta_1 - \zeta_p} \frac{C_p^{1/p}}{C_1} A_1,$$

where $\zeta_p = p/3 + \tau_p = p/9 + 2(1 - (2/3)^{p/3})$ are the Kolmogorov-Obukhov-She-Leveque, scaling exponents of the structure functions of turbulence, see [8], and l^*

is a small constant. The C_p s are the Kolmogorov-Obukhov scaling coefficients, computed below. The sub-Gaussian (they lie below the values expected from a Gaussian distribution) behavior of the A_p s is caused by the KOSL scaling. These authors also computed the probability density functions of the fluctuations in the inertial range. These PDFs turn out to be Generalized Hyperbolic Distributions multiplied by a discrete measure, see the forthcoming article [10] and Sect. 5. In the viscous range the PDFs are skewed Gaussians. The boundary layer theory allows us to determine the statistical theory in boundary layer flow, we will not give the details below but refer to the forthcoming article [11].

However, even if we know the statistics of the flow of the air in a boundary layer, we still do not know how the sand is carried with the turbulent flow. To do this we have to develop the Lagrangian Theory of turbulence for the boundary layer. The reason is that the grains of sand have their own characteristics and are carried with the airflow and interact with it. Some grains creep along the surface others fly and when the land they can send the grains they land on flying. This is called saltation, see [24]. In Lagrangian flow you move along the air flow and observe the sand as it is carried along this flow. The details are complicated and it is this theory that we have to work out before, we can say that Ole's conjecture that the turbulence in the air is carrying and sorting the sand can be proven. This is the missing 1/3 of the story.

2 The Deterministic Navier-Stokes Equations

A general incompressible fluid flow satisfies the Navier-Stokes Equation

$$u_t + u \cdot \nabla u = \nu \Delta u - \nabla p, \quad u(x, 0) = u_0(x)$$

with the incompressibility condition $\nabla \cdot u = 0$.¹ Eliminating the pressure using the incompressibility condition gives

$$u_t + u \cdot \nabla u = \nu \Delta u + \nabla \Delta^{-1} \text{trace}(\nabla u)^2, \quad u(x, 0) = u_0(x).$$

The turbulence is quantified by the dimensionless Taylor-Reynolds number $Re_\lambda = \frac{U\lambda}{\nu}$ [20]. We will impose periodic boundary conditions on the small scales below.

2.1 Reynolds Decomposition

Following the classical Reynolds decomposition [21], we decompose the velocity into mean flow U and the fluctuations u . Then the velocity is written as $U + u$, where U describes the large scale flow and u describes the small scale turbulence. We must

¹ $\Delta u = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2}$ and $\nabla \cdot u = \frac{\partial u}{\partial x} + \frac{\partial u}{\partial y} + \frac{\partial u}{\partial z}$.

also decompose the pressure into mean pressure P and the fluctuations p , then the equation for the large scale flow can be written as

$$U_t + U \cdot \nabla U = \nu \Delta U - \nabla P - \nabla \cdot (\overline{u \otimes u}), \quad (3)$$

where in coordinates $\nabla \cdot (\overline{u \otimes u}) = \frac{\partial \overline{u_i u_j}}{\partial x_j}$, that is ∇ is dotted with the rows of $\overline{u_i u_j}$ and $R_{ij} = \overline{u \otimes u}$ is the Reynolds stress, see [5]. The Reynolds stress has the interpretation of a turbulent momentum flux and the last term in (3) is also known as the eddy viscosity. It describes how the small scales influence the large scales. In addition we get divergence free conditions for U , and u

$$\nabla \cdot U = 0, \quad \nabla \cdot u = 0.$$

Together, (3) and the divergence free condition on U give Reynolds Averaged Navier-Stokes (RANS) that forms the basis for most contemporary simulations of turbulent flow.

Finding a constitutive law for the Reynolds stress $\overline{u \otimes u}$ is the famous closure problem in turbulence and we will solve that by writing down a stochastic equation for the small scale velocity u . The hypothesis is that the large scales influence the small scales directly, through the fluid instabilities and the noise in fully developed turbulence. An example of these mechanics, how the instabilities magnify the tiny ambient noise to produce large noise, is given in [6], see also Chap. 1 in [8].

The consequence of the above hypothesis is that the small scale velocity u in turbulence is a stochastic process that is determined by a stochastic partial differential equation (SPDE). This is the Eq. (4) and it is the Navier-Stokes equation driven by noise. This is the point of view taken by Kolmogorov in [15–17], but the question we have to answer is: what is the form of the noise? There is a large literature on this question, trying to trace the form of the noise back to the fluid instabilities, but these attempts have proven to be unsuccessful. Any memory of the fluid instabilities is quickly forgotten in fully-developed turbulence and the noise seems to be of a general form. Thus it makes sense to try to put generic noise into the Navier-Stokes equations and see how the Navier-Stokes evolution colors generic noise. Below we will answer what generic noise in the Navier-Stokes equation must look like, see [8] for more details.

Now consider the inertial range in turbulence. In Fourier space this is the range of wave numbers k : $\frac{1}{L} \leq |k| \leq \frac{1}{\eta}$, where $\eta = (\nu^3/\varepsilon)^{1/4}$ is the Kolmogorov length scale, ε is the energy dissipation and L the size of the largest eddies, see [8]. If we assume that dissipation takes place on all length scales in the inertial range then the form of the dissipation processes are determined by the fundamental theorems of probability. Namely, if we impose periodic boundary conditions (different boundary conditions correspond to different basis vectors), then the central limit theorem and the large deviation principle stipulate that the additive noise in the Navier-Stokes equation for the small scale must be of the form:

$$\sum_{k \neq 0} c_k^{\frac{1}{2}} db_t^k e_k(x) + \sum_{k \neq 0} d_k |k|^{1/3} dt e_k(x),$$

where $e_k(x) = e^{2\pi i k \cdot x}$ are the Fourier coefficient and $c_k^{\frac{1}{2}}$ and d_k are coefficients that ensure the series converge in 3 dimensions. The first term describes the mean of weakly coupled dissipation processes given by the central limit theorem and the second term describes the large deviations of that mean, given by the large deviation principle, see [8]. Thus together the two terms give a complete description of the mean of the dissipation process similar to the mean of many processes in probability. The factor $|k|^{1/3}$ implies that the mean dissipation has only one scaling. Notice that we do not impose any convergence rate on the coefficients $c_k^{1/2}$ and d_k beyond that the above term must make sense in $L^2(\mathbb{T}^3)$, the function space of square integrable functions where we want the SPDE to make sense. $\mathbb{T}^3$ is the three-dimensional torus because of the periodic boundary conditions on u . Thus any scaling must come from the Navier-Stokes evolutions acting on the noise. The Fourier coefficients of the first series contain independent Brownian motions b_t^k and thus the noise is white in time in the infinitely many directions in function space. The noise cannot be white in space, hence the decaying coefficients $c_k^{1/2}$ and d_k , because if it was the small scale velocity u would be discontinuous in 3 dimension, see [7]. This is contrary to what is observed in nature.

However, the noise must also have another multiplicative term. This part of the noise, in fully developed turbulence, models the noise associated with the excursion (jumps) in the velocity gradient or vorticity concentrations. It is known from simulations that such jumps in the velocity gradient of the fluid are present in fully-developed turbulence. If we let N_t^k denote the integer number of velocity excursion, associated with k th wavenumber, that have occurred at time t , so that the differential $dN^k(t) = N^k(t + dt) - N^k(t)$ denotes the number of excursions in the time interval $(t, t + dt]$, then the process $df_t^3 = \sum_{k \neq 0}^M \int_{\mathbb{R}} h_k(t, z) \bar{N}^k(dt, dz)$, gives the multiplicative noise term. One can show that any noise corresponding to jumps in the velocity gradients must have this multiplicative noise to leading order, see [7]. A detailed derivation of both the noise terms can be found in [7, 8].

Adding the additive noise and the multiplicative noise we get the stochastic Navier-Stokes equations describing the small scales in fully developed turbulence

$$\begin{aligned} du = & (\nu \Delta u - u \cdot \nabla u + \nabla \Delta^{-1} \text{tr}(\nabla u)^2) dt + \sum_{k \neq 0} c_k^{\frac{1}{2}} db_t^k e_k(x) + \sum_{k \neq 0} d_k |k|^{1/3} dt e_k(x) \\ & + u \left(\sum_{k \neq 0}^M \int_{\mathbb{R}} h_k \bar{N}^k(dt, dz) \right), \quad u(x, 0) = u_0(x), \end{aligned} \quad (4)$$

where we have used the divergence free condition $\nabla \cdot u = 0$ to eliminate the small scale pressure p . Each Fourier component e_k comes with its own Brownian motion b_t^k and a deterministic bound $|k|^{1/3} dt$.

2.2 Solution of the Stochastic Navier-Stokes

The next step is to figure out how the generic noise interacts with the Navier-Stokes evolution. This is determined by the integral form of the Eq. (4),

$$u = e^{Kt} e^{\int_0^t dq} M_t u^0 + \sum_{k \neq 0} \int_0^t e^{K(t-s)} e^{\int_s^t dq} M_{t-s} (c_k^{1/2} d\beta_s^k + d_k |k|^{1/3} ds) e_k(x), \quad (5)$$

where K is the operator $K = \nu \Delta + \nabla \Delta^{-1} \text{tr}(\nabla u \nabla)$, and we have omitted the terms $-U \cdot \nabla u - u \cdot \nabla U$ in (4), to simplify the exposition. We solve (4) using the Feynmann-Kac formula, and the Cameron-Martin formula (or Girsanov's Theorem) from probability theory, see [8], to get (5). The Cameron-Martin formula gives the Martingale $M_t = \exp\{-\int_0^t u(B_s, s) \cdot dB_s - \frac{1}{2} \int_0^t |u(B_s, s)|^2 ds\}$. The Feynmann-Kac formula gives the exponential of a sum of terms of the form $\int_s^t dq^k = \int_0^t \int_{\mathbb{R}} \ln(1 + h_k) N^k(dt, dz) - \int_0^t \int_{\mathbb{R}} h_k m^k(dt, dz)$, see [7] or [8] Chap. 2 for details. The form of the processes

$$e^{\int_0^t \int_{\mathbb{R}} \ln(1+h_k) N^k(dt, dz) - \int_0^t \int_{\mathbb{R}} h_k m^k(dt, dz)} = e^{N_t^k \ln \beta + \gamma \ln |k|} = |k|^\gamma \beta^{N_t^k} \quad (6)$$

was found by She and Leveque [22], for $h_k = \beta - 1$. It was pointed out by She and Waymire [23] and by Dubrulle [13] that they are log-Poisson processes. The upshot of this computation will be that we see the Navier-Stokes evolution acting on the additive noise to give the Kolmogorov-Obukhov'41 scaling, and the Navier-Stokes evolution acting on the multiplicative noise to produce the intermittency corrections through the Feynmann-Kac formula. Together these two scaling then combine to give the scaling of the structure functions in turbulence. This will become clear when we consider the structure functions and invariant measure of the stochastic Navier-Stokes equation below.

3 The Kolmogorov-Obukhov-She-Leveque Scaling

The structure functions in turbulence are defined to be

$$S_p(|x - y|, t) = E(|\delta u|^p) = E(|u(x, t) - u(y, t)|^p),$$

$p \in \mathbb{N}$, where E is the expectation, that is substituted by an ensemble average in simulations and experiments. $l = |x - y|$ is called the lag variable. The structure functions are a better probe of turbulence than the usual moments of the velocity, since the latter exhibit (skewed) Gaussian behavior and do not reveal the scaling found in turbulent fluids.

3.1 Computation of the Structure Functions

The structure functions are non-negative so it suffices to estimate them from above, when the lag variable is small the estimates are sharp.

Lemma 3.1 (The Kolmogorov-Obukhov-She-Leveque scaling) *The scaling of the structure functions is*

$$S_p \sim C_p |x - y|^{\zeta_p}, \quad \zeta_p = \frac{p}{3} + \tau_p = \frac{p}{9} + 2(1 - (2/3)^{p/3}).$$

$\frac{p}{3}$ being the Kolmogorov scaling and τ_p the intermittency corrections, for $l = |x - y|$ small. The scaling of the structure functions is consistent with Kolmogorov's 4/5 law, $S_3 = -\frac{4}{5}\varepsilon|x - y|$, to leading order, where $\varepsilon = -\frac{dE}{dt}$ is the energy dissipation.

Here S_3 is the structure function of the velocity differences, without the absolute value.

3.2 The First Few Structure Functions

The first structure functions is estimated by

$$S_1(x, y, \infty) \leq \frac{2}{C} \sum_{k \in \mathbb{Z}^3 \setminus \{0\}} \frac{|d_k|(1 - e^{-\lambda_k t})}{|k|^{\zeta_1}} |\sin(\pi k \cdot (x - y))|.$$

We get a stationary state as $t \rightarrow \infty$, and for $|x - y|$ small, $S_1(x, y, \infty) \sim \frac{2\pi\epsilon_1}{C} \sum_{k \in \mathbb{Z}^3 \setminus \{0\}} |d_k| |x - y|^{\zeta_1}$, where $\zeta_1 = 1/3 + \tau_1 \approx 0.37$. Similarly, $S_2(x, y, \infty) \sim \frac{4\pi\epsilon_2}{C^2} \sum_{k \in \mathbb{Z}^3 \setminus \{0\}} [d_k^2 + (\frac{C}{2})c_k] |x - y|^{\zeta_2}$, when $|x - y|$ is small, where $\zeta_2 = 2/3 + \tau_2 \approx 0.696$, and $S_3(x, y, \infty) \sim \frac{2^3\pi}{C^3} \sum_{k \in \mathbb{Z}^3 \setminus \{0\}} [|d_k|^3 + 3(C/2)c_k |d_k|] |x - y|$. For the p th structure functions, we get that S_p is estimated by

$$S_p \leq \frac{2^p}{C^p} \sum_{k \in \mathbb{Z}^3 \setminus \{0\}} \frac{\sigma^p 2^{p/2} \Gamma((1+p)/2) {}_1F_1\left(-\frac{1}{2}p, \frac{1}{2}, -\frac{1}{2}(M/\sigma)^2\right)}{\sqrt{\pi}|k|^{\zeta_p}} |\sin^p(\pi k \cdot (x - y))|,$$

where ${}_1F_1$ is the hypergeometric function, $M = |d_k|(1 - e^{-\lambda_k t})$ and $\sigma = \sqrt{(C/2)c_k(1 - e^{-2\lambda_k t})}$. The details of these estimates are given in [7, 8]. These values are consistent with the values found in high-Reynolds number experiments, see [12, 25].

The integral equation can be considered to be an infinite-dimensional Ito process, see [8]. This means that we can find the associated Kolmogorov backward equation for the Ito diffusion associated with the Eq. (5) and this equation that determines the invariant measure of turbulence, see [7], is linear. This was first attempted by Hopf

[14] wrote down a functional differential equation for the characteristic function of the invariant measure of the deterministic Navier-Stokes equation. The Kolmogorov-Hopf (backward) equation for (4) is

$$\frac{\partial \phi}{\partial t} = \frac{1}{2} \text{tr}[P_t C P_t^* \Delta \phi] + \text{tr}[P_t \bar{D} \nabla \phi] + \langle K(z) P_t, \nabla \phi \rangle, \quad (7)$$

see [7, 8] Chap. 3, where $\bar{D} = (|k|^{1/3} D_k)$, $\phi(z)$ is a bounded function of z , $P_t = e^{-\int_0^t \nabla u \, dr} M_t \prod_k^m |k|^{2/3} (2/3)^{N_t^k}$. The variance and drift are defined to be

$$Q_t = \int_0^t e^{K(s)} P_s C P_s^* e^{K^*(s)} ds, \quad E_t = \int_0^t e^{K(s)} P_s \bar{D} ds. \quad (8)$$

3.3 The Invariant Measure of the Stochastic Navier-Stokes

In distinction to the nonlinear Navier-Stokes equation (4) that cannot be solved explicitly, the linear equation (7) can be solved. The solution of the Kolmogorov-Hopf equation (7) is

$$R_t \phi(z) = \int_H \phi(e^{K_t} P_t z + E_t I + y) \mathcal{N}_{(0, Q_t)} * \mathbb{P}_{N_t}(dy),$$

$\mathbb{P}_{N_t}$ being the law of the log-Poisson process (6). $\mathcal{N}_{(E, Q)}$ is the law of an infinite-dimensional Gaussian process with mean E and variance Q and $*$ denotes convolution. The invariant measure of turbulence that appears in the last equation can now be expressed explicitly,

Theorem 3.1 *The invariant measure of the Navier-Stokes equation on $H_c = H^{3/2+}(\mathbb{T}^3)$ is,*

$$\mu(dx) = e^{\langle Q^{-1/2} E I, Q^{-1/2} x \rangle - \frac{1}{2} |Q^{-1/2} E I|^2} \mathcal{N}_{(0, Q)}(dx) \sum_k \delta_{k,l} \prod_{j \neq l}^m \delta_{N_t^j} \sum_{j=0}^{\infty} p_{m_l}^j \delta_{(N_l - j)}$$

where $Q = Q_\infty$, $E = E_\infty$, $m_k = \ln |k|^{2/3}$ is the mean of the log-Poisson processes (6) and $p_{m_k}^j = \frac{(m_k)^j e^{-m_k}}{j!}$ is the probability of $N_\infty^k = N_k$ having exactly j jumps, $\delta_{k,l}$ is the Kroncker delta function and $\delta_{N_t^j}$ is the Dirac delta function concentrated at N_t^j .

This shows that the invariant measure of turbulence is simply a product of two measures, one an infinite-dimensional Gaussian that gives the Kolmogorov-Obukhov scaling and the other a discrete Poisson measure that gives the She-Leveque intermittency corrections. Together they produce the scaling of the structure functions in Lemma 1. This makes it clear how the Navier-Stokes evolution operating on the additive and the multiplicative noise in Eq.(4) produces the scaling. Miraculously, the

infinite dimensional Gaussian part (Kolmogorov-Obukhov) and the intermittency part (She-Leveque) separate in the invariant measure on the infinite-dimensional functions space, one represented by a continuous measure the other one by a discrete measure. It is when we project this measure onto measures of experimental or simulated quantities, see below, that these two parts get mixed.

3.4 The Differential Equation for the PDF

A quantity that can be compared directly to experiments and simulations is the probability density function (PDF) for the velocity differences δu . We take the trace of the Kolmogorov-Hopf equation (7), see [8] Chap. 3, to compute the differential equation satisfied by the PDF for δu . First we do this ignoring the intermittency corrections τ_p in Lemma 1, see [9] for details. The stationary equation satisfied by the PDF without intermittency corrections is

$$\frac{1}{2} \phi + \frac{1 + |c|}{r} \phi = \frac{1}{2} \phi. \quad (9)$$

The probability density functions (PDF) is going to be a solution of this equation, if the intermittency corrections τ_p are ignored, the details of the computation are given in [7, 8].

4 The Probability Density Function

4.1 The Generalized Hyperbolic Distributions

Lemma 4.1 *The PDF, without intermittency corrections, is a Generalized Hyperbolic Distribution (GHD) of Barndorff-Nielsen [3]:*

$$f(x) = \frac{(\gamma/\delta)^\lambda}{\sqrt{2\pi} K_\lambda(\delta\gamma)} \frac{K_{\lambda-\frac{1}{2}}\left(\alpha\sqrt{\delta^2 + (x-\mu)^2}\right) e^{\beta(x-\mu)}}{\left(\sqrt{\delta^2 + (x-\mu)^2}/\alpha\right)^{\lambda-\frac{1}{2}}} \quad (10)$$

where K_λ is the modified Bessel's function of the second kind with index λ , $\gamma = \sqrt{\alpha^2 - \beta^2}$. α , β , δ and μ are parameters.

Equation (10) is the solution of (9), see [7, 8] for details of the proof, and the PDF that can be compared a large class of experimental data.

5 The PDF of Turbulence

The PDF of δu becomes more complicated when the intermittency is included. Then it becomes impossible to have a single continuous PDF for all the different moments and instead one has to have a distribution that is a product of a discrete and continuous distributions, just as the invariant measure of turbulence Theorem 3.1 itself. One can put in the intermittency correction in the Eq. (9) defining the PDF and get a different continuous PDF for each moment, this is done in [9]. In this paper we will take a different approach and project the invariant measure Theorem 3.1 to a PDF that is a product of a continuous and a discrete measure analogous to the invariant measure itself. The continuous part of the PDF will be the Generalized Hyperbolic Distribution (10).

We start with the log-Poisson process $|x| \left(\frac{2}{3}\right)^{N_t^k}$ and the mean $m_k = \ln(|x|^{-6})$ of the associated Poisson distribution. Now the mixed continuous and discrete distribution is given by:

$$\bar{\mu}(\cdot) = \int_{-\infty}^{\infty} \sum_{j=0}^{\infty} \frac{(\ln(|x|^{-6}))^j}{j!} |x|^6 \delta_{N_t^k - j}(\cdot) f(x) dx, \quad (11)$$

where $\bar{\mu}$ denotes the projection of the measure. We assume that the velocity is a Hölder continuous function of Hölder index $1/3$, see [8]. Then evaluating the measure on the p th moment of the velocity differences gives,

$$\begin{aligned} \int_{-\infty}^{\infty} \sum_{j=0}^{\infty} \frac{(\ln(|x|^{-6}))^j}{j!} |x|^6 \delta_{N_t^k - j}(|x| \left(\frac{2}{3}\right)^{N_t^k})^{\frac{p}{3}} f(x) dx &= \int_{-\infty}^{\infty} |x|^{\frac{p}{3}} |x|^{6(1-(2/3)^{\frac{p}{3}})} f(x) dx \\ &= \int_{-\infty}^{\infty} |x|^{p+3\tau_p} f(x) dx = \int_{-\infty}^{\infty} |x|^{3\zeta_p} f(x) dx, \end{aligned}$$

where

$$\zeta_p = \frac{p}{3} + \tau_p$$

is the scaling exponent of the p th structure function, with the intermittency correction τ_p . The upshot is that the discrete part of the PDF adds the intermittency correction $|x|^{3\tau_p}$ to the p th moment and

$$\bar{\mu}(|\delta u|^p) = \int_{-\infty}^{\infty} |x|^{p+3\tau_p} f(x) dx, \quad (12)$$

where δu are the velocity differences and the intermittency corrections are $\tau_p = -\frac{2p}{9} + 2(1 - (2/3)^{\frac{p}{3}})$.

We have shown that the PDF (11) of the velocity differences in turbulence consists of a continuous distribution convolved with a discrete one. This is a direct consequence of the similar structure of the invariant measure Theorem 3.1 of

turbulence. The continuous part of the PDF corresponds to the Kolmogorov-Obukhov's theory represented by the infinite-dimensional Gaussian measure in Theorem 3.1, whereas the discrete part corresponds to the She-Leveque intermittency correction just as it does in Theorem 3.1. Remarkably, all of this follows from the Navier-Stokes equations when generic noise is added to it (4).

Acknowledgments This work was supported by a grant from the UC Santa Barbara Academic Senate and a Chair of Excellence at Universidad de Carlos III, Madrid, whose support is gratefully acknowledged.

References

1. Bagnold, R.A.: *The Physics of Blown Sand and Desert Dunes*. Methuen, London (1954)
2. Barndorff-Nielsen, O.: Models for non-gaussian variation, with applications to turbulence. *Proc. R. Soc. Lond. A. Math. Phys. Sci.* **368**(1735), 501–520 (1979)
3. Barndorff-Nielsen, O.E.: Exponentially decreasing distributions for the logarithm of the particle size. *Proc. R. Soc. Lond. A* **353**, 401–419 (1977)
4. Barndorff-Nielsen, O.E., Blaesild, P., Schmiegel, Jörgen: A parsimonious and universal description of turbulent velocity increments. *Eur. Phys. J. B* **41**, 345–363 (2004)
5. Bernard, P.S., Wallace, J.M.: *Turbulent Flow*. Wiley, Hoboken (2002)
6. Birnir, B.: Turbulence of a unidirectional flow. In: *Proceedings of the Conference on Probability, Geometry and Integrable Systems*, MSRI, Dec. 2005 MSRI Publications, vol. 55. Cambridge University Press (2007). <http://repositories.cdlib.org/cnls/>
7. Birnir, B.: The Kolmogorov-Obukhov statistical theory of turbulence. *J. Nonlinear Sci.* (2013). doi:[10.1007/s00332-012-9164-z](https://doi.org/10.1007/s00332-012-9164-z)
8. Birnir, B.: *The Kolmogorov-Obukhov Theory of Turbulence*. Springer, New York (2013)
9. Birnir, B.: The Kolmogorov-Obukhov-She-Leveque scaling in turbulence. *Commun. Pure Appl. Anal.* **13**(5) (2014)
10. Birnir, B.: The probability density function of turbulence. CNLS preprint to appear (2015)
11. Birnir, B., Chen, X., Liu, C.: Moments and probability density functions in turbulent boundary layers. CNLS preprint to appear (2015)
12. Chen, S.Y., Dhruva, B., Kurien, S., Sreenivasan, K.R., Taylor, M.A.: Anomalous scaling of low-order structure functions of turbulent velocity. *Journ. Fluid Mech.* **533**, 183–192 (2005)
13. Dubrulle, B.: Intermittency in fully developed turbulence: in log-Poisson statistics and generalized scale covariance. *Phys. Rev. Lett.* **73**(7), 959–962 (1994)
14. Hopf, E.: Statistical hydrodynamics and functional calculus. *J. Rat. Mech. Anal.* **1**(1), 87–123 (1953)
15. Kolmogorov, A.N.: Dissipation of energy under locally isotropic turbulence. *Dokl. Akad. Nauk SSSR* **32**, 16–18 (1941)
16. Kolmogorov, A.N.: The local structure of turbulence in incompressible viscous fluid for very large Reynolds number. *Dokl. Akad. Nauk SSSR* **30**, 9–13 (1941)
17. Kolmogorov, A.N.: A refinement of previous hypotheses concerning the local structure of turbulence in a viscous incompressible fluid at high Reynolds number. *J. Fluid Mech.* **13**, 82–85 (1962)
18. Obukhov, A.M.: On the distribution of energy in the spectrum of turbulent flow. *Dokl. Akad. Nauk SSSR* **32**, 19 (1941)
19. Obukhov, A.M.: Some specific features of atmospheric turbulence. *J. Fluid Mech.* **13**, 77–81 (1962)
20. Reynolds, O.: An experimental investigation of the circumstances which determine whether the motion of water shall be direct or sinuous, and the law of resistance in parallel channels. *Phil. Trans. Roy. Soc. Lond.* **174**(11), 935–982 (1883)

21. Reynolds, O.: On the dynamical theory of incompressible viscous fluids and the determination of the criterion. *Phil. Trans. Roy. Soc. Lond.* **186A**, 123–164 (1885)
22. She, Z.-S., Leveque, E.: Universal scaling laws in fully developed turbulence. *Phys. Rev. Lett.* **72**(3), 336–339 (1994)
23. She, Z.-S., Waymire, E.: Quantized energy cascade and log-poisson statistics in fully developed turbulence. *Phys. Rev. Lett.* **74**(2), 262–265 (1995)
24. Sørensen, Michael: On the rate of aeolian sand transport. *Geomorphology* **59**(1), 53–62 (2004)
25. Sreenivasan, K.R., Dhruva, B.: Is there scaling in high- Reynolds-number turbulence? *Prog. Theor. Phys. Suppl.* 103–120 (1998)

Modelling Turbulent Time Series by *BSS*-Processes

José Ulises Márquez and Jürgen Schmiegel

Abstract Brownian semi-stationary processes have been proposed as a class of stochastic models for time series of the turbulent velocity field. We show, by detailed comparison, that these processes are able to reproduce the main characteristics of turbulent data. Furthermore, we present an algorithm that allows to estimate the model parameters from second and third order statistics. As an application we synthesise a turbulent time series measured in a helium jet flow.

Keywords Turbulence · Brownian semi-stationary processes · Ambit fields

1 Introduction

Stochastic modelling of the turbulent velocity field, understood as an explicit stochastic approach (in contrast to an implicit set up in terms of governing equations and/or in terms of related quantities like velocity increments or velocity derivatives), is generally speaking not well developed in the literature. Most of the existing literature on stochastic turbulence modelling deals with models for derived quantities like velocity increments, energy dissipation and accelerations.

Early attempts to model the rapid variation of the turbulent velocity field include [4, 10, 11, 18, 25, 36] (among many others). Such phenomenological approaches are sometimes called “synthetic turbulence” and can be divided into two classes. The first direction starts from modelling the velocity field and derives the model for the energy dissipation by taking squared small scale increments. The second line of investigation focuses on modelling the energy dissipation field and derives the velocity field by various, partly ad hoc, manipulations. The approach presented here

J.U. Márquez · J. Schmiegel (✉)

Department of Mathematics, Aarhus University, Ny Munkegade 118,
8000 Aarhus C, Denmark
e-mail: schmiegel@imf.au.dk

J.U. Márquez

e-mail: ulises@math.au.dk

conceives the energy dissipation as the fundamental field, entering directly the model for the velocity field and obeying the physical interpretation of the energy dissipation as the squared small scale fluctuations.

In [36], an iterative, geometric multi-affine model for the one-dimensional velocity process is constructed and some of the basic, global statistical quantities of the energy dissipation field are derived. However, this discrete, dyadic approach does not allow to give explicit expressions for more specific statistical quantities.

Another dyadic, iterative approach for the construction of the velocity field is discussed in [10]. Their model is based on a wavelet decomposition of the velocity field combined with a multiplicative cascading structure for the wavelet coefficients. As discussed in [18], such wavelet approaches are superior over discrete geometric approaches as they allow to model stationarity in a mathematical more rigorous way. The approach discussed here does not suffer from problems related to mathematical rigour and no iterative limit arguments are needed for the construction. A related and interesting wavelet-based approach is discussed in [11], which allows for a sequential construction of the field. A further wavelet-based approach [4] builds on random functions and their orthogonal wavelet transform. The authors show that to each such random function there is an associated cascade on a dyadic tree of wavelet coefficients.

The models [10, 36] fail to incorporate skewness for the velocity increments [25], a basic property of turbulent fields. As an alternative approach, [25] proposes a combination of a multiplicative cascade for the energy dissipation, the use of Kolmogorov's refined similarity hypothesis [26] and an appropriate summation rule for the increments to construct the velocity field. Here, again, only discrete iterative procedures are employed which make analytical statistical statements very difficult.

The stochastic models discussed in the present paper, called Brownian semi-stationary processes, have been proposed to be potentially suitable for turbulence modelling in [7, 8]. These processes define the turbulent velocity field explicitly and as such allow for analytic calculations and identification of the parameters of the model with physical quantities.

In [7] it has been shown that Brownian semi-stationary processes are able to qualitatively reproduce some aspects of turbulence statistics like the evolution of the densities of velocity increments across scales and the conditional statistics of the so-called Kolmogorov variable. Here we will extend and quantify in detail the comparison of the model with empirical data by including more stylised features of turbulent data. Our goal is to estimate the parameters entering Brownian semi-stationary processes from a given turbulent data set. Based on this estimation, a numerical simulation of the model is then compared in great detail with the turbulent data set at hand, including statistical properties not used for the estimation procedure.

The paper is organised as follows. In Sect. 2 we list the main stylised features of turbulent time series we use to validate the model. Brownian semi-stationary processes as models for the turbulent velocity field along with cascades processes as

models for the energy dissipation are presented in Sect. 3. Section 4 addresses the estimation procedure for the parameters of the model and briefly outlines the numerics behind the simulations. Finally, Sect. 5 concludes and summarises the results.

2 Stylised Features of Turbulent Time Series

In general, turbulence concerns the dynamics in a fluid flow of the three-dimensional velocity vector $\mathbf{v}(\mathbf{r}, t) = (v_x(\mathbf{r}, t), v_y(\mathbf{r}, t), v_z(\mathbf{r}, t))$ as a function of position $\mathbf{r} = (x, y, z)$ and time t . A derived quantity is the energy dissipation, defined as

$$\varepsilon(\mathbf{r}, t) \equiv \frac{\nu}{2} \sum_{i,j=x,y,z} (\partial_i v_j(\mathbf{r}, t) + \partial_j v_i(\mathbf{r}, t))^2. \quad (1)$$

The energy dissipation describes the loss of kinetic energy due to friction forces characterised by the viscosity ν .

A pedagogically valuable illustration of a turbulent flow can be gained from the Kolmogorov cascade [20]. In this representation kinetic energy is injected into the flow at large scales through large scale forcing. Non-linear effects redistribute the kinetic energy towards smaller scales. This cascade of energy stops at small scales where dissipation transforms kinetic energy into heat. It is traditional to call the large scale l of energy input the integral scale and the small scale η of dominant dissipation the dissipation scale or Kolmogorov scale. With increasing Reynolds number the fraction l/η increases, giving space for the so called inertial range $\eta \ll l \ll I$ where turbulent statistics are expected to have some universal character. A more precise definition defines the inertial range as the range of scales $1/k$ where the spectrum $E(k)$ (the Fourier transform of the correlation function of the velocity field) displays a power law $E(k) \propto k^{-5/3}$ [20].

The resolution of all dynamically active scales in experiments is at present not achievable for the full three-dimensional velocity vector. Most experiments measure a time series of one component v (in direction of the mean flow) of the velocity vector at a fixed single location $\mathbf{r}_0$. Based on this restriction one defines the temporal (or surrogate) energy dissipation for stationary, homogeneous and isotropic flows

$$\varepsilon_t(\mathbf{r}_0) \equiv \frac{15\nu}{\bar{v}^2} \left(\frac{dv(\mathbf{r}_0, t)}{dt} \right)^2, \quad (2)$$

where $\bar{v}$ denotes the mean velocity (in direction of the mean flow).

The temporal energy dissipation (2) is expected to approximate basic statistical properties of the true energy dissipation (1) for stationary, homogeneous and isotropic flows. For other flow conditions, the temporal energy dissipation still contains important statistical information about the turbulent velocity field.

The transformation of the spatial derivatives in (1) to the temporal derivative in (2) is performed under the assumption of a stationary, homogeneous and isotropic flow and the assumption of Taylor's Frozen Flow Hypothesis [35] which states that spatial structures of the flow are predominantly swept by the mean velocity $\bar{v}$ without relevant distortion. Under this hypothesis, widely used in analysing turbulent time series, spatial increments along the direction of the mean flow (in direction x) are expressed in terms of temporal increments

$$v(x, y, z, t + s) - v(x, y, z, t) = v(x - \bar{v}s, y, z, t) - v(x, y, z, t). \quad (3)$$

In the present paper, we only deal with homogeneous, isotropic and stationary turbulence. Furthermore, we restrict ourselves to temporal statistics at a fixed position in space and refer to the inertial range as the temporal counterpart of the spatial inertial range defined by time scales s where $\eta/\bar{v} \ll s \ll I/\bar{v}$. Time scales $s \lesssim \eta/\bar{v}$ are called dissipation time scales and time scales $s \gtrsim I/\bar{v}$ are called integral time scales.

In what follows, the notion energy dissipation refers to the temporal energy dissipation, unless otherwise stated. We also skip reference to the spatial location $\mathbf{r}_0$ in (2) and write ε_t for $\varepsilon_t(\mathbf{r}_0)$.

The most striking feature of time series of the energy dissipation is the strong variability with localised and clustered outbursts of different size and duration. This strongly fluctuating behaviour, which is far away from what might be expected in a Gaussian framework, is called the *intermittency* of the energy dissipation.

The traditional characterisation of the intermittent behaviour of the energy dissipation refers to the coarse grained field amplitude over a time horizon T

$$\bar{\varepsilon}(T, t) = \frac{1}{T} \int_{t-T/2}^{t+T/2} \varepsilon_s ds. \quad (4)$$

It has been shown in numerous publications that the moments of the coarse grained energy dissipation follow a scaling law in the inertial range for large Reynolds numbers

$$\mathbb{E} \left\{ \bar{\varepsilon}(T, t)^n \right\} \propto T^{-\xi(n)} \quad (5)$$

where the positive multifractal scaling exponents $\xi(n)$ are expected to be universal in the limit of very large Reynolds number (cf. e.g. [20, 32] and references therein). The term multifractality refers to the non-linear dependence of the scaling exponents $\xi(n)$ on the order n . The notion of a Reynolds number refers to the time-wise defined Taylor micro-scale Reynolds number [20]

$$R = \frac{\text{Var}\{v\}}{v\sqrt{\mathbb{E}\{\varepsilon_t\}}} \quad (6)$$

where Var denotes the variance.

An immediate consequence of the scaling relation (5) in second order $n = 2$ is scaling of correlators $c_{p,q}$ of order $(p, q) = (1, 1)$. These correlators are defined as

$$c_{p,q}(s) \equiv \frac{\mathbb{E} \{ \varepsilon_t^p \varepsilon_{t+s}^q \}}{\mathbb{E} \{ \varepsilon_t^p \} \mathbb{E} \{ \varepsilon_{t+s}^q \}}. \quad (7)$$

The empirical analysis of $c_{p,q}$ revealed the existence of a range of scales s where

$$c_{p,q}(s) \propto s^{-\tau(p,q)} \quad (8)$$

and $\tau(1, 1) = \xi(2)$ [2, 14–16, 21, 28–31].

Intermittency of the velocity field refers to the fact that fluctuations around the mean velocity occur in clusters and are more violent than expected from Gaussian statistics. Furthermore, the frequency of large fluctuations increases with increasing resolution. In terms of moments of temporal velocity increments

$$u_s(t) \equiv v_{s+t} - v_t, \quad s > 0 \quad (9)$$

intermittency is usually described by (approximate) multifractal scaling of structure functions (e.g. [1, 27])

$$S_n(s) = \mathbb{E} \{ u_s(t)^n \} \propto s^{\tau(n)}. \quad (10)$$

Here, v_t is one component of the velocity (usually along the mean flow) at time t and at a fixed position and the time scale s is within the inertial range. When appropriate, we write u_s instead of $u_s(t)$ in (9) since we are only dealing with stationary time series.

Multifractal scaling of structure functions is assumed to hold in the limit of infinite Reynolds number [20]. However, experiments show that the scaling behaviour (10) might be poor, even for large Reynolds numbers [3, 32]. Furthermore, even if the scaling relation (10) holds, the inertial range still covers only part of the accessible scales where intermittency is observed.

From a probabilistic point of view, (10) expresses a scaling relation for the moments of the probability density function (pdf) of velocity increments. A proper estimation of higher-order moments requires an accurate estimation of the tails of the pdf. Thus it may be advantageous to directly work with the pdf. In terms of the pdf, intermittency refers to the increase of the non-Gaussian behaviour of the pdf of velocity increments with decreasing time scale.

A typical scenario is characterised by an approximate Gaussian shape for the large scales (larger than scales at the inertial range), turning to exponential tails within the inertial range and stretched exponential tails for dissipation scales (below the inertial range). This change of shape across all scales clearly reveals the inadequacy of a characterisation of intermittency solely via multifractal scaling of structure functions (which is observed only within the inertial range). In [6, 9] it is shown that

normal inverse Gaussian (NIG) distributions are well adapted to accurately describe the densities of velocity increments at all scales and for a wide range of Reynolds numbers.

In 1962, Kolmogorov [26] published two hypotheses (usually referred to as K62) about a quantity V that combines velocity increments, being a large scale quantity, and the energy dissipation, being a small scale quantity. The first hypothesis states that the pdf of the stochastic variable

$$V_r = \frac{\Delta v_t(r)}{(r\varepsilon_r)^{1/3}} \quad (11)$$

depends, for $r \ll L$, only on the local Reynolds number

$$\text{Re}_r = r(r\varepsilon_r)^{1/3}/\nu. \quad (12)$$

Here,

$$\Delta v_t(r) = v_t(x + r/2, y, z) - v_t(x - r/2, y, z) \quad (13)$$

denotes the increment of the component v of the velocity vector in direction of the mean flow (the x -direction) at scale r and $r\varepsilon_r$ is the integrated energy dissipation (1) over a domain of linear size r

$$\varepsilon_r = \frac{1}{r} \int_{x-r/2}^{x+r/2} \varepsilon(\sigma, y, z, t) d\sigma. \quad (14)$$

The second hypothesis states that, for $\text{Re}_r \gg 1$, the pdf of V_r does not depend on Re_r , either, and is therefore universal.

Although, for small r , an additional r dependence of the pdf of V_r has been observed [33], the validity of several aspects of K62 has been verified experimentally and by numerical simulation of turbulence [24, 33, 34, 37]. In particular it has been shown that the conditional densities $p(V_r|r\varepsilon_r)$ become independent of $r\varepsilon_r$ for a certain range of scales r within the inertial range. However, the universality of the distribution of V has not been verified in the literature. In this respect, it is important to note that the experimental verification of the Kolmogorov hypotheses is, with reasonable resolution of scales, restricted to temporal statistics and as such relies on the use of the temporal energy dissipation (2) instead of the true energy dissipation (1). In the time domain, the Kolmogorov variable V is defined as

$$V_{t,s} = \frac{u_s(t - s/2)}{(\bar{v}s\bar{\varepsilon}(s, t))^{1/3}} \quad (15)$$

where $u_s(t)$ denotes the temporal velocity increment (9) at time scale s , $\bar{v}$ is the mean velocity and $\bar{\varepsilon}(s, t)$ the coarse grained temporal energy dissipation (4).

3 Modelling Framework

In this Section we present the stochastic framework for modelling turbulent time series. One of the main ingredients of the model is the surrogate energy dissipation which, in our approach, will be modelled as a continuous cascade process. We briefly discuss cascade models in Sect. 3.1. Section 3.2 presents the model for the temporal turbulent velocity field along with its most relevant properties.

3.1 The Cascade Model for the Turbulent Energy Dissipation

Our model for the turbulent velocity field takes as an input the temporal energy dissipation which is modelled as a continuous cascade model [23]. In [23] it is shown that this Lévy based approach is able to reproduce the main stylised features of the energy dissipation observed for a wide range of data sets, including the data we analyse in the present paper.

Specifically, we model the temporal energy dissipation ε as the exponential of an integral with respect to a homogeneous Lévy basis L on $\mathbb{R}^2$,

$$\varepsilon_t = \exp \left(\int_{A(t)} L(dy, ds) \right) = \exp(L(A(t))), \quad t \in \mathbb{R}, \quad (16)$$

where $A(t) = A + (0, t)$ for a bounded set $A \subset \mathbb{R}^2$. The set $A(t)$ is called the *ambit* set. From the homogeneity of L it follows that (16) is a stationary stochastic process. For details about Lévy bases and the derivation of some of the properties of (16), we refer to [23] and the references therein.

The ambit set A is given as

$$A = \{(x, t) : 0 \leq t \leq T, -f(t) \leq x \leq f(t)\}, \quad (17)$$

where $T > 0$. For $T > 0$, $k > 1$, and $\theta > 0$, the function f is defined as

$$f(t) = \left(\frac{1 - (t/T)^\theta}{1 + (k \cdot t/T)^\theta} \right)^{1/\theta} \quad 0 \leq t \leq T. \quad (18)$$

This specification of the ambit set is adapted to reproduce the empirically observed scaling of correlators.

In [23] it is shown that the density of the logarithm of the energy dissipation is well described by a normal inverse Gaussian distribution, i.e. $\log \varepsilon_t \sim \text{NIG}(\alpha, \beta, \mu, \delta)$. For the correlators $c_{p,q}$ (7) to exist it is necessary to assume that ε_t has exponential moments of order $p + q$ leading to the condition $p + q < \alpha - \beta$. As discussed in [23], for a realistic modelling, it is enough to require existence of $c_{p,q}$ up to order $p + q = 4.5$. Furthermore, we set $E\{\varepsilon\} = 1$ for convenience. Under these

Table 1 Parameters for the data set analysed in this paper

Data				Ambit set			NIG-law			
R	η	I	$\hat{f}$	T	k	θ	α	β	μ	δ
985	0.21	443.9	367500	880	10000	2.20	2.50	-2.00	2.42	3.06

R denotes the Taylor micro-scale Reynolds number, η is the Kolmogorov scale (in units of the finest resolution), I denotes the integral scale (in units of the finest resolution) and $\hat{f}$ is the sampling frequency. The parameters T (in units of the finest resolution), k and θ characterise the ambit set (17). The parameters α , β , μ and δ specify the NIG-law of L in (16)

constraints, the parameters of the underlying NIG-laws and the corresponding ambit sets have been estimated in [23] for a number of turbulent data sets including the one we use in our analysis. Table 1 lists these parameters.

3.2 A Stochastic Model for Turbulent Velocity Time Series

Brownian semi-stationary (BSS) processes, introduced in [8] as potential models for turbulent velocity time series, are stochastic processes of the form

$$Z_t = \mu + \int_{-\infty}^t g(t-s) \sigma_s dW_s + \int_{-\infty}^t q(t-s) a_s ds, \quad t \in \mathbb{R}, \quad (19)$$

where μ is a constant, $(W_t)_{t \in \mathbb{R}}$ is a standard Brownian motion, g and q are nonnegative deterministic functions on $\mathbb{R}_+$ and $(\sigma_t)_{t \in \mathbb{R}}$ and $(a_t)_{t \in \mathbb{R}}$ are càdlàg processes. When (σ, a) is stationary and independent of W , then Z is stationary.

In general, BSS processes are not necessarily semimartingales. However, in our modelling application, the choice of the ingredients of the model (19) ensures the semimartingale property. For that reason, we focus on the special case where (19) constitutes a semimartingale, keeping in mind that many of the arguments in the sequel are equally true for the general case. A sufficient condition for Z to be a semimartingale is that σ and a have finite second moments, $g, q \in L^1(\mathbb{R}_+) \cap L^2(\mathbb{R}_+)$, $g' \in L^2(\mathbb{R}_+)$ and $g(0+) < \infty$ (see [17]).

Following [7], we model the time series of the main component v_t of the turbulent velocity field as a BSS process of the specific form

$$v_t = v_t(g, \sigma, \beta) = \int_{-\infty}^t g(t-s) \sigma_s dW_s + \beta \int_{-\infty}^t g(t-s) \sigma_s^2 ds \equiv R_t + \beta S_t, \quad t \in \mathbb{R}, \quad (20)$$

where $g \in L^1(\mathbb{R}_+) \cap L^2(\mathbb{R}_+)$ is a non-negative function, σ is a stationary process independent of W with $E[\sigma^6] < \infty$, and β is a constant.

It is well-known that for any semimartingale X the limit

$$[X]_t = \lim_{n \rightarrow \infty} \sum_{j=1}^n \left(X_{j \frac{t}{n}} - X_{(j-1) \frac{t}{n}} \right)^2 \quad (21)$$

exists as a limit in probability. The derived process $[X]$ expresses the cumulative quadratic variation exhibited by X and is called quadratic variation. For the case where (20) is a semimartingale, using Itô algebra, we get

$$(dv_t)^2 = g^2(0+) \sigma_t^2 dt$$

and

$$[v]_t = \int_0^t (dv_s)^2 = g^2(0+) \int_0^t \sigma_s^2 ds. \quad (22)$$

In this setting, the quantity $(dv_t)^2 / dt$ is the natural analogue of the squared first order derivative of v which in the classical formulation is taken to express the temporal energy dissipation (2). Consequently, the quadratic variation $[v]$ is the stochastic analogue of the integrated energy dissipation and σ^2 can be identified with the temporal energy dissipation. We will therefore assume that $\sigma^2 = \varepsilon$, where ε is the process given by (16). Note that in this set-up the Kolmogorov variable V (15) is given as

$$V_{t,s} = \frac{u_s(t - s/2)}{(\bar{v} [v]_s^t)^{1/3}} \quad (23)$$

and the conditional independence refers to the independence of $p(V_{t,s} | [v]_s^t)$ on $[v]_s^t$. Here $[v]_s^t = [v]_{t+s/2} - [v]_{t-s/2}$ denotes the quadratic variation over the time horizon $[t - s/2, t + s/2]$.

The limit in (21) may not exist in the non-semimartingale case. However, even in the general case, σ^2 can still be identified, up to a normalisation, with the surrogate energy dissipation (see [17, Theorem 3.1]).

Note that $[v]$ in (22) is independent of the second term in (20). This second term determines the skewness of the density of velocity increments $u_t = v_t - v_0$. For this reason we refer to β as the skewness parameter.

For the specification of the kernel g we start the discussion following [22] where a convolution of gamma kernels was proposed to model the second order statistics of turbulent velocity time series. The gamma kernel is defined as

$$h(x; a, \nu, \lambda) = a \cdot x^{\nu-1} \exp(-\lambda x) 1_{(0, \infty)}(x), \quad (24)$$

with $a > 0$, $\nu > 0$ and $\lambda > 0$.

The convolution of two gamma kernels, $h(x; a_1, v_1, \lambda_1)$ and $h(x; a_2, v_2, \lambda_2)$, can be expressed as

$$\begin{aligned} g(x; a_1, v_1, \lambda_1, a_2, v_2, \lambda_2) &= h(x; a_1, v_1, \lambda_1) * h(x; a_2, v_2, \lambda_2) \\ &= a_1 a_2 x^{v_1+v_2-1} \mathbb{1}_{\mathbb{R}_+}(x) \int_0^1 e^{-x(\lambda_1 u + \lambda_2(1-u))} u^{v_1-1} (1-u)^{v_2-1} du. \end{aligned} \quad (25)$$

The relevant parameters are $(a, v_1, \lambda_1, v_2, \lambda_2)$ with $a = a_1 a_2$. We say that a function is a 2-gamma kernel if it can be written as the convolution of two gamma kernels.

In the following g will denote a 2-gamma kernel with parameters $(a, v_1, \lambda_1, v_2, \lambda_2)$, $\lambda_1 < \lambda_2$. In [22] it is shown that the sdf $\widehat{r}_v(\omega; g, \sigma, \beta)$ of (20) is then given as

$$\widehat{r}_v(\omega; g, \sigma, \beta) = a^2 (1 + \beta^2 \widehat{r}_{\sigma^2}(\omega)) \left(1 + \left(\frac{2\pi\omega}{\lambda_1} \right)^2 \right)^{-v_1} \left(1 + \left(\frac{2\pi\omega}{\lambda_2} \right)^2 \right)^{-v_2}, \quad (26)$$

where $\widehat{r}_{\sigma^2}$ is the sdf of the process σ^2 . Referring to the parameters of g , we also write $\widehat{r}_v(\omega, a, v_1, \lambda_1, v_2, \lambda_2, \sigma, \beta)$ for $\widehat{r}_v(\omega; g, \sigma, \beta)$.

Ignoring the skewness term in (20), i.e. $\beta = 0$, the sdf $\widehat{r}_v$ of the velocity field v behaves as

$$\widehat{r}_v(\omega; a, v_1, \lambda_1, v_2, \lambda_2, \sigma, 0) \propto \begin{cases} 1 & \omega \ll \lambda_1/2\pi \\ \omega^{-2v_1} & \lambda_1/2\pi \ll \omega \ll \lambda_2/2\pi \\ \omega^{-2(v_1+v_2)} & \omega \gg \lambda_2/2\pi. \end{cases} \quad (27)$$

Thus, for $v_1 = 5/6$, λ_1 denotes the frequency where the inertial range starts and λ_2 denotes the frequency where the inertial range ends. The value $v_1 = 5/6$ reflects Kolmogorov's 5/3rd law [20] and $2(v_1 + v_2)$ gives the slope within the dissipation range (large frequencies). For the general case, i.e. $\beta \neq 0$, the previous interpretation remains essentially true since, for the data set we analysed, $\beta^2 \widehat{r}_{\sigma^2}(\omega) \ll 1$ for $\lambda_1/2\pi \ll \omega$.

From (25) it follows that $v(g, \sigma, \beta)$ is a semimartingale for $v_1 + v_2 > 3/2$. Kolmogorov's 5/3rd law implies that $v_1 \approx 5/6$. Combining this with the estimated value of v_2 (by fitting the empirical sdf at large frequencies) shows that $v_1 + v_2 > 3/2$ for the data set we analysed (see Sect. 4.2). It is for this reason that our focus is on the semimartingale case.

In the semimartingale case, the convolution of gamma kernels does not allow to identify the process σ^2 with the energy dissipation since $g(0+) = 0$. Therefore, we propose for the kernel g a shifted 2-gamma kernel

$$g(x; x_0, a_1, a_2, v_1, v_2, \lambda_1, \lambda_2) = (h(\cdot; a_1, v_1, \lambda_1) * h(\cdot; a_2, v_2, \lambda_2))(x + x_0) \mathbb{1}_{\mathbb{R}_+}(x), \quad (28)$$

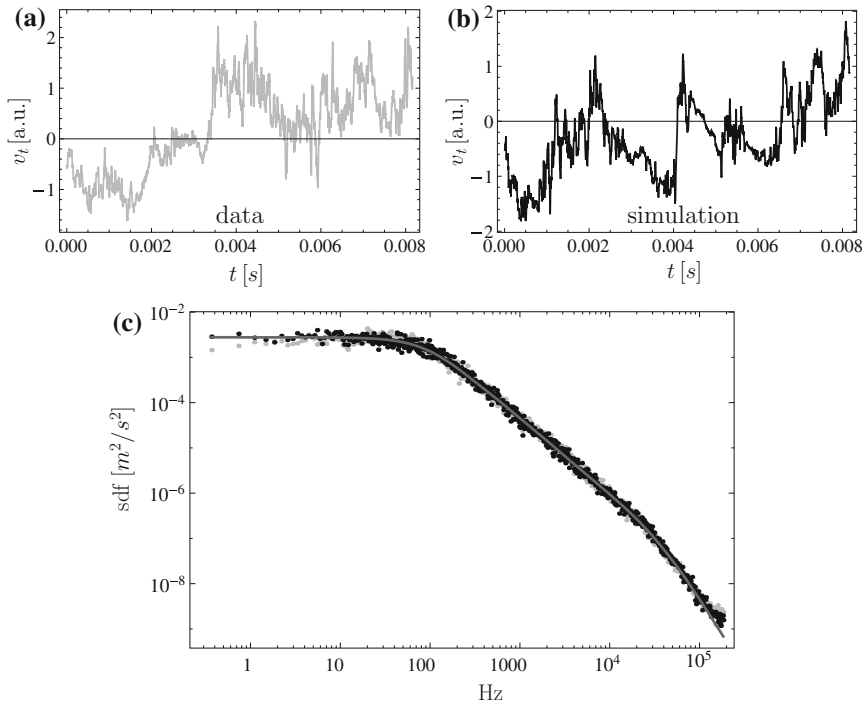


Fig. 1 **a** Excerpt of the empirical time series (in arbitrary units). **b** Excerpt of the simulated time series (in arbitrary units) using the model (20). **c** Comparison of the sdf from the data (*grey dots*) and from the simulation of the model (20) (*black dots*). The solid line corresponds to the sdf obtained from the simulation with $x_0 = 0$ and all other parameters unchanged

where $h(x; a_1, v_1, \lambda_1)$ and $h(x; a_2, v_2, \lambda_2)$ are gamma kernels and x_0 is a positive constant. We say that a function is a $(2, x_0)$ -gamma kernel if it can be expressed as (28).

Figure 1 shows, as an illustrating example, an excerpt of the empirical velocity time series and from a simulation of the model (20) using a $(2, 10^{-7})$ -gamma kernels for g , a cascade model for σ^2 and the estimated β (see Sect. 4.2). The sdf for the data and the sdf from the simulation are compared in Fig. 1c. We can identify three characteristic regimes in the empirical sdf: a flat part at small frequencies, a scaling regime with approximate exponent $-5/3$ and a steeper part at the large frequencies. The central part reflects Kolmogorov's 5/3rd law [20], which is expected to hold in the inertial range. The influence of the shift x_0 on the sdf of v_t can be expected to be negligible for frequencies $\omega \ll x_0^{-1}$. To confirm this conjecture we include in Fig. 1c the sdf of the model with $x_0 = 0$ and all other parameters unchanged. Differences only arise at frequencies around 10^5 . This implies that for not too large frequencies the interpretation of the parameters of the model according to (27) remains valid. The inclusion of the shift x_0 bends the sdf away from the scaling $\omega^{-2(v_1+v_2)}$ at the very large frequencies.

4 Simulation Results

In this Section we compare, in detail, the statistical properties of the model (20) to the stylised features described in Sect. 2. We also briefly mention some aspects of the numerics behind the simulation and discuss how the skewness parameter β and the $(2, x_0)$ -gamma kernel g can be estimated from empirical time series.

The data set we analysed consists of one-point time records of the longitudinal (along the mean flow) velocity component in a gaseous helium jet flow. We refer to [13] for more information about the data set. In Table 1 we list the Taylor Reynolds number R , the Kolmogorov scale η , the integral scale I and the sampling frequency $\hat{f}$.

4.1 Model Performance

The performance of (20) for modelling turbulent time series is illustrated by comparing the marginal distributions of velocity increments, the structure functions and the conditional independence of the Kolmogorov variable. The estimation of the model parameters is based on the analysis of the sdf and the third order structure function and on the marginal distribution and the correlators of the derived energy dissipation.

Figure 1a, b shows examples of time series of the velocity from data and from the simulation, respectively. The similarity between the characteristics of both time series is clearly present. A first quantitative result is given in Fig. 1c displaying the corresponding sdf. The model reproduces the empirical sdf for the whole range of observed frequencies.

The excellent agreement for the sdf translates directly to the corresponding second order structure functions that are shown in Fig. 2a. Note that the sdf (or equivalently the second order structure function) is the basic observable that determines the parameters of the $(2, x_0)$ -gamma kernel g used for the simulation. The excellent agreement for the sdf (and S_2) strongly indicates that the parametric choice of a $(2, x_0)$ -gamma kernel is appropriate.

The estimation of the skewness parameter β is essentially based on fitting the third order structure function S_3 (see Fig. 2b). Taking into account the notorious uncertainty for the estimation of S_3 from turbulent data, the model captures well the details of $S_3(t)$.

Examples of higher order structure functions are shown in Fig. 2c, d. Again the model shows excellent agreement. Only for S_6 some small systematic deviation is observed which is due to an amplification of small errors not visible for S_2 and slightly visible for S_3 . It is important to note that the model is completely specified from the energy dissipation statistics and the structure functions S_2 and S_3 with no adjustable parameter for tuning the behavior of S_4 and S_6 .

Figure 3 shows the densities of velocity increments u_s for various time lags s . The densities evolve from semi-heavy tails at small time scales towards a Gaussian shape at the large time scales. NIG distributions fit these densities very well for all

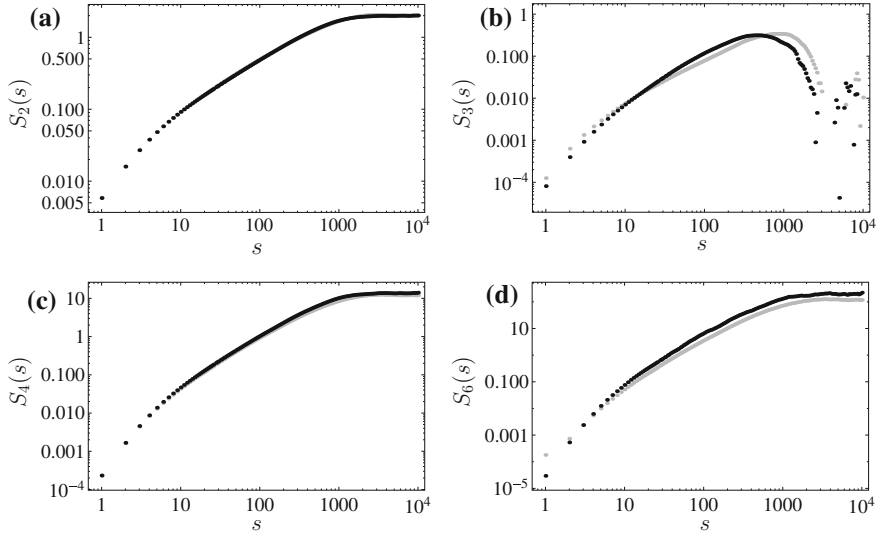


Fig. 2 Comparison of the structure functions S_n (10), $n = 2, 3, 4, 6$ from the simulation of (20) (black) and the structure functions estimated from the data (grey). The time lags s are measured in units of the finest resolution of the empirical data

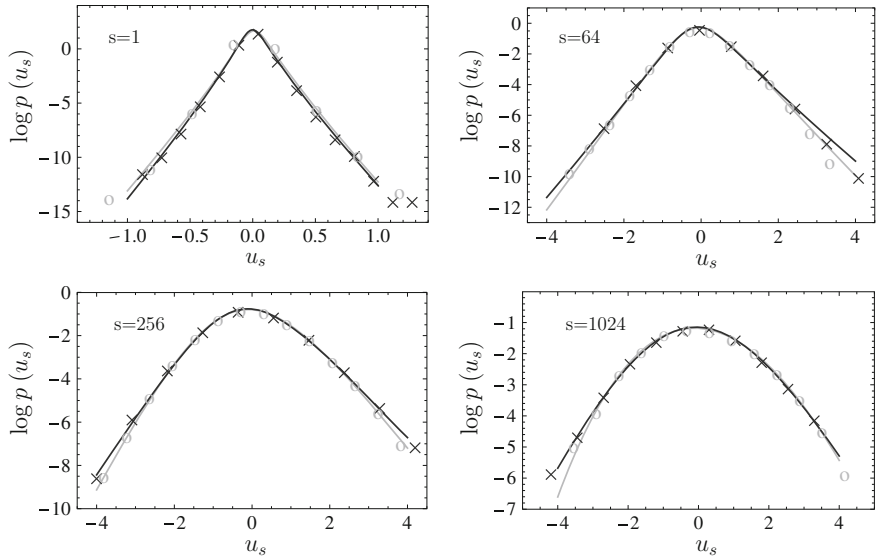


Fig. 3 Comparison of the densities of velocity increments $p(u_s)$, $s = 1, 64, 256, 1024$ from data (grey circles) and from the simulation of (20) (black crosses). The solid lines correspond to fitted NIG-distributions based on maximum likelihood estimation

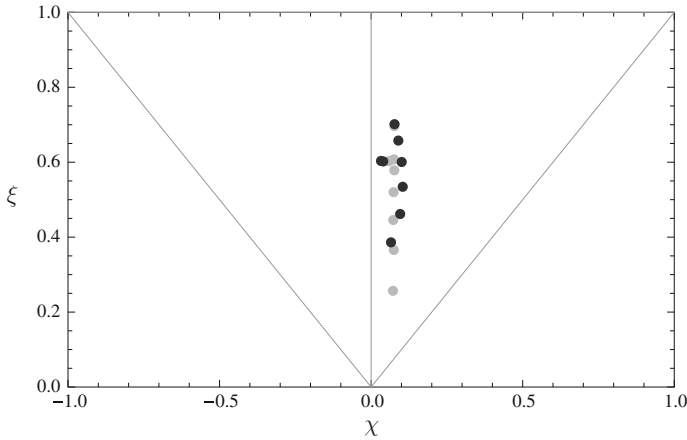


Fig. 4 NIG-shape triangle for the evolution of the pdf of velocity increments across lags for the data (grey) and for the simulation of (20) (black). Each point corresponds to a different time lag $s = 1, 4, 16, 32, 64, 128, 256, 512, 1024$, increasing from top to bottom

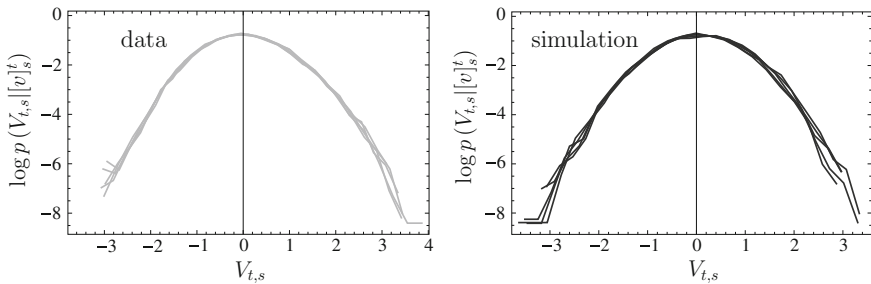


Fig. 5 Comparison of the conditional densities $p(V_{t,s} | [v]_s^t)$ of the Kolmogorov variable (23) from the data and from the simulation of (20) for $t = 128$ (in units of the finest resolution) and values $[v]_s^t = 0.8, 0.9, 1, 1.1, 1.2$

scales and all amplitudes in full agreement with the results reported in [6, 9]. The corresponding steepness and asymmetry parameters are shown in the NIG shape triangle in Fig. 4. Again, simulation and data show a good agreement.

Figure 5 illustrates the performance of the model concerning the conditional independence of the densities of the Kolmogorov variable. For t within the inertial range, $p(V_{t,s} | [v]_s^t)$ is independent of $[v]_s^t$. The values of $[v]_s^t$ cover the core of the distribution of $[v]_s^t$ for which a sufficient sample size is ensured. Figure 6 shows a direct comparison of $p(V_{t,s} | [v]_s^t = 1)$ for the data and the simulation, showing the strong similarity of the distributions.

Finally, Fig. 7 shows the correlators of order (1, 1) and (1, 2) of the energy dissipation estimated from the empirical velocity time series and from the simulation of (20). Besides small scale scatter, data and simulations show (nearly) perfect agreement.

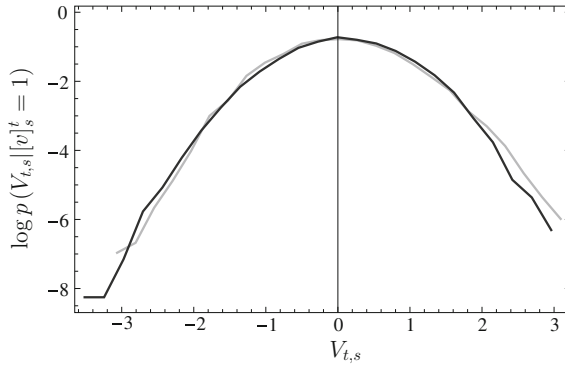


Fig. 6 Comparison of the conditional density $p(V_{t,s} | [v]_s^t = 1)$ from the data (grey) and from the simulation of (20) (black)

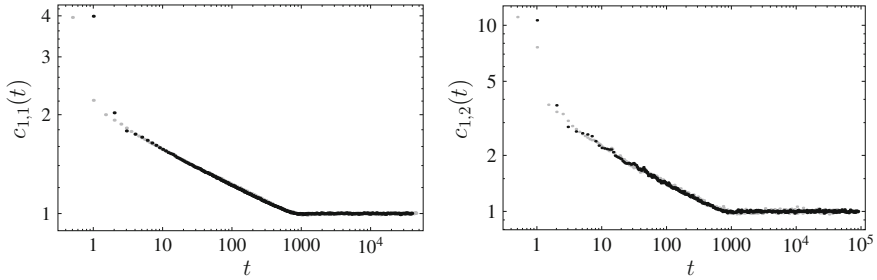


Fig. 7 Comparison of the correlators $c_{p,q}(t)$, $(p, q) = (1, 1), (1, 2)$, of the energy dissipation estimated from the data (grey) and from the simulation of (20) (black). The time lags t are measured in units of the finest resolution of the data

4.2 Simulation of BSS Processes

In this section we briefly describe how we simulate the BSS-process (20) using time series of σ^2 . The algorithm is based on

$$v_t(g, \sigma, \beta) | \sigma \sim N \left(\beta \int_0^\infty g(s) \sigma_{t-s}^2 ds, \int_0^\infty g^2(s) \sigma_{t-s}^2 ds \right) \quad (29)$$

which provides the conditional distribution for $v_t(g, \sigma, \beta) | \sigma^2$. In principle, to reproduce $v_t(g, \sigma, \beta) | \sigma^2$ a complete path for σ^2 is required. However, using a sufficiently small mesh, a linear interpolation on σ^2 gives the desired accuracy.

Let $\delta > 0$ and $N \in \mathbb{N}$. Assume that we know the values $\sigma_{i\delta}^2$ for $i = 0, \dots, N$. The linear interpolation for σ^2 is given by the formula

$$\tilde{\sigma}_s^2 = \sum_{i=1}^{N-1} \left\{ \left(\frac{\sigma_{i+1}^2 - \sigma_i^2}{\delta} \right) (s - i \cdot \delta) + \sigma_i^2 \right\} 1_{[i \cdot \delta, (i+1) \cdot \delta]}(s).$$

Thus,

$$v_t(g, \tilde{\sigma}, \beta) \sim N \left(\beta \int_0^\infty g(s) \tilde{\sigma}_{t-s}^2 ds, \int_0^\infty g^2(s) \tilde{\sigma}_{t-s}^2 ds \right).$$

Assuming that N is large enough, and since $g \in L_1(\mathbb{R}_+) \cap L_2(\mathbb{R}_+)$, we can, for $\varepsilon > 0$, choose $q \gg 0$ such that

$$\left| \int_0^\infty g(s) \tilde{\sigma}_{t-s}^2 ds - \int_0^q g(s) \tilde{\sigma}_{t-s}^2 ds \right| < \varepsilon.$$

Thus we replace the kernel g by the approximation $\tilde{g} = g \cdot 1_{[0, q]}$. Let $q = n \cdot \delta$, for $n \in \mathbb{N}$. Expanding and applying a change of variable, we have that

$$\int_0^\infty \tilde{g}(s) \tilde{\sigma}_{t-s}^2 ds = \delta \sum_{k=0}^n \int_0^1 \tilde{g}(\delta\{y+k\} + t) (\sigma_{j-k+1}^2 (1-y) + \sigma_{j-k}^2 y) dy, \quad (30)$$

$$\int_0^\infty \tilde{g}^2(s) \tilde{\sigma}_{t-s}^2 ds = \delta \sum_{k=0}^n \int_0^1 \tilde{g}^2(\delta\{y+k\} + t) (\sigma_{j-k+1}^2 (1-y) + \sigma_{j-k}^2 y) dy, \quad (31)$$

for $t \geq \delta \cdot n$. Therefore, $v_{j \cdot \delta}(\tilde{g}, \tilde{\sigma}, \beta) | \tilde{\sigma}, j = n, \dots, N-1$, can be obtained by simulating a normal random variable with mean (30) and variance (31). Thus, by interpolation, we obtain a path for $v_t(g, \sigma, \beta) | \sigma$ on the interval $[\delta \cdot n, N \cdot \delta]$ through the approximation $v_t(\tilde{g}, \tilde{\sigma}, \beta) | \tilde{\sigma}$.

4.3 Estimation Procedure

Our modelling framework (20) has three degrees of freedom, the energy dissipation σ^2 , the kernel g and the skewness parameter β . The energy dissipation can be estimated from velocity increments at the smallest time scale. This has been done in [23] for the data set in the present study. Following [23], we use an NIG Lévy basis in (16) and the ambit set (17). The relevant parameters are listed in Table 1.

It remains to estimate β and the kernel g within the class of $(2, x_0)$ -gamma kernels. Given a value of β , our estimators for the parameters of g are those that minimise the distance, in the sense of least squares, between the empirical sdf and the sdf of (20) using $x_0 = 0$. As a consequence of the physical interpretation derived from (27), the minimisation is performed restricting λ_1 to values around the initial frequencies of the inertial range, and restricting λ_2 to values near the end of the inertial range.

These constraints have proven to produce good approximations to the empirical sdf. The shift x_0 is obtained by fitting the very large frequency behaviour of the sdf. It is important to note that the scatter of the data at large frequencies does not allow to estimate a precise value of x_0 . We choose $x_0 = 10^{-7}$ by visual inspection.

For the estimation of the skewness parameter β , we consider the third order structure function S_3 of (20) rewritten as

$$S_3(l) = 3\beta E\{(\Delta_l R)^2 (\Delta_l S)\} + \beta^3 E\{(\Delta_l S)^3\}, \quad (32)$$

where $\Delta_l R = R_l - R_0$ and $\Delta_l S = S_l - S_0$, for $l > 0$. Given paths of R and S , our estimator for β is the value that minimises the distance, in the sense of least squares, between the empirical third order structure function and (32) for a suitable range of scales l (between the smallest scale and the location of the peak, see Fig. 2b).

The complete estimation procedure can be described as follows. We first neglect the skewness parameter β and we estimate the parameters of g under this restriction from the sdf. Then, having a simulation of the process σ , we perform a simulation of (20). Using this simulation, we estimate β as described above. Next, we re-estimate the kernel g using the empirical sdf and the current value of β . We perform this algorithm until we observe stabilisation of β . This algorithm has proven to stabilise after 7 iterations. Figure 8 shows the parameters of g and β obtained after each iteration. The resulting function g is depicted in Fig. 9.

The algorithm described above produces similar kernels in each iteration. Therefore it is reasonable to assume that the L^2 -distance between these kernels is small.

The following Lemma provides some bound for the convergence of (20).

Lemma 1 *Consider the model (20). Let $\beta, \beta_1 > 0$ and $g, g_1 \in L_1(\mathbb{R}) \cap L_2(\mathbb{R})$. Assume that:*

1. $c \equiv E[\sigma^4] < \infty$.
2. $\max\{\|g_1 - g\|_{L_2(\mathbb{R})}, \|g_1 - g\|_{L_1(\mathbb{R})}\} < \frac{\varepsilon}{(1+c^{1/4}\beta_1)c^{1/4}}$, for $\varepsilon > 0$.

Then, we have that

$$\|v_t(g_1, \sigma, \beta_1) - v_t(g, \sigma, \beta)\|_{L_2(\Omega)} \leq c^{1/2}|\beta_1 - \beta| \|g\|_{L_1(\mathbb{R})} + \varepsilon.$$

Proof Let

$$\begin{aligned} P_t &= \int_{-\infty}^t \{g_1(t-s) - g(t-s)\} \sigma_s dB_s, \\ Q_t &= \int_{-\infty}^t \{\beta_1 g_1(t-s) - \beta g(t-s)\} \sigma_s^2 ds \end{aligned}$$

for $t > 0$. Observe that

$$Q_t = \beta_1 \int_{-\infty}^t \{g_1(t-s) - g(t-s)\} \sigma_s^2 ds + (\beta_1 - \beta) \int_{-\infty}^t g(t-s) \sigma_s^2 ds.$$

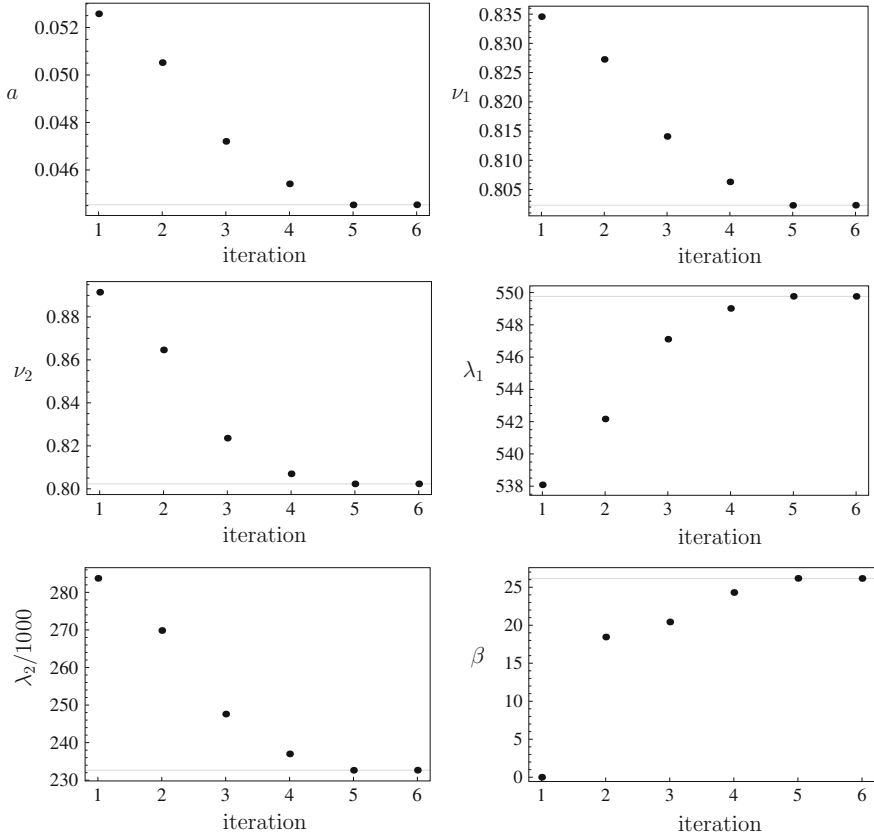
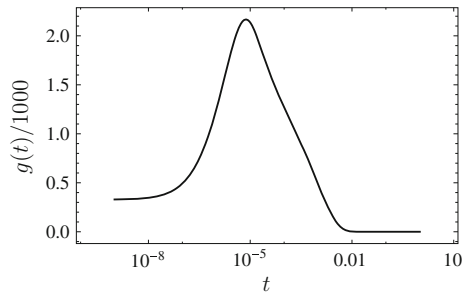


Fig. 8 Estimated values of the parameters of the $(2, x_0)$ -gamma kernel g with $x_0 = 10^{-7}$ and estimated value of the skewness parameters β for each of the iterations performed

Fig. 9 The $(2, 10^{-7})$ -gamma kernel g corresponding to the estimated parameters in Fig. 8 (iteration 7). Time t is measured in units of the finest resolution of the data



Then

$$\begin{aligned} \|v_t(g_1, \sigma, \beta_1) - v_t(g, \sigma, \beta)\|_{L_2(\Omega)} &\leq \|P_t\|_{L_2(\Omega)} + |\beta_1 - \beta| \left\| \int_{-\infty}^t g(t-s) \sigma_s^2 ds \right\|_{L_2(\Omega)} \\ &\quad + \beta_1 \left\| \int_{-\infty}^t \{g_1(t-s) - g(t-s)\} \sigma_s^2 ds \right\|_{L_2(\Omega)}. \end{aligned}$$

The Cauchy-Schwarz inequality implies that

$$E[\sigma_s^2 \sigma_r^2] \leq E[\sigma_s^4]^{1/2} E[\sigma_r^4]^{1/2} = E[\sigma^4] = c, \quad (33)$$

for any $(s, r) \in \mathbb{R}^2$. Thus,

$$\left\| \int_{-\infty}^t g(t-s) \sigma_s^2 ds \right\|_{L_2(\Omega)}^2 = E \left[\left(\int_{-\infty}^t g(t-s) \sigma_s^2 ds \right)^2 \right] \leq c \|g\|_{L_1(\mathbb{R})}^2$$

and

$$\left\| \int_{-\infty}^t \{g_1(t-s) - g(t-s)\} \sigma_s^2 ds \right\|_{L_2(\Omega)}^2 \leq c \|g_1 - g\|_{L_1(\mathbb{R})}^2.$$

The Itô isometry and (33) imply that

$$\|P_t\|_{L_2(\Omega)}^2 \leq c^{1/2} \|g_1 - g\|_{L_2(\mathbb{R})}^2.$$

Thus,

$$\begin{aligned} \|v_t(g_1, \sigma, \beta_1) - v_t(g, \sigma, \beta)\|_{L_2(\Omega)} &\leq c^{1/4} \|g_1 - g\|_{L_2(\mathbb{R})} + c^{1/2} |\beta_1 - \beta| \|g\|_{L_1(\mathbb{R})} \\ &\quad + c^{1/2} \beta_1 \|g_1 - g\|_{L_1(\mathbb{R})} \\ &\leq \varepsilon + c^{1/2} |\beta_1 - \beta| \|g\|_{L_1(\mathbb{R})}, \end{aligned}$$

which concludes the proof.

Under certain assumptions, we can ensure that the estimation procedure gives reasonable estimators for the skewness parameter β . Assume that $v_t(g, \sigma, \beta)$ is the ideal model with g in the class of $(2, x_0)$ -gamma kernels. Let $\{g_n\}_{n \in \mathbb{N}}$ be a sequence of $(2, x_0)$ -gamma kernels and $\{\beta_n\}_{n \in \mathbb{N}}$ a sequence of positive numbers, and assume that $E[\sigma^6] < \infty$, $g_n \rightarrow g$ pointwise and g_n bounded by an integrable and square-integrable function.

Following the notation in (20), for each $n \in \mathbb{N}$, we decompose $v_t(g_n, \sigma, \beta_n) = R_t(g_n, \sigma) + \beta_n S_t(g_n, \sigma)$. For $l > 0$, let

$$\begin{aligned}
a_l^n &= E \left[(\Delta_l R(g_n, \sigma))^2 (\Delta_l S(g_n, \sigma)) \right], \\
a_l &= E \left[(\Delta_l R(g, \sigma))^2 (\Delta_l S(g, \sigma)) \right], \\
b_l^n &= E \left[(\Delta_l S(g_n, \sigma))^3 \right], \\
b_l &= E \left[(\Delta_l S(g, \sigma))^3 \right], \\
f_n(l; \beta) &= 3(\beta a_l - \beta_n a_l^n) + (\beta^3 b_l - \beta_n^3 b_l^n),
\end{aligned}$$

where $\Delta_l R(\cdot, \cdot) = R_l(\cdot, \cdot) - R_0(\cdot, \cdot)$ and $\Delta_l S(\cdot, \cdot) = S_l(\cdot, \cdot) - S_0(\cdot, \cdot)$. Observe that f_n is the difference of the third order structure functions of $v(g, \sigma, \beta_n)$ and $v(g_n, \sigma, \beta)$.

Lemma 2 $a_l^n \rightarrow a_l$ and $b_l^n \rightarrow b_l$ for any $l > 0$.

Proof Define

$$\begin{aligned}
\phi_l^n(s) &:= g_n(l-s) - g_n(-s) \mathbf{1}_{(-\infty, 0]}(s) \\
\phi_l(s) &:= g(l-s) - g(-s) \mathbf{1}_{(-\infty, 0]}(s).
\end{aligned}$$

$E[\sigma^6] < \infty$ implies that $(s_1, s_2, s_3) \mapsto E[\sigma_{s_1}^2 \sigma_{s_2}^2 \sigma_{s_3}^2]$ is a bounded function in $\mathbb{R}^3$. Thus, from the Dominated Convergence Theorem,

$$\begin{aligned}
b_l^n &= E[(\Delta_l S(g_n, \sigma))^3] \\
&= \int_{-\infty}^l \int_{-\infty}^l \int_{-\infty}^l \phi_l^n(s_1) \phi_l^n(s_2) \phi_l^n(s_3) E[\sigma_{s_1}^2 \sigma_{s_2}^2 \sigma_{s_3}^2] ds_1 ds_2 ds_3 \\
&\rightarrow \int_{-\infty}^l \int_{-\infty}^l \int_{-\infty}^l \phi_l(s_1) \phi_l(s_2) \phi_l(s_3) E[\sigma_{s_1}^2 \sigma_{s_2}^2 \sigma_{s_3}^2] ds_1 ds_2 ds_3 = b_l.
\end{aligned}$$

On the other hand, from the Itô isometry and the Dominated Convergence Theorem,

$$\begin{aligned}
a_l^n &= E[(\Delta_l R(g_n, \sigma))^2 (\Delta_l S(g_n, \sigma))] \\
&= E[\Delta_l S(g_n, \sigma) E[(\Delta_l R(g_n, \sigma))^2 | \sigma]] \\
&= E\left[\int_{-\infty}^l \phi_l^n(s) \sigma_s^2 ds \cdot E\left[\left(\int_{-\infty}^l \phi_l^n(r) \sigma_r dB_r\right)^2 \middle| \sigma\right]\right] \\
&= \int_{-\infty}^l \int_{-\infty}^l \phi_l^n(s) (\phi_l^n(r))^2 E[\sigma_s^2 \sigma_r^2] ds dr \\
&\rightarrow \int_{-\infty}^l \int_{-\infty}^l \phi_l(s) (\phi_l(r))^2 E[\sigma_s^2 \sigma_r^2] ds dr = a_l.
\end{aligned}$$

This finishes the proof.

The next Proposition ensures that a converging sequence β_n converges to the right value.

Proposition 1 Assume there is some $l_0 > 0$ such that $\lim_{n \rightarrow \infty} f_n(l_0; \beta) = 0$ and that β_n converges. Then, $\beta_n \rightarrow \beta$.

Proof Assume that $l_0 > 0$ satisfies $\lim_{n \rightarrow \infty} f_n(l_0; \beta) = 0$. Then,

$$\lim_{n \rightarrow \infty} \{3(\beta a_{l_0} - \beta_n a_{l_0}^n) + (\beta^3 b_{l_0} - \beta_n^3 b_{l_0}^n)\} = 0.$$

This implies

$$3(\beta - \beta_*) a_{l_0} + (\beta^3 - \beta_*^3) b_{l_0} = 0,$$

where $\beta_* = \lim \beta_n$. Therefore,

$$(\beta - \beta_*) (3a_{l_0} + (\beta^2 + \beta\beta_* + \beta_*^2) b_{l_0}) = 0,$$

Thus, necessarily, $\beta_* = \beta$.

5 Conclusion

The analysis performed in this paper clearly demonstrates that Brownian semi-stationary processes are well adapted to reproduce key characteristics of turbulent time series. The parameters of the model are solely estimated from the marginal distribution and the correlator $c_{1,1}$ of the energy dissipation [23] and from second and third order structure functions of velocity increments. This has been done under the specific model specification (20) with a $(2, x_0)$ -gamma kernel g . The use of a $(2, x_0)$ -gamma kernel is motivated by its ability to reproduce the empirical sdf. The fact that, starting from second order and third order structure function, higher order structure functions, the evolution of the densities of velocity increments across scales and the essential statistics of the Kolmogorov variable are also reproduced clearly indicates the appropriateness of the semi-parametric model (20).

In [12, 19] a similar approach for modelling turbulent velocity time series is suggested. They propose a causal continuous-time moving average of the form

$$Y_t = \int_{-\infty}^t g(t-s) dL_s \quad (34)$$

where L is a Lévy process with zero mean and finite second moment. A non-parametric estimation of the kernel g from second order statistics of turbulent data shows the same qualitative behaviour as the kernel estimated in the current study. The extraction of the driving noise L from velocity time series is addressed in [19] showing that the autocorrelation of the energy dissipation resulting from the model agrees well with empirical findings. The performance of (34) for turbulence modelling beyond second order statistics is, however, not discussed. It would be interesting to

compare the two approaches in more detail, including more of the stylised features discussed in the present paper.

The data set analysed here has a relatively high Reynolds number, with a visible inertial range. It is important to investigate how the model performs for lower Reynolds numbers, where inertial range scaling is not observed. Concerning the model for the energy dissipation, this has been done in [23] where it was shown that continuous cascades are equally suitable for a wide range of Reynolds numbers. For the velocity field itself this is work in progress.

Acknowledgments The authors are much indebted to B. Chabaud for allowing the use of the data set. The first author is grateful to the Mexican National Council of Science and Technology CONACYT for its support through the Ph.D. grant 217040. Finally, the authors are very thankful to Emil Hedeang for the many stimulating discussions and suggestions in the preparation of this work. The authors are also grateful to the referee of this paper for helpful comments.

Appendix

The normal inverse Gaussian (NIG) distribution is a four-parameter family of continuous probability distributions whose probability density function is given by

$$f_{\text{NIG}}(x; \alpha, \beta, \mu, \delta) = \frac{\alpha e^{\delta\gamma}}{\pi} e^{\beta(x-\mu)} \frac{K_1\left(\delta\alpha q\left(\frac{x-\mu}{\delta}\right)\right)}{q\left(\frac{x-\mu}{\delta}\right)}, \quad x \in \mathbb{R}, \quad (35)$$

where $\gamma = \alpha^2 - \beta^2$, $q(x) = \sqrt{1+x^2}$ and K_1 denotes the modified Bessel function of the second kind with index 1. The domain of variation of the parameters is given by $\mu \in \mathbb{R}$, $\delta \in \mathbb{R}_+$, and $0 \leq |\beta| < \alpha$. The parameters α and β are shape parameters, μ determines the location, and δ determines the scale. The distribution is denoted by $\text{NIG}(\alpha, \beta, \mu, \delta)$.

The normal inverse Gaussian distribution is a subclass of the generalised hyperbolic distribution. These distributions were introduced by Barndorff-Nielsen [5] to describe the law of the logarithm of the size of sand particles.

The cumulant function $K(z; \alpha, \beta, \mu, \delta) = \log E[\exp\{zX\}]$ of a random variable X with distribution $\text{NIG}(\alpha, \beta, \mu, \delta)$ is given by

$$K(z; \alpha, \beta, \mu, \delta) = z\mu + \delta \left(\gamma - \sqrt{\alpha^2 - (\beta + z)^2} \right). \quad (36)$$

It follows immediately from this that the normal inverse Gaussian distribution is infinitely divisible. Namely, if $X_i \sim \text{NIG}(\alpha, \beta, \mu_i, \delta_i)$, $i = 1, 2$, are independent random variables, then we have $X_1 + X_2 \sim \text{NIG}(\alpha, \beta, \mu_1 + \mu_2, \delta_1 + \delta_2)$.

It is often of interest to consider alternative parametrisations of the normal inverse Gaussian laws. In particular, letting $\bar{\alpha} = \delta\alpha$ and $\bar{\beta} = \delta\beta$, we have that $\bar{\alpha}$ and $\bar{\beta}$ are invariant under location-scale changes.

Sometimes it is useful to represent NIG distributions in the so-called shape triangle. Consider the alternative asymmetry and steepness parameters χ and ξ defined by

$$\xi = (1 + \bar{\gamma})^{-1/2}, \quad \chi = \rho\xi,$$

where $\rho = \beta/\alpha$ and $\bar{\gamma} = \delta\gamma = \delta\sqrt{\alpha^2 - \beta^2}$. These parameters are invariant under location-scale changes. Their range defines the NIG shape triangle

$$\{(\chi, \xi) : 0 < \xi < 1, -\xi < \chi < \xi\}.$$

When $\chi = 0$ the NIG distribution is symmetric. Values $\chi > 0$ indicate a positively skewed distribution and $\chi < 0$ a negatively skewed law. The steepness parameter ξ measures the heaviness of the tails of the NIG distribution. The limiting case $\xi = 0$ corresponds to a normal distribution.

References

1. Anselmet, F., Gagne, Y., Hopfinger, E.J., Antonia, R.A.: Higher-order velocity structure functions in turbulent shear flows. *J. Fluid Mech.* **140**, 63–89 (1984)
2. Antonia, R.A., Satyaprakash, B.R., Hussain, A.K.M.F.: Statistics of fine-scale velocity in turbulent plane and circular jets. *J. Fluid Mech.* **119**, 55–89 (1982)
3. Arneodo, A., Baudet, C., Belin, F., Benzi, R., Castaing, B., Chavarria, R., Ciliberto, S., Camussi, R., Chillà, F., Dubrulle, B., Gagne, Y., Hébral, B., Herweijer, J., Marchand, M., Maurer, J., Muzy, J.F., Naert, A., Noullez, A., Peinke, J., Roux, F., Tabeling, P., van de Water, W., Willaime, H.: Structure functions in turbulence, in various flow configurations, at Reynolds numbers between 30 and 5000, using extended self-similarity. *Europhys. Lett.* **34**, 411–416 (1996)
4. Arneodo, A., Bacry, E., Muzy, J.F.: Random cascades on wavelet dyadic trees. *J. Math. Phys.* **39**, 4142–4164 (1998)
5. Barndorff-Nielsen, O.E.: Exponentially decreasing distributions for the logarithm of particle size. *Proc. R. Soc. Lond. A* **353**, 401409 (1977)
6. Barndorff-Nielsen, O.E., Schmiegel, J.: Time change and universality in turbulence. Technical report, 2005–15, Thiele Centre, University of Aarhus, Denmark (2006)
7. Barndorff-Nielsen, O.E., Schmiegel, J.: A stochastic differential equation framework for the timewise dynamics of turbulent velocities. *Theory Probab. Appl.* **52**, 372–388 (2008)
8. Barndorff-Nielsen, O.E., Schmiegel, J.: Brownian semistationary processes and volatility/intermittency. In: Albrecher, H., Runggaldier, W., Schachermayer, W. (eds.) *Advanced Financial Modelling*, pp. 1–25. Walter de Gruyter, Berlin (2009)
9. Barndorff-Nielsen, O.E., Blæsild, P., Schmiegel, J.: A parsimonious and universal description of turbulent velocity increments. *Eur. Phys. J. B* **41**, 345–363 (2004)
10. Benzi, R., Biferale, L., Crisanti, A., Paladin, G., Vergassola, M., Vulpiani, A.: A random process for the construction of multifractal fields. *Physica D* **65**, 352–358 (1993)
11. Biferale, L., Boffetta, G., Celani, A., Crisanti, A., Vulpiani, A.: Mimicking a turbulent signal: sequential multifractal processes. *Phys. Rev. E* **57**, 6261–6264 (1998)
12. Brockwell, P.J., Ferrazzano, V., Klüppelberg, C.: High-frequency sampling and kernel estimation for continuous-time moving average processes. *J. Time Ser. Anal.* **34**, 385–404 (2013)
13. Chanal, O., Chebaud, B., Castaing, B., Hébral, B.: Intermittency in a turbulent low temperature gaseous helium jet. *Eur. Phys. J. B* **17**, 309–317 (2000)
14. Cleve, J., Greiner, M., Pearson, B.R., Sreenivasan, K.R.: Intermittency exponent of the turbulent energy cascade. *Phys. Rev. E* **69**, 066316 (2004)

15. Cleve, J., Dziekan, T., Schmiegél, J., Barndorff-Nielsen, O.E., Pearson, B.R., Sreenivasan, K.R., Greiner, M.: Finite-size scaling of two-point statistics and the turbulent energy cascade generators. *Phys. Rev. E* **71**, 026309 (2005)
16. Cleve, J., Schmiegél, J., Greiner, M.: To be and not to be: scale correlations in random multifractal processes. Technical report, 2006–7, Thiele Centre, University of Aarhus, Denmark (2006)
17. Corcuera, J.M., Hedevang, E., Pakkanen, M.S., Podolskij, M.: Asymptotic theory for brownian semi-stationary processes with application to turbulence. *Stoch. Proc. Appl.* **123**, 2552–2574 (2013)
18. Elliot, F., Majda, A., Hornthrop, D., McLaughlin, R.: Hierarchical Monte Carlo methods for fractal random fields. *J. Stat. Phys.* **81**, 717–736 (1995)
19. Ferrazzano, V., Klüppelberg, C.: Turbulence modelling by time-series methods. *ArXiv e-prints* **1205**, 6614 (2012)
20. Frisch, U.: *Turbulence: The Legacy of A.N. Kolmogorov*. Cambridge University Press, Cambridge (1995)
21. Greiner, M., Cleve, J., Schmiegél, J., Sreenivasan, K.R.: Data-driven stochastic processes in fully developed turbulence. In: Waymire, E.C., Duan, J. (eds.) *IMA Volume 140: Probability and Partial Differential Equations in Modern Applied Mathematics*, pp. 137–150. Springer, New York (2004)
22. Hedevang, E.: *Stochastic modelling of turbulence: with applications to wind energy*. PhD Thesis, Aarhus University (2012)
23. Hedevang, E., Schmiegél, J.: A causal continuous-time stochastic model for the turbulent energy cascade in a helium jet flow. *J. Turbul.* **14**, 1–26 (2013)
24. Hosokawa, I., Van Atta, C.W., Thoroddsen, S.T.: Experimental study of the kolmogorov refined similarity variable. *Fluid Dyn. Res.* **13**, 329–333 (1994)
25. Juneja, A., Lathrop, D., Sreenivasan, K.R., Stolovitzky, G.: Synthetic turbulence. *Phys. Rev. E* **49**, 5179–5194 (1994)
26. Kolmogorov, A.N.: A refinement of previous hypotheses concerning the local structure of turbulence in a viscous incompressible fluid at high reynolds number. *J. Fluid Mech.* **13**, 82–85 (1962)
27. Maurer, J., Tabeling, P., Zocchi, G.: Statistics of turbulence between two counter-rotating disks in low temperature helium gas. *Europhys. Lett.* **26**, 31–36 (1994)
28. Schmiegél, J.: Self-scaling of turbulent energy dissipation correlators. *Phys. Lett. A* **337**, 342–353 (2005)
29. Schmiegél, J., Cleve, J., Eggers, H.C., Pearson, B.R., Greiner, M.: Stochastic energy-cascade model for $(1 + 1)$ -dimensional fully developed turbulence. *Phys. Lett. A* **320**, 247–253 (2004)
30. Schmiegél, J., Barndorff-Nielsen, O.E., Eggers, H.C.: A class of spatio-temporal and causal stochastic processes, with application to multiscaling and multifractality. *S. Afr. J. Sci.* **101**, 513–519 (2005)
31. Sreenivasan, K.R., Kailasnath, P.: An update on the intermittency exponent in turbulence. *Phys. Fluids A* **5**, 512–514 (1992)
32. Sreenivasan, K.R., Antonia, R.A.: The phenomenology of small-scale turbulence. *Ann. Rev. Fluid Mech.* **29**, 435–472 (1997)
33. Stolovitzky, G., Kailasnath, P., Sreenivasan, K.R.: Kolmogorov’s refined similarity hypothesis. *Phys. Rev. Lett.* **69**, 1178–1181 (1992)
34. Stolovitzky, G., Sreenivasan, K.R.: Kolmogorov’s refined similarity hypotheses for turbulence and general stochastic processes. *Rev. Mod. Phys.* **66**, 229–239 (1994)
35. Taylor, G.I.: The spectrum of turbulence. *Proc. R. Soc. Lond. A* **164**, 476–490 (1938)
36. Vicssek, T., Barabási, A.L.: Multi-affine model for the velocity distribution in fully turbulent flows. *J. Phys. A* **24**, L845 (1991)
37. Zhu, Y., Antonia, R.A., Hosokawa, I.: Refined similarity hypotheses for turbulent velocity and temperature fields. *Phys. Fluids* **7**, 1637–1648 (1995)

Associated Natural Exponential Families and Elliptic Functions

G rard Letac

To Ole Barndorff-Nielsen for his 80th birthday.

Abstract This paper studies the variance functions of the natural exponential families (NEF) on the real line of the form $(Am^4 + Bm^2 + C)^{1/2}$ where m denoting the mean. Surprisingly enough, most of them are discrete families concentrated on $\lambda\mathbb{Z}$ for some constant λ and the Laplace transform of their elements are expressed by elliptic functions. The concept of association of two NEF is an auxiliary tool for their study: two families F and G are associated if they are generated by symmetric probabilities and if the analytic continuations of their variance functions satisfy $V_F(m) = V_G(m\sqrt{-1})$. We give some properties of the association before its application to these elliptic NEF. The paper is completed by the study of NEF with variance functions $m(Cm^4 + Bm^2 + A)^{1/2}$. They are easier to study and they are concentrated on $a\mathbb{N}$.

Keywords Variance functions · Exponential dispersion models · Function $\wp$ of weierstrass

1 Foreword

Ole and I met for the first time in the Summer School of Saint Flour in 1986. Having been converted to statistics by V. Seshadri 2 years before, I had learnt about exponential families through Ole's book (1978) and I had fought with cuts and steepness. Marianne Mora had just completed her thesis in Toulouse and was one of the

G. Letac (✉)

Equipe de Statistique et Probabilit s, Universit  de Toulouse,

118 Route de Narbonne, 31062 Toulouse, France

e-mail: gerard.letac@math.univ-toulouse.fr; gerard.letac@alsatis.net

  Springer International Publishing Switzerland 2016

M. Podolskij et al. (eds.), *The Fascination of Probability,*

Statistics and their Applications, DOI 10.1007/978-3-319-25826-3_4

Saint Flour participants: she was the first from Toulouse to make the pilgrimage to Aarhus the year after, followed by many others researchers from Toulouse: Evelynne Bernadac, Célestin Kokonendji, Dhafer Malouche, Muriel Casalis, Abdelhamid Hassairi, Angelo Koudou and myself. Over the years all of us were in contact with the everflowing ideas of Ole. During these Aarhus days (and Ole's visits to Toulouse) we gained a better understanding of the Lévy processes, of generalized inverse Gaussian distributions and their matrix versions, of differential geometry applied to statistics. Among all the topics which have interested Ole, the choice today is the one for which he may be the least enthusiastic (see the discussion of Letac [1]), namely the classification of exponential families through their variance functions: Ole thought correctly that although the results were satisfactory for the mind, one could not see much real practical applications: in the other hand Mendeleviev is universally admired for its prophetic views of chemical elements which had not been yet discovered. Descriptions of natural exponential families with more and more sophisticated variance functions V have been done: when V is a second degree polynomial in the mean [2], a power [3, 4], a third degree polynomial [5], the Babel class $P + Q\sqrt{R}$ where polynomials P, Q, R have degrees not bigger than 2, 1, 2 [6, 7]. This is for univariate NEF: even more important works have been done for multivariate NEF, but the present paper will confine to one dimensional distributions only.

Having forgotten variance functions during the last 20 years and having turned to random matrices and Bayesian theory, our interest for the topic has been rejuvenated by the paper by Bar-Lev and Van de Duyn Schouten [8]. The authors consider exponential dispersion models G such that one of the transformations

$$T(P)(dx) = \frac{xP(dx)}{\int_{\mathbb{R}} xP(dx)}, \quad T^2(P)(dx) = \frac{x^2P(dx)}{\int_{\mathbb{R}} x^2P(dx)}.$$

maps G into itself or one of its translates. For $T(P)$ they obtain exactly the exponential dispersion models concentrated on the positive line with quadratic variance functions: gamma, Poisson, negative binomial and binomial and no others. For $T^2(P)$ they obviously obtain the previous ones, but they observe that new variance functions appear, in particular $(m^4 + c^2)^{1/2}$, without being able to decide whether these natural exponential families (NEF) exist after all (it should be mentioned here that their formula (11) is not correct and this fact greatly invalidates their paper). As a result, our initial motivation was to address this particular question: is $(m^4 + c^2)^{1/2}$ a variance function? As we shall see, the answer is yes for a discrete set of c . To solve this particular problem, we have to design methods based on elliptic functions, and these methods appear to have a wider domain of applicability. For this reason, the aim of the present paper is the classification of the variance functions of the form $(Am^4 + Bm^2 + C)^{1/2}$ and their reciprocals in the sense of Letac and Mora [5], namely the variance functions of the form $m(Cm^4 + Bm^2 + A)^{1/2}$.

Section 2 recalls general facts and methods for dealing with NEF's. Section 3 opens a long parenthesis on pairs of associated NEF: if F and G are NEF generated by symmetric probabilities, we say that they are associated roughly if we can write

$V_F(m) = f(m^2)$ and $V_G(m) = f(-m^2)$. This definition seems to be a mere curiosity of distribution theory, but appears to be illuminating when applied to our elliptic NEF. Since this concept of association has several interesting aspects, I have provided here several detailed examples (Sect. 3.1) that the reader should skip if he is only interested in elliptic NEF.

Section 4 rules out trivial values of the parameters for elliptic NEF. Sections 5–7 investigate the various cases according to the parameters (A, B, C) , Sect. 8 considers the reciprocal families of the previous elliptic NEF: these reciprocal NEF are interesting distributions of the positive integers. Section 9 makes brief comments on the variance functions $(\alpha m + \beta)\sqrt{P(m)}$ where P is an arbitrary polynomial of degree ≤ 4 , actually a complete new field of research. While the statements of the present paper are understandable without knowledge of elliptic functions, the proofs of Sects. 5–7 make heavy use of them, and we shall constantly refer to the magnificent book by Sansone and Gerretsen [9] that we frequently quote by SG.

2 Retrieving an NEF from Its Variance Function

The concept of exponential family is obviously the backbone of Ole's book (Barndorff-Nielsen [10]) or of Brown [20], but the notations for the simpler object called natural exponential family are rather to be found in Morris [2], Jørgensen [3] and Letac and Mora [5].

If μ is a positive non Dirac Radon measure on $\mathbb{R}$ consider its Laplace transform

$$L_\mu(\theta) = \int_{-\infty}^{\infty} e^{\theta x} \mu(dx) \leq \infty.$$

Assume that the interior $\Theta(\mu)$ of the interval $D(\mu) = \{\theta \in \mathbb{R}; L_\mu(\theta) < \infty\}$ is not empty. The set of positive measures μ on $\mathbb{R}$ such that $\Theta(\mu)$ is not empty and such that μ is not concentrated on one point is denoted by $\mathcal{M}(\mathbb{R})$. We denote by $\mathcal{M}_1(\mathbb{R})$ the set of probabilities μ contained in $\mathcal{M}(\mathbb{R})$. In other terms, the elements of $\mathcal{M}_1(\mathbb{R})$ are the probability laws on $\mathbb{R}$ which have a non trivial Laplace transform.

Write $k_\mu = \log L_\mu$. Then the family of probabilities

$$F = F(\mu) = \{P(\theta, \mu); \theta \in \Theta(\mu)\}$$

where

$$P(\theta, \mu)(dx) = e^{\theta x - k_\mu(\theta)} \mu(dx)$$

is called the natural exponential family generated by μ . Two basic results are

$$m = k'_\mu(\theta) = \int_{-\infty}^{\infty} x P(\theta, \mu)(dx)$$

and the fact that k'_μ is increasing (or that k_μ is convex). The set $k'_\mu(\Theta(\mu)) = M_F$ is called the domain of the means. We denote by $\psi_\mu : M_F \rightarrow \Theta(\mu)$ the reciprocal function of $m = k'_\mu(\theta)$. Thus $F = F(\mu)$ can be parametrized by M_F by the map from M_F to F which is

$$m \mapsto P(\psi_\mu(m), \mu) = P(m, F).$$

In other terms an element of $F(\mu)$ can be identified by the value of the mean m rather than by the value of the canonical parameter θ . One can prove that the variance $V_F(m)$ of $P(m, F)$ is

$$V_F(m) = k''_\mu(\psi_\mu(m)) = \frac{1}{\psi'_\mu(m)}. \quad (1)$$

The map $m \mapsto V_F(m)$ from M_F to $(0, \infty)$ is called the variance function and characterizes F . The Jørgensen set of μ is the set $\Lambda(\mu)$ of positive numbers t such that there exists a positive measure μ_t such that $\Theta(\mu_t) = \Theta(\mu)$ and such that $L_{\mu_t} = (L_\mu)^t$. Obviously, $\Lambda(\mu)$ is an additive semigroup which contains all positive integers. If $t \in \Lambda(\mu)$ we denote $F_t = F(\mu_t)$ and it is easily checked that $M_{F_t} = tM_F$ and that

$$V_{F_t}(m) = tV_F\left(\frac{m}{t}\right). \quad (2)$$

The union $G = G(\mu) = \cup_{t \in \Lambda(\mu)} F(\mu_t)$ is called the exponential dispersion model generated by μ .

If F is a NEF and if $h(x) = ax + b$ (with $a \neq 0$) then the family $h(F)$ of images of elements of F by h is still a NEF with $M_{h(F)} = h(M_F)$ and

$$V_{h(F)}(m) = a^2 V_F\left(\frac{m-b}{a}\right). \quad (3)$$

In spite of the similarity between (2) and (3), the last formula is much more useful for dealing with a NEF F which is known only by its variance function: the reason is that the Jørgensen set $\Lambda(F)$ of F is unknown in many circumstances. In fact $\Lambda(F)$ is a closed additive semi group of $[0, \infty)$ which can be rather complicated (see Letac et al. [11] for an example). In the other hand an affinity is always defined, and this fact can be use to diminish the number of parameters of a family of variance functions. For instance, if we consider the variance function $\sqrt{Am^4 + Bm^2 + C}$ such that $C > 0$, without loss of generality we could assume that $C = 1$ by using the dilation $x \mapsto x/\sqrt{C}$.

An important fact for the sequel is that V_F is real analytic, that means that for any $m_0 \in M_F$ there exists a positive number r such that for $m_0 - r < m < m_0 + r$ we have

$$V_F(m) = \sum_{n=0}^{\infty} \frac{(m - m_0)^n}{n!} V_F^{(n)}(m_0)$$

which implies that V_F is analytically extendable to a connected open set of the complex plane containing the real segment M_F . If $\mu \in \mathcal{M}_1(\mathbb{R})$, the Laplace transform L_μ defined on the open interval $\Theta(\mu)$ is extendable analytically in a unique way to the strip $\Theta(\mu) + i\mathbb{R}$ of the complex plane. This extension is also denoted L_μ and $\theta \mapsto L_\mu(i\theta)$ is the Fourier transform of the probability μ . The function $k_\mu(\theta) = \log L_\mu(i\theta)$ could be also extendable to an analytic function on the same strip, but it would be a multivalued function if $L_\mu(\theta)$ has zeros in the strip.

To conclude this section, recall the four steps allowing us to pass from the variance function V_F of a NEF F to a measure μ such that $F = F(\mu)$.

1. Writing $d\theta = \psi'_\mu(m)dm = \frac{dm}{V_F(m)}$, we compute $\theta = \psi_\mu(m)$ as a function of m by a quadrature;
2. we deduce from this the parameter m as a function $m = k'_\mu(\theta)$ (this is generally a difficult point);
3. we compute $k_\mu(\theta)$ by a second quadrature and obtain $L_\mu = e^{k_\mu}$;
4. we use dictionary, creativity, or inversion Fourier formulas to retrieve μ from its Laplace transform.

We keep these four steps in mind for dealing with $V_F(m) = \sqrt{Am^4 + Bm^2 + C}$ in the sequel. It is worthwhile to sketch here an example with

$$V(m) = \sqrt{1 + 4m^4}.$$

For $0 < w < 1$ we do the change of variable and perform the first step:

$$m = \frac{\sqrt{1-w^4}}{2w}, \quad w^2 = -2m^2 + \sqrt{1+4m^4}, \quad d\theta = \frac{dw}{\sqrt{1-w^4}}, \quad \theta = \int_{w(m)}^1 \frac{dw}{\sqrt{1-w^4}}.$$

The second step introduces a function $C(\theta)$ defined on the interval $[0, K]$ where $K = \int_0^1 \frac{dw}{\sqrt{1-w^4}} = 1.3098..$ as

$$\theta = \int_{C(\theta)}^1 \frac{dw}{\sqrt{1-w^4}}. \quad (4)$$

Since $w(m) = C(\theta)$, up to the knowledge of $C(\theta)$, and taking derivative of both sides of (4), the second step is performed since

$$m = k'(\theta) = \frac{\sqrt{1-C(\theta)^4}}{2C(\theta)} = -\frac{C'(\theta)}{2C(\theta)}.$$

The third step is easy and we get $k(\theta) = -\frac{1}{2} \log C(\theta)$ and the Laplace transform $L(\theta) = \frac{1}{\sqrt{C(\theta)}}$. The fourth step needs to be explicit about $C(\theta)$ and the theory of elliptic functions becomes necessary: details about this particular example are in Theorem 4.1 when doing $k^2 = -1$. The function L will be the Laplace transform of

a discrete distribution concentrated on set on numbers of the form n/a where n is a relative integer and where a is the complicated number $\frac{2K}{\pi} = 0.8338\dots$ If we use formula (3) we get the following surprising result: the function $\sqrt{a^4 + 4m^4}$ is the variance function of a NEF concentrated on $\mathbb{Z}$.

3 Associated Natural Exponential Families

The source of this concept is the pair of identities (5) below: if

$$\mu(dx) = \frac{dx}{2 \cosh \frac{\pi x}{2}}$$

and $\nu = \frac{1}{2}(\delta_{-1} + \delta_1)$ is the symmetric Bernoulli distribution then

$$\int_{-\infty}^{+\infty} e^{\theta x} \mu(dx) = \frac{1}{\cosh \theta} \quad (\text{for } |\theta| < \pi/2), \quad \int_{-\infty}^{+\infty} e^{i\theta x} \nu(dx) = \cos \theta, \quad (5)$$

which could be as well presented by reversing the roles of Fourier and Laplace transforms:

$$\int_{-\infty}^{+\infty} e^{i\theta x} \mu(dx) = \frac{1}{\cosh \theta}, \quad \int_{-\infty}^{+\infty} e^{\theta x} \nu(dx) = \cosh \theta. \quad (6)$$

This is an example of what we are going to call an associated pair (μ, ν) of probabilities on $\mathbb{R}$. Here is the definition:

Definition 3.1 Let μ and ν be in $\mathcal{M}_1(\mathbb{R})$ such that μ and ν are symmetric. We say that (μ, ν) is an associated pair if for all $\theta \in \Theta(\mu)$ the Fourier transform of ν is $1/L_\mu(\theta)$. In other terms for $\theta \in \Theta(\mu)$ we have

$$\int_{-\infty}^{+\infty} e^{\theta x} \mu(dx) = L_\mu(\theta), \quad \int_{-\infty}^{+\infty} e^{i\theta x} \nu(dx) = \frac{1}{L_\mu(\theta)}. \quad (7)$$

The corresponding natural exponential families $F = F(\mu)$ and $G = F(\nu)$ are also said to be associated.

We describe now the easy consequences of this definition:

Proposition 3.1 Let (μ, ν) in $\mathcal{M}_1(\mathbb{R})$ be an associated pair. Then

1. (Symmetry) The pair (ν, μ) is also associated;
2. (Uniqueness) If (μ, ν_1) is also associated, then $\nu_1 = \nu$.
3. (Convolution) If (μ', ν') is an associated pair then $(\mu * \mu', \nu * \nu')$ is also an associated pair.

4. (Zeros) Denote $z_\mu = \inf\{\theta > 0 ; L_\mu(i\theta) = 0\}$. Then $\Theta(\nu) = (-z_\mu, z_\mu)$.
5. (Variance functions) Consider the associated pair $F = F(\mu)$ and $G = F(\nu)$ of NEF. If V_F and V_G are extended as analytic functions to the complex plane in a neighborhood of zero, then $V_F(m) = V_G(im)$.

Comments

1. Clearly, since the Fourier transform $\frac{1}{L_\mu(\theta)}$ is real, the probability ν must be symmetric.
2. Symmetry of ν implies that $\Theta(\nu)$ is a symmetric interval, as well as the mean domain of $F(\nu)$.
3. Because of the uniqueness of Part 2, we shall also write μ^* for indicating that (μ, μ^*) is an associated pair. In this case μ^* is called the associated probability to μ (when it exists). We also observe that

$$(\mu^*)^* = \mu, \quad (\mu * \mu')^* = \mu^* * (\mu')^*.$$

4. It is not correct to think that if μ is in $\mathcal{M}_1(\mathbb{R})$ then μ^* always exists. An example is given by the first Laplace distribution (also called the bilateral exponential distribution)

$$\mu(dx) = \frac{1}{2}e^{-|x|}dx, \quad \Theta(\mu) = (-1, 1), \quad L_\mu(\theta) = \frac{1}{1 - \theta^2}.$$

Suppose that $\nu = \mu^*$ exists. Then its Fourier transform on $(-1, 1)$ is $1 - \theta^2$. This implies that its Laplace transform is $L_\nu(\theta) = 1 + \theta^2$ and therefore $\Theta(\nu) = \mathbb{R}$. But if $k_\nu = \log L_\mu$ it is easy to see that the sign of $k_\nu''(\theta)$ is the sign of $1 - \theta^2$, which implies that k_ν is not convex, a contradiction.

A more complicated example is given by

$$\mu_t(dx) = \frac{2^{t-2}}{\pi \Gamma(t)} \left| \Gamma\left(\frac{t+ix}{2}\right) \right|^2 dx \quad (8)$$

for $t > 0$. We will see in Sect. 3.1 that $\mu_t \in \mathcal{M}_1(\mathbb{R})$ satisfies $\Theta(\mu_t) = (-\frac{\pi}{2}, \frac{\pi}{2})$ and

$$\frac{2^{t-2}}{\pi \Gamma(t)} \int_{-\infty}^{+\infty} e^{\theta x} \left| \Gamma\left(\frac{t+ix}{2}\right) \right|^2 dx = \frac{1}{(\cos \theta)^t}. \quad (9)$$

If t is not an integer, then $\mu_t^* = \nu$ does not exist (Proposition 3.2). An obvious case is $t = 1/2$: if X, Y are iid such that $\Pr(X + Y = \pm 1) = 1/2$ then $\Pr(X = \pm 1/2) = 1/4$ and $\Pr(X + Y = 0) \geq 1/16 > 0$, a contradiction.

5. In Definition 3.1, suppose that we relax the constraint on ν to have a Laplace transform. Consider the example

$$\mu(dx) = \frac{dx}{2 \cosh(\pi x)/2}$$

with Laplace transform $1/\cos \theta$ defined on $\Theta(\mu) = (-\frac{\pi}{2}, \frac{\pi}{2})$. A possible associated ν is the Bernoulli $\frac{1}{2}(\delta_{-1} + \delta_1)$ which satisfies $\int_{-\infty}^{+\infty} e^{i\theta x} \nu(dx) = \cos \theta$ in particular on $|\theta| < \pi/2$. However it is not excluded that there exists other probabilities ν fulfilling the same property on $|\theta| < \pi/2$. Imposing $\nu \in \mathcal{M}_1(\mathbb{R})$ rules out this phenomenon, from Part 2 of Proposition 3.1.

6. Here is the simplest example illustrating Part 5 of Proposition 3.1. We use once more the associated pair (5). In this case $M_F = \mathbb{R}$, $a_F = \infty$, $V_F(m) = 1 + m^2$, $M_G = (-1, 1)$, $a_G = 1$, $V_G(m) = 1 - m^2$. See Proposition 3.5 below.
7. SELF ASSOCIATED PAIRS AND NEF: A trivial example is $\mu = N(0, 1)$ since $\mu = \mu^*$. More generally, V_F is a function of m^4 if and only if $\mu = \mu^*$. An other important example will be found in Theorem 5.1 below, which is $V_F(m) = \sqrt{1 + 4m^4}$. Note that the symmetry of μ is essential: if $V_F(m) = m^4$, with $M_F = (0, \infty)$, we have $V_F(m) = V_F(im)$ but the concept of association does not make sense here.
8. This Part 5 provides also a way to decide quickly from the examination of the variance function that μ^* does not exist. For instance, if $\mu \sim X - Y$ where X and Y are iid with the Poisson distribution of mean 1, then $F = F(\mu)$ has variance function $V_F(m) = \sqrt{1 + m^2}$. Would $G = F(\mu^*)$ exist, its variance function would be $V_G(m) = \sqrt{1 - m^2}$. The domain of the mean of G would be $(-1, 1)$, from the principle of analytic continuation of variance functions (Theorem 3.1 in Letac and Mora [5]). However on around the point $m = 1$, the function V_G would be equivalent to $2(1 - m)^{1/2}$. This is forbidden by the principle of Jørgensen et al. [12]: this principle says that if $M_G = (a, b)$ with $b < \infty$ and if

$$V_G(m) \sim_{m \rightarrow b} A \times (b - m)^p \quad (10)$$

then $p \notin (0, 1)$.

Similarly consider the variance function $V_F(m) = (1 + m^2)^{3/2}$ defined on $M_F = \mathbb{R}$. One can consult Letac [1] Chap. 5 Example 1.2 for a probabilistic interpretation. It is generated by a μ such that $\Theta(\mu) = (-1, 1)$ and $k_\mu = \sqrt{1 - \theta^2} - 1$. For seeing that $V_G(m) = (1 - m^2)^{3/2}$ cannot be a variance function we observe the following. If $\nu = \mu^*$ exists then

$$k_\nu(\theta) = \sqrt{1 + \theta^2} - 1.$$

Therefore, by using the principle of maximal analytic continuation (see Proposition 3.2 below), we have $\Theta(\nu) = \mathbb{R}$. As a consequence $L_\nu(\theta) = e^{\sqrt{1 + \theta^2} - 1}$ is an entire function, which is clearly impossible.

Proof of Proposition 3.1 (1) Suppose that ν is in $\mathcal{M}_1(\mathbb{R})$. Then the knowledge of the Fourier transform of ν on the interval $\Theta(\mu)$ gives the knowledge of the Laplace transform L_ν on $\Theta(\nu)$. Now the Fourier transform of μ restricted to $\Theta(\nu)$ is $L_\mu(i\theta) = 1/L_\nu(i\theta)$ from the relation (7) extended by analyticity.

(2) If ν_1 exists, its Fourier transform coincides with the Fourier transform of ν on the interval $\Theta(\mu)$. By analyticity, the two coincide everywhere and $\nu = \nu_1$.

(3) is obvious.

(4) Since the Fourier transform of ν restricted to $\Theta(\mu)$ is $1/L_\mu(\theta)$ then in a neighborhood of $\theta = 0$, the Laplace transform of ν satisfies $L_\nu(\theta) = 1/L_\mu(i\theta)$. Now we use the following result:

Proposition 3.2 (Principle of maximal analyticity) *If $\nu \in \mathcal{M}(\mathbb{R})$ and if $\Theta(\nu) = (a, b)$ suppose that there exists $(a_1, b_1) \supset (a, b)$ and a real analytic function f on (a_1, b_1) which is strictly positive and such that $f(\theta) = L_\nu(\theta)$ for $a < \theta < b$. Then $a = a_1$ and $b = b_1$.*

Proof Use the method of proof of Theorem 3.1 of Letac and Mora [5] or Kawata [13], Chap. 7.

We now return to the proof of Proposition 3.1, Part (4). Write $\Theta(\nu) = (-b, b)$. Clearly $b > z_\mu$ is impossible since it would imply that $L_\nu(z_\mu)$ would be finite, a contradiction with $L_\mu(z_\mu) = 0$. We apply Proposition 2.2 to the present ν , to $(a_1, b_1) = (-z_\mu, z_\mu)$ and to the positive analytic function on this interval $f(\theta) = 1/L_\mu(\theta)$. As a consequence $b = b_1 = z_\mu$ and the result (4) is proved.

(5) Consider the functions L_μ and L_ν . They are analytic on the strips $\Theta(\mu) \times i\mathbb{R}$ and $\Theta(\nu) \times i\mathbb{R}$, and from Part (4) $\Theta(\mu) + i\Theta(\nu)$ is the open square with vertices $\pm z_\nu \pm iz_\nu$. Let Z be the set of zeros of the analytic function $\theta \mapsto L_\mu(\theta)$ restricted to the square $\Theta(\mu) + i\Theta(\nu)$. From the principle of isolated zeros, Z contains only a finite number of points in the compact set $[-a, a] \times [-b, b]$ when $a < z_\nu$ and $b < z_\mu$. Also Z has no zeros on the set $S = (-z_\nu, z_\nu) \cup (-iz_\mu, iz_\mu)$. Consider now the part Z_{++} contained in the first quadrant, and its closed convex hull C_{++} . Similarly consider $C_{\pm, \pm}$, the closed set $C = C_{++} \cup C_{+-} \cup C_{-+} \cup C_{--}$ and the open set $U = \Theta(\mu) + i\Theta(\nu) \setminus C$. Then U is a simply connected set and is a neighborhood of S .

We are in position to define $\log L_\mu = k_\mu$ on the open set U as an analytic function. On this set U we have

$$k_\mu(\theta) = -k_\nu(i\theta), \quad k'_\mu(\theta) = -ik_\nu(i\theta), \quad k''_\mu(\theta) = k''_\nu(i\theta). \quad (11)$$

Since

$$V_F(k'_\mu(\theta)) = k''_\mu(\theta), \quad V_G(k'_\nu(\theta)) = k''_\nu(\theta)$$

we get finally

$$V_G(ik'_\mu(i\theta)) = k''_\mu(i\theta)$$

and this is saying that for m in the open set $k'_\mu(U)$ we have $V_F(m) = V_G(im)$, which is the desired result.

Proposition 3.3 (Convolution of Bernoulli's) *Let $(a_n)_{n=1}^\infty$ be a real sequence such that $\sum_{n=1}^\infty a_n^2 < \infty$. Let $(X_n)_{n=1}^\infty$ and $(Y_n)_{n=1}^\infty$ be two iid sequences such that*

$$X_n \sim \frac{1}{2 \cosh(\pi x/2)}, \quad Y_n \sim \frac{1}{2}(\delta_{-1} + \delta_1).$$

Then the distributions μ of $\sum_{n=1}^{\infty} a_n X_n$ and ν of $\sum_{n=1}^{\infty} a_n Y_n$ are associated.

Proof Easy, from (5) and Part (3) of Proposition 2.1. Note that for $a_n = 1/3^n$ then ν is the purely singular Cantor distribution on $(-1/2, 1/2)$, while μ has a density.

3.1 Examples of Associated Probabilities

Here are 3 groups of examples. It can be observed that they offer three different generalizations of (5). We start with the classical formula for $t > 0$ correct for $\theta \in (-t, t)$:

$$\int_{-\infty}^{+\infty} e^{x\theta} \frac{dx}{(\cosh x)^t} = \frac{2^{t-1}}{\Gamma(t)} \Gamma\left(\frac{t+\theta}{2}\right) \Gamma\left(\frac{t-\theta}{2}\right) \quad (12)$$

with $\Theta(\mu_t) = (-t, t)$. In particular using the duplication formula $\sqrt{\pi} \Gamma(t) = 2^{t-1} \Gamma(\frac{t}{2}) \Gamma(\frac{t+1}{2})$ we get the Laplace transform of the probability α_t below:

$$\alpha_t(dx) = \frac{\Gamma(\frac{t+1}{2})}{\sqrt{\pi} \Gamma(\frac{t}{2})} \times \frac{dx}{(\cosh x)^t}, \quad L_{\alpha_t}(\theta) = \frac{1}{\Gamma(\frac{t}{2})^2} \times \Gamma\left(\frac{t+\theta}{2}\right) \Gamma\left(\frac{t-\theta}{2}\right) \quad (13)$$

with $\Theta(\alpha_t) = (-t, t)$. It is worthwhile to mention that if X and Y are iid with distribution

$$\beta\left(\frac{t}{2}, 1\right)(dx) = \frac{t}{2} x^{\frac{t}{2}-1} \mathbf{1}_{(0,1)}(x) dx$$

and if $U = \sqrt{X/Y}$ then $\log U \sim \alpha_t$.

Formula (12) is easily proven by the change of variable $u = e^{2x}$ and the formula $\int_0^\infty \frac{u^{p-1} du}{(1+u)^{p+q}} = B(p, q)$ for $p, q > 0$. The Fourier version of (12) is

$$\int_{-\infty}^{+\infty} e^{ix\theta} \frac{dx}{(\cosh x)^t} = \frac{2^{t-1}}{\Gamma(t)} \left| \Gamma\left(\frac{t+i\theta}{2}\right) \right|^2 \quad (14)$$

leading by Fourier inversion to

$$\frac{2^{t-1}}{2\pi \Gamma(t)} \int_{-\infty}^{+\infty} e^{i\theta x} \left| \Gamma\left(\frac{t+ix}{2}\right) \right|^2 dx = \frac{1}{(\cosh \theta)^t} \quad (15)$$

and by analyticity to (9). For a while, let us specialize these formulas to $t = 2p-1$ and to $t = 2p$ where p is a positive integer. From the complements formula $\Gamma(z) \Gamma(1-z) = \pi / \sin(\pi z)$ and $\Gamma(z+1) = z \Gamma(z)$ we have for $t = 1, 2$

$$\Gamma\left(\frac{1+\theta}{2}\right)\Gamma\left(\frac{1-\theta}{2}\right) = \frac{\pi}{\cos \frac{\pi\theta}{2}}, \quad \Gamma\left(1+\frac{\theta}{2}\right)\Gamma\left(1-\frac{\theta}{2}\right) = \frac{\pi\theta}{2 \sin \frac{\pi\theta}{2}}$$

and more generally

$$\Gamma\left(\frac{2p+1+\theta}{2}\right)\Gamma\left(\frac{2p+1-\theta}{2}\right) = \frac{1}{2^p}(1-\theta^2)(9-\theta^2)\dots((2p-1)^2-\theta^2) \times \frac{\pi}{\cos \frac{\pi\theta}{2}}, \quad (16)$$

$$\Gamma\left(p+\frac{\theta}{2}\right)\Gamma\left(p-\frac{\theta}{2}\right) = \frac{1}{2^p}(4-\theta^2)(16-\theta^2)\dots(4p^2-\theta^2) \times \frac{\pi\theta}{2 \sin \frac{\pi\theta}{2}}. \quad (17)$$

Proposition 3.4 *If α_t is defined by (13) then α_t^* exists if and only if $t \geq 1$. In particular $\alpha_1^* = \frac{1}{2}(\delta_{-\pi/2} + \delta_{\pi/2})$ is a Bernoulli distribution and for $t > 1$ we have*

$$\alpha_t^*(dx) = \frac{\Gamma(t)}{\sqrt{\pi}\Gamma(\frac{t-1}{2})}(\cos x)^{t-2}1_{(-\pi/2, \pi/2)}(x)dx.$$

In particular for $t = 2p + 1$ and $t = 2p + 2$ where p is a non negative integer, then (16) and (17) give $(\varphi_t)^{-1}$ when φ_t is the Fourier transform of α_t^* .

Comments. For this example, the explicit calculation of the variance functions of $F = F(\alpha_t)$ and $G = F(\alpha_t^*)$ is not possible. For instance if $t = 2$ the probability α_2^* is the uniform distribution on the segment $(-\pi/2, \pi/2)$. In this case $L_{\alpha_2^*}(\theta) = \frac{\sinh(\pi t/2)}{\pi t/2}$: no way to compute $\theta = \psi_{\alpha_2^*}(m)$ in a close formula when

$$m = k'_{\alpha_2^*}(\theta) = \frac{\pi}{2} \left(\coth\left(\frac{\pi\theta}{2}\right) - \frac{2}{\pi\theta} \right).$$

Shanbhag [14] and, in their Proposition 4, Barlev and Letac [15], have other proofs of the ‘only if’ condition of existence of α_t^* .

Proof For $t > 1$ we just rely on entry 3.631, 9 of Gradshteyn and Ryzhik [16]. If $t < 1$ we show that α_t^* does not exist by showing that $\theta \mapsto k''_{\alpha_t}(i\theta)$ is not positive. We obtain

$$k''_{\alpha_t}(i\theta) = \sum_{n=0}^{\infty} \frac{(n + \frac{t}{2})^2 - \frac{\theta^2}{4}}{[(n + \frac{t}{2})^2 + \frac{\theta^2}{4}]^2}.$$

and a careful calculation shows that

$$\lim_{\theta \rightarrow \infty} \theta^2 k''_{\alpha_t}(i\theta) = 2(t-1)$$

If $t < 1$ then $\theta \mapsto k''_{\alpha_t}(i\theta)$ cannot be positive for all $\theta \in \mathbb{R}$, and this ends the proof.

Proposition 3.5 *If μ_t is defined by (9) then μ_t^* exists if and only if t is a positive integer N . In this case μ_N^* is the image of the binomial distribution $B(N, 1/2)$ by $x \mapsto 2x - N$.*

Comments. The most interesting particular case corresponds to $t = 2$ since in this case we meet the uniform distribution on a segment with the associated pair

$$\mu_2(dx) = \frac{x}{4 \sinh(\pi x/2)} dx, \quad (\mu_2)^*(dy) = \frac{1}{2} 1_{(-1,1)}(y) dy.$$

This is also an illustration of Proposition 2.3 applied to $a_n = 1/2^n$ since $\sum_{n=1}^{\infty} \frac{Y_n}{2^n}$ is uniform on $(-1,1)$ when $(Y_n)_{n=0}^{\infty}$ is an iid sequence of symmetric Bernoulli random variables. For this example, the explicit calculation of the variance functions of $F = F(\mu_t)$ and $G = F(\mu_t^*)$ gives

$$V_F(m) = t + \frac{m^2}{t}, \quad V_G(m) = N - \frac{m^2}{N}.$$

Proof of Proposition 3.5. $\Leftarrow$ is obvious. To prove $\Rightarrow$ suppose that there exists a positive integer n_0 such that $n_0 - 1 < t < n_0$ and suppose that μ_t^* exists. Taking the image τ of μ_t^* by the map $x \mapsto x' = x - t$, choosing $\theta > 0$ and denoting $z = e^{-2\theta}$ we get

$$\int_{-\infty}^{+\infty} e^{\theta x'} \tau(dx') = \int_{-\infty}^{+\infty} e^{\theta(x-t)} \mu_t^*(dx) = \frac{1}{2^t} \sum_{n=0}^{\infty} \frac{t(t-1) \dots (t-n+1)}{n!} z^n.$$

Since $t(t-1) \dots (t-n+1) < 0$ when $n = n_0 + 1$ this shows that $\tau(\{-2n_0 - 2\}) < 0$, a contradiction.

The third example is obtained by considering the Babel class of NEF, namely the set of exponential families such that the variance function has the form $V_F = P\Delta + Q\sqrt{\Delta}$ where Δ , P and Q are polynomials with respective degrees less or equal to 2, 1, 2. Looking for possible pairs (F, G) in this class such that $V_F(m) = V_G(im)$ and such that F and G are generated by associated distributions (μ, ν) -and therefore symmetric- implies that $\Delta(m) = Am^2 + C$, P is a constant and $Q(m) = A'm^2 + C'$. The case $C = 0$ is excluded since the domain of the mean M_F and M_G are symmetric interval and V_F and V_G are real analytic on them. As a consequence either F or G must be such that $\Delta(m) = 1 - m^2$ (up to affinities). But there is only one type of NEF in the Babel class such that $\Delta(m) = 1 - m^2$ and it is generated by the trinomial distributions defined for $0 < a < 1$ by

$$\mu_a = \frac{1}{a+1} (a\delta_0 + \frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_1) \quad (18)$$

and their entire powers of convolution. Of course the limit cases are related to Bernoulli, since

$$\mu_0 = \frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_1, \quad \mu_1 = \left(\frac{1}{2}\delta_{-1/2} + \frac{1}{2}\delta_{1/2}\right) * \left(\frac{1}{2}\delta_{-1/2} + \frac{1}{2}\delta_{1/2}\right).$$

Proposition 3.6 *If μ_a is defined by (18) with $a \in (0, 1)$ then μ_a^* exists and is*

$$\mu_a^* = \tau_b * \tau_{-b}.$$

where $a = \cos 2b$ with $0 < b < \pi/4$ and

$$\tau_{\pm b}(x) = \frac{\cos b}{\cosh \frac{\pi x}{4}} e^{\pm bx} dx.$$

Proof We have

$$L_{\mu_a}(\theta) = \frac{a + \cosh \theta}{a + 1}, \quad V_{F(\mu_a)} = \frac{1}{1 - a^2} - m^2 - \frac{a}{\sqrt{1 - a^2}} \sqrt{\frac{1}{1 - a^2} - m^2}.$$

Therefore, if μ_a^* does exist it must satisfy

$$L_{\mu_a^*}(\theta) = \frac{a + 1}{a + \cos \theta}, \quad V_{F(\mu_a^*)} = \frac{1}{1 - a^2} + m^2 - \frac{a}{\sqrt{1 - a^2}} \sqrt{\frac{1}{1 - a^2} + m^2}$$

with $\Theta(\mu_a)^* = (-z_{\mu_a}, z_{\mu_a})$ where z_{μ_a} is the smallest positive solution of $\cos \theta = -a$. Such a μ_a^* actually exists. To see this we write $a = \cos 2b$ with $0 < b < \pi/4$ and by simple trigonometry and the help of formula (6):

$$\frac{\cos 2b + 1}{\cos 2b + \cos \theta} = \frac{\cos b}{\cos(\frac{\theta}{2} - b)} \times \frac{\cos b}{\cos(\frac{\theta}{2} + b)} = L_{\tau_b}(\theta) L_{\tau_{-b}}(\theta)$$

where

$$\tau_{\pm b}(x) = \frac{\cos b}{\cosh \frac{\pi x}{4}} e^{\pm bx} dx.$$

4 Discussion and Easy Cases for $(Am^4 + Bm^2 + C)^{1/2}$

In this section we recall known and not so well known results about a few particular cases. The cases where only one of the three numbers A, B, C is not zero are classical: we get respectively the gamma, Poisson or normal case. We now investigate three more interesting particular cases (they are all described in Letac [7] as elements of the Babel class).

4.1 The Case $A = 0$

The useful results are contained in the following proposition:

Proposition 4.1 *Let $t > 0$. Let N_1 and N_2 be two independent standard Poisson random variables with expectation $t/2$. Then the exponential family F_t with domain of the means $\mathbb{R}$ and variance function $(m^2 + t^2)^{1/2}$ exists and is generated by the distribution μ_t of $N_1 - N_2$. Furthermore*

$$\mu_t(dx) = \sum_{n \in \mathbb{Z}} e^{-t} I_{|n|}(t) \delta_n(dx)$$

where

$$I_x(t) = \sum_{n=0}^{\infty} \frac{1}{n! \Gamma(n+x+1)} \left(\frac{t}{2}\right)^{2n+x}.$$

Proof Since $\mathbb{E}(e^{\theta(N_1 - N_2)}) = e^{t(\cosh \theta - 1)}$ we get that $\Theta(\mu_t) = \mathbb{R}$ and that

$$k_{\mu_t}(\theta) = t(\cosh \theta - 1), \quad k'_{\mu_t}(\theta) = t \sinh \theta, \quad k''_{\mu_t}(\theta) = t \cosh \theta = (k'_{\mu_t}(\theta))^2 + t^2)^{1/2}.$$

Thus $V_{F(\mu_t)}(m) = (m^2 + t^2)^{1/2}$ as desired, and the domain of the means is $\mathbb{R}$. A consequence of this proposition and of (3) and (2) is that $(Bm^2 + C)^{1/2}$ is always a variance function for B and $C > 0$.

4.2 The Case $C = 0$

s

Proposition 4.2 *Let $t > 0$. Then the exponential family F_t with domain of the means $\mathbb{R}$ and variance function $m(1 + \frac{m^2}{t^2})^{1/2}$ exists. In particular F_1 is generated by $\mu_1 = \delta_0 + 2 \sum_{n=1}^{\infty} \delta_n$. More specifically, P is in F_1 if and only if there exists $q \in (0, 1)$ such that P is the convolution of the Bernoulli distribution $\frac{1}{1+q} \delta_0 + \frac{q}{1+q} \delta_1$ with the geometric distribution $(1 - q) \sum_{n=0}^{\infty} q^n \delta_n$.*

Proof Writing for $\theta < 0$ $L_{\mu_1}(\theta) = \frac{1+e^\theta}{1-e^\theta}$ it is easily seen that it generates a natural exponential family with domain of the means $(0, \infty)$ and variance function $m(1 + m^2)^{1/2}$. The only non trivial point of the proposition is the fact that the elements of F_1 are infinitely divisible. For this we write

$$k_{\mu_1}(\theta) = \sum_{n=1}^{\infty} \frac{1}{n} (1 + (-1)^n) e^{n\theta}.$$

Since the coefficient $\frac{1}{n}(1 + (-1)^n)$ of $e^{n\theta}$ is ≥ 0 the result is proved (although it is difficult to compute μ_t explicitly when t is not an integer).

A consequence of this proposition is that $(Am^4 + Bm^2)^{1/2}$ is a variance function for A and $B > 0$ with domain of the means $(0, \infty)$.

4.3 The Case $B^2 - 4AC = 0$

Here is a well known fact (see Morris [2]):

Proposition 4.3 *Let $t > 0$. The natural exponential family F_t with domain of the means $\mathbb{R}$ and variance function $t(1 + \frac{m^2}{t^2})$ is generated by the probability μ_t defined by (8).*

This rules the case $B^2 - 4AC = 0$ such that $Ax^2 + Bx + C$ has a negative double root with $A > 0$.

Proposition 4.4 *Let $t > 0$. The natural exponential family F_t with domain of the means (t, ∞) and variance function $\frac{t}{2}(\frac{m^2}{t^2} - 1)$ exists. In particular F_1 is generated by $\mu_1 = \sum_{n=1}^{\infty} n\delta_n$.*

Proof We do not give the details about μ_1 which are standard. Since the elements of F_1 are negative binomial distributions shifted by 1, they are still infinitely divisible and F_t does exist for all $t > 0$.

This rules out the case $B^2 - 4AC = 0$ such that $Ax^2 + Bx + C$ has a positive double root x_0 with $A > 0$ and domain of the means $(\sqrt{x_0}, \infty)$.

Proposition 4.5 *Let $N > 0$ be an integer. The natural exponential family F_t with domain of the means $(-N, N)$ and variance function $\frac{N}{2}(1 - \frac{m^2}{N^2})$ exists. It is generated by $(\delta_1 + \delta_{-1})^{*N}$.*

Proof This is an easy and classical fact.

4.4 $Ax^2 + Bx + C$ Cannot Have Simple Roots on $(0, \infty)$

We discard some values of (A, B, C) . Suppose that $Ax^2 + Bx + C$ has a positive simple root $x_0 > 0$. Then $(Am^4 + Bm^2 + C)^{1/2}$ cannot be a variance function. For by the principle of maximal analyticity, the domain of the means will have $m_0 = \sqrt{x_0}$ has boundary point. Since x_0 is a simple root, then the variance function around m_0 will be equivalent to $k|m - m_0|^{1/2}$ for some positive constant k . But this is forbidden by the principle of Jørgensen et al. [12] mentioned in (10).

4.5 The Splitting of the Elliptic Variances in Three Cases

The only cases that we are left to consider in order to have a classification of the variance functions of the form $(Am^4 + Bm^2 + C)^{1/2}$ are now the cases where $Ax^2 + Bx + C$ is strictly positive on $[0, \infty)$ and has no double negative root. Of course this implies that $A > 0$ and $C > 0$. To simplify the matters, we choose $C = 1$ and we introduce the function $V(m) = (Am^4 + Bm^2 + 1)^{1/2}$ and, for $t > 0$, the function $V_t(m) = tV(m/t)$. A simple analysis shows that $Ax^2 + Bx + 1$ has no positive roots and no double negative roots if and only if there exists a non zero real number a and a positive number b such that

$$Ax^2 + Bx + 1 = (1 + ax)^2 + 2b^2x.$$

Let us insist of the fact that a can be negative. Finally we introduce a complex number k through its square in order to use the standard notations of elliptic functions:

$$k^2 = 1 + \frac{2a}{b^2}$$

This leads to three cases

1. The case $-1 \leq k^2 < 0$. It corresponds to the fact that $P(x) = (1 + ax)^2 + 2b^2x$ has no roots and that the minimum of P on $[0, \infty)$ is reached on 0.
2. The case $k^2 < -1$. Here P has no roots and reaches its minimum on $[0, \infty)$ at $-b^2(a + b^2)/a^2$.
3. The case $k^2 > 0$. Here P has two distinct negative real roots. Taking $A = 1$ instead of $C = 1$ and $P(x) = (x + a^2)(x + b^2)$ is convenient.

We investigate these cases in the next three sections.

5 The Elliptic Cases: The Case $-1 \leq k^2 < 0$

We write $k^2 = -1 + p$ with $0 \leq p < 1$ and we introduce the following two constants:

$$\begin{aligned} K &= \int_0^1 (1 - x^2)^{-1/2} (2 - p - x^2)^{-1/2} dx \\ K' &= \int_0^1 (1 - x^2)^{-1/2} (1 + (1 - p)x^2)^{-1/2} dx. \end{aligned} \tag{19}$$

Here is our first serious result:

Theorem 5.1 Suppose that $k^2 = 1 + \frac{2a}{b^2} = -1 + p \in [-1, 0)$. For $b = \sqrt{2}$ and $a = -2 + p$ there exists a natural exponential family G_t with domain of the means $\mathbb{R}$ and variance function

$$t\sqrt{(1 + a\frac{m^2}{t^2})^2 + 2b^2\frac{m^2}{t^2}}$$

when t is a multiple of a . It is concentrated on $\frac{\pi}{2K}\mathbb{Z}$. The family $G_{|a|}$ is generated by a symmetric probability measure $\mu_{|a|}$ which is the convolution of the Bernoulli distribution $\frac{1}{2}(\delta_{-\frac{\pi}{2K}} + \delta_{\frac{\pi}{2K}})$ by an infinitely divisible distribution $\alpha_{|a|}$ concentrated on $\frac{\pi}{K}\mathbb{Z}$. We denote $q = e^{-\pi K'/K}$ and for a positive integer v we denote

$$c_v = c_{-v} = \frac{q^v - (-1)^v q^{2v}}{1 - q^{2v}} > 0.$$

Then the Laplace transform of α_t is

$$\int_{-\infty}^{\infty} e^{\theta x} \alpha_t(dx) = \exp\left(\frac{t}{|a|} \sum_{v \in \mathbb{Z} \setminus \{0\}} c_v (e^{\frac{v\pi\theta}{K}} - 1)\right).$$

Finally the characteristic function of $\mu_{2|a|}$ is $\frac{1}{\wp(s+K) - \frac{p}{3}}$ where $\wp$ is the elliptic Weierstrass function satisfying

$$\wp'^2 = 4(\wp - 1 + \frac{2p}{3})(\wp - \frac{p}{3})(\wp + 1 - \frac{p}{3})$$

which is doubly periodic with primitive periods $2K$ and $2iK'$. In particular it has zeros and G_t cannot be infinitely divisible.

Comments. Doing $b = \sqrt{2}$ is not really a restriction. Using the formula $a^2 V_F(m/a)$ for the image of F by $x \mapsto ax$ gives the description of F for an arbitrary $b > 0$.

Proof We apply the standard procedure for computing the Laplace transform of a generating measure when the variance function is given. We shall use the following change of variable $u^2 = (1 + am^2)^2 + 2b^2m^2$ for $u \geq 1$. This implies that

$$m^2 = \frac{1}{a^2}[-a - b^2 + \sqrt{b^4 + 2ab^2 + a^2u^2}].$$

We consider now the new change of variable

$$u = \frac{1}{2}((2 + \frac{b^2}{a})w^2 - \frac{b^2}{aw^2}) = \frac{b^2}{2a}(k^2w^2 - \frac{1}{w^2})$$

with $0 < w < 1$. This choice is designed in order to have $b^4 + 2ab^2 + a^2u^2 = b^4k^2 + a^2u^2$ transformed in a perfect square of a rational function of w :

$$\sqrt{b^4 + 2ab^2 + a^2u^2} = \frac{a}{2} \left(\left(2 + \frac{b^2}{a}\right)w^2 + \frac{b^2}{aw^2} \right) = \frac{b^2}{2} \left(k^2w^2 + \frac{1}{w^2} \right)$$

This leads to

$$m^2 = \frac{b^2}{2a^2w^2} (1 - w^2)(1 - k^2w^2) \quad (20)$$

but also to a surprising result

$$\begin{aligned} a + b^2 + a^2m^2 &= \left(a + \frac{b^2}{2}\right)w^2 + \frac{b^2}{2w^2} = \frac{b^2}{2} \left(k^2w^2 + \frac{1}{w^2}\right) \\ du &= \left[\left(a + \frac{b^2}{2}\right)w^2 + \frac{b^2}{2w^2}\right] \frac{2}{aw} dw = \frac{b^2}{a} \left(k^2w^2 + \frac{1}{w^2}\right) \frac{dw}{w} \\ \frac{du}{a + b^2 + a^2m^2} &= \frac{2}{aw} dw \end{aligned} \quad (21)$$

Recall that $a < 0$ and that $w \mapsto u$ is decreasing. Thus we get, gathering (21) and (20)

$$\begin{aligned} d\theta &= \frac{1}{V(m)} \times dm = \frac{1}{u} \times \frac{2udu}{4m(a + b^2 + a^2m^2)} = \frac{du}{2m(a + b^2 + a^2m^2)} = \frac{dw}{awm} \\ &= -\frac{\sqrt{2}}{b} \frac{dw}{\sqrt{(1 - w^2)(1 - k^2w^2)}} \end{aligned} \quad (22)$$

We introduce the function $\theta \mapsto C(\theta)$ by

$$\theta = \int_{C(\theta)}^1 \frac{dw}{\sqrt{(1 - w^2)(1 - k^2w^2)}}$$

Thus $C(0) = 1$ and the function C is defined on $[0, K']$. Actually, we have $C(\theta) = \operatorname{sn}(K' - \theta)$. In $(0, K')$ it satisfies $C'(\theta) = -(1 - C(\theta)^2)^{1/2}(1 - k^2C(\theta)^2)^{1/2}$. Now we can write

$$\theta = \int_0^m \psi'_\mu(x) dx = \int_{w(m)}^1 \frac{dw}{(1 - w^2)^{1/2}(1 - k^2w^2)^{1/2}}.$$

Thus $w(m) = C(\theta)$ and from (20)

$$m = k'_{\mu_i}(\theta) = m(C(\theta)) = \frac{1}{|a|C(\theta)} (1 - C(\theta)^2)^{1/2} (1 - k^2C(\theta)^2)^{1/2} = \frac{C'(\theta)}{aC(\theta)}.$$

Thus finally we get the Laplace transform of μ_t as

$$L_{\mu_t}(\theta) = \frac{1}{(C(\theta))^{1/|a|}}.$$

We observe that $\theta \mapsto C(\theta)$ has an analytic continuation to the whole complex plane. We now consider its restriction $c(s) = C(is)$ to the imaginary line. It satisfies the differential equation

$$c'(s)^2 = (c(s)^2 - 1)(1 - k^2 c(s)^2) \quad (23)$$

with the initial condition $c(0) = 1$. Now introduce the function $s \mapsto f(s) = -k^2 c^2(s)$. It satisfies the differential equation

$$f'(s)^2 = 4f^3(s) + 4(1 + k^2)f^2(s) + 4f(s) = 4(f(s) + 1)f(s)(f(s) + k^2)$$

(just multiply (23) by c^2 to reach this result). From now it is convenient to write

$$k^2 = 1 + a = p - 1$$

with $p \in [0, 1)$. Then writing $f(s) = -\frac{p}{3} + h(s)$ we get $h'(s)^2 = 4h(s)^3 - g_2 h(s) - g_3$ with

$$g_2 = 4(1 - p + \frac{2p^2}{3}), \quad g_3 = -\frac{4p}{3}(1 - p + \frac{2p^2}{9})$$

Thus h satisfies the differential equation of the $\wp$ function of Weierstrass for the parameters g_2 and g_3 (see SG 247). We can also write $h^2(s) = 4(h - e_1)(h - e_2)(h - e_3)$ with $e_1 = 1 - \frac{2p}{3} > e_2 = \frac{p}{3} > e_3 = -1 + \frac{p}{3}$ with discriminant

$$\Delta = g_2^3 - 27g_3^2 = [4(1 - p)(2 - p)]^2.$$

Thus (see SG page 279 and page 283) the function $\wp$ has periods $2K = 2\omega$ and $2iK' = 2\omega'$

$$\begin{array}{ccccc} 2iK' & \text{---} & K + 2iK' & \text{---} & 2K + 2iK' \\ | & & | & & | \\ iK' & \text{---} & K + iK' & \text{---} & 2K + iK' \\ | & & | & & | \\ 0 & \text{---} & K & \text{---} & 2K \end{array}$$

$$\omega = \frac{1}{\sqrt{2-p}} \int_0^1 \frac{dw}{(1-w^2)^{1/2} (1 - \frac{1-p}{2-p} w^2)^{1/2}},$$

$$\omega' = \frac{i}{\sqrt{2-p}} \int_0^1 \frac{dw}{(1-w^2)^{1/2} (1 - \frac{1-p}{2-p} w^2)^{1/2}} = iK'.$$

The last equality $\omega' = iK'$ is obtained from the changes of variable $w = u^{1/2}$, $u = 1 - v$ and $v = t^2$. We have $h(s) = \wp(s + C)$ for some constant C . Now, since the variance function $t(1 + a\frac{m^2}{t^2})^{1/2} + \frac{4m^4}{t^2})^{1/2}$ is symmetric, there exists a symmetric measure which generates it and without loss of generality we assume that the characteristic function $s \mapsto \frac{1}{f(s)^{t/2|a|}}$ is real. Thus we have to take C such that $c(0) = 1$ or $f(0) = 1 - p$ or $h(0) = 1 - \frac{2p}{3} = e_1$ or $\wp(C) = e_1$. Hence from SG page 279 $C = K$.

Since $s \mapsto \wp(s)$ has periods $2K$ and $2iK'$ and since $\wp(K) = 1 - \frac{2p}{3} = e_1$ we have

$$f(s) = \wp(s + K) - \frac{p}{3}.$$

Note that $s \mapsto \wp(s)$ has no real zeros, only poles on multiples of $2K$ and is periodic. See the picture in SG page 280. Thus

$$s \mapsto \left(\frac{1-p}{f(s)} \right)^{t/2|a|}$$

is $2K$ periodic and has zeros on odd multiples of K . Since it is $2K$ periodic, this implies that it is the characteristic function of a symmetric probability concentrated on multiples of π/K .

$$\frac{1-p}{f(s)}^{t/2|a|} = \sum_{v \in \mathbb{Z}} p_v(t) e^{iv \frac{\pi}{K}} = p_0(t) + 2 \sum_{v=1}^{\infty} p_v(t) \cos \frac{\pi v}{K}.$$

We are going to consider $-\frac{t}{2|a|} \log f(s)$ and to compute its Fourier series. For this we use formula 5.8–22 in SG page 263 applied to $\alpha = 2$ which gives here since $e_2 = p/3$

$$f(2Kz)^{1/2} = \sqrt{\wp(2Kz + K) - \frac{p}{3}} = \frac{\vartheta_3(z + \frac{1}{2})}{C \vartheta_1(z + \frac{1}{2})} \quad (24)$$

where C is some constant and where the q occurring in the Theta functions is given by SG page 261 by $q = e^{i\pi\tau}$ with $\tau = iK'/K$. Thus $q = e^{-\frac{K'}{K}\pi}$ here. Forgetting the factor $t/|a|$ we have

$$-\frac{1}{2} \log f(s) = \log C + \log \vartheta_1\left(\frac{s}{2K} + \frac{1}{2}\right) - \log \vartheta_3\left(\frac{s}{2K} + \frac{1}{2}\right).$$

Consider the derivative of this function:

$$-\frac{1}{2}(\log f(s))' = \frac{1}{2K} \frac{\vartheta_1'(\frac{s}{2K} + \frac{1}{2})}{\vartheta_1(\frac{s}{2K} + \frac{1}{2})} - \frac{1}{2K} \frac{\vartheta_3'(\frac{s}{2K} + \frac{1}{2})}{\vartheta_3(\frac{s}{2K} + \frac{1}{2})}.$$

Now we use formulas about ϑ_j'/ϑ_j given in SG page 274. They are

$$\begin{aligned} \frac{\vartheta_1'(z)}{\vartheta_1(z)} &= \pi \frac{\cos \pi z}{\sin \pi z} + 4\pi \sum_{v=1}^{\infty} \frac{q^{2v}}{1 - q^{2v}} \sin 2v\pi z \\ \frac{\vartheta_3'(z)}{\vartheta_3(z)} &= 4\pi \sum_{v=1}^{\infty} \frac{(-1)^v q^v}{1 - q^{2v}} \sin 2v\pi z \end{aligned}$$

In these expressions we replace z by $\frac{s}{2K} + \frac{1}{2}$ and we get

$$\begin{aligned} \frac{\vartheta_1'(\frac{\pi s}{2K} + \frac{\pi}{2})}{\vartheta_1(\frac{\pi s}{2K} + \frac{\pi}{2})} &= -\pi \frac{\sin \frac{\pi s}{2K}}{\cos \frac{\pi s}{2K}} + 4\pi \sum_{v=1}^{\infty} \frac{(-1)^v q^{2v}}{1 - q^{2v}} \sin \frac{\pi s}{K} \\ \frac{\vartheta_3'(\frac{\pi s}{2K} + \frac{\pi}{2})}{\vartheta_3(\frac{\pi s}{2K} + \frac{\pi}{2})} &= 4\pi \sum_{v=1}^{\infty} \frac{q^v}{1 - q^{2v}} \sin \frac{\pi s}{K} \end{aligned}$$

Thus

$$\begin{aligned} -\frac{K}{\pi}(\log f(s))' &= -\frac{\sin \frac{\pi s}{2K}}{\cos \frac{\pi s}{2K}} + 4 \sum_{v=1}^{\infty} \frac{(-1)^v q^{2v} - q^v}{1 - q^{2v}} \sin 2\frac{\pi s}{2K}, \\ -\frac{1}{2} \log f(s) &= C_1 + \log \cos \frac{\pi s}{2K} + 2 \sum_{v=1}^{\infty} \frac{q^v - (-1)^v q^{2v}}{1 - q^{2v}} \cos 2v \frac{\pi s}{2K}, \\ -\frac{1}{2} \log f(s) &= C_1 + \log \cos \frac{\pi s}{2K} + \sum_{v \in \mathbb{Z} \setminus \{0\}} \frac{q^{|v|} - (-1)^v q^{2|v|}}{1 - q^{2|v|}} \exp 2vi\pi \frac{s}{2K}, \end{aligned}$$

where C_1 is a constant such that $f(0) = 1 - p$. Thus

$$\frac{1}{c(s)^{t/|a|}} = C_2 (\cos \frac{\pi s}{2K})^{t/|a|} \exp \left[\sum_{v \in \mathbb{Z} \setminus \{0\}} \frac{q^{|v|} - (-1)^v q^{2|v|}}{1 - q^{2|v|}} \exp 2vi\pi \frac{s}{2K} \right]$$

The theorem is proved.

6 The Elliptic Cases: $k^2 < -1$

This case is more complicated when treated by the retrieving method of Sect. 2. A reason is the fact that the function $m \mapsto V(m)$ is not convex. More specifically, if $P(x) = (1+ax)^2 + 2b^2x$ is used to define $V(m) = \sqrt{P(m^2)}$ the case $k^2 = 1 + \frac{2a}{b^2} < -1$ correspond to the case where P' has a positive root. Here we shall rather use the method of associated NEF, but no new interesting distributions will appear, as shown by the following result:

Theorem 6.1 *If $k^2 = 1 + \frac{2a}{b^2} < -1$ then $\sqrt{(1+am^2)^2 + 2b^2m^2}$ is not a variance function.*

Proof Suppose that $\sqrt{(1+am^2)^2 + 2b^2m^2}$ is the variance function of some NEF F_1 . Let us assume first that the associated F_2 exists. As a consequence the variance function of F_2 is

$$V_{F_2}(m) = \sqrt{(1-am^2)^2 - 2b^2m^2} = \sqrt{(1+am^2)^2 + 2(2a-b^2)m^2}.$$

Like in Theorem 5.1 without loss of generality we may assume that $2a-b^2 = 2$, and Theorem 5.1 gives us a detailed description of F_2 . If μ_2 is the symmetric probability generating F_2 we have seen that $L_{\mu}(\theta) = C(\theta)^{-1/|a|}$ where $C(\theta) = \text{sn}(K' - \theta)$ with K and K' defined by (19). As a consequence, if μ_1 is the symmetric probability generating F_1 , then from Proposition 3.1 its Fourier transform is

$$\varphi_{\mu_1}(s) = C(s)^{1/|a|}.$$

Now we use the fact that the function C is doubly periodic with periods $2K$ and $2iK'$. This implies that the Fourier transform $\varphi_{\mu_1}(s)$ has period $2K'$ which means that μ_2 is concentrated on a coset of the group $\mathbb{Z}/(2K')$. We are going to use this to see that $s \mapsto C(s)^{-1/|a|}$ is also a Fourier transform of a probability and this will obviously contradict the existence of μ_2 . For this, we have to understand the $2K'$ periodicity of C by coming back to formula (24) which shows that $C(2Ks)$ is the power of a function of the form

$$\frac{\vartheta_3(z + \frac{1}{2})}{C\vartheta_1(z + \frac{1}{2})}.$$

Now we use the Jacobi imaginary transformation (see SG pages 269–272). In our particular case this Jacobi transformation is the following. Denote $\tau = iK'/K$. Then $-1/\tau = iK/K'$. The theta functions ϑ_1 and ϑ_3 are implicitly functions of τ and it is sometimes useful to write $\vartheta_i(z|\tau)$ instead of $\vartheta_i(z)$. Formulas linking $\vartheta_i(z|\tau)$ and $\vartheta_i(z|-1/\tau)$ are known (see 5.10–9 in SG). These formulas show, by the magic of the Jacobi transformation, that $C(s)^{-1/|a|}$ is the Fourier transform of the probability on $\mathbb{Z}/K'$ which is obtained from Theorem 5.1 simply by exchanging the roles of K and K' . We therefore obtain the desired contradiction.

The last task is to get rid of the hypothesis that F_1 has an associated NEF. If $F_1 = F(\mu_1)$ exists with $V_{F_1}(m) = \sqrt{(1 + am^2)^2 + 2b^2m^2}$ then playing with the affine and the Jørgensen transformations (2) and (3) it is possible to find a positive number t such that $\sqrt{(1 - a\frac{m^2}{t})^2 - 2b^2\frac{m^2}{t}}$ is the variance function of some NEF F_2 . This $F_2 = F(\mu_2)$ is necessarily of the type considered in Theorem 5.1 (namely with $1 < k^2 < 0$). We have seen that in this case the Laplace transform of μ_2 is a negative power of $C(\theta) = \text{sn}(\theta - K')$. Therefore the Fourier transform of μ_1 is a positive power of $C(s)$. However, a negative power of $C(s)$ was also the Fourier transform of a probability: we get a contradiction and this ends the proof.

7 The Elliptic Cases: $k^2 > 0$

In this section we study the variances of the form $V(m) = \sqrt{(m^2 + a^2)(m^2 + b^2)}$ where $0 < a < b$.

Theorem 7.1 *Let $0 < a < b$ such that $b^2 - a^2 = 2^{4/3}$. There exists an natural exponential family with variance function $2\sqrt{(\frac{m^2}{4} + a^2)(\frac{m^2}{4} + b^2)}$. It is generated by a symmetric discrete distribution concentrated on the group $\frac{\pi}{K}\mathbb{Z}$ where K is a constant given below by (25).*

Proof We have $\theta = \int_0^m \frac{dt}{\sqrt{(t^2 + a^2)(t^2 + b^2)}}$. We do the changes of variable $t \mapsto u \mapsto v \mapsto w$ defined for $u > ab$, $v > \frac{2ab}{b^2 - a^2}$, $w > \sqrt{\frac{b+a}{b-a}}$.

$$u^2 = (t^2 + a^2)(t^2 + b^2), \quad u = \frac{b^2 - a^2}{2}v, \quad v = \frac{1}{2}(w^2 - \frac{1}{w^2}).$$

Since $udu = 2t(2t^2 + b^2 + a^2)dt$ and $dv = (w^2 + \frac{1}{w^2})\frac{dw}{w}$ we get

$$\theta = \frac{b^2 - a^2}{4} \int_{\frac{2ab}{b^2 - a^2}}^{\frac{2\sqrt{(m^2 + a^2)(m^2 + b^2)}}{b^2 - a^2}} \frac{dv}{t(2t^2 + b^2 + a^2)}.$$

We have also

$$\begin{aligned} t^2 &= -\frac{b^2 + a^2}{2} + \frac{b^2 - a^2}{2}\sqrt{v^2 + 1} = -\frac{b^2 + a^2}{2} + \frac{b^2 - a^2}{4}(w^2 + \frac{1}{w^2}) \\ &= \frac{b^2 - a^2}{4w^2}(w^2 - \frac{b+a}{b-a})(w^2 - \frac{b-a}{b+a}). \end{aligned}$$

Thus $2t^2 + a^2 + b^2 = \frac{b^2 - a^2}{2}(w^2 + \frac{1}{w^2})$ which implies that $\frac{dv}{2t^2 + b^2 + a^2} = \frac{b^2 - a^2}{2} \frac{dw}{w}$. Denoting for simplification $r = \frac{b+a}{b-a} > 1$ we get

$$\theta = \frac{(b^2 - a^2)^{3/2}}{4} \int_{\sqrt{r}}^{w(m)} \frac{dw}{\sqrt{(w^2 - r)(w^2 - r^{-1})}}$$

where $m \mapsto w(m) > \sqrt{r}$ is defined by

$$m^2 = \frac{b^2 - a^2}{4w^2(m)} (w^2(m) - r)(w^2(m) - r^{-1}).$$

Now for simplification let us assume that $\frac{(b^2 - a^2)^{3/2}}{4} = 1$. Introduce the function $\theta \mapsto C(\theta)$ from $(0, \infty)$ to $(\sqrt{r}, \infty)$ defined by

$$\theta = \int_{\sqrt{r}}^{C(\theta)} \frac{dw}{\sqrt{(w^2 - r)(w^2 - r^{-1})}}.$$

Thus we have $w(m) = w(k'(\theta)) = C(\theta)$. This function C satisfies the differential equation

$$C' = \sqrt{(C^2 - r)(C^2 - r^{-1})}, \quad \frac{C'(\theta)}{C(\theta)} = \frac{1}{C(\theta)} \sqrt{(C^2(\theta) - r)(C^2(\theta) - r^{-1})} = \frac{2}{\sqrt{b^2 - a^2}} k'(\theta).$$

Thus the Laplace transform of μ is $L(\theta) = C(\theta)^c$ where $c = \frac{\sqrt{b^2 - a^2}}{2}$. We now imitate the procedure used in Theorem 4.1: we consider the Fourier transform $c(s) = C(is)$ for $s \in \mathbb{R}$, which satisfies

$$c'(s)^2 = -(c^2(s) - r)(c^2(s) - r^{-1}),$$

then $f(s) = c^2(s)$ which satisfies $f'^2 = 4c^2 c'^2 = -4f(f - r)(f - r^{-1})$, then $h(s) = \frac{1}{3}(r + r^{-1}) - f(s)$ which satisfies

$$\begin{aligned} h'(s)^2 &= 4h(s)^3 - g_2 h(s) - g_3 \\ &= 4[h(s) - \frac{1}{3}(r + r^{-1})][h(s) - \frac{1}{3}(r - 2r^{-1})][h(s) - \frac{1}{3}(r^{-1} - 2r)] \\ &= 4(h(s) - e_1)(h(s) - e_2)(h(s) - e_3) \end{aligned}$$

with $g_2 = \frac{4}{3}(r^2 + r^{-2} - 1)$, $g_3 = \frac{4}{27}(3r + 3r^{-1} - 2r^3 - 2r^{-3})$ and

$$e_1 = \frac{1}{3}(r + r^{-1}) > e_2 = \frac{1}{3}(r - 2r^{-1}) > e_3 = \frac{1}{3}(r^{-1} - 2r)$$

Hence for some complex constant C we have $h(s) = \wp(s + C)$ with periods $2K$ and $2iK'$ defined by

$$K = \int_{e_1}^{\infty} \frac{dx}{\sqrt{4x^3 - g_2x - g_3}}, \quad K' = \int_{-\infty}^{e_3} \frac{dx}{\sqrt{-4x^3 + g_2x + g_3}} \quad (25)$$

(see Whittaker and Watson [17] Example 1 p. 444). Now to determine the constant C one observes that $f(s) = \frac{1}{3}(r + r^{-1}) - \wp(s + C)$ is real since this is the Fourier transform of a symmetric measure. Furthermore $f(0) = c(0)^2 = r$. Thus $\wp(C) = \frac{1}{3}(r^{-1} - 2r) = e_3$ which implies $C = iK'$. Now we use the formula (see SG formula 5.8–22)

$$\sqrt{\wp(z) - e_j} = \frac{\pi}{2K} \left[\frac{\vartheta_1'(0)\vartheta_{j+1}(\frac{z}{2K})}{\vartheta_{j+1}(0)\vartheta_1(\frac{z}{2K})} \right]$$

that we shall use for writing

$$c^2(s) = f(s) = \frac{1}{3}(r + r^{-1}) - h(s) = e_1 - \wp(s + iK') = \frac{\pi^2}{4K^2} \left| \frac{\vartheta_1'(0)\vartheta_2(\frac{s}{2K} + i\frac{K'}{2K})}{\vartheta_2(0)\vartheta_1(\frac{s}{2K} + i\frac{K'}{2K})} \right|^2.$$

Let us introduce the notation $q = e^{-\pi K/K'}$. With it, ϑ_1 and ϑ_2 are given by

$$\begin{aligned} \vartheta_1(z) &= 2Cq^{1/4} \sin \pi z \prod_{\nu=1}^{\infty} (1 - 2q^{2\nu} \cos 2\pi z + q^{4\nu}) \\ \vartheta_2(z) &= 2Cq^{1/4} \cos \pi z \prod_{\nu=1}^{\infty} (1 + 2q^{2\nu} \cos 2\pi z + q^{4\nu}) \end{aligned}$$

where $C = \prod_{\nu=1}^{\infty} (1 - q^{2\nu})$ (see SG pages 268–9). Let us give a simpler presentation of $c^2(s)$: using $z = \frac{s}{2K} + i\frac{K'}{2K}$ and $u = e^{\frac{i\pi s}{2K}}$ we introduce the following symbols for $\nu = 1, 2, \dots$

$$\varphi_\nu(u) = \frac{u^4 + u^{-4}}{q^{2\nu} + q^{-2\nu} + q^2 + q^{-2}}, \quad \varphi_0(u) = \frac{u^2 + u^{-2}}{q + q^{-1}}.$$

We get

$$\left| \frac{\cos \pi z}{\sin \pi z} \right|^2 = \frac{1 + \varphi_0(u)}{1 - \varphi_0(u)}, \quad \left| \frac{1 + 2q^{2\nu} \cos 2\pi z + q^{4\nu}}{1 - 2q^{2\nu} \cos 2\pi z + q^{4\nu}} \right|^2 = \frac{1 + \varphi_\nu(u)}{1 - \varphi_\nu(u)}.$$

and finally the elegant formula

$$c^2(s) = f(s) = C \prod_{\nu=0}^{\infty} \frac{1 + \varphi_\nu(u)}{1 - \varphi_\nu(u)}$$

where the constant C is such that $f(0) = 1$. The last step is the formula, correct for $|Z| < 1$:

$$\frac{1+Z}{1-Z} = 1 + 2 \sum_{n=1}^{\infty} Z^n = \sum_{n \in \mathbb{Z}} Z^{|n|}.$$

Replacing Z by $\varphi_v(u)$ we see that $f_v(s) = C_v \frac{1+\varphi_v(u)}{1-\varphi_v(u)}$ where C_v is such that $f_v(0) = 1$ is the characteristic function of a probability distribution concentrated on the additive group $\frac{2\pi}{K}\mathbb{Z}$ for $v \geq 1$ and on the additive group $\frac{\pi}{K}\mathbb{Z}$ for $v = 0$. As a result f is the characteristic function of a symmetric discrete distribution on the group $\frac{\pi}{K}\mathbb{Z}$.

Comments. Of course the restriction $b^2 - a^2 = 2^{4/3}$ is not important and can be generalized by a dilation. In the other hand, finding the Jørgensen set of these families is a difficult question. It should also be mentioned that the characteristic function $f_v(s)$ above has the form $\frac{1-c_v}{1-c_v} \frac{1+c_v \cos s}{1-c_v \cos s}$ after dilation $x \mapsto \frac{2\pi}{K}x$ if $v \neq 0$ or $x \mapsto \frac{\pi}{K}x$. If the Poisson kernel distribution on $\mathbb{Z}$ of parameter $r \in (0, 1)$ is defined by $p_n = \frac{1-r}{1+r} r^{|n|}$, then $\frac{1-c_v}{1-c_v} \frac{1+c_v \cos s}{1-c_v \cos s}$ can be seen as a mixing of a Dirac mass on zero and of Poisson kernel distribution with parameter $r = c_v$.

8 The Family F

Theorem 8.1 *Let $x > 0$. The NEF F_x with domain of the means $(0, \infty)$ and variance function*

$$V_{F_x}(m) = m(1 + \frac{4m^4}{x^4})^{1/2}$$

is generated by a positive measure on $\mathbb{N}$ which is $v_x(dt) = \sum_{n=0}^{\infty} \frac{p_n(x)}{n!} \delta_n(dt)$ with generating function

$$f_x(z) = \sum_{n=0}^{\infty} \frac{p_n(x)}{n!} z^n = e^{x \int_0^z \frac{dw}{(1-w^4)^{1/2}}}. \quad (26)$$

which satisfies

$$(1 - z^4)f_x''(z) - 2z^3f_x'(z) - x^2f_x(z) = 0.$$

The total mass of v_x is $\exp(x \frac{1}{4} B(\frac{1}{2}, \frac{1}{4}))$. The polynomials p_n are given by $p_n(x) = x^n$ for $n = 0, 1, 2, 3, 4$, $p_5(x) = x^5 + 12x$ and for $n \geq 2$

$$p_{n+2}(x) = x^2 p_n(x) + n(n-1)^2(n-2)p_{n-2}(x).$$

Proof The proof of the first formula is a routine calculation for exponential families concentrated on $\mathbb{N}$, but we give details. We use successively the change of variables $u = 2m^2/x^2$ and $u = \sinh v$.

$$d\theta = \frac{dm}{V(m)} = \frac{4mdm}{4m^2\sqrt{1+4\frac{m^4}{x^4}}} = \frac{du}{u\sqrt{1+u^2}} = \frac{dv}{2\sinh v} = \frac{1}{2}\left(\frac{1}{e^v-1} - \frac{1}{e^v+1}\right)e^v dv$$

Denoting $z = e^\theta$ we get

$$z^2 = \frac{e^v - 1}{e^v + 1}, \quad e^v = \frac{1 + z^2}{1 - z^2}, \quad e^{-v} = \frac{1 - z^2}{1 + z^2}, \quad u = \sinh v = \frac{2z^2}{1 - z^4}, \quad m = \frac{xz}{\sqrt{1 - z^4}}.$$

Thus $k'_\mu(\theta) = x \frac{e^\theta}{\sqrt{1 - e^{4\theta}}}$ and this leads to the result (26).

The trick to obtain the differential equation for f_x is to write $(1 - z^4)^{1/2} f'_x = x f_x$, then to differentiate with respect to z and then to multiply both sides of the result by $(1 - z^4)^{1/2}$. Then the differential equation leads to the equality

$$\begin{aligned} \sum_{n \geq 0} (n+1)(n+2) \frac{p_{n+2}(x)}{(n+2)!} z^n - \sum_{n \geq 4} (n-3)(n-2) \frac{p_{n-2}(x)}{(n-2)!} z^n \\ - 2 \sum_{n \geq 3} (n-2) \frac{p_{n-2}(x)}{(n-2)!} z^n - x^2 \sum_{n \geq 0} \frac{p_n(x)}{n!} z^n = 0. \end{aligned}$$

Using $f_x(0) = p_0(x) = 1$ and $f'_x(0) = p_1(x) = x$ we get $p_n(x) = x^n$ for $0 \leq n \leq 4$ and for $n \geq 4$ we have

$$\begin{aligned} (n+1)(n+2) \frac{p_{n+2}(x)}{(n+2)!} - (n-3)(n-2) \frac{p_{n-2}(x)}{(n-2)!} - 2(n-2) \frac{p_{n-2}(x)}{(n-2)!} - x^2 \frac{p_n(x)}{n!} = 0, \\ (n+1)(n+2) \frac{p_{n+2}(x)}{(n+2)!} - (n-1)(n-2) \frac{p_{n-2}(x)}{(n-2)!} - x^2 \frac{p_n(x)}{n!} = 0. \end{aligned}$$

Now we multiply both sides by $n!$ and we use the definition of p_n for getting

$$p_{n+2}(x) = x^2 p_n(x) + n(n-1)^2 (n-2) p_{n-2}(x).$$

Checking the correctness of this equality for $n = 2, 3$ is easy.

Remarks It is easy to check that if $n = 4q + r$ with $r = 0, 1, 2, 3$ then there exists a monic polynomial $P_{q,r}$ of degree q such that

$$p_n(x) = x^r P_{q,r}(x^4).$$

For instance $P_{0,r}(z) = 1$, $P_{1,0}(z) = z$, $P_{1,1}(z) = z + 12$, $P_{1,2}(z) = z + 72$, $P_{1,3}(z) = z + 252$, $P_{2,0}(z) = z^2 + 672z$, $P_{2,1}(z) = z^2 + 1512z + 1260$.

We now extend Theorem 8.1 to a more general variance function, without being so specific about calculation of the corresponding distribution. This variance function for $x = 1$ is the reciprocal variance function for $t = 1$ of V_{G_t} where

$$V_{G_t}(m) = \sqrt{1 + 2p \frac{m^2}{t^2} + (2-p)^2 \frac{m^4}{t^4}}.$$

Theorem 8.2 *Let $p \in [0, 1)$. Let $x > 0$. The NEF F_x with domain of the means $(0, \infty)$ and variance function*

$$V_{F_x}(m) = m \left[1 + 2p \frac{m^2}{x^2} + \frac{(2-p)^2 m^4}{x^4} \right]^{1/2}$$

is generated by a positive measure on $\mathbb{N}$ with generating function

$$f_x(z) = \exp \left[x \sqrt{\frac{2}{2-p}} \int_0^z (1 + qw^2)^{1/2} (1 - w^4)^{-1/2} dw \right]. \quad (27)$$

where $q = p/(2-p)^2$.

Proof It is convenient to denote $c = p/(2-p)$ and to observe that

$$0 \leq c < 1, \quad \sqrt{1-c^2} = \frac{2\sqrt{1-p}}{2-p}, \quad \frac{c \pm 1}{\sqrt{1-c^2}} = \pm \frac{1}{\sqrt{1-p}}.$$

We use successively the change of variables $u = (2-p)m^2/x^2$ and $u = \sqrt{1-c^2} \sinh v - c$.

$$\begin{aligned} d\theta &= \frac{dm}{V(m)} = \frac{4mdm}{4m^2 \sqrt{1 + 2p \frac{m^2}{x^2} + \frac{(2-p)^2 m^4}{x^4}}} = \frac{du}{u \sqrt{1 + 2cu + u^2}} = \frac{dv}{2\sqrt{1-c^2} \sinh v - 2c} \\ &= \frac{e^v dv}{\sqrt{1-c^2} e^{2v} - 2c e^v - \sqrt{1-c^2}} = \frac{1}{2} \left[\frac{1}{e^v - \frac{1}{\sqrt{1-p}}} - \frac{1}{e^v + \frac{1}{\sqrt{1-p}}} \right] e^v dv. \end{aligned}$$

Denoting $z = e^\theta$ we get

$$z^2 = \frac{e^v - \frac{1}{\sqrt{1-p}}}{e^v + \frac{1}{\sqrt{1-p}}}, \quad e^v = \frac{1}{\sqrt{1-p}} \frac{1+z^2}{1-z^2}, \quad e^{-v} = \sqrt{1-p} \frac{1-z^2}{1+z^2}$$

$$\sqrt{1-c^2} \sinh v = \frac{4z^2 + p(1-z^2)^2}{(2-p)(1-z^4)}, \quad u = \frac{2z^2}{1-z^4} (1 + cz^2), \quad m^2 = \frac{2}{2-p} \frac{z^2}{1-z^4} (1 + qz^2).$$

Thus

$$m = k'_\mu(\theta) = x \sqrt{\frac{2}{2-p}} \frac{(1 + qe^{2\theta})^{1/2}}{(1 - e^{4\theta})^{1/2}} e^\theta$$

and this leads to the result (27).

It remains to prove that the Taylor expansion of $z \mapsto f_x(z)$ defined by (27) has positive coefficients. For this it is enough to prove that the argument of the exponential

$$z \mapsto \int_0^z (1 + qw^2)^{1/2} (1 - w^4)^{-1/2} dw$$

has positive coefficients. It is enough to prove that $z \mapsto (1 + qz^2)^{1/2} (1 - z^4)^{-1/2}$ has positive coefficients. It is enough to prove that $z \mapsto (1 + qz)^{1/2} (1 - z^2)^{-1/2}$ has positive coefficients. It is enough to prove that

$$z \mapsto \log[(1 + qz)(1 - z^2)^{-1}] = \sum_{n=1}^{\infty} a_n z^n$$

has positive coefficients. But this very last point is easy to check since $0 \leq q < 1$ and since a_n is computable: for odd n then $a_n = q^n/n > 0$ and for even $n = 2p$ we have

$$a_n = \frac{1}{p} - \frac{q^{2p}}{2p} > 0.$$

The theorem is proved.

9 Conclusion: General Elliptic Variances

It seems that the present paper is only scratching the surface of an interesting theory. Indeed, consider the set of variance functions of the form

$$V_F(m) = (\alpha m + \beta) \sqrt{P(m)} \quad (28)$$

where P is a polynomial with degree ≤ 4 . The present paper has considered only the cases $P(m) = Am^4 + Bm^2 + C$. Recall a definition appearing in Hassairi [18] and Barlev et al. [19]. We say that two NEF F_1 and F_2 on the real line belong to the same orbit if there exists a Moebius transformation $y = (ax + b)/(cx + d)$ such that $ad - bc = 1$ and such that on a suitable interval for m we have

$$V_{F_1}(m) = (cm + d)^3 V_{F_2} \left(\frac{am + b}{cm + d} \right).$$

The most celebrated pair (F_1, F_2) is the set of normal distributions with variance 1 and the set of inverse Gaussian distributions, with variance m^3 on $(0, \infty)$. The pair $(\sqrt{4 + m^4}, m\sqrt{1 + 4m^4})$ offers another example. Roughly saying that F_1 and F_2 belong to the same orbit means the following: suppose that F_1 and F_2 are generated by μ_1 and μ_2 and let us draw in $\mathbb{R}^2$ the curves C_1 and C_2 which are the representative curves of the convex functions k_{μ_1} and k_{μ_2} (in the case of the pair normal-inverse

Gaussian, they are a parabola and a piece of parabola). Then F_1 and F_2 are in the same orbit if and only if there exists an affine transformation of the plane $\mathbb{R}^2$ which maps a part of C_1 onto a part of C_2 . This affine transformation can be described in terms of the coefficients (a, b, c, d) of the Moebius transformation. A very satisfactory fact observed in Hassairi [18] is that the quadratic and cubic NEF are split in 4 orbits, respectively generated by the normal, the Poisson, the hyperbolic and the Ressel-Kendall distributions. Now we remark that if F has the form (28) and if G is in the orbit of F then necessarily $V_G(m) = (\alpha_1 m + \beta_1) \sqrt{P_1(m)}$ where the polynomial P_1 has also degree ≤ 4 . Therefore we are facing the problem of a whole classification of this set (28) of variance functions into orbits. This implies a mastering of the elliptic curves $y^2 = P(x)$ and the use of beautiful mathematics. The theory of exponential families expanded by Ole Barndorff Nielsen 40 years ago is still hiding many secrets.

References

1. Letac, G.: The classification of the natural exponential families by their variance functions. Proc. ISI Cairo Congr. **3**, 271–298 (1991)
2. Morris, C.N.: Natural exponential families with quadratic variance functions. Ann. Statist. **10**, 65–80 (1982)
3. Jørgensen, B.: Exponential dispersion models. J. Roy. Statist. Soc. Ser. B **49**, 127–162 (1987)
4. Tweedie, M.C.K.: An index which distinguishes between some important exponential families. In: Ghosh, J.K., Roy, J. (eds.) Statistics: Applications and New Directions. Proceedings Indian Institute Golden Jubilee International Conference, 579–604. Indian Statistical Institute, Calcutta (1984)
5. Letac, G., Mora, M.: Natural exponential families with cubic variances. Ann. Statist. **18**, 1–37 (1990)
6. Jørgensen, B.: The Theory of Dispersion Models. Chapman and Hall, London (1997)
7. Letac, G.: Lectures on natural exponential families and their variance functions. 128 pages, Monografias de matemática, **50**, IMPA, Rio de Janeiro (1992)
8. Bar-Lev, S.K., Van de Duyn Shouten, F.A.: A note on exponential dispersion models which are invariant under length-biased sampling. Statist. Probab. Lett. **70**, 275–284 (2004)
9. Sansone, G., Gerretsen, J.: Lectures on the Theory of Functions of a Complex variable, vol. 1. P. Noordhof Ltd, Groningen (1960)
10. Barndorff-Nielsen, O.: Information and Exponential Families in Statistical Theory. Wiley, New York (1978)
11. Letac, G., Malouche, D., Maurer, S.: The real powers of the convolution of a negative binomial and a Bernoulli distribution. Proc. Amer. Math. Soc. **130**, 2107–2114 (2002)
12. Jørgensen, B., Martinez, J.R., Tsao, M.: Asymptotic behavior of the variance function. Scand. J. Statist. **21**, 223–243 (1993)
13. Kawata, T.: Fourier Analysis in Probability Theory. Academic Press, New York (1972)
14. Shanbhag, D.N.: Diagonality of the Bhattacharyya matrix as a characterization. Theo. Prob. Appl. **23**, 430–433 (1979)
15. Bar-Lev, S.K., Letac, G.: Increasing hasard rate of mixtures for natural exponential families. Adv. Appl. Probab. **44**, 373–390 (2012)
16. Gradshteyn, I.S., Ryzhik, I.M.: Table of Integrals, Series and Products. In: Jeffrey, A., Zwillinger, D. (eds.), 7th edn. Academic Press, New York (2007)
17. Whittaker, E.T., Watson, G.N.: A Course in Modern Analysis, 4th edn. Cambridge University Press (1927)

18. Hassairi, A.: La classification des familles exponentielles naturelles sur $\mathbb{R}^n$ par l'action du groupe linéaire de $\mathbb{R}^{n+1}$. C.R. Acad. Sci. Paris Série I. **315**, 207–210 (1992)
19. Bar-Lev, S.K., Enis, P., Letac, G.: Sampling models which admit a given exponential family as conjugate family of priors. Ann. Statist. **22**, 1555–1586 (1994)
20. Brown, L.D.: Fundamentals of Statistical Exponential Families with Applications in Statistical Decision Theory. Institute of Mathematical Statistics, Hayward (1978)

Cumulants and Bartlett Identities in Cox Regression

Per Aslak Mykland and Jianming Ye

Abstract Expressions are found for the cumulants needed to analyze and improve the accuracy of Cox regression up to order $O(n^{-3/2})$, and consistent estimators for these are given. In addition, the Bartlett identities are shown to hold for the Cox partial likelihood (and, in a broad sense, for partial likelihood in general), so that the cumulant estimators found can be used to adjust statistics in accordance with formulas from likelihood theory. Numerical results are also given.

Keywords Bartlett identities · Bartlett correction · Partial likelihood · Proportional hazards model · Survival analysis

1 Introduction

In statistics, *accuracy* refers to the closeness of an (often asymptotic) approximation to the true distribution of a statistic. Frequently, accuracy can be improved by adjusting normal approximations with the help of expansions, saddlepoint approximations, and ancillary statistics. Good accuracy is crucial for correct inference.¹

¹Accuracy is sometimes confused with efficiency. The two are not the same, and both are important. In the case of unconditional inference, bootstrapping [11, 15] is also a popular approach to the accuracy problem, see [21] and the references therein.

P.A. Mykland (✉)
Department of Statistics, The University of Chicago, 5734 University Avenue,
Chicago 60637, USA
e-mail: mykland@pascal.uchicago.edu

J. Ye
Zicklin School of Business, Baruch College, One Bernard Baruch Way,
New York 10010, USA
e-mail: jianmng.ye@baruch.cuny.edu

J. Ye
School of Management, Xiamen University,
Baoxinliying Building, Xiamen 361005, China

The study of accuracy is one of the main contributions of Ole Barndorff-Nielsen. His work stands out with creative approaches like the p^* formula [3] and the r and r^* statistics [4, 5], and with thoughtful views on the philosophy of statistical inference, as in, for example, [6].

One path to improved accuracy is to start with the an asymptotic (limiting) distribution, and to modify it with the help of Edgeworth-type correction terms or factors. In the case of likelihood inference, this approach involves the Bartlett identities for log likelihood derivatives [9, 10], and the ensuing Bartlett correction factor for the likelihood ratio statistic and its signed square root [7, 8, 28] (see Sect. 2).

The purpose of the current paper is to show that the same approach works for partial likelihood inference [14, 40], and specifically in Cox regression, and to operationalize the finding with an estimator for the correction factor.

Inference in Cox regression [1, 13, 14, 39], see also [2, 16], is usually based on asymptotic (limiting) distributions. There is thus a *prima facie* case for trying the correction methods that work in ordinary likelihood.

For the Cox model, this issue has been addressed in [18, 19, 22, 30, 32]. Our aim in this paper is to continue this research by finding expressions for the cumulants needed to describe the distributions of estimators and test statistics up to order $O(n^{-3/2})$. This provides a formula for the Bartlett correction factor. The cumulants will be given in a form that lends itself to estimation by empirical substitution, and we also give the resulting estimators.

We also show that the critical Bartlett identities [9, 10] hold quite generally for the Cox model (Theorem 1), thus enabling also other likelihood properties, such as those discussed in [33].

In the following, we begin by explaining the Bartlett identities for (ordinary) likelihood, and their implications for the Bartlett correction the likelihood ratio statistic (Sect. 2). We then discuss the question of Bartlett identities for partial likelihoods (Sect. 3). The Cox model and partial likelihood is reviewed in Sect. 4, with theorems about Bartlett identities and correction for this criterion function. We then present expressions and estimators for the cumulants (Sects. 5 and 6). A few technical arguments are gathered in an appendix.

2 The Bartlett Identities, the Bartlett Factor, and Other Likelihood Quantities

Let $L(\theta, t)$ be a log likelihood function, and set

$$L_{\alpha_1, \dots, \alpha_p}(\theta, t) = \partial^p L(\theta, t) / \partial \theta^{\alpha_1} \dots \partial \theta^{\alpha_p} \quad (1)$$

to the extent that the derivatives exist. We drop θ as an argument when dealing with the true value (in the notation only, the log likelihood derivatives depend on θ).

Since $\exp\{L(\theta, T) - L(\theta_0, T)\}$ is a likelihood ratio (Radon-Nikodym derivative), it follows that $E \exp\{L(\theta, T) - L(\theta_0, T)\} = 1$. Assuming that one can interchange differentiation and expectation, one can thus take derivatives with respect to θ_α , obtaining $E_\theta L_\alpha(\theta, T) \exp\{L(\theta, T) - L(\theta_0, T)\} = 0$, $E_\theta (L_{\alpha_1, \alpha_2}(\theta, T) + L_{\alpha_1}(\theta, T) L_{\alpha_2}(\theta, T)) \exp\{L(\theta, T) - L(\theta_0, T)\} = 0$, and so on. Setting $\theta = \theta_0$, we obtain $EL_\alpha(T) = 0$ (meaning $E_{\theta_0} L_{\alpha_1}(\theta_0, T) = 0$), $EL_{\alpha_1, \alpha_2}(T) + EL_{\alpha_1}(T)L_{\alpha_2}(T) = 0$. These two identities are crucial for the central limit theorem for the maximum likelihood estimator (MLE). One can, however, continue the process, obtaining

$$\begin{aligned} EL_{\alpha_1, \alpha_2, \alpha_3}(T) + EL_{\alpha_1, \alpha_2}(T)L_{\alpha_3}(T)[3] + EL_{\alpha_1}(T)L_{\alpha_2}(T)L_{\alpha_3}(T) &= 0, \\ EL_{\alpha_1, \alpha_2, \alpha_3, \alpha_4}(T) + EL_{\alpha_1, \alpha_2, \alpha_3}(T)L_{\alpha_4}(T)[4] + EL_{\alpha_1, \alpha_2}(T)L_{\alpha_3, \alpha_4}(T)[3] \\ + EL_{\alpha_1, \alpha_2}L_{\alpha_3}(T)L_{\alpha_4}(T)[6] + EL_{\alpha_1}(T)L_{\alpha_2}(T)L_{\alpha_3}(T)L_{\alpha_4}(T) &= 0, \quad (2) \end{aligned}$$

and so on by taking further derivatives of $E \exp\{L(\theta, T) - L(\theta_0, T)\} = 1$ with respect to θ_α s. The “[3]” in the first line of (2) means that we sum over the three terms that arise when recombining α_1 , α_2 and α_3 , and similarly for “[3]”, “[4]”, and “[6]” in the second and third line. These are the four first Bartlett identities [9, 10]. The “[.]” notation, and the identities are further discussed in [28]. The Bartlett identities extend to martingales and can heuristically be derived from the likelihood identities [29].

By a combinatorial argument, the Bartlett identities also hold for *cumulants* [28]. Cumulants are moment-like objects that are additive for independent sums. See Chap. 2 of [28] for a description and discussion. The second cumulant, for example, is the covariance. The cumulant concept seems to have been invented independently by [38] (translated to English in the Appendix of [20]) and [17], see, for example, [20, 27, 37].

To write the Bartlett identities on cumulant form, one simply replaces moments by their corresponding cumulants. If we let “cum” denote cumulants (under probability P_{θ_0}), then cumulants κ of $L_{\alpha_1, \dots, \alpha_p}(T)$ are defined as in [28], Chap. 7.2; for example, $\kappa_{\alpha, \beta} = \text{Cov}(L_\alpha(T), L_\beta(T))$, $\kappa_{\alpha\beta, \gamma\delta} = \text{Cov}(L_{\alpha\beta}(T), L_{\gamma\delta}(T))$ and $\kappa_{\alpha, \beta, \gamma\delta} = \text{cum}(L_\alpha(T), L_\beta(T), L_{\gamma\delta}(T))$. The cumulant form of the Bartlett identities then becomes

$$\begin{aligned} \kappa_\alpha &= 0, \\ \kappa_{\alpha_1 \alpha_2} + \kappa_{\alpha_1, \alpha_2} &= 0, \\ \kappa_{\alpha_1 \alpha_2 \alpha_3} + \kappa_{\alpha_1 \alpha_2, \alpha_3}[3] + \kappa_{\alpha_1, \alpha_2, \alpha_3} &= 0, \\ \kappa_{\alpha_1 \alpha_2 \alpha_3 \alpha_4} + \kappa_{\alpha_1 \alpha_2 \alpha_3, \alpha_4}[4] + \kappa_{\alpha_1 \alpha_2, \alpha_3 \alpha_4}[3] + \kappa_{\alpha_1 \alpha_2, \alpha_3, \alpha_4}[6] + \kappa_{\alpha_1, \alpha_2, \alpha_3, \alpha_4} &= 0, \quad (3) \end{aligned}$$

and so on.

The first four Bartlett identities yield the following. In the absence of nuisance parameters, the likelihood ratio statistic $w = 2(L(\hat{\theta}, T) - L(\theta, T))$ has a Bartlett correction factor [7, 8, 28] $b(\theta)$ given by (using the summation convention)

$$\begin{aligned}
npb(\theta) = & (\kappa_{\alpha\beta\gamma\delta} + 4\kappa_{\alpha,\beta\gamma\delta} + 4(\kappa_{\alpha,\gamma,\beta\delta} + \kappa_{\alpha\gamma,\beta\delta}))\kappa^{\alpha,\beta}\kappa^{\gamma,\delta}/4 \\
& + \lambda_\alpha\lambda_\beta\kappa^{\alpha,\beta}/4 \\
& + (\kappa_{\alpha\beta\gamma}\kappa_{\delta\epsilon\zeta} + 6\kappa_{\alpha\beta,\gamma}\kappa_{\delta\epsilon,\zeta})\kappa^{\alpha,\delta}\kappa^{\beta,\epsilon}\kappa^{\gamma,\zeta}/6
\end{aligned}$$

where $\lambda_\alpha = -(2\kappa_{\alpha\beta,\gamma} + \kappa_{\alpha\beta\gamma})\kappa^{\beta,\gamma}$ (cf. Chap. 7 of [28]). We have here put the expression in a form where it can immediately be estimated by the methods developed in the previous section.

The actual corrected likelihood ratio statistic is $w/(1 + b(\theta)/n)$, which is well approximated by χ_p^2 under a variety of regularity conditions.

The improvement remains strong in the tails (as compared to usual Edgeworth expansion based results). This is presumably due to the connection between Bartlett correction and the saddlepoint approximation, [8, 35]. Similar results (both additive and multiplicative) apply to the signed square root statistic r [7, 8], see also [28].

It should be emphasized that the Bartlett factor is not the only quantity that can be found from the Bartlett identities. Higher order identities (beyond four) give rise to improved approximation in the tails [33]. This can also be obtained by the elegant r^* approach developed by [4, 5, 26] and others. For the connection of this work to the Bartlett identities, see [34].

Also, for example, the expressions for bias, variance and third cumulant given on p. 209 in [28] carry over directly to any criterion function that satisfies the Bartlett identities. In our notation,

$$E\hat{\theta}^\alpha = \theta - \frac{1}{n} \sum_{\beta,\gamma,\delta} \kappa^{\alpha,\beta}\kappa^{\alpha,\delta}(\kappa_{\beta,\gamma,\delta} + \kappa_{\beta,\gamma\delta})/2 + O(n^{-3/2}),$$

and so on ($\kappa^{\alpha,\beta}$ is the inverse of $\kappa_{\alpha,\beta}$).

3 Bartlett Identities for Partial Likelihood

The Bartlett identities hold quite generally in partial likelihood problems. To see this, let $\theta = (\theta^1, \theta^2, \dots)$ be the parameter appearing in the partial likelihood, and let θ_0 be the true value of θ . Also let P be the true probability distribution. Set up a dummy inference problem by $Q_{\theta_0} = P$, $dQ_\theta/dQ_{\theta_0} =$ the partial likelihood ratio between θ and θ_0 . The Bartlett identities are then satisfied (subject to regularity conditions) for the derivatives of the log partial likelihood provided Q_θ is a probability measure, i.e., if the partial likelihood ratio integrates to 1 w.r.t. P .

This is obviously true for the partial likelihood formulation in [14, 40]. It is also true for the counting process formulation of the Cox model with which we shall be working in view of Theorem II.7 (p. 93) of [2], see also [23].

It should be emphasized that there is no need for Q_θ to be the actual distribution associated with regression coefficients θ . Thus, for example, there is no need for any

assumption that the censoring is noninformative (in the sense of, e.g., Definition 4.3.1 (p. 139) of [16]). If censoring is informative, one does, of course, lose efficiency by using the partial likelihood, but one does not lose the Bartlett identities.

A conceptually important point relating to the Bartlett identities is that there are two sets of them. In addition to the ordinary ones discussed above, there are also conditional variation identities. For example, if L is the partial log likelihood in a one parameter problem, $\dot{L}$, $[\dot{L}, \dot{L}] + \ddot{L}$, $[\dot{L}, \dot{L}, \dot{L}] + 3[\ddot{L}, \dot{L}] + \ddot{\ddot{L}}$, etc., are martingales (the $[\cdot, \cdot]$ -notation is as in stochastic calculus (cf. [24], for example); the higher order $[\cdot]$ -notation is discussed in [29]). These latter identities do not require the setting up of a dummy inference problem.

The two first conditional variation identities are, of course, at the heart of partial likelihood [13, 14, 40]. Since $\text{Var } \dot{L} = E[\dot{L}, \dot{L}]$, they also relate closely to the two first ordinary Bartlett identities. We make the conjecture that a similar relationship also holds for higher order identities, by exploiting the Bartlett identities for martingales [29]. We have not explored this matter further, however, as the current Theorem 1 follows without any technicalities whatsoever.

4 The Cox Model

Following [1], the model can be described by point processes $N^1(t), \dots, N^n(t)$, $0 \leq t \leq T < \infty$, where $N^i(t)$ jumps from 0 to 1 when patient no. i dies, provided he or she does so under observation. $N^i(t)$ has intensity $\lambda_0(t)\lambda^i(\theta, t)$, where

$$\lambda^i(\theta, t) = Y_i(t) \exp\left(\sum_{\alpha} \theta^{\alpha} Z_{\alpha,i}(t)\right),$$

where $Y_i(t)$ is a predictable process which is 0 or 1 according to whether the i 'th patient is under observation, $Z_{\alpha,i}(t)$ is a predictable covariate process, θ^{α} is an unknown parameter, and $\lambda_0(t)$ is a baseline intensity. All remaining patients are assumed to be censored at the nonrandom time T .

The partial log likelihood at time t is given by

$$L(\theta, t) = \sum_{\alpha,i} \int_0^t \theta^{\alpha} Z_{\alpha,i}(s) dN^i(s) - \int_0^t f(\theta, s) d\bar{N}(s)$$

where $\bar{N} = N^1 + \dots + N^n$, $\bar{\lambda} = \lambda^1 + \dots + \lambda^n$ and $f(\theta, t) = \ln \bar{\lambda}(\theta, t) + \ln \lambda_0(t)$ when $\bar{\lambda}(\theta, t) \neq 0$, and $= 0$ otherwise. Let $f_{\alpha_1, \dots, \alpha_p}(\theta, t)$ and $\bar{\lambda}_{\alpha_1, \dots, \alpha_p}(\theta, t)$ be derivatives with respect to components of θ , defined in analogy with (1).

Note that the way we define it, the partial log likelihood does depend on the baseline hazard $\lambda_0(t)$. This is done for notational convenience. However, the dependence

on λ_0 washes out both in the likelihood ratio statistic and in the derivatives of $L(\theta, t)$. The same occurs in the estimators of the cumulants (formula (7) below).

Expressions and estimators for the cumulants are particularly useful because of the following fact (explained in Sect. 3):

Theorem 1 *Assume that the covariates are bounded, and that $\int_0^T \lambda_0(t)dt < \infty$. Then the Cox partial likelihood satisfies the Bartlett identities.*

The importance of this theorem is that there are a number of results relating to likelihood inference which depend only on the likelihood structure through the Bartlett identities, as discussed in Sect. 2.

In the following, the cumulants are found under the assumption that patients are independent of each other. For a given patient, however, censoring can be arbitrarily dependent on the history of the patient, and covariates can be random (predictable) and time varying. Specifically, we shall be using the framework of [1].

Condition 1 *We shall assume that $\lambda_0(t)$ is non-random and that if $\mathcal{F}_i = \sigma(Z_{\alpha,i}(t), Y_i(t), N^i(t), \text{all } \alpha, t)$, then the $\mathcal{F}_i$ are independent for all i (i.e., the patients are independent). An implication of this assumption is that a decision to censor one patient cannot be dependent on what happens to another patient. A patient can, however, be censored on basis of his or her own medical history.*

In addition, we use the technical assumption that the (absolute value of the) covariates are bounded, and that there are nonrandom constants λ_0^- and λ_0^+ so that $0 < \lambda_0^- \leq \lambda_0(t) \leq \lambda_0^+ < \infty$.

In order to show that these corrections actually do improve the asymptotic accuracy of rejection and coverage probabilities, one also needs to show the existence of relevant Edgeworth expansions. We do not do this in this paper, but a result to this effect is stated in [19]. Alternatively, we can use the results in [31], which can be extended to likelihood ratio statistics in much the same way as the proof of Theorem 2 in [30]. The conditions in [19] are stronger than in [30], but so are the conclusions (in terms of topology of convergence).

To state a specific result, we shall borrow the expansion of [19].

Theorem 2 *Assume Condition 1, and also that there is only one covariate; that the patients (in terms of death, censoring and covariate) are i.i.d.; and that for each patient, death and censoring are conditionally independent given the covariate. Then the distribution of the Bartlett corrected Cox partial likelihood ratio statistic (the Bartlett factor being given in Sect. 2) is $\chi_1^2 + O(n^{-3/2})$.*

Note that the additional conditions imposed relative to those in Sect. 4 is what is needed to make the result in [19] hold, cf. Sect. 1 of that paper. The result follows from [19] and (our) Theorem 1 by the same method as used in [12, 25].

In a sense, the results in this paper are orthogonal to those of [19, 31]. Whereas these two papers provide conditions for expansions to exist, the issue here is what the expansions look like once they have been shown to exist. For example, it is clear from our results that the coefficient κ_4 in [19] is zero (cf. the proof of Theorem 3).

5 Cumulants for Partial Likelihood

Our basic result is the following theorem giving expressions for such cumulants of degree up to 4. In writing these expressions, we use v to denote any unpartitioned index set of more than one element (i.e., $\{\beta\gamma\}$, $\{\beta\gamma\delta\}$, etc.). Also, (i) means that patient no. i has been removed from the computation of the expression, so that, e.g., $f_{\alpha}^{(i)}$ denotes f_{α} calculated on the basis of $n - 1$ patients with patient i excluded.

Theorem 3 *Under Condition 1,*

$$\begin{aligned}\kappa_v &= -E \int_0^T f_v(t) \bar{\lambda}(t) \lambda_0(t) dt \text{ when } |v| \geq 2, \text{ and } = 0 \text{ when } |v| = 1, \\ \kappa_{\alpha,v} &= -\sum_i E \int_0^T f_v^{(i)}(t) \bar{\lambda}^{(i)}(t) \lambda_0(t) dt \int_0^t (Z_{\alpha,i}(s) - f_{\alpha}(s)) \lambda^i(s) \lambda_0(s) ds, \\ \kappa_{\alpha,\beta,\gamma\delta} + \kappa_{\alpha\beta,\gamma\delta} &= \sum_i E \left(\int_0^T f_{\gamma\delta}^{(i)}(t) \bar{\lambda}^{(i)}(t) \lambda_0(t) dt \right. \\ &\quad \left. \times \int_0^t [f_{\alpha\beta}(s) - (Z_{\alpha,i}(s) - f_{\alpha}(s))(Z_{\beta,i}(s) - f_{\beta}(s))] \lambda^i(s) \lambda_0(s) ds \right),\end{aligned}$$

and

$$\begin{aligned}\kappa_{\alpha\beta,\gamma\delta} &= E \int_0^T f_{\alpha\beta}(t) f_{\gamma\delta}(t) \bar{\lambda}(t) \lambda_0(t) dt \\ &\quad + \sum_i E \int_0^T \left[f_{\alpha\beta}^{(i)}(t) \bar{\lambda}^{(i)}(t) \lambda_0(t) f_{\alpha\beta}(t) \bar{\lambda}(t) \right] dt \int_0^t f_{\gamma\delta}(s) \lambda^i(s) \lambda_0(s) ds \quad [2] \\ &\quad + \int_0^T dt \int_0^T \text{Cov}(f_{\alpha\beta}(t) \bar{\lambda}(t) \lambda_0(t), f_{\gamma\delta}(s) \bar{\lambda}(s) \lambda_0(s)) ds.\end{aligned}$$

The [2] denotes that we sum over the two permutations of $\alpha\beta$ and $\gamma\delta$. The cumulants $\kappa_{\alpha,\beta}$, $\kappa_{\alpha,\beta,\gamma}$ and $\kappa_{\alpha,\beta,\gamma,\delta}$ can be found from the above theorem with the help of the Bartlett identities for cumulants (which follow from the identities for moments in view of [36] or [28], ex. 7.1 (p. 222)).

6 Approximation and Estimation of the Cumulants

We focus on the three first quantities in Theorem 3, as all first order stable cumulant combinations of degree up to four can be expressed as a linear combination of these (in particular, this is true for the quantities appearing in the Bartlett correction factor, cf. Chaps. 7 and 8 of [28]). Consider first $\kappa_{\alpha,v}$ and $\kappa_{\alpha,\beta,\gamma\delta} + \kappa_{\alpha\beta,\gamma\delta}$. These are both on the form $E K(\theta_0)$, where

$$K(\theta) = - \sum_i \int_0^T (f_v(\theta, t) \bar{\lambda}(\theta, t) - f_v^{(i)}(\theta, t) \bar{\lambda}^{(i)}(\theta, t)) \lambda_0(t) dt \int_0^t g_{i,v}(\theta, s) \lambda^i(\theta, s) \lambda_0(s) ds \quad (4)$$

where $g_{i,v}$ is $Z_{\alpha,i} - f_\alpha$ or $f_{\alpha\beta} - (Z_{\alpha,i} - f_\alpha)(Z_{\beta,i} - f_\beta)$, respectively. This is because $\sum_i g_i(s) \lambda^i(s) = 0$ in both the cases considered.

Theorem 4 *Under Condition 1,*

$$\begin{aligned} \text{case } \kappa_{\alpha,v} : \quad K(\theta) &= - \sum_i \int_0^T \lambda_v^i(\theta, t) \lambda_0(t) dt \int_0^t g_{i,v}(\theta, s) \lambda^i(\theta, s) \lambda_0(s) ds \\ \text{case } \kappa_{\alpha,\beta,\gamma\delta} + \kappa_{\alpha\beta,\gamma\delta} : \quad K(\theta) &= \sum_i \int_0^T \frac{\bar{\lambda}_\gamma(\theta, t) \bar{\lambda}_\delta(\theta, t)}{\bar{\lambda}(\theta, t)} \lambda_0(t) dt \int_0^t g_{i,\gamma\delta}(\theta, s) \lambda^i(\theta, s) \lambda_0(s) ds + o_p(n) \end{aligned} \quad (5)$$

uniformly on compact sets of θ s.

Hence $K(\theta)$ is, to first order, a functional of independent sums of bounded terms, and it is easy to see that the remainder term is $o(n)$ in expectation. By the law of large numbers and by uniform integrability, it follows that

$$K(\theta_0) = EK(\theta_0) + o(n). \quad (6)$$

Also, this representation, along with Theorem 4, shows that consistent estimation is possible. A natural estimator is

$$\hat{K} = \sum_i \int_0^T f_v^{(i)}(\tilde{\theta}, t) \bar{\lambda}^{(i)}(\tilde{\theta}, t) d\hat{\Lambda}_0(t) \int_0^t g_i(\tilde{\theta}, s) \lambda^i(\tilde{\theta}, s) d\hat{\Lambda}_0(s), \quad (7)$$

where $\tilde{\theta}$ and $\hat{\Lambda}_0$ are consistent estimators of θ and the baseline hazard, respectively. By the same type of reasoning as in Theorem 4,

$$\hat{K} = K + o_p(n), \quad (8)$$

whence consistent estimation of EK/n can be carried out. Note that we do not assume that EK/n has a limit as $n \rightarrow \infty$, and so do not require i.i.d. assumptions. (If EK/n only has cluster points, one can do the arguments with subsequences.)

To state a formal result summarizing the above, we assume that $\tilde{\theta}$ is the maximum partial likelihood estimator, either unrestricted or under a null hypothesis (in particular, it can be the true parameter). We also let $\hat{\Lambda}_0$ be the estimator of the baseline hazard given on p. 1103 of [1]:

$$\hat{\Lambda}_0(t) = \int_0^t \bar{\lambda}(\tilde{\theta}, s)^{-1} d\bar{N}(s)$$

Theorem 5 *Assume Condition 1. Then (6) and (8) hold.*

As far as estimation of the κ_v s is concerned, a natural choice is the observed derivatives of the log likelihood,

$$\hat{\kappa}_v = - \int_0^T f_v(\tilde{\theta}, t) d\bar{N}(t),$$

and, obviously,

$$\hat{\kappa}_v = \kappa_v + o_p(n).$$

Estimating $\kappa_{\alpha\beta, \gamma\delta}$ is different from what is discussed above only in so far as the term involving the covariance is concerned. This covariance, however, can be asymptotically approximated by a sum of generalized cumulants of sums of independent random variables, and this approximation is easily used to specify a consistent estimator.

7 Conclusion

The Bartlett identities and Bartlett correction are important tools for improving the accuracy of likelihood inference (see Sect. 2 for a brief review). We here argue that the technology carries over to partial likelihood (Sect. 3), and then provide specific conditions for the identities (Theorem 1) and the correction (Theorem 2) to hold in the case of Cox regression (Sect. 4). The identities and the correction involve cumulants of low order derivatives of the log partial likelihood, and these are given explicit expression in Theorem 3 in Sect. 5. In order to apply the Bartlett correction in practice, the relevant cumulants need to have consistent estimators, and these are found in Sect. 6. The proposed correction really does improve the accuracy, as documented in Tables 1, 2 and 3.

Table 1 Baseline failure distribution: exponential (1); censoring: uniform (0,1); # patients: 20; # simulations: 10000

Classes	Nominal	Uncorrected LR	Corrected LR
Mean	1	1.081	1.045
Variance	2	2.364	2.059
χ^2 tests,	1 %	1.3 %	1.0 %
Nominal and	5 %	6.0 %	5.4 %
Actual rejection	10 %	11.2 %	10.6 %
Probabilities	50 %	52.0 %	52.0 %

Table 2 Baseline failure distribution: exponential (1); censoring: uniform (0,2); # patients: 10; # simulations: 15000

Classes	Nominal	Uncorrected LR	Corrected LR
Mean	1	1.150	1.050
Variance	2	2.635	2.099
χ^2 tests,	1 %	1.5 %	0.8 %
Nominal and	5 %	6.9 %	5.1 %
Actual rejection	10 %	12.8 %	10.8 %
Probabilities	50 %	53.9 %	53.0 %

Table 3 Baseline failure distribution: exponential (1); censoring: uniform (0,=2); # patients: 10; # simulations: 20000

Classes	Nominal	Uncorrected LR	Corrected LR
Mean	1	1.141	1.025
Variance	2	2.298	1.813
χ^2 tests,	1 %	1.3 %	0.8 %
Nominal and	5 %	6.6 %	4.4 %
Actual rejection	10 %	12.6 %	9.8 %
Probabilities	50 %	54.1 %	53.7 %

Acknowledgments We would like to thank Ole Barndorff-Nielsen for your leadership and inspiring work, both on likelihood, and on high frequency data. We would also like to thank Almut Veraart and the referee for feedback that has substantially improved the paper. The authors are also grateful to the (U.S.) National Science Foundation for support under several grants, most recently DMS-14-07812.

Appendix: Proofs of Theorems

In addition to the definitions in Sect. 4, let $M^i(t) = N^i(t) - \int_0^t \lambda^i(s)ds$, with $\bar{M}(t) = M^1(t) + M^2(t) + \dots + M^n(t)$.

Proof of Theorem 3 The formula for κ_v is obvious, as $L_v(T) = - \int_0^T f_v(t)d\bar{N}(t)$ for $|v| \geq 2$.

Next,

$$\begin{aligned}
 \kappa_{\alpha,v} &= -EL_{\alpha}(T) \int_0^T f_v(t)\bar{\lambda}(t)\lambda_0(t)dt \\
 &\quad - EL_{\alpha}(T) \int_0^T f_v(t)(d\bar{N}(t) - \bar{\lambda}(t)\lambda_0(t)dt) \\
 &= - \int_0^T EL_{\alpha}(t) f_v(t)\bar{\lambda}(t)\lambda_0(t)dt
 \end{aligned}$$

since $L_\alpha(t)$ is a martingale, and since $\langle L_\alpha, \int_0^\cdot f_v(t)(d\bar{N}(t) - \bar{\lambda}(t)\lambda_0(t)dt) \rangle_T = \int_0^T (Z_{\alpha,i}(s) - f_\alpha(s))f_v(s)\bar{\lambda}^i(s)\lambda_0(s)ds = 0$. On the other hand, $\lambda^i(t)$ is zero after the jump of $N^i(t)$, hence

$$\kappa_{\alpha,v} = - \sum_i \int_0^T E \left[\int_0^t (Z_{\alpha,i}(s) - f_\alpha(s)) dN^i(s) \right] f_v^{(i)}(t) \bar{\lambda}^{(i)}(t) \lambda_0(t) dt$$

which yields the desired result since $dM^i(s) = dN^i(s) - \lambda^i(s)\lambda_0(s)ds$ is the differential of a martingale which is independent of $f_v^{(i)}(t)\bar{\lambda}^{(i)}(t)\lambda_0(s)$ (so that the latter quantity can be moved inside an integral with respect to this martingale).

Turning to $\kappa_{\alpha,\beta,\gamma\delta} + \kappa_{\alpha\beta,\gamma\delta}$, note that both the martingales

$L_\alpha(t)L_\beta(t) - [L_\alpha, L_\beta]_t = \int_0^t L_\alpha(s)dL_\beta(s)[2]$ and $L_{\alpha\beta}(t) + [L_\alpha, L_\beta]_t = \int_0^t [(Z_{\alpha,i}(s) - f_\alpha(s))(Z_{\beta,i}(s) - f_\beta(s)) - f_{\alpha\beta}(s)]dN^i(s)$ are orthogonal to $L_{\gamma\delta}(t) + \langle L_\gamma, L_\delta \rangle_t = - \int_0^t f_{\gamma\delta}(s)d\bar{M}(s)$, in the former case because $\langle L_\beta, \bar{M} \rangle_t \equiv 0$ and in the latter case for the same reason that makes $L_{\alpha\beta}(t) + [L_\alpha, L_\beta]_t$ a martingale. It follows that $\text{Cov}(L_\alpha(t)L_\beta(t) + L_{\alpha\beta}(t), L_{\gamma\delta}(t) + \langle L_\gamma, L_\delta \rangle_t) = 0$, whence

$$\begin{aligned} \kappa_{\alpha,\beta,\gamma\delta} + \kappa_{\alpha\beta,\gamma\delta} &= \text{Cov}(L_\alpha(T)L_\beta(T) + L_{\alpha\beta}(T), \langle L_\gamma, L_\delta \rangle_T) \\ &= \int_0^T E \left[(L_\alpha(t)L_\beta(t) + L_{\alpha\beta}(t)) f_{\gamma\delta}(t) \bar{\lambda}(t) \lambda_0(t) \right] dt \end{aligned}$$

since $L_\alpha(t)L_\beta(t) + L_{\alpha\beta}(t)$ is a martingale. The stated formula then follows by the same reasoning as for $\kappa_{\alpha,v}$.

Finally, for $\kappa_{\alpha\beta,\gamma\delta}$, note that $\text{Cov}(L_{\alpha\beta}(T) + \langle L_\alpha, L_\beta \rangle_T, L_{\gamma\delta}(T) + \langle L_\gamma, L_\delta \rangle_T) = E \int_0^T f_{\alpha\beta}(t)f_{\gamma\delta}(t)dt$, and that, by the method used for $\kappa_{\alpha,v}$,

$$\begin{aligned} &\text{Cov}(\langle L_\alpha, L_\beta \rangle_T, L_{\gamma\delta}(T) + \langle L_\gamma, L_\delta \rangle_T) \\ &= - \sum_i E \int_0^T \left[f_{\alpha\beta}^{(i)}(t) \bar{\lambda}^{(i)}(t) - f_{\alpha\beta}(t) \bar{\lambda}(t) \right] \lambda_0(t) dt \int_0^t f_{\gamma\delta}(s) \lambda^i(s) \lambda_0(s) ds. \end{aligned}$$

The expression for $\kappa_{\alpha\beta,\gamma\delta}$ then follows.

Proof of Theorem 4 The first case follows since $f_v(\theta, t)\bar{\lambda}(\theta, t) = \bar{\lambda}_v(\theta, t)$ and $f_v^{(i)}(\theta, t)\bar{\lambda}^{(i)}(\theta, t) = \bar{\lambda}_v^{(i)}(\theta, t)$. In the second case, $f_{\gamma\delta}(\theta, t)\bar{\lambda}(\theta, t) = (\bar{\lambda}_{\gamma\delta}(\theta, t) - \bar{\lambda}_\gamma(\theta, t)\bar{\lambda}_\delta(\theta, t))/\bar{\lambda}(\theta, t)$, and similarly when removing patient i . By Condition 1 and by Taylor expansion, it is then easy to see that $\sum_i (f_v(\theta, t)\bar{\lambda}(\theta, t) - f_v^{(i)}(\theta, t)\bar{\lambda}^{(i)}(\theta, t)) = -\bar{\lambda}_\gamma(\theta, t)\bar{\lambda}_\delta(\theta, t)/\bar{\lambda}(\theta, t) + o_p(n)$, uniformly in $t \in [0, T]$ and in a compact set of θ s.

Proof of Theorem 5 We use the result in Theorem 4. Equation (6) is obvious in view of our assumptions, while (8) follows by the same reasoning as in the proofs of Lemma 3.1 (p. 1105) and Theorem 3.4 (p. 1108) of [1]. Note that in view of

Condition 1, $\liminf_{n \rightarrow \infty} \sum_{i=1}^n EY_i(t)/n$ is nonzero in an interval in t containing 0, which is sufficient (in view of Condition 1) to ensure that Conditions A–D on p. 1105 of [1] are satisfied, whence $\hat{\theta}$ is consistent and asymptotically normal, and similarly for $\hat{\Lambda}_0$.

References

1. Andersen, P.K., Gill, R.D.: Cox's regression model for counting processes: a large sample study. *Ann. Statist.* **10**, 1100–1120 (1982)
2. Andersen, P.K., Borgan, Ø., Gill, R.D., Keiding, N.: *Statistical Models Based on Counting Processes*. Springer, New York (1993)
3. Barndorff-Nielsen, O.E.: Conditionality resolutions. *Biometrika* **67**, 293–310 (1980)
4. Barndorff-Nielsen, O.E.: Inference on full or partial parameters based on the standardized signed log likelihood ratio. *Biometrika* **73**, 307–322 (1986)
5. Barndorff-Nielsen, O.E.: Modified signed log likelihood ratio. *Biometrika* **78**, 557–563 (1991)
6. Barndorff-Nielsen, O.E.: Diversity of evidence and Birnbaum's theorem [with comments and reply]. *Scand. J. Statist.* **22**, 513–522 (1995)
7. Barndorff-Nielsen, O.E., Cox, D.R.: Edgeworth and saddle-point approximations with statistical applications (with discussion). *J. Roy. Statist. Soc. Ser. B.* **41**, 279312 (1979)
8. Barndorff-Nielsen, O.E., Cox, D.R.: Bartlett adjustments to the likelihood ratio statistic and the distribution of the maximum likelihood estimator. *J. Roy. Statist. Soc. Ser. B.* **46**, 483–495 (1984)
9. Bartlett, M.S.: Approximate confidence intervals. *Biometrika* **40**, 12–19 (1953)
10. Bartlett, M.S.: Approximate confidence intervals. II. More than one unknown parameter. *Biometrika* **40**, 306–317 (1953)
11. Beran, R.J.: Estimated sampling distributions: the bootstrap and its competitors. *Ann. Statist.* **10**, 212–225 (1982)
12. Cordeiro, G.M.: On the corrections to the likelihood ratio statistics. *Biometrika* **74**, 265–274 (1987)
13. Cox, D.R.: Regression models and life tables (with discussion). *J. Roy. Statist. Soc. Ser. B.* **34**, 187–220 (1972)
14. Cox, D.R.: Partial likelihood. *Biometrika* **62**, 269–276 (1975)
15. Efron, B.: The Jackknife, the bootstrap and other resampling plans. In: *CBMS-NSF Regional Conference Series in Applied Mathematics*. Society for Industrial and Applied Mathematics, Philadelphia (1982)
16. Fleming, T.R., Harrington, D.P.: *Counting Processes and Survival Analysis*. Wiley, New York (1991)
17. Fisher, R.A.: Moments and product moments of sampling distributions. *Proc. Lond. Math. Soc. Ser. 2*(30), 199–238 (1929)
18. Gu, M.: On the Edgeworth expansion and bootstrap approximation for the Cox regression model under random censorship. *Canad. J. Statist.* **20**, 399–414 (1992)
19. Gu, M., Zheng, Z.: On the Bartlett adjustment for the partial likelihood ratio test in the Cox regression model. *Statistica Sinica* **3**, 543–555 (1993)
20. Hald, A.: The early history of the cumulants and the Gram-Charlier series. *Inter. Statist. Rev.* **68**, 137–153 (2000)
21. Hall, P.: *The Bootstrap and Edgeworth Expansion*. Springer, New York (1992)
22. Hjort, N.L.: Bootstrapping Cox's regression model. Technical Report No. 241, Department of Statistics, Stanford University (1985)
23. Jacod, J.: Multivariate point processes: predictable projection, Radon-Nikodym derivatives, representation of martingales. *Zeitschr. Wahr. verw. Geb.* **31**, 235–253 (1975)

24. Jacod, J., Shiryaev, A.N.: Limit Theorems for Stochastic Processes, 2nd edn. Springer, Berlin (2003)
25. Jensen, J.L.: A historical sketch and some new results on the improved log likelihood ratio statistic. *Scand. J. Statist.* **20**, 1–15 (1993)
26. Jensen, J.L.: A simple derivation of r^* for curved exponential families. *Scand. J. Statist. Scand. J. Statist.* **24**, 33–46 (1997)
27. Lauritzen, S.L.: Thiele - Pioneer in Statistics. OUP, Oxford (2002)
28. McCullagh, P.: Tensor Methods in Statistics. Chapman and Hall, London (1987)
29. Mykland, P.A.: Bartlett identities for martingales. *Ann. Statist.* **22**, 21–38 (1994)
30. Mykland, P.A.: Dual likelihood. *Ann. Statist.* **23**, 396–421 (1995)
31. Mykland, P.A.: Embedding and asymptotic expansions for martingales. *Probab. Theory Rel. Fields* **103**, 475–492 (1995)
32. Mykland, P.A.: Martingale expansions and second order inference. *Ann. Statist.* **23**, 707–731 (1995)
33. Mykland, P.A.: Bartlett identities and large deviations in likelihood theory. *Ann. Statist.* **27**, 1105–1117 (1999)
34. Mykland, P.A.: Likelihood computations without Bartlett identities. *Bernoulli* **7**, 473–485 (2001)
35. Reid, N.: Saddlepoint methods and statistical inference. *Statist. Sci.* **2**, 213–227 (1988)
36. Skovgaard, I.: A note on the differentiation of cumulants of log likelihood derivatives. *Int. Statist. Rev.* **54**, 29–32 (1986)
37. Stigler, S.M.: Discussion for ISI history session - IMP3. <http://www.stat.fi/isi99/proceedings/arkisto/varasto/stig1042.pdf> Cited 12Aug 1999
38. Thiele, T.N.: Om iagttagelseslærens halvinvarianter. *Kgl. danske Videnskabernes Selskaps Forhandlinger* **3**, 135–141 (1899)
39. Tsiadis, A.A.: A large sample study of Cox's regression model. *Ann. Statist.* **9**, 93–108 (1981)
40. Wong, W.H.: Theory of partial likelihood. *Ann. Statist.* **14**, 88–123 (1986)

Exchangeability and Infinite Divisibility

Martin Drapatz and Alexander Lindner

Abstract We characterize exchangeability of infinitely divisible distributions in terms of the characteristic triplet. This is applied to stable distributions and self-decomposable distributions, and a connection to Lévy copulas is made. We further study general mappings between classes of measures that preserve exchangeability and give various examples which arise from discrete time settings, such as stationary distributions of AR(1) processes, or from continuous time settings, such as Ornstein–Uhlenbeck processes or Upsilon-transforms.

Keywords Exchangeability · Exchangeability preserving transformation · Infinitely divisible distribution · Lévy copula · Ornstein-Uhlenbeck process · Random recurrence equation · Upsilon-transform

1 Introduction

Throughout, let $d \in \{2, 3, 4, \dots\}$ be an integer. The set of all *permutations* π of $\{1, \dots, d\}$, i.e. of all bijections $\pi : \{1, \dots, d\} \rightarrow \{1, \dots, d\}$, is denoted by $[d]$. For fixed d , exactly $d!$ permutations of $\{1, \dots, d\}$ exist. Corresponding to a permutation $\pi \in [d]$ we define the *permutation matrix* of π by

$$P_\pi = (e_{\pi(1)}, e_{\pi(2)}, \dots, e_{\pi(d)}) \in \mathbb{R}^{d \times d}$$

where e_i is the i -th unit (column) vector in $\mathbb{R}^d$. A permutation matrix P_π is an orthogonal matrix, i.e. $P_\pi P_\pi^T = \text{Id}_d$, where Id_d is the identity matrix in $\mathbb{R}^{d \times d}$. Thus, $P_\pi^T = P_\pi^{-1} = P_{\pi^{-1}}$. A distribution $\mu = \mathcal{L}(X)$ of a random vector $X = (X_1, \dots, X_d)^T$ is *exchangeable*, if

M. Drapatz · A. Lindner (✉)

Institute for Mathematical Finance, Ulm University, Helmholtzstraße 18,
89081 Ulm, Germany
e-mail: alexander.lindner@uni-ulm.de

M. Drapatz

e-mail: martin.drapatz@uni-ulm.de

$$\mathcal{L}(X) = \mathcal{L}((X_{\pi(1)}, \dots, X_{\pi(d)})^T) \quad \forall \pi \in [d],$$

and we shall also say that the *random vector* X is *exchangeable*. Since

$$P_\pi(X_1, \dots, X_d)^T = (X_{\pi^{-1}(1)}, \dots, X_{\pi^{-1}(d)})^T$$

and since the distribution of $P_\pi(X_1, \dots, X_d)^T$ is the image measure $P_\pi\mu$, defined by

$$(P_\pi\mu)(B) = \mu(P_\pi^{-1}(B)) \quad \forall B \in \mathcal{B}_d,$$

where $\mathcal{B}_d$ denotes the σ -algebra of Borel sets in $\mathbb{R}^d$, this is equivalent to saying that $P_\pi\mu = \mu$ for all $\pi \in [d]$. This can be extended to general (positive) measures on $(\mathbb{R}^d, \mathcal{B}_d)$:

Definition 1 A measure μ on $(\mathbb{R}^d, \mathcal{B}_d)$ is *exchangeable*, if $P_\pi\mu = \mu$ for all permutations $\pi \in [d]$, where $P_\pi\mu$ denotes the image measure of μ under P_π .

Exchangeable probability distributions have various applications, e.g. permutation tests [13]. The dependence structure of exchangeable random vectors is limited to some extent. This can be used for deriving inequalities for tail probabilities of sums of exchangeable random variables, see e.g. [27] for large deviation results or [11] for a concentration of measure result. Furthermore, an important class of copulas are exchangeable Archimedian copulas, see e.g. Nelsen [19], Chap. 4.

In this paper, we are interested in infinitely divisible distributions which are exchangeable and in transformations of (infinitely divisible) distributions and Lévy measures that preserve exchangeability. Recall that a distribution μ on $(\mathbb{R}^d, \mathcal{B}_d)$ is *infinitely divisible*, if for every $n \in \mathbb{N}$ there exists a distribution μ_n on $(\mathbb{R}^d, \mathcal{B}_d)$, such that the n -fold convolution of μ_n with itself is equal to μ . By the Lévy-Khintchine formula (e.g. Sato [22, Theorem 8.1]), a distribution μ on $(\mathbb{R}^d, \mathcal{B}_d)$ is infinitely divisible if and only if its characteristic function $\mathbb{R}^d \ni z \mapsto \widehat{\mu}(z) = \int_{\mathbb{R}^d} e^{ix^T z} \mu(dx)$ can be represented in the form

$$\begin{aligned} \widehat{\mu}(z) &:= \exp\{\Psi_\mu(z)\}, \quad \text{with} \\ \Psi_\mu(z) &= -\frac{1}{2}z^T A z + i\gamma^T z + \int_{\mathbb{R}^d} (e^{ix^T z} - 1 - ix^T z \mathbb{1}_{\mathbb{D}}(x)) \nu(dx) \quad \forall z \in \mathbb{R}^d, \end{aligned}$$

where $\gamma = (\gamma_1, \dots, \gamma_d) \in \mathbb{R}^d$, $A \in \mathbb{R}^{d \times d}$ is symmetric and nonnegative definite and ν is a measure on $(\mathbb{R}^d, \mathcal{B}_d)$ such that $\nu(\{0\}) = 0$ and $\int_{\mathbb{R}^d} (|x|^2 \wedge 1) \nu(dx) < \infty$. Here,

$$\mathbb{D} := \{x \in \mathbb{R}^d : |x|^2 \leq 1\}$$

denotes the unit ball in $\mathbb{R}^d$ and $|x| = \sqrt{x^T x}$ the Euclidean norm, where x^T denotes the transpose of $x = (x_1, \dots, x_d)^T \in \mathbb{R}^d$ which we will throughout understand as column vectors. The quantities Ψ_μ and (A, ν, γ) are unique and called the *characteristic exponent* and *characteristic triplet* of μ , respectively. The matrix A is called the

Gaussian covariance matrix of μ and ν the *Lévy measure* of μ . Conversely, to every triplet (A, ν, γ) with these properties there exists a unique infinitely divisible distribution having this characteristic triplet. Infinitely divisible distributions are closely connected to *Lévy processes*, i.e. to $\mathbb{R}^d$ -valued stochastic processes $L = (L_t)_{t \geq 0}$ that have stationary and independent increments, start in 0 and have almost surely right-continuous sample paths with finite left-limits: for every Lévy process L , the distribution $\mathcal{L}(L_t)$ is infinitely divisible for every $t \geq 0$, and conversely, to every infinitely divisible distribution μ on $(\mathbb{R}^d, \mathcal{B}_d)$, there exists a Lévy process L , unique in law, such that $\mathcal{L}(L_1) = \mu$. See Sato [22] for further information regarding Lévy processes and infinitely divisible distributions.

Exchangeability of multivariate Poisson distributions has been considered by Griffiths and Milne [17], but to our knowledge no systematic study of exchangeability of infinitely divisible distributions has been carried out. In the next section, we shall characterize exchangeability of infinitely divisible distributions and apply it to stable distributions. We also give a general construction criterion for exchangeable measures and give relations between exchangeable Lévy measures and their Lévy copula. Then, in Sect. 3, we summarise some results about exchangeability preserving linear transformations due to Dean and Verducci [15] and Commenges [13]. In Sect. 4, we study general exchangeability preserving transformations and give a result when the inverse of such a transformation preserves exchangeability. This is then applied in Sect. 5 to various transformations based on time series, such as the mapping which transforms a (noise) distribution to the stationary distribution of an associated autoregressive process of order 1, moving average processes with random coefficients, or random recurrence equations. Finally, in Sect. 6 we study mappings of the form $\mu \mapsto \mathcal{L}(\int_0^\infty f(t) dL_t^\mu)$, where f is some suitable (mostly deterministic) function and $(L_t^\mu)_{t \geq 0}$ a Lévy process with distribution μ at time 1. This is in particular applied to self-decomposable distributions and their background driving Lévy process, or to Upsilon-transforms.

Throughout, we use the following notation: the characteristic function of a random vector X will be denoted by φ_X and the Fourier transform of a measure μ by $\widehat{\mu}$, so that $\varphi_X(z) = \mathbb{E}(e^{iX^T z})$ and $\widehat{\mu}(z) = \int e^{iz^T x} \mu(dx)$ for $z \in \mathbb{R}^d$. When we speak of a measure μ on $\mathbb{R}^d$, we always mean a measure on $\mathbb{R}^d$ with the corresponding Borel- σ -algebra which we denoted by $\mathcal{B}_d$. Sometimes we will consider a Borel- σ -algebra on a subset F of $\mathbb{R}^d$ which we will denote by $\mathcal{B}(F)$. Equality in distribution of random vectors will be denoted by “ $\stackrel{d}{=}$ ”, and convergence in distribution by “ $\stackrel{d}{\rightarrow}$ ”. We write $\log^+(x) = \log(\max\{x, 1\})$ for $x \in \mathbb{R}$.

2 Infinitely Divisible Exchangeable Distributions

In this section we shall characterise exchangeability of infinitely divisible distributions in terms of their characteristic triplet. For that, we need the following definition:

Definition 2 A matrix $A \in \mathbb{R}^{d \times d}$ commutes with permutations if

$$P_\pi A = A P_\pi \quad \forall \pi \in [d].$$

Commenges [13] calls matrices that commute with permutations *exchangeable matrices*, but we shall stick to our notation. In the next section in Theorem 5 we shall summarise some known results about matrices that commute with permutations, in particular give an explicit description for them, but for the moment we shall confine ourselves with the fact that a feasible characterization of these matrices exists.

Remark 1 Let X be a normal random vector in $\mathbb{R}^d$ with mean m and covariance matrix Σ . Then $P_\pi X$ is $N(P_\pi m, P_\pi \Sigma P_\pi^T)$ distributed, and from that it is easy to see that X is exchangeable if and only if $m = (m_1, \dots, m_1)^T$ for some $m_1 \in \mathbb{R}$ and Σ commutes with permutations.

More generally, we can characterise when an infinitely divisible distribution is exchangeable:

Theorem 1 (Exchangeable infinitely divisible distributions) *Let μ be an infinitely divisible distribution on $\mathbb{R}^d$ with characteristic exponent Ψ_μ and characteristic triplet (A, ν, γ) . Then the following are equivalent:*

- (i) μ is exchangeable.
- (ii) $\Psi_\mu(P_\pi z) = \Psi_\mu(z)$ for all $z \in \mathbb{R}^d$ and $\pi \in [d]$.
- (iii) The Gaussian covariance matrix A commutes with permutations, the Lévy measure ν is exchangeable and $\gamma_i = \gamma_j$ for all $i, j \in \{1, \dots, d\}$, where γ_i denotes the i 'th component of γ .

Proof Let $X = (X_1, \dots, X_d)^T \in \mathbb{R}^d$ be a random vector with infinitely divisible distribution $\mu = \mathcal{L}(X)$. Since $\mathbb{E}e^{i(P_\pi X)^T z} = \widehat{\mu}(P_\pi^{-1} z) = e^{\Psi_\mu(P_\pi^{-1} z)}$ for all $\pi \in [d]$ and since $z \mapsto \Psi_\mu(P_\pi^{-1} z)$ defines indeed the characteristic exponent of the infinitely divisible distribution $\mathcal{L}(P_\pi X)$ (as a consequence of Lemma 7.6 and Proposition 11.10 in [22]), the equivalence of (i) and (ii) follows.

To see the equivalence of (i) and (iii), observe that since μ is uniquely described by its characteristic triplet, it is exchangeable if and only if the characteristic triplets of $\mathcal{L}(X)$ and $\mathcal{L}(P_\pi X)$ coincide for all permutations π on $[d]$. Denote the characteristic triplet of $\mathcal{L}(P_\pi X)$ by $(A_\pi, \nu_\pi, \gamma_\pi)$. Then by Proposition 11.10 of [22] the characteristic triplet of $\mathcal{L}(P_\pi X)$ is given by

$$A_\pi = P_\pi A P_\pi^T = P_\pi A P_{\pi^{-1}}, \quad \nu_\pi = P_\pi \nu, \quad \gamma_\pi = P_\pi \gamma, \quad (1)$$

where we used that P_π is orthogonal and hence $(P_\pi \nu)(\{0\}) = 0$, and $P_\pi x \in \mathbb{D}$ if and only if $x \in \mathbb{D}$. The equivalence of (i) and (iii) then follows from (1).

Corollary 1 *Let $(L_t)_{t \geq 0}$ be a Lévy process in $\mathbb{R}^d$ and $s > 0$. Then $\mathcal{L}(L_s)$ is exchangeable if and only if $\mathcal{L}(L_1)$ is exchangeable.*

Proof This is immediate from Theorem 1, since $\Psi_{\mathcal{L}(L_s)}(z) = s\Psi_{\mathcal{L}(L_1)}(z)$.

As an application of Theorem 1, let us characterize all exchangeable stable distributions. Recall that a distribution μ on $\mathbb{R}^d$ is *stable* if and only if it is Gaussian (2-stable case), or if it is infinitely divisible with Gaussian covariance matrix A being 0 and such that there are $\alpha \in (0, 2)$ and a finite measure λ on $\mathbb{S} := \{x \in \mathbb{R}^d : |x| = 1\}$ such that the Lévy measure ν of μ can be represented in the form

$$\nu(B) = \int_{\mathbb{S}} \lambda(d\xi) \int_0^\infty \mathbb{1}_B(r\xi) \frac{dr}{r^{1+\alpha}}, \quad B \in \mathcal{B}(\mathbb{R}^d), \quad (2)$$

see Sato [22], Theorems 13.15, 14.1 and 14.3. We speak of α -stable distributions in this case. The measure λ is unique and called *spectral measure* or *spherical part* of ν .

Exchangeability of 2-stable, i.e. normal distributions, has been already settled in Remark 1, so we shall restrict to $\alpha \in (0, 2)$. By extending the spherical measure to a measure on $\mathbb{R}^d$ by setting it 0 outside $\mathbb{S}$, we can speak of exchangeability of a spherical measure.

Theorem 2 (Exchangeable stable distributions) *Let μ be an α -stable distribution with characteristic triplet $(0, \nu, (\gamma_1, \dots, \gamma_d)^T)$, where $\alpha \in (0, 2)$. Then μ is exchangeable if and only if the spherical part λ of ν is exchangeable and if $\gamma_i = \gamma_j$ for all $i, j \in \{1, \dots, d\}$.*

Proof By Theorem 1, it is enough to show that ν is exchangeable if and only if λ is exchangeable. Observe that each P_π maps $\mathbb{S}$ bijectively onto $\mathbb{S}$.

Suppose ν is exchangeable, i.e. for all permutations π on $[d]$ and $B \in \mathcal{B}_d$ we have $\nu(B) = \nu(P_\pi(B))$. Now consider the system of subsets $\mathcal{A} := \{(b, \infty)C : b > 0, C \in \mathcal{B}(\mathbb{S})\}$ of $\mathbb{R}^d$, where $(b, \infty)C := \{x \in \mathbb{R}^d : |x| \in (b, \infty), \frac{x}{|x|} \in C\}$. By (2) the Lévy measure of $(b, \infty)C \in \mathcal{A}$ is

$$\nu((b, \infty)C) = \lambda(C)\alpha^{-1}b^{-\alpha}. \quad (3)$$

Combining exchangeability of ν with Eq.(3) yields $\lambda(C) = \lambda(P_\pi(C))$ for all permutations π on $[d]$ and all $C \in \mathcal{B}(\mathbb{S})$. Thus, λ is exchangeable.

For the converse, assume now that λ is exchangeable. Using Eq.(3) we get for all $b > 0$ and $C \in \mathcal{B}(\mathbb{S})$

$$\nu((b, \infty)C) = \lambda(C)\alpha^{-1}b^{-\alpha} = \lambda(P_\pi(C))\alpha^{-1}b^{-\alpha} = \nu((b, \infty)P_\pi(C)) = \nu(P_\pi((b, \infty)C)).$$

The system $\mathcal{A}$ is a generator of $\mathcal{B}(\mathbb{R}^d \setminus \{0\})$ and a π -system. Furthermore $\mathcal{D} = \{B \in \mathcal{B}(\mathbb{R}^d \setminus \{0\}) : \nu(B) = (P_\pi \nu)(B)\}$ is a λ -system containing $\mathcal{A}$ and this implies by Theorem 1.3.2 in [12] that $\mathcal{D} = \mathcal{B}(\mathbb{R}^d \setminus \{0\})$.

Let us mention another result regarding exchangeable stable distributions of Nguyen [20], who showed that a random vector $(X_1, \dots, X_d)^T$ of identically distributed random variables and with α -stable conditional margin $X_1 | X_2 = x_2, \dots, X_d = x_d$, is α -multivariate stable distributed, if $(X_1, \dots, X_d)^T$ is exchangeable.

In view of Theorems 1 and 2 it is interesting to know how to construct exchangeable measures, in particular exchangeable Lévy measures or exchangeable spherical measures, which is the contents of the next result:

Theorem 3 (Construction of exchangeable measures) *Let*

$$F := \{(x_1, \dots, x_d) \in \mathbb{R}^d : x_1 \geq x_2 \geq \dots \geq x_d\}$$

and $F_\pi := P_\pi(F)$ for $\pi \in [d]$. Then $\cup_{\pi \in [d]} F_\pi = \mathbb{R}^d$ and the following two statements hold:

- (i) A measure ν on $(\mathbb{R}^d, \mathcal{B}_d)$ is exchangeable if and only if for all $\pi \in [d]$ and all $C \in \mathcal{B}_d$

$$\nu(C \cap F_\pi) = \nu(P_\pi^{-1}(C) \cap F). \quad (4)$$

In particular, an exchangeable measure is uniquely determined by its values on $\mathcal{B}(F)$.

- (ii) Let $\tilde{\nu}$ be a measure on $(F, \mathcal{B}(F))$. Then the measure ν on $(\mathbb{R}^d, \mathcal{B}_d)$, defined by

$$\nu(C \cap F_\pi) := \tilde{\nu}(P_\pi^{-1}(C) \cap F), \quad C \in \mathcal{B}_d, \quad \pi \in [d],$$

is well-defined, exchangeable and satisfies $\nu|_{\mathcal{B}(F)} = \tilde{\nu}$. Further, ν is finite if and only if $\tilde{\nu}$ is finite, and ν is a Lévy measure on $\mathbb{R}^d$ if and only if $\tilde{\nu}$ is a Lévy measure on F , the latter meaning that $\tilde{\nu}(\{0\}) = 0$ and $\int_F (|x|^2 \wedge 1) \tilde{\nu}(dx) < \infty$.

Proof That $\cup_{\pi \in [d]} F_\pi = \mathbb{R}^d$ is clear.

- (i) If ν is exchangeable, i.e. $\nu = P_\pi \nu$ for all $\pi \in [d]$, then for all $\pi \in [d]$ and all $C \in \mathcal{B}_d$

$$\nu(C \cap F_\pi) = (P_\pi \nu)(C \cap F_\pi) = \nu(P_\pi^{-1}(C \cap F_\pi)) = \nu(P_\pi^{-1}(C) \cap F),$$

i.e. Eq.(4) holds. For the converse suppose Eq.(4) is satisfied. For $\pi, \sigma \in [d]$ we define $\tau = \pi^{-1} \circ \sigma$. Then for every $C \in \mathcal{B}_d$ we have

$$\begin{aligned} (P_\pi \nu)(C \cap F_\sigma) &= \nu(P_\pi^{-1}(C) \cap P_\pi^{-1}(F_\sigma)) \\ &\stackrel{(4)}{=} \nu(P_\tau^{-1}(P_\pi^{-1}(C)) \cap F) \\ &= \nu(P_\sigma^{-1}(C) \cap F) \stackrel{(4)}{=} \nu(C \cap F_\sigma), \end{aligned}$$

i.e. $\nu|_{\mathcal{B}(F_\sigma)} = (P_\pi \nu)|_{\mathcal{B}(F_\sigma)}$, and $\nu = P_\pi \nu$ follows since $\cup_{\sigma \in [d]} F_\sigma = \mathbb{R}^d$.

- (ii) If ν is well-defined, it is clear that ν is a finite, respectively a Lévy measure, if and only if $\tilde{\nu}$ is, and by (i) ν will be exchangeable. Further, $\nu|_{\mathcal{B}(F)} = \tilde{\nu}$. Hence we

only have to show that ν is well-defined. Let $C \in \mathcal{B}_d$ with $C \subset F_{\pi_1} \cap F_{\pi_2}$ for some $\pi_1, \pi_2 \in [d]$. We have to show that

$$\tilde{\nu}(P_{\pi_1}^{-1}(C)) = \tilde{\nu}(P_{\pi_2}^{-1}(C)). \quad (5)$$

Define $C_1 = P_{\pi_1}^{-1}(C)$ and $C_2 = P_{\pi_2}^{-1}(C)$ and notice that $C_1, C_2 \subset F$. Furthermore, let $\pi_3 = \pi_2^{-1} \circ \pi_1$ such that $P_{\pi_3}(C_1) = C_2$. If now $x = (x_1, \dots, x_d)^T \in C_1$, then $x_1 \geq x_2 \geq \dots \geq x_d$ since $C_1 \subset F$. Since $P_{\pi_3}x = (x_{\pi_3^{-1}(1)}, \dots, x_{\pi_3^{-1}(d)})^T \in C_2 \subset F$ we conclude $x_{\pi_3^{-1}(1)} \geq x_{\pi_3^{-1}(2)} \geq \dots \geq x_{\pi_3^{-1}(d)}$. Hence $x_i = x_{\pi_3^{-1}(i)}$ for all $i = 1, \dots, d$ and $P_{\pi_3}x = x$. Thus, $P_{\pi_3}(C_1) = C_1$ and $C_1 = C_2$, which gives Eq. (5).

As an example, define for $d = 2$ a measure λ on $(\mathbb{S} \cap F, \mathcal{B}(\mathbb{S} \cap F))$ by

$$\lambda(\{e_1\}) = w_1, \quad \lambda(\{\frac{e_1 - e_2}{\sqrt{2}}\}) = w_2, \quad \lambda((\mathbb{S} \cap F) \setminus \{e_1, \frac{e_1 - e_2}{\sqrt{2}}\}) = 0, \quad (6)$$

where $e_1 = (1, 0)^T$ and $e_2 = (0, 1)^T$ denote the unit vectors in $\mathbb{R}^2$ and $w_1, w_2 > 0$. Then

$$\tilde{\nu}(B) = \int_{\mathbb{S}} \lambda(d\xi) \int_0^\infty \mathbb{1}_B(r\xi) \frac{e^{-r^2\lambda(\{\xi\})} dr}{r}, \quad B \in \mathcal{B}(F),$$

defines a Lévy measure on $(F, \mathcal{B}(F))$. By Eq. (6),

$$\tilde{\nu}(B) = w_1 \int_0^\infty \mathbb{1}_B(re_1) \frac{e^{-r^2w_1} dr}{r} + w_2 \int_0^\infty \mathbb{1}_B(r \frac{e_1 - e_2}{\sqrt{2}}) \frac{e^{-r^2w_2} dr}{r}, \quad B \in \mathcal{B}(F).$$

Denote by σ the permutation such that $\sigma(1) = 2$ and $\sigma(2) = 1$. By Theorem 3,

$$\nu(C \cap F_\pi) := \tilde{\nu}(P_\pi^{-1}(C) \cap F), \quad C \in \mathcal{B}_2, \quad \pi \in [2],$$

defines an exchangeable measure on $(\mathbb{R}^2, \mathcal{B}_2)$. One easily concludes for $C \in \mathcal{B}_2$

$$\nu(C \cap F_\sigma) = w_1 \int_0^\infty \mathbb{1}_C(rP_\sigma e_1) \frac{e^{-r^2w_1} dr}{r} + w_2 \int_0^\infty \mathbb{1}_C(rP_\sigma \frac{e_1 - e_2}{\sqrt{2}}) \frac{e^{-r^2w_2} dr}{r},$$

and hence with $f_A(rx) := \mathbb{1}_A(rx) + \mathbb{1}_A(rP_\sigma x)$ for $A \in \mathcal{B}_2, r \geq 0$ and $x \in \mathbb{R}^2$,

$$\begin{aligned} \nu(A) &= \nu(A \cap F_\sigma) + \nu(A \cap F) \\ &= w_1 \int_0^\infty f_A(re_1) \frac{e^{-r^2w_1} dr}{r} + w_2 \int_0^\infty f_A(r \frac{e_1 - e_2}{\sqrt{2}}) \frac{e^{-r^2w_2} dr}{r}, \quad A \in \mathcal{B}_2, \end{aligned}$$

is an exchangeable measure on $(\mathbb{R}^2, \mathcal{B}_2)$.

It is well known that a distribution $\mu = \mathcal{L}(X)$ is exchangeable if and only if its one-dimensional margins are equal and if it admits a copula which is exchangeable, see e.g. Nelsen [19, Theorem 2.7.4]. It is natural to ask if the same result holds for the relation between Lévy measures and Lévy copulas. Lévy copulas have been introduced by Tankov [26] and Cont and Tankov [14] for Lévy measures concentrated on $[0, \infty)^d$ and by Kallsen and Tankov [18] for Lévy measures on $\mathbb{R}^d$. The concept for measures on $\mathbb{R}^d$ is more complicated and we shall restrict ourselves to Lévy measures on $[0, \infty)^d$ and call the corresponding Lévy copula a positive Lévy copula. In order to define it, let ν be a Lévy measure concentrated on $[0, \infty)^d$. Then its *tail integral* U_ν is defined as the function $U_\nu : [0, \infty]^d \rightarrow [0, \infty]$ given by

$$U_\nu(x_1, \dots, x_d) := \begin{cases} \nu([x_1, \infty) \times \dots \times [x_d, \infty)), & (x_1, \dots, x_d) \neq (0, \dots, 0), \\ \infty, & (x_1, \dots, x_d) = (0, \dots, 0). \end{cases}$$

Its *marginal tail integrals* U_{ν_i} are defined as the tail integrals of the (one-dimensional) marginal Lévy measures $\nu_1, \dots, \nu_d$, i.e.

$$U_{\nu_i}(x_i) = U_\nu(0, \dots, 0, x_i, 0, \dots, 0) = \begin{cases} \nu_i([x_i, \infty)), & x_i \in (0, \infty], \\ \infty, & x_i = 0. \end{cases}$$

It is clear that a Lévy measure on $[0, \infty)^d$ is uniquely determined by its tail integral. Now a *positive Lévy copula* is a function $C : [0, \infty]^d \rightarrow [0, \infty]$ such that $C(x_1, \dots, x_d) = 0$ if at least one of the x_i is zero, such that

$$C(\infty, \dots, \infty, x_i, \infty, \dots, \infty) = x_i \quad \forall x_i \in [0, \infty], \quad i = 1, \dots, d,$$

such that $C(x_1, \dots, x_d) \neq \infty$ unless $x_1 = \dots, x_d = \infty$, and such that C is a d -increasing function (cf. Cont and Tankov [14, Definition 5.11]). Similar to copulas, Lévy copulas allow to separate the margins and the dependence structure of Lévy measures. More precisely, for every Lévy measure ν on $[0, \infty)^d$ there exists a positive Lévy copula C such that

$$U_\nu(x_1, \dots, x_d) = C(U_{\nu_1}(x_1), \dots, U_{\nu_d}(x_d)) \quad \forall x_1, \dots, x_d \in [0, \infty]. \quad (7)$$

The Lévy copula is uniquely determined on $U_{\nu_1}([0, \infty]) \times \dots \times U_{\nu_d}([0, \infty])$ and we shall call every positive Lévy copula C satisfying (7) a positive Lévy copula *associated with* ν . Conversely, if $\nu_1, \dots, \nu_d$ are one-dimensional Lévy measures on $[0, \infty)$ and if C is a positive Lévy copula, then the right-hand side of (7) defines the tail-integral of a Lévy measure ν on $[0, \infty)^d$ with margins $\nu_1, \dots, \nu_d$ and associated Lévy copula C ; see Cont and Tankov [14, Theorem 5.6].

Barndorff-Nielsen and Lindner [1, Theorem 6] showed that there is a one-to-one correspondence between (d -dimensional) positive Lévy copulas and Lévy measures on $[0, \infty)^d$ with “unit” 1-stable margins, i.e. Lévy measures with marginal tail integrals $[0, \infty] \rightarrow [0, \infty]$, $x_k \mapsto x_k^{-1}$ (the tail integral of a one-dimensional stable Lévy

measure is $x_k \mapsto ax_k^{-1}$ for some $a > 0$, and by “unit” we mean that $a = 1$). More precisely, if C is a positive Lévy copula, then there exists a unique Lévy measure ν_C on $[0, \infty)^d$ such that

$$\nu_C([x_1^{-1}, \infty) \times \cdots \times [x_d^{-1}, \infty)) = C(x_1, \dots, x_d) \quad \forall x_1, \dots, x_d \in [0, \infty] \quad (8)$$

which then has unit 1-stable margins (more precisely, is the Lévy measure of an infinitely divisible distribution with unit 1-stable margins). Conversely, to any Lévy measure ν_C on $[0, \infty)^d$ with unit 1-stable margins, the left-hand side of (8) defines a positive Lévy copula.

Definition 3 A positive Lévy copula $C : [0, \infty]^d \rightarrow [0, \infty]$ is *exchangeable*, if

$$C(x_1, \dots, x_d) = C(x_{\pi(1)}, \dots, x_{\pi(d)}) \quad \forall x_1, \dots, x_d \in [0, \infty], \pi \in [d].$$

We now give a connection between exchangeable Lévy copulas and exchangeable Lévy measures:

Theorem 4 (Exchangeability and Lévy copulas)

- (i) A positive Lévy copula C is exchangeable if and only if the Lévy measure ν_C with unit 1-stable margins defined by (8) is exchangeable.
- (ii) Let ν be a Lévy measure on $[0, \infty)^d$ with marginal Lévy measures $\nu_1, \dots, \nu_d$. If $\nu_1 = \cdots = \nu_d$ and if an associated positive Lévy copula C exists which is exchangeable, then ν is exchangeable. Conversely, if ν is exchangeable and $U_{\nu_1}([0, \infty]) = [0, \infty]$ (i.e. ν_1 has no atoms and is infinite), then $\nu_1 = \cdots = \nu_d$ and the unique associated positive Lévy copula C is exchangeable.

It seems very likely that the result in (ii) can be extended to the case when $U_{\nu_1}([0, \infty]) \neq [0, \infty]$, to a general statement of the form that ν is exchangeable if and only if $\nu_1 = \cdots = \nu_d$ and if it admits an exchangeable associated positive Lévy copula C (observe that then the positive Lévy copula is not necessarily unique any more). For simplicity, we have not pursued this issue further.

Proof First observe that as in the proof of Theorem 2, a Lévy measure ν on $[0, \infty)^d$ is exchangeable if and only if

$$(P_\pi \nu)([x_1, \infty) \times \cdots \times [x_d, \infty)) = \nu([x_1, \infty) \times \cdots \times [x_d, \infty)) \quad \forall x_1, \dots, x_d \in [0, \infty], \pi \in [d],$$

i.e. if and only if

$$U_{P_\pi \nu} = U_\nu \quad \forall \pi \in [d]. \quad (9)$$

- (i) For $(x_1, \dots, x_d) \neq (0, \dots, 0)$ we have

$$U_{P_\pi \nu_C}(x_1, \dots, x_d) = \nu_C([x_{\pi(1)}, \infty) \times \cdots \times [x_{\pi(d)}, \infty)) = C(x_{\pi(1)}^{-1}, \dots, x_{\pi(d)}^{-1})$$

by (8), hence we conclude from (9) that ν_C is exchangeable if and only if C is exchangeable.

(ii) Suppose first that $\nu_1 = \dots = \nu_d$ and that ν admits an exchangeable associated Lévy copula C . Then $U_{\nu_1} = \dots = U_{\nu_d}$ and it follows from (7) that

$$\begin{aligned} & (P_\pi \nu)([x_1, \infty) \times \dots \times [x_d, \infty)) \\ &= U_\nu(x_{\pi(1)}, \dots, x_{\pi(d)}) \\ &= C(U_{\nu_1}(x_{\pi(1)}), \dots, U_{\nu_d}(x_{\pi(d)})) \\ &= C(U_{\nu_1}(x_1), \dots, U_{\nu_1}(x_d)) \\ &= U_\nu(x_1, \dots, x_d) = \nu([x_1, \infty) \times \dots \times [x_d, \infty)) \quad \forall x_1, \dots, x_d \in [0, \infty], \pi \in [d]. \end{aligned}$$

Hence, ν is exchangeable.

Conversely, suppose that ν is an exchangeable Lévy measure. Then it is easy to see that $\nu_1 = \dots = \nu_d$, hence $U_{\nu_1} = \dots = U_{\nu_d}$. If now additionally $U_{\nu_1}([0, \infty]) = [0, \infty]$, then for every $y \in [0, \infty]$ there exists some $U_{\nu_1}^\leftarrow(y) \in [0, \infty]$ such that $U_{\nu_1}(U_{\nu_1}^\leftarrow(y)) = y$. For $y_1, \dots, y_d \in [0, \infty]$ we then have necessarily by (7)

$$C(y_1, \dots, y_d) = U_\nu(U_{\nu_1}^\leftarrow(y_1), \dots, U_{\nu_1}^\leftarrow(y_d)),$$

and exchangeability of C follows from (9).

Many positive Lévy copulas used in practice are *Archimedean Lévy copulas*. They are of the form

$$C(x_1, \dots, x_d) = \phi^{-1}(\phi(x_1) + \dots + \phi(x_d))$$

for some strictly decreasing function $\phi : [0, \infty] \rightarrow [0, \infty]$ (the *generator*) such that $\phi(0) = \infty$, $\phi(\infty) = 0$, and ϕ^{-1} has derivatives up to order d on $(0, \infty)$ with alternating signs, see Cont and Tankov [14, Proposition 5.7]. It is evident that Archimedean Lévy copulas are exchangeable. We hence have

Corollary 2 *Let ν be a Lévy measure on $[0, \infty)^d$ which admits an Archimedean Lévy copula. Then ν is exchangeable if and only if $\nu_1 = \dots = \nu_d$.*

Proof This is immediate from Theorem 4 and its proof.

3 Matrices that Preserve Exchangeability

In this section we summarise some results of Dean and Verducci [15] and Commenges [13] about matrices that preserve (second order) exchangeability, which will be needed later. Denote by $J_d \in \mathbb{R}^{d \times d}$ the matrix with all entries equal to 1 and recall that $\text{Id}_d \in \mathbb{R}^{d \times d}$ denotes the identity matrix.

Theorem 5 (Commenges [13])

- (i) A matrix $A \in \mathbb{R}^{d \times d}$ commutes with permutations, if and only if there are $a, b \in \mathbb{R}$ such that

$$A = a \text{Id}_d + b J_d.$$

- (ii) Suppose $A, B \in \mathbb{R}^{d \times d}$ commute with permutations. Then $A + B$ and AB commute with permutations.
- (iii) If $A \in \mathbb{R}^{d \times d}$ commutes with permutations and $\det A \neq 0$, then A^{-1} commutes with permutations.
- (iv) Let $C \in \mathbb{R}^{d \times d}$. Then $\mathcal{L}(CX)$ is exchangeable for every exchangeable normal distribution $\mathcal{L}(X)$ on $\mathbb{R}^d$ if and only if C can be represented in the form $C = A Q$, where $A \in \mathbb{R}^{d \times d}$ commutes with permutations and $Q = (q_{ij})_{i,j=1,\dots,d} \in \mathbb{R}^{d \times d}$ is an orthogonal matrix that satisfies $\sum_{j=1}^d q_{ij} = \sum_{j=1}^d q_{1j}$ for all $i \in \{1, \dots, d\}$.

Commenges [13] calls property (iv) above the *preservation of second moment exchangeability*. For later use, we mention the following consequence.

Corollary 3 Let $A \in \mathbb{R}^{d \times d}$ commute with permutations. Then also e^A commutes with permutations. If A is additionally positive definite, then the converse also holds.

Proof Since $e^A = \sum_{k=0}^{\infty} \frac{1}{k!} A^k$, the first statement is immediate from Theorem 5(ii). For the converse suppose that $e^A \in \mathbb{R}^{d \times d}$ commutes with permutations. Then for every $\pi \in [d]$

$$\exp(A) = P_{\pi} \exp(A) P_{\pi^{-1}} = \sum_{k=0}^{\infty} \frac{P_{\pi} A^k P_{\pi^{-1}}}{k!} = \sum_{k=0}^{\infty} \frac{(P_{\pi} A P_{\pi^{-1}})^k}{k!} = \exp(P_{\pi} A P_{\pi^{-1}}).$$

Thus, we proved $\exp(A) = \exp(P_{\pi} A P_{\pi^{-1}})$. By positive definiteness of A , Proposition 11.2.9 of [7] implies $A = P_{\pi} A P_{\pi^{-1}}$.

While matrices that preserve exchangeability of *normal* random vectors have been characterized in Theorem 5, the corresponding question which linear transformations preserve exchangeability of any random vector has been solved already earlier by Dean and Verducci [15].

Definition 4 Let $A \in \mathbb{R}^{d \times d}$. Then A is said to be *exchangeability preserving*, if $\mathcal{L}(AX)$ is exchangeable for every exchangeable random vector X in $\mathbb{R}^d$. Denote

$$\begin{aligned} \mathcal{E}_d &= \{A \in \mathbb{R}^{d \times d} : A \text{ is exchangeability preserving}\}, \\ \mathcal{E}_d^0 &= \{A \in \mathbb{R}^{d \times d} : A \text{ is exchangeability preserving and } \det(A) \neq 0\}. \end{aligned}$$

Dean and Verducci [15, Theorem 4, Corollaries 1.2, 1.3] gave the following characterization of exchangeability preserving matrices:

Theorem 6 (Dean and Verducci) *A matrix $A \in \mathbb{R}^{d \times d}$ is exchangeability preserving if and only if for every $\pi \in [d]$ there exists $\pi' \in [d]$ such that*

$$P_\pi A = A P_{\pi'}.$$

Further, $\mathcal{E}_d^0$ can be characterized as

$$\mathcal{E}_d^0 = \{A \in \mathbb{R}^{d \times d} : \exists a, b \in \mathbb{R}, a \neq 0, a \neq -db, \pi \in [d] \text{ such that } A = a P_\pi + b J_d\}.$$

Notice that for $d = 2$ the set $\mathcal{E}_2^0$ of invertible exchangeability preserving matrices is precisely the set of invertible matrices which commute with permutations. Dean and Verducci [15, Corollary 1.1] also gave an explicit expression for $\mathcal{E}_d$, which is however more complicated to formulate. However, it is easy to see that

$$\{A \in \mathbb{R}^d : \exists a, b \in \mathbb{R}, \pi \in [d] : A = a P_\pi + b J_d\} \subset \mathcal{E}_d.$$

4 Exchangeability Preserving Transformations

In this section we study more general transformations that preserve exchangeability, not necessarily linear ones. We define:

Definition 5 Let $\mathcal{M}_1$ and $\mathcal{M}_2$ be two classes of measures on $\mathbb{R}^d$ and $G : \mathcal{M}_1 \rightarrow \mathcal{M}_2$ a mapping. We say that G

- (i) *is exchangeability preserving* if $G(\mu)$ is exchangeable whenever μ is exchangeable,
- (ii) *commutes weakly with permutations*, if for every $\mu \in \mathcal{M}_1$ and $\pi \in [d]$ there exists $\pi' \in [d]$ such that $P_{\pi'} \mu \in \mathcal{M}_1$ and

$$P_\pi G(\mu) = G(P_{\pi'} \mu),$$

- (iii) *commutes with permutations* if $P_\pi \mu \in \mathcal{M}_1$ for all $\mu \in \mathcal{M}_1$ and $\pi \in [d]$, and

$$P_\pi G(\mu) = G(P_\pi \mu) \quad \forall \mu \in \mathcal{M}_1, \quad \pi \in [d].$$

When $\mathcal{M}_1 = \mathcal{M}_2$ is the class of all probability distributions on $\mathbb{R}^d$, and G is induced by a linear mapping $A : \mathbb{R}^d \rightarrow \mathbb{R}^d$, i.e. $G(\mu) := A\mu$, the image measure of μ under A , then G commutes with permutations if and only if A does in the sense of Definition 2, and G is exchangeability preserving if and only if A is in the sense of Definition 4. Further, by Theorem 6, in this case G is exchangeability preserving if and only if it commutes weakly with permutations. For general mappings that are not necessarily linear this is no longer true, as the following simple example shows:

Example 1 (i) If the set $\mathcal{M}_1$ does not contain any exchangeable measure, then any $G : \mathcal{M}_1 \rightarrow \mathcal{M}_2$ will be exchangeability preserving, but it will commute weakly with permutations only in special cases.

(ii) Let $\mathcal{M}_1 = \mathcal{M}_2$ be the class of all probability distributions on $\mathbb{R}^d$, and define

$$G(\mu) := \begin{cases} \delta_{(0, \dots, 0)^T}, & \text{if } \mu \text{ is exchangeable,} \\ \delta_{(1, 0, \dots, 0)^T}, & \text{if } \mu \text{ is not exchangeable,} \end{cases}$$

where δ_x denotes the Dirac measure at x . Then G is exchangeability preserving, but does not commute weakly with permutations, since $P_\pi \delta_{(1, 0, \dots, 0)^T}$ is not in the range of G for π a permutation different from the identity.

(iii) If $\mathcal{M}_2$ contains only exchangeable measures, then any $G : \mathcal{M}_1 \rightarrow \mathcal{M}_2$ is exchangeability preserving and it even commutes weakly with permutations, since for all $\pi \in [d]$ and $\mu \in \mathcal{M}_1$ we can choose the identity for π' and obtain $P_\pi G(\mu) = G(\mu) = G(P_{\pi'} \mu)$.

So we have seen that there are mappings that are exchangeability preserving but do not commute weakly with permutations. On the other hand, it is easy to see that any transformation that commutes weakly with permutations is exchangeability preserving. More precisely, we have:

Proposition 1 *Let $\mathcal{M}_1$ and $\mathcal{M}_2$ be two classes of measures on $\mathbb{R}^d$. Then every mapping $G : \mathcal{M}_1 \rightarrow \mathcal{M}_2$ that commutes with permutations also commutes weakly with permutations, and every mapping that commutes weakly with permutations is exchangeability preserving.*

Proof That every mapping which commutes with permutations commutes weakly with permutations is clear. Now let $G : \mathcal{M}_1 \rightarrow \mathcal{M}_2$ commute weakly with permutations and let $\mu \in \mathcal{M}_1$ be exchangeable. Let $\pi \in [d]$. Then there exists $\pi' \in [d]$ such that

$$P_\pi G(\mu) = G(P_{\pi'} \mu) = G(\mu),$$

which shows that $G(\mu)$ is exchangeable.

Dean and Verducci [15, Corollary 1.3, Theorem 4] showed that an invertible matrix $A \in \mathbb{R}^{d \times d}$ is exchangeability preserving if and only if its inverse A^{-1} is exchangeability preserving. This then trivially transfers to the mappings G and G^{-1} induced by A and A^{-1} . For general transformations G , such a result is not true:

Example 2 Let $G : \mathcal{M}_1 \rightarrow \mathcal{M}_2$ be bijective and assume that $\mathcal{M}_2$ consists only of exchangeable measures, while $\mathcal{M}_1$ contains at least one non-exchangeable measure. Then G commutes weakly with permutations by Example 1(iii), but the inverse G^{-1} is not exchangeability preserving, in particular it cannot commute weakly with permutations.

When G commutes with permutations and G is bijective, then however also the inverse commutes with permutations:

Theorem 7 *Let $\mathcal{M}_1$ and $\mathcal{M}_2$ be two classes of measures on $\mathbb{R}^d$ and $G : \mathcal{M}_1 \rightarrow \mathcal{M}_2$ an injective mapping that commutes with permutations. Then its inverse $G^{-1} : G(\mathcal{M}_1) \rightarrow \mathcal{M}_1$ also commutes with permutations, in particular G^{-1} is exchangeability preserving.*

Proof Let G be injective. Let $\pi \in [d]$ and $\nu \in G(\mathcal{M}_1)$. Let $\mu := G^{-1}(\nu)$. Then $P_\pi \mu \in \mathcal{M}_1$ and

$$P_\pi \nu = P_\pi G(\mu) = G(P_\pi \mu),$$

showing that $P_\pi \nu \in G(\mathcal{M}_1)$. Applying G^{-1} to the equation above gives

$$G^{-1}(P_\pi \nu) = P_\pi \mu = P_\pi G^{-1}(\nu).$$

Hence G^{-1} commutes with permutations.

In the next two sections we give some examples of exchangeability preserving mappings.

5 Exchangeability Preserving Transformations Based on Discrete Time

In this section we discuss some transformations related to time series analysis and their exchangeability preserving property.

Example 3 (Convolution with an exchangeable distribution) Let $\rho = \mathcal{L}(X)$ be an exchangeable distribution on $\mathbb{R}^d$ and let $\mathcal{M}_1 = \mathcal{M}_2$ be the class of all probability distributions on $(\mathbb{R}^d, \mathcal{B}_d)$. Then the mapping

$$G_\rho : \mathcal{M}_1 \rightarrow \mathcal{M}_2, \quad \mu \mapsto \mu * \rho$$

commutes with permutations, in particular is exchangeability preserving, since for $\pi \in [d]$ and $B \in \mathcal{B}_d$ we have

$$\begin{aligned} (P_\pi(\mu * \rho))(B) &= \int_{\mathbb{R}^d} \mu(P_\pi^{-1}(B) - x) \rho(dx) \\ &= \int_{\mathbb{R}^d} (P_\pi \mu)(B - P_\pi x) \rho(dx) \\ &= \int_{\mathbb{R}^d} (P_\pi \mu)(B - y) (P_\pi \rho)(dy) = ((P_\pi \mu) * \rho)(B). \end{aligned}$$

- (i) Now assume that the characteristic function $z \mapsto \widehat{\rho}(z)$ is different from 0 for z from a dense subset of $\mathbb{R}^d$, e.g. if ρ is infinitely divisible (cf. [22, Lemma 7.5]). Since $\widehat{G_\rho(\mu)}(z) = \widehat{\rho}(z) \widehat{\mu}(z)$, it follows that G_ρ is injective and hence the

inverse $G_\rho^{-1} : G_\rho(\mathcal{M}_1) \rightarrow \mathcal{M}_2$ commutes with permutations by Theorem 7. Hence $\rho * \mu$ is exchangeable if and only if μ is exchangeable, provided ρ is exchangeable and $\widehat{\rho}$ does not vanish on a dense set. Similarly, using the Laplace transform to establish injectivity, if ρ and μ are both concentrated on $[0, \infty)^d$ and ρ is exchangeable, then $\rho * \mu$ is exchangeable if and only if μ is exchangeable.

- (ii) Let X_1, X_2, X_3, Y_1 be four independent one-dimensional random variables such that

$$\mathcal{L}(X_1) = \mathcal{L}(X_2) = \mathcal{L}(X_3) \neq \mathcal{L}(Y_1),$$

but

$$\mathcal{L}(X_1) * \mathcal{L}(Y_1) = \mathcal{L}(X_1) * \mathcal{L}(X_1).$$

Examples of such distributions can be found in Feller [16, p. 506]. Now let $d = 2$ and consider the two-dimensional distributions

$$\rho := \mathcal{L}((X_1, X_2)^T), \quad \text{and} \quad \mu := \mathcal{L}((Y_1, X_3)^T).$$

Then ρ and

$$\rho * \mu = \mathcal{L}((X_1 + Y_1, X_2 + X_3)^T) = \rho * \rho$$

are exchangeable, but μ is not exchangeable. Hence without extra assumptions on the exchangeable ρ (as done e.g. in (i)), it is not true that $\rho * \mu$ is exchangeable if and only if μ is exchangeable.

- (iii) It is worth noting that the convolution of two non-exchangeable distributions can be exchangeable. To see this, let $d = 2$ and let X_1, X_2, Y_1, Y_2 be independent one-dimensional random variables such that $\mathcal{L}(X_1) = \mathcal{L}(Y_2) \neq \mathcal{L}(Y_1) = \mathcal{L}(X_2)$. Define the distributions $\mu = \mathcal{L}((X_1, X_2)^T)$ and $\rho = \mathcal{L}((Y_1, Y_2)^T)$. Then $\mu * \rho = \mathcal{L}((X_1 + Y_1, X_2 + Y_2)^T)$ is exchangeable, although neither μ nor ρ are exchangeable.

Next we shall consider distributions that arise from infinite moving average processes. We first note that exchangeability is closed under weak convergence:

Lemma 1 *Let $(X_n)_{n \in \mathbb{N}}$ be a sequence of exchangeable random vectors in $\mathbb{R}^d$ that converges in distribution to a random vector Y as $n \rightarrow \infty$. Then Y is exchangeable.*

Proof The convergence $X_n \xrightarrow{d} Y$ for $n \rightarrow \infty$ implies $P_\pi X_n \xrightarrow{d} P_\pi Y$. Since $P_\pi X_n \stackrel{d}{=} X_n$ by assumption the two limits coincide.

Corollary 4 (Two sided moving average processes) *Let $(Z_t)_{t \in \mathbb{Z}}$ be an i.i.d. sequence of exchangeable random vectors in $\mathbb{R}^d$ and $(A_t)_{t \in \mathbb{Z}}$ be an i.i.d. sequence, independent of $(Z_t)_{t \in \mathbb{Z}}$, taking values in $\mathcal{E}_d$, the space of exchangeability preserving $(d \times d)$ -matrices. Assume that $\sum_{j=-\infty}^{\infty} A_j Z_{t-j}$ converges almost surely (equivalently, in distribution as a sum with independent increments). Then $\sum_{j=-\infty}^{\infty} A_j Z_{t-j}$ is exchangeable.*

Proof Let ρ be the distribution of A_j . Conditioning on $A_j = M$ for $M \in \mathcal{E}_d$, we have for the characteristic function

$$\varphi_{P_\pi A_j Z_{t-j}}(z) = \int_{\mathcal{E}_d} \varphi_{P_\pi M Z_{t-j}}(z) \rho(dM) = \int_{\mathcal{E}_d} \varphi_{M Z_{t-j}}(z) \rho(dM) = \varphi_{A_j Z_{t-j}}(z)$$

for all $z \in \mathbb{R}^d$ and $\pi \in [d]$, hence $A_j Z_{t-j}$ is exchangeable. Since $(A_j Z_{t-j})_{j \in \mathbb{Z}}$ is independent, Lemma 1 together with Example 3 shows that $\sum_{j=-\infty}^{\infty} A_j Z_{t-j}$ is exchangeable.

Under conditions which guarantee the existence of mean and covariance, we give a necessary condition for the stationary distribution of an infinite moving average process to be exchangeable:

Proposition 2 *Let $(Z_t)_{t \in \mathbb{Z}}$ be an i.i.d. sequence of $\mathbb{R}^d$ -valued random vectors with finite variance and $(A_k)_{k \in \mathbb{Z}}$ be an $\mathbb{R}^{d \times d}$ -valued deterministic sequence such that $(A_k^{ij})_{k \in \mathbb{Z}}$ is absolutely summable for each $i, j \in \{1, \dots, d\}$, where A_k^{ij} denotes the (i, j) -component of A_k . Then $\sum_{j=-\infty}^{\infty} A_j Z_{t-j}$ converges almost surely absolutely and in L^2 , and necessary conditions for $\sum_{j=-\infty}^{\infty} A_j Z_{t-j}$ to be exchangeable are that $(\sum_{j=-\infty}^{\infty} A_j) \mathbb{E} Z_0 = (\gamma_1, \dots, \gamma_1)^T$ for some $\gamma_1 \in \mathbb{R}$ and that $\sum_{k=-\infty}^{\infty} A_k \text{Cov}(Z_0) A_k^T$ commutes with permutations, where $\text{Cov}(Z_0) = \mathbb{E}(Z_0 Z_0^T) - \mathbb{E}(Z_0) \mathbb{E}(Z_0^T)$ denotes the covariance matrix of Z_0 .*

Proof Almost sure and L^2 -convergence under the stated conditions is well known, and $X := \sum_{j=-\infty}^{\infty} A_j Z_{t-j}$ has mean $(\sum_{j=-\infty}^{\infty} A_j) \mathbb{E} Z_0$ and covariance matrix

$$\sum_{k=-\infty}^{\infty} A_k \text{Cov}(Z_0) A_k^T.$$

If X is exchangeable, then $P_\pi X$ and X must share the same mean and covariance matrix for all $\pi \in [d]$, from which follows that $\mathbb{E} X = (\gamma_1, \dots, \gamma_1)^T$ for some $\gamma_1 \in \mathbb{R}$ and $P_\pi \text{Cov}(X) P_\pi^T = \text{Cov}(X)$, which is the claim.

Now consider the multivariate AR model of first order

$$Y_t - \Phi Y_{t-1} = Z_t, \quad t \in \mathbb{Z}, \quad (10)$$

where $\Phi \in \mathbb{R}^{d \times d}$ and $(Z_t)_{t \in \mathbb{Z}}$ is an i.i.d. sequence of d -dimensional random vectors. Necessary and sufficient conditions for the existence of a strictly stationary solution to this equation have been derived by Brockwell et al. [10, Theorem 1]. For simplicity, we shall assume that all eigenvalues of Φ lie in the open unit ball $\{z \in \mathbb{C} : |z| < 1\}$. Then a sufficient condition for the existence of a strictly stationary solution of (10) is $\mathbb{E} \log^+ |Z_0| < \infty$, in which case the stationary solution is unique and given by

$$Y_t = \sum_{k=0}^{\infty} \Phi^k Z_{t-k}, \quad t \in \mathbb{Z}, \quad (11)$$

where the right-hand side converges almost surely absolutely. If Φ is additionally invertible, then the condition $\mathbb{E} \log^+ |Z_0| < \infty$ is also necessary for the existence of a strictly stationary solution, see [10, Corollary 1].

Theorem 8 (Stationary solution of AR(1) equation) *Let $\Phi \in \mathbb{R}^{d \times d}$ such that all eigenvalues of Φ lie in $\{z \in \mathbb{C} : |z| < 1\}$. Let $\mathcal{M}_1$ be the set of all probability distributions $\mathcal{L}(X)$ on $\mathbb{R}^d$ with $\mathbb{E} \log^+ |X| < \infty$ (i.e. with finite log-moment) and $\mathcal{M}_2$ be the set of all probability distributions on $\mathbb{R}^d$. Consider the mapping*

$$G_\Phi : \mathcal{M}_1 \rightarrow \mathcal{M}_2, \quad \mathcal{L}(Z_0) \mapsto \mathcal{L}\left(\sum_{k=0}^{\infty} \Phi^k Z_{-k}\right), \quad t \in \mathbb{Z},$$

where $(Z_{-k})_{k \in \mathbb{N}_0}$ is an i.i.d. sequence with distribution $\mathcal{L}(Z_0)$, so that G_Φ associates to each $\mathcal{L}(Z_0)$ the distribution of the corresponding stationary solution of the AR(1) Eq. (10). We then have:

- (i) If Φ is exchangeability preserving (commutes with permutations), then G_Φ is exchangeability preserving (commutes with permutations), so that Y_t given by (11) is exchangeable whenever $\mathcal{L}(Z_0)$ is exchangeable.
- (ii) Let $\mathcal{M}'_1$ be the subset of all infinitely divisible $\mu \in \mathcal{M}_1$, and denote by G'_Φ the restriction of G_Φ to $\mathcal{M}'_1$. Assume that Φ commutes with permutations. Then G'_Φ commutes with permutations, G'_Φ is injective, and the inverse $(G'_\Phi)^{-1} : G_\Phi(\mathcal{M}'_1) \rightarrow \mathcal{M}'_1$ commutes with permutations. In particular, for $\mathcal{L}(Z_0) \in \mathcal{M}'_1$, $\mathcal{L}(\sum_{k=0}^{\infty} \Phi^k Z_{-k})$ is exchangeable if and only if $\mathcal{L}(Z_0)$ is exchangeable.
- (iii) Suppose that $(Z_t)_{t \in \mathbb{Z}}$ is i.i.d. $N(0, \Sigma)$ -distributed. Then a necessary and sufficient condition for exchangeability of $\mathcal{L}(\sum_{k=0}^{\infty} \Phi^k Z_{-k})$ is that $\sum_{k=0}^{\infty} \Phi^k \Sigma (\Phi^T)^k$ commutes with permutations.

Proof (i) If Φ is exchangeability preserving this is an immediate consequence of Corollary 4, since with Φ obviously also Φ^k is exchangeability preserving for each $k \in \mathbb{N}_0$. In the case when $P_\pi \Phi = \Phi P_\pi$ for all $\pi \in [d]$, we have also $P_\pi \Phi^k = \Phi^k P_\pi$ for all $k \in \mathbb{N}_0$, hence

$$P_\pi \sum_{k=0}^{\infty} \Phi^k Z_{-k} = \sum_{k=0}^{\infty} \Phi^k P_\pi Z_{-k},$$

which shows that G_Φ commutes with permutations.

(ii) That G'_Φ commutes with permutations is easy to see from (i). Let $Y_t := \sum_{k=0}^{\infty} \Phi^k Z_{t-k}$. Then Y_t is infinitely divisible and hence its characteristic function φ_Y has no zeros. By (10),

$$\varphi_Y(z) = \varphi_Y(\Phi^T z) \varphi_Z(z) \quad \forall z \in \mathbb{R}^d,$$

where φ_Z is the characteristic function of Z_t . Hence G'_Φ is injective. The rest follows from Theorem 7.

(iii) By (11), we have $\mathcal{L}(Y_t) = N(0, \sum_{k=0}^{\infty} \Phi^k \Sigma (\Phi^T)^k)$. The result then follows from Remark 1.

The following example shows that some conditions on $\mathcal{L}(Z_0)$ are needed in order for the assertion in Theorem 8(ii) to hold:

Example 4 Let $(U_t)_{t \in \mathbb{Z}}$, $(V_t)_{t \in \mathbb{Z}}$ and $(W_t)_{t \in \mathbb{Z}}$ be one-dimensional independent i.i.d. sequences such that U_0 and V_0 have characteristic function

$$\varphi_{U_0}(z) = \varphi_{V_0}(z) = \begin{cases} 1 - |z|, & |z| \leq 1, \\ 0, & |z| > 1, \end{cases}$$

and $\varphi_{W_0}(z) = \varphi_{U_0}(z)$ for $|z| \leq 1$ and otherwise φ_{W_0} being periodic with period 4. These functions are indeed characteristic functions of random variables, see Feller [16, p. 506]. Observe that U_0 has density $\mathbb{R} \rightarrow \mathbb{R}$, $x \mapsto \pi^{-1}(1 - \cos(x))x^{-2}$ [16, p. 503], hence $\mathbb{E} \log^+ |U_0| < \infty$ and it follows that $\sum_{k=0}^{\infty} 2^{-k} U_{-k}$ converges almost surely. Hence $\prod_{k=0}^{\infty} \varphi_{U_0}(2^{-k} z)$ converges pointwise to some characteristic function, and it is easy to see that $\prod_{k=0}^{\infty} \varphi_{U_0}(2^{-k} z) = \prod_{k=0}^{\infty} \varphi_{W_0}(2^{-k} z)$ for $z \in \mathbb{R}$. By Lévy's continuity theorem, $\sum_{k=0}^{\infty} 2^{-k} W_{-k}$ converges in distribution, hence almost surely. In particular, $\mathbb{E} \log^+ |W_0| < \infty$ and $\mathcal{L}(\sum_{k=0}^{\infty} 2^{-k} U_{-k}) = \mathcal{L}(\sum_{k=0}^{\infty} 2^{-k} W_{-k})$. Now let $d = 2$,

$$\Phi = \begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix}, \quad Z_t = (U_t, W_t)^T, \quad \text{and} \quad X_t = (U_t, V_t)^T.$$

Then $\mathcal{L}(\sum_{k=0}^{\infty} \Phi^k Z_{-k}) = \mathcal{L}(\sum_{k=0}^{\infty} \Phi^k X_{-k})$ is exchangeable and Φ is a diagonal matrix, but $\mathcal{L}(Z_0)$ is not exchangeable.

Observe that exchangeability of $\mathcal{L}(Y_t)$ and $\mathcal{L}(Z_t)$ in (11) does not imply that Φ is exchangeability preserving, as can be seen in the next example.

Example 5 Let $(Z_t)_{t \in \mathbb{Z}}$ be an i.i.d sequence of two-dimensional $N(0, \text{Id}_2)$ -distributed random vectors and let $\Phi \in \mathbb{R}^{2 \times 2}$ be of the form $\Phi = \begin{pmatrix} a & b \\ 0 & c \end{pmatrix}$ with $0 < a < c < 1$, and let Y_t be defined by (11). Then easy calculations show that

$$\Phi^k = \begin{pmatrix} a^k & b \frac{a^k - c^k}{a - c} \\ 0 & c^k \end{pmatrix} \quad \text{and} \quad \sum_{k=0}^{\infty} \Phi^k (\Phi^T)^k = \sum_{k=0}^{\infty} \begin{pmatrix} a^{2k} + b^2 \left(\frac{a^k - c^k}{a - c} \right)^2 & c^k b \frac{a^k - c^k}{a - c} \\ c^k b \frac{a^k - c^k}{a - c} & c^{2k} \end{pmatrix}.$$

By Theorem 8(iii) and Theorem 5(i) we conclude that $\mathcal{L}(Y_t)$ is exchangeable if and only if $\sum_{k=0}^{\infty} \left(a^{2k} + b^2 \left(\frac{a^k - c^k}{a - c} \right)^2 \right) = \sum_{k=0}^{\infty} c^{2k}$. Therefore, setting

$$b^2 := (a - c)^2 \frac{\sum_{k=0}^{\infty} c^{2k} - \sum_{k=0}^{\infty} a^{2k}}{\sum_{k=0}^{\infty} (a^k - c^k)^2} = \frac{1 - ac}{1 + ac} (c^2 - a^2)$$

gives exchangeability of $\mathcal{L}(Y_t)$. However, from Theorem 6 it is easy to see that Φ is not exchangeability preserving (observe that Φ is invertible).

Now we consider random recurrence equations. Let $(A_t, Z_t)_{t \in \mathbb{Z}}$ be an $\mathbb{R}^{d \times d} \times \mathbb{R}^d$ -valued i.i.d. sequence. Suppose that

$$E \log^+ |Z_1| < \infty, \quad E \log^+ \|A_1\| < \infty, \quad \text{and} \quad \gamma := \inf_{n \in \mathbb{N}} \left\{ \frac{1}{n} \mathbb{E} \log \|A_1 A_2 \cdots A_n\| \right\} < 0, \quad (12)$$

where $\|A\| := \sup_{|x|=1} |Ax|$ for $A \in \mathbb{R}^{d \times d}$ denotes the matrix norm induced by the Euclidean vector norm $|x| = \sqrt{x^T x}$, $x \in \mathbb{R}^d$. The quantity γ is called the *top Lyapunov exponent* of the sequence $(A_t)_{t \in \mathbb{Z}}$. Under these conditions, there exists a unique strictly stationary solution $(X_t)_{t \in \mathbb{Z}}$ of the random recurrence equation

$$X_t = A_t X_{t-1} + Z_t, \quad t \in \mathbb{Z}, \quad (13)$$

and it is given by

$$X_t = \sum_{k=0}^{\infty} A_t A_{t-1} \cdots A_{t-k+1} Z_{t-k}, \quad (14)$$

where the sum converges almost surely absolutely. In particular, the unique strictly stationary solution is non-anticipative in the sense that $(X_s)_{s \leq t}$ is independent of $(A_s, Z_s)_{s \geq t+1}$. Further, for any $\mathbb{R}^d$ -valued random variable V_0 on the same probability space, let $(V_n)_{n \in \mathbb{N}}$ be defined recursively by

$$V_n = A_n V_{n-1} + Z_n, \quad n \in \mathbb{N}. \quad (15)$$

Then V_n converges in distribution to $\mathcal{L}(X_0)$ as $n \rightarrow \infty$, where X_0 is given by (14), see Brandt [9, Theorem 1] for the one-dimensional case and Bougerol and Picard [8, Theorem 1.1] and Stelzer [25, Theorem 4.1] for the multivariate case. In [8, Theorem 2.5], under finite log-moment conditions and certain irreducibility conditions, a characterization for the existence of strictly stationary non-anticipative solutions of (13) in terms of negativity of the top Lyapunov exponent is achieved.

Under the above conditions, we obtain a mapping $G : \mathcal{L}(Z_0) \mapsto \mathcal{L}(X_0)$ as described below, where X_0 is given by (14).

Theorem 9 *Let $(A_t)_{t \in \mathbb{Z}}$ be an i.i.d. sequence in $\mathbb{R}^{d \times d}$ such that (12) is satisfied (i.e. $E \log^+ \|A_1\| < \infty$ and $\gamma < 0$). Let $\mathcal{M}_1$ be the set of all probability distributions $\mathcal{L}(Z_0)$ with $\mathbb{E} \log^+ |Z_0| < \infty$, $\mathcal{M}_2$ be the set of all probability distributions, and let $G : \mathcal{M}_1 \rightarrow \mathcal{M}_2$ be defined by*

$$\mathcal{L}(Z_0) \mapsto \mathcal{L}(X_0)$$

where $(Z_t)_{t \in \mathbb{Z}}$ is an i.i.d. sequence with distribution $\mathcal{L}(Z_0)$, independent of $(A_t)_{t \in \mathbb{Z}}$, and X_0 is given by (14).

- (i) If $(A_t)_{t \in \mathbb{Z}}$ takes only values in the set $\mathcal{E}_d$ of exchangeability preserving matrices, then G is exchangeability preserving, i.e. $\mathcal{L}(X_0)$ is exchangeable provided that $\mathcal{L}(Z_0)$ is exchangeable.
- (ii) If $(A_t)_{t \in \mathbb{Z}}$ takes only values in the set of all matrices that commute with permutations, then G commutes with permutations.
- (iii) Suppose that $(A_t)_{t \in \mathbb{Z}}$ takes only values in the set of matrices that commute with permutations and that have only non-negative entries. Let $\mathcal{M}'_1$ be the set of all $\mathcal{L}(Z_0) \in \mathcal{M}_1$ with distribution concentrated on $[0, \infty)^d$, and denote by G' the restriction of G to $\mathcal{M}'_1$. Then G' is injective, and its inverse $(G')^{-1} : G'(\mathcal{M}'_1) \rightarrow \mathcal{M}_1$ commutes with permutations. In particular, in this case, $\mathcal{L}(X_0)$ is exchangeable if and only if $\mathcal{L}(Z_0)$ is exchangeable.

Proof (i) Suppose that $\mathcal{L}(Z_0) \in \mathcal{M}_1$ is exchangeable. Let $V_0 := Z_0$ and define V_n recursively by (15). Since $\mathcal{L}(V_n)$ converges in distribution to $\mathcal{L}(X_0)$, by Lemma 1 it is enough to show that each V_n is exchangeable. We do that by induction on n . Observe that V_0 is exchangeable and independent of $(A_t, Z_t)_{t \geq 1}$. Suppose that V_{n-1} is proved to be exchangeable. Observe that V_{n-1} is independent of (A_n, Z_n) . Since A_n takes values in the space of exchangeability preserving matrices, by conditioning on A_n we see that $A_n V_{n-1}$ is exchangeable, similar to the proof of Corollary 4. Since Z_n is independent of $A_n V_{n-1}$, we conclude that $V_n = A_n V_{n-1} + Z_n$ is exchangeable by Example 3. This gives the claim.

(ii) By (14), for every $\pi \in [d]$ we have

$$P_\pi X_t = \sum_{k=0}^{\infty} P_\pi A_t \cdots A_{t-k+1} Z_{t-k} = \sum_{k=0}^{\infty} A_t \cdots A_{t-k+1} P_\pi Z_{t-k},$$

and $(P_\pi Z_t)_{t \in \mathbb{Z}}$ is i.i.d. with $\mathcal{L}(P_\pi Z_1) \in \mathcal{M}_1$. Hence G commutes with permutations.

(iii) Under the given conditions, it follows from (14) that also X_t is concentrated on $[0, \infty)^d$. Then also $A_t X_{t-1}$ is concentrated on $[0, \infty)^d$ and is independent of Z_t by (14). Hence we may take Laplace transforms in (13) and obtain

$$\mathbb{L}_{X_1}(u) = \mathbb{L}_{A_1 X_0}(u) \mathbb{L}_{Z_1}(u) \quad \forall u \in [0, \infty)^d,$$

where $\mathbb{L}_{X_1}(u) = \mathbb{E} e^{-X_1^T u}$ denotes the Laplace transform of X_1 for $u \in [0, \infty)^d$. From this it follows that G' is injective, and it obviously commutes with permutations by (ii). The claim then follows from Theorem 7.

6 Exchangeability Preserving Transformations Based on Continuous Time

In this section we consider stochastic integrals of the form

$$\int_0^T f(t) dL_t \quad \text{and} \quad \int_0^\infty f(t) dL_t := \text{dlim}_{T \rightarrow \infty} \int_0^T f(t) dL_t,$$

where $L = (L_t)_{t \geq 0}$ is a Lévy process in $\mathbb{R}^d$, f is an $\mathbb{R}^{d \times d}$ -valued stochastic process with càglàd paths, and dlim denotes the limit in distribution (provided it exists). Here, we assume implicitly that there is an underlying filtration $\mathbb{F} = (\mathcal{F}_t)_{t \geq 0}$ satisfying the usual conditions such that L is a semimartingale with respect to $\mathbb{F}$ and that $(f_t)_{t \geq 0}$ is adapted with respect to $\mathbb{F}$; if the processes f and L are independent, such a filtration always exists, see Protter [21, Theorem VI.2], so that $\int_0^T f(t) dL_t$ always exists in this case. See [21] also for the definition and facts about stochastic integration. When f is deterministic, then $T \mapsto \int_0^T f(t) dL_t$ has independent increments, and it follows that it converges in distribution as $T \rightarrow \infty$ if and only if it converges in probability (or even almost surely). For deterministic f , the integral coincides with integrals with respect to the induced independently scattered random measure as defined in Sato [23, Proposition 2.11, Example 2.12, Definitions 2.16, 2.20], and $V := \int_0^\infty f(t) dL_t$ is infinitely divisible (provided it exists); denoting by Ψ_L and Ψ_V the characteristic exponents of L_1 and V , we then have

$$\Psi_V(z) = \lim_{T \rightarrow \infty} \int_{[0, T]} \Psi_L(f(s)^T z) ds \quad \forall z \in \mathbb{R}^d, \quad (16)$$

see [23, Proposition 2.17]. A relation between the characteristic triplets of L and V can also be established in this case, as well as a characterization when $\int_0^\infty f(t) dL_t$ actually exists as a limit in probability; see [23, Proposition 2.17, Corollary 2.19]. If f is not deterministic but independent of L , then $\int_0^\infty f(t) dL_t$ needs not to be infinitely divisible, but its characteristic function can be calculated from (16) by conditioning on f .

Theorem 10 *Let $f = (f_t)_{t \geq 0}$ be an $\mathbb{R}^{d \times d}$ -valued stochastic process with càglàd paths. Let $\mathcal{M}_1$ be a set of infinitely divisible distributions μ , such that $\int_0^\infty f(t) dL_t^\mu$ is definable as a limit in distribution for each $\mu \in \mathcal{M}_1$, where $(L_t^\mu)_{t \geq 0}$ is a d -dimensional Lévy process with $\mathcal{L}(L_1^\mu) = \mu$, independent of f (possibly defined on a suitably enlarged probability space). Denote by $\mathcal{M}_2$ the set of all probability distributions on $\mathbb{R}^d$. This gives a mapping*

$$G : \mathcal{M}_1 \rightarrow \mathcal{M}_2, \quad \mu \mapsto \mathcal{L} \left(\int_0^\infty f(t) dL_t^\mu \right).$$

- (i) Suppose that f takes only values in the set $\mathcal{E}_d$ of exchangeability preserving matrices. Then G is exchangeability preserving, i.e. exchangeability of μ implies exchangeability of $G(\mu)$.
- (ii) Suppose that f takes only values in the set of matrices that commute with permutations. Then $\int_0^\infty f(t) dL_t^\mu$ exists as a limit in distribution if and only if $\int_0^\infty f(t) d(P_\pi L_t^\mu)$ exists as a limit in distribution for all $\pi \in [d]$, so that (by possibly enlarging $\mathcal{M}_1$) we can assume that $P_\pi \mu \in \mathcal{M}_1$ for all $\pi \in [d]$ and $\mu \in \mathcal{M}_1$. Then G commutes with permutations. If additionally G is injective, then the inverse $G^{-1} : G(\mathcal{M}_1) \rightarrow \mathcal{M}_1$ also commutes with permutations, so that in this case, $\int_0^\infty f(t) dL_t^\mu$ is exchangeable if and only if μ is exchangeable.

Proof (i) Let $V = \int_0^\infty f(t) dL_t^\mu$. Suppose first that f is deterministic and has compact support in $[0, T]$. Since $f(s)$ takes values in $\mathcal{E}_d$, for every $\pi \in [d]$ and $s \geq 0$ there exists $\pi'(s) \in [d]$ such that $f(s)^T P_\pi = P_{\pi'(s)} f(s)^T$. Hence

$$\Psi_V(P_\pi z) = \int_0^T \Psi_{L^\mu}(f(s)^T P_\pi z) ds = \int_0^T \Psi_{L^\mu}(P_{\pi'(s)} f(s)^T z) ds \quad \forall z \in \mathbb{R}^d.$$

Now if μ is exchangeable, then $\Psi_{L^\mu}(P_{\pi'(s)} z) = \Psi_{L^\mu}(z)$, and we conclude

$$\Psi_V(P_\pi z) = \int_0^T \Psi_{L^\mu}(f(s)^T z) ds = \Psi_V(z) \quad \forall z \in \mathbb{R}^d,$$

so that G is exchangeability preserving.

Next assume that f is not deterministic but has support in $[0, T]$ and its distribution is ρ . Then conditioning on the paths of f , and since f and L^μ are independent, we obtain

$$\begin{aligned} \varphi_{P_\pi \int_0^T f(t) dL_t^\mu}(z) &= \int_{\mathcal{E}_d} \varphi_{P_\pi \int_0^T M(t) dL_t^\mu}(z) \rho(dM) \\ &= \int_{\mathcal{E}_d} \varphi_{\int_0^T M(t) dL_t^\mu}(z) \rho(dM) = \varphi_{\int_0^T f(t) dL_t^\mu}(z), \quad z \in \mathbb{R}, \end{aligned}$$

where the second equality follows from the previous case, so that G must also be exchangeability preserving in this case.

Finally, the case when f is random (but independent) and does not have compact support follows from Lemma 1 since

$$\begin{aligned} \varphi_{P_\pi \int_0^\infty f(t) dL_t^\mu} &= \lim_{T \rightarrow \infty} \varphi_{P_\pi \int_0^T f(t) dL_t^\mu}(z) \\ &= \lim_{T \rightarrow \infty} \varphi_{\int_0^T f(t) dL_t^\mu}(z) = \varphi_{\int_0^\infty f(t) dL_t^\mu}(z), \quad z \in \mathbb{R}. \end{aligned}$$

(ii) Suppose that $f(t)$ commutes with permutations for each t . Consider first the case that f is deterministic and has support in $[0, T]$. For each $\pi \in [d]$, we then have

$$\begin{aligned}\Psi_{\int_0^T f(t) d(P_\pi L_t^\mu)}(z) &= \int_0^T \Psi_{P_\pi L_t^\mu}(f(s)^T z) ds = \int_0^T \Psi_{L_t^\mu}(P_\pi^T f(s)^T z) ds \\ &= \int_0^T \Psi_{L_t^\mu}(f(s)^T P_\pi^T z) ds = \Psi_{\int_0^T f(t) dL_t^\mu}(P_\pi^T z) \\ &= \Psi_{P_\pi \int_0^T f(t) dL_t^\mu}(z), \quad z \in \mathbb{R}^d.\end{aligned}$$

Now if f is stochastic with support in $[0, T]$ but independent of L^μ , then by conditioning on f we see that

$$\varphi_{\int_0^T f(t) d(P_\pi L_t^\mu)}(z) = \varphi_{\int_0^T f(t) dL_t^\mu}(P_\pi^T z) = \varphi_{P_\pi \int_0^T f(t) dL_t^\mu}(z), \quad z \in \mathbb{R}^d. \quad (17)$$

Since $\int_0^T f(t) d(P_\pi L_t^\mu)$ converges in distribution if and only if $\varphi_{\int_0^T f(t) d(P_\pi L_t^\mu)}$ converges pointwise to a characteristic function as $T \rightarrow \infty$, it follows that $\int_0^T f(t) d(P_\pi L_t^\mu)$ converges in distribution if and only if $\int_0^T f(t) dL_t^\mu$ does. Hence we can assume that $P_\pi \mu \in \mathcal{M}_1$ for all $\pi \in [d]$ and $\mu \in \mathcal{M}_1$. Taking the limit as $T \rightarrow \infty$ in (17) then shows that G commutes with permutations. The rest follows from Theorem 7.

Remark 2 (i) When restricting to deterministic f , we can use integration theory with respect to independently scattered random measures as in Sato [23], and the proof above carries easily over to deterministic, but not necessarily càglàd f , as long it is integrable with respect to L^μ .

(ii) Another approach to prove Theorem 10(ii) is to use that

$$P_\pi \int_0^T f(t) dL_t^\mu = \int_0^T f(t) d(P_\pi L_t^\mu),$$

as a consequence of approximating the stochastic integrals by Riemann sums (e.g. [21, Theorem II.21]) and the fact that $f(t)$ commutes with permutations.

Theorem 10 will be mostly used in the case when $f(t) = g(t)\text{Id}_d$ for a deterministic scalar valued càglàd function g . Observe that in that case, $f(t)$ obviously commutes with permutations. As a first application, we consider self-decomposable distributions.

Recall that a distribution μ on $\mathbb{R}^d$ is *self-decomposable* if and only if for each $b > 1$ there exists a probability measure ρ_b on $\mathbb{R}^d$ such that

$$\widehat{\mu}(z) = \widehat{\mu}(b^{-1}z)\widehat{\rho_b}(z) \quad \forall z \in \mathbb{R}^d$$

(e.g. [22, Definition 15.1]). It is well known that self-decomposable distributions constitute exactly the class of stationary distributions of Ornstein–Uhlenbeck processes.

More precisely, given $c > 0$, a distribution σ is self-decomposable if and only if there exists a Lévy process $L^\mu = (L_t^\mu)_{t \geq 0}$ with $\mu = \mathcal{L}(L_1^\mu)$ such that $\mathbb{E} \log^+ |L_1^\mu| < \infty$ (i.e. $\int_{\mathbb{R}} \log^+ |x| \mu(dx) < \infty$) and

$$\sigma = \mathcal{L} \left(\int_0^\infty e^{-ct} dL_t^\mu \right),$$

where the integral converges almost surely (equivalently, in distribution). It is known that $\mathbb{E} \log^+ |L_1^\mu| < \infty$ is a necessary and sufficient condition for convergence of $\int_0^\infty e^{-ct} dL_t^\mu$, and that the mapping

$$G : \mathcal{M}_1 \rightarrow \mathcal{M}_2, \quad \mu \mapsto \mathcal{L} \left(\int_0^\infty e^{-ct} dL_t^\mu \right), \quad (18)$$

where $\mathcal{M}_1$ is the class of all infinitely distributions μ on $\mathbb{R}^d$ with $\mathbb{E} \log^+ |L_1^\mu| < \infty$, and $\mathcal{M}_2$ is the class of all self-decomposable distributions, is a bijection; cf. Sato [22, Theorem 17.5]. For self-decomposable σ , the unique (in distribution) Lévy process L^μ such that $G(\mathcal{L}(L_1^\mu)) = \sigma$ is then called the *background driving Lévy process* of σ , or μ the *background driving infinitely divisible distribution* of σ .

The following is now immediate:

Corollary 5 (Exchangeable self-decomposable distributions) *With the notations above, the mapping $G : \mathcal{M}_1 \rightarrow \mathcal{M}_2$ defined in (18) as well as its inverse commute with permutations. In particular, a self-decomposable distribution is exchangeable if and only if its background driving infinitely divisible distribution is exchangeable (for fixed $c > 0$).*

Let us give yet another proof that a self-decomposable distribution can only be exchangeable if the background driving infinitely divisible distribution is exchangeable, without referring to the general result Theorem 7. For that, we need the following Lemma, which is interesting in its own right. It is well known in one dimension (e.g. Barndorff-Nielsen and Shephard [4, Eq. (4.15)]) and proved similarly in higher dimensions, but since we were not able to find a ready reference we provide a proof:

Lemma 2 *Let $c > 0$, $(L_t)_{t \geq 0}$ a Lévy process with $\mathbb{E} \log^+ |L_1| < \infty$ and $V = \int_0^\infty e^{-cs} dL_s$. Denote by Ψ_V and Ψ_L the characteristic exponents of V respectively of L_1 . Then for $\xi := \frac{z}{|z|}$, $z \in \mathbb{R}^d \setminus \{0\}$*

$$\frac{\partial \Psi_V(z)}{\partial \xi} = \frac{\Psi_L(z)}{c|z|}.$$

Proof We have $\Psi_V(z) = \int_0^\infty \Psi_L(e^{-cs} z) ds$ for $z \in \mathbb{R}^d$ as a limit by (16), but it even holds $\int_0^\infty |\Psi_L(e^{-cs} z)| ds < \infty$ (cf. [22], proof of Theorem 17.5). Letting $\xi := z/|z|$ for $z \neq 0$ and substituting $x = e^{-cs} |z|$, we obtain

$$\Psi_V(z) = \int_0^\infty \Psi_L(e^{-cs}|z|\xi) ds = \int_0^{|z|} \Psi_L(x\xi) \frac{dx}{cx}.$$

Differentiation with respect to ξ gives

$$\frac{\partial \Psi_V(z)}{\partial \xi} = \lim_{h \rightarrow 0} h^{-1} \int_{|z|}^{|z|+h} \Psi_L(x\xi) \frac{dx}{cx} = \frac{\Psi_L(z)}{c|z|},$$

which finishes the proof of Lemma 2.

Let us now show again that $V := \int_0^\infty e^{-ct} dL_t^\mu$ can only be exchangeable if μ is exchangeable. So suppose that V is exchangeable. Then $\Psi_V(z) = \Psi_V(P_\pi z)$ for all $\pi \in [d]$ and $z \in \mathbb{R}^d \setminus \{0\}$. Therefore for $\xi_\pi := P_\pi \xi = \frac{P_\pi z}{|z|}$,

$$\frac{\partial \Psi_V(z)}{\partial \xi} = \lim_{h \rightarrow 0} \frac{\Psi_V(z + \xi h) - \Psi_V(z)}{h} = \lim_{h \rightarrow 0} \frac{\Psi_V(P_\pi z + \xi_\pi h) - \Psi_V(P_\pi z)}{h} = \frac{\partial \Psi_V(P_\pi z)}{\partial \xi_\pi},$$

which implies by the previous lemma $\Psi_{L_\mu}(z) = \Psi_{L_\mu}(P_\pi z)$ for all $z \in \mathbb{R}^d$ and $\pi \in [d]$, showing that μ is exchangeable.

Next we generalise our results on self-decomposable distributions to A -decomposable distributions. Let $A \in \mathbb{R}^{d \times d}$ such that all eigenvalues of A have strictly positive real part. Then $\int_0^\infty e^{-As} dL_s^\mu$ is definable for a Lévy process L^μ if and only if $\mathbb{E} \log^+ |L_1^\mu| < \infty$. By Theorem 4.1 in Sato and Yamazato [24], the mapping

$$G : \mathcal{M}_1 \rightarrow \mathcal{M}_2, \quad \mu \mapsto \mathcal{L} \left(\int_0^\infty e^{-As} dL_s^\mu \right) \quad (19)$$

defines a bijection from the class $\mathcal{M}_1$ of all infinitely divisible distributions on $\mathbb{R}^d$ with finite log-moment to the class $\mathcal{M}_2$ of all A -decomposable distributions; here, a distribution σ on $\mathbb{R}^d$ is A -decomposable, if for every $t > 0$ there exists a probability measure ρ_t on $\mathbb{R}^d$ such that

$$\widehat{\sigma}(z) = \widehat{\sigma}(e^{-tA^T} z) \widehat{\rho}_t(z), \quad z \in \mathbb{R}^d.$$

The distribution $G(\mu)$ is then the unique stationary distribution of the Ornstein–Uhlenbeck process

$$dX_t(\omega) = dL_t^\mu(\omega) - AX_{t-}(\omega)dt, \quad t \geq 0.$$

All this can be found in [24]. If A commutes with permutations, then also e^{-As} commutes with permutations for all $s \geq 0$ by Corollary 3. Hence, the following result is immediate from Theorem 10:

Corollary 6 (Exchangeable A -decomposable distributions) *Let $A \in \mathbb{R}^{d \times d}$ such that all eigenvalues of A have strictly positive real part and that A commutes with per-*

mutations. With the notations above, the mapping $G : \mathcal{M}_1 \rightarrow \mathcal{M}_2$ defined in (19) as well as its inverse commute with permutations. In particular, $\int_0^\infty e^{-As} dL_s^\mu$ is exchangeable if and only if μ is exchangeable.

The distribution $\int_0^\infty e^{-As} dL_s$ can be exchangeable without A being exchangeability preserving, as follows from the following result applied to $A = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}$.

Proposition 3 *Suppose $(L_t)_{t \geq 0}$ is a standard Brownian motion, and A a normal matrix (i.e. $AA^T = A^T A$) such that all eigenvalues of A have only strictly positive real parts. Then $\int_0^\infty e^{-As} dL_s$ is exchangeable if and only if $A + A^T$ commutes with permutations.*

Proof Since A is normal, it follows that $e^{-As}e^{-A^Ts} = e^{-A^Ts}e^{-As} = e^{-(A+A^T)s}$, see [7, Proposition 11.2.8]. Therefore, the infinitely divisible distribution $\mu = \mathcal{L}(\int_0^\infty e^{-As} dL_s)$ is normal distributed with mean 0 and variance

$$B = \int_0^\infty e^{-sA} e^{-sA^T} ds = \int_0^\infty e^{-s(A+A^T)} ds = -(A + A^T)^{-1}.$$

By Remark 1, μ is exchangeable if and only if B commutes with permutations, which by Theorem 5(iii) is equivalent to the fact that $A + A^T$ commutes with permutations.

Let us consider some other mappings. Barndorff-Nielsen and Thorbjørnsen ([5] and others) have introduced the Upsilon-transform in dimension one, and this has been generalised by Barndorff-Nielsen et al. [2] to a multivariate setting. For an infinitely divisible distribution μ on $\mathbb{R}^d$ with associated Lévy process L^μ , the Upsilon-transform is defined by

$$\Upsilon(\mu) = \mathcal{L} \left(\int_0^1 \log \frac{1}{t} dL_t^\mu \right).$$

Observe that the integrand $t \mapsto \log t^{-1}$ has a singularity at 0 and hence is not càglàd, but Theorem 10 carries over by Remark 2. As shown in [2, Theorem A], Υ defines a bijection from the class of all infinitely divisible distributions to the Goldie-Steuel-Bondesson class $B(\mathbb{R}^d)$. By Theorem 10, we then have:

Theorem 11 *The Υ -transform as defined above commutes with permutations, as does its inverse. In particular, $\Upsilon(\mu)$ is exchangeable if and only if μ is exchangeable.*

By restricting Υ to the class of self-decomposable distributions, one obtains the Thorin class [2, Theorem B], and by composing Υ with the mapping G of (18), one obtains a bijection from the class of infinitely divisible distributions with finite log-moment to the Thorin class [2, Theorem C]. Results similar to Theorem 11 can then be stated for this composition.

We note that the Upsilon-transform has been generalised in various directions, sometimes acting on infinitely divisible distributions, sometimes acting directly

on Lévy measures. We just mention the general Upsilon-transforms of Barndorff-Nielsen et al. [3]: let ρ be a σ -finite measure on $(0, \infty)$ and for each σ -finite measure ν on $\mathbb{R}^d$ with $\nu(\{0\}) = 0$ define Υ_ρ as the positive measure on $\mathbb{R}^d$ given by

$$[\Upsilon_\rho(\nu)](B) = \int_0^\infty \nu(x^{-1}B) \rho(dx), \quad B \in \mathcal{B}_d;$$

Υ_ρ is called the *Upsilon transformation with dilation measure ρ* . Restricting the domain of Υ_ρ to the set of all σ -finite measures ν such that $\Upsilon_\rho(\nu)$ is a Lévy measure, it is easy to see that Υ_ρ commutes with permutations. The domain has been derived in [3, Sect. 3] in various cases. In [3, Sect. 6], the injectivity property of Υ_ρ has been further studied, and shown to be equivalent to the *cancellation property* of the multiplicative convolution; see [3, Eq. (6.1)] for details. It is clear that Theorem 10 applies for injective Υ_ρ .

Finally, we mention that a natural continuous time analogue of random recurrence equations with iid coefficients is the multivariate generalized Ornstein-Uhlenbeck process, as introduced in Behme and Lindner [6]. It would be interesting to investigate conditions under which an analogue of Theorem 9 holds for these processes, but we leave this for future research.

Acknowledgments We would like to thank the anonymous referee for valuable suggestions and for pointing out the Ref. [20] to us.

References

1. Barndorff-Nielsen, O.E., Lindner, A.: Lévy copulas: dynamics and transforms of Upsilon type. *Scand. J. Stat.* **34**, 298–316 (2007)
2. Barndorff-Nielsen, O.E., Maejima, M., Sato, K.: Some classes of multivariate infinitely divisible distributions admitting stochastic integral representations. *Bernoulli* **12**, 1–33 (2006)
3. Barndorff-Nielsen, O.E., Rosinski, J., Thorbjørnsen, S.: General Υ -transformations. *ALEA* **4**, 131–165 (2008)
4. Barndorff-Nielsen, O.E., Shephard, N.: Modelling by Lévy processes for financial econometrics. In: Barndorff-Nielsen, O.E., Mikosch, T., Resnick, S.R. (eds.) *Lévy Processes: Theory and Applications*. Birkhäuser, Boston (2001)
5. Barndorff-Nielsen, O.E., Thorbjørnsen, S.: A connection between free and classical infinite divisibility. *Infin. Dimens. Anal. Quantum Probab. Relat. Top.* **7**, 573–590 (2004)
6. Behme, A., Lindner, A.: Multivariate generalized Ornstein-Uhlenbeck processes. *Stoch. Process. Appl.* **122**, 1487–1518 (2012)
7. Bernstein, D.S.: *Matrix Mathematics : Theory, Facts, and Formulas*, 2nd edn. Princeton University Press, Princeton (2009)
8. Bougerol, P., Picard, N.: Strict stationarity of generalized autoregressive processes. *Ann. Probab.* **20**, 1714–1730 (1992)
9. Brandt, A.: The stochastic equation $Y_{n+1} = A_n Y_n + B_n$ with stationary coefficients. *Adv. Appl. Probab.* **18**, 211–220 (1986)
10. Brockwell, P., Lindner, A., Vollenbröker, B.: Strictly stationary solutions of multivariate ARMA equations with i.i.d. noise. *Ann. Inst. Stat. Math.* **64**, 1089–1119 (2012)

11. Chobanyan, S., Levental, S.: Contraction principle for tail probabilities of sums of exchangeable random vectors with multipliers. *Stat. Probab. Lett.* **83**, 1720–1724 (2013)
12. Chow, Y.S., Teicher, H.: *Probability Theory - Independence, Interchangeability, Martingales*, 3rd edn. Springer, New York (1997)
13. Commenges, D.: Transformations which preserve exchangeability and application to permutation tests. *J. Nonparametr. Stat.* **15**(2), 171–185 (2003)
14. Cont, R., Tankov, P.: *Financial Modelling with Jump Processes*. Chapman & Hall/CRC, Boca Raton (2004)
15. Dean, A.M., Verducci, J.S.: Linear transformations that preserve majorization, Schur concavity, and exchangeability. *Linear Algebra Appl.* **127**, 121–138 (1990)
16. Feller, W. *An Introduction to Probability Theory and its Applications*. Vol. II. 2nd ed., John Wiley & Sons, New York (1971)
17. Griffiths, R.C., Milne, R.K.: Structure of exchangeable infinitely divisible sequences of Poisson random vectors. *Stoch. Process. Appl.* **22**, 145–160 (1986)
18. Kallsen, J., Tankov, P.: Characterization of dependence of multidimensional Lévy processes using Lévy copulas. *J. Multivar. Anal.* **97**, 1551–1572 (2006)
19. Nelsen, R.B.: *An Introduction to Copulas*, 2nd edn. Springer, Berlin (2007)
20. Nguyen, T.T.: Conditional distributions and characterizations of multivariate stable distributions. *J. Multivar. Anal.* **53**, 181–193 (1995)
21. Protter, P.E.: *Stochastic Integration and Differential Equations*, 2nd edn. Springer, Berlin (2004)
22. Sato, K.: *Lévy Processes and Infinitely Divisible Distributions*. Cambridge University Press, Cambridge (1999)
23. Sato, K.: Additive processes and stochastic integrals. III. *J. Math.* **50**, 825–851 (2006)
24. Sato, K., Yamazato, M.: Operator-selfdecomposable distributions as limit distributions of processes of Ornstein-Uhlenbeck type. *Stoch. Process. Appl.* **17**, 73–100 (1984)
25. Stelzer, R.J. *On Markov-Switching Models – Stationarity and Tail Behaviour*. Diploma thesis, TU München (2005)
26. Tankov, P. *Dependence structure of Lévy processes with applications in risk management*. Rapport Interne 502, CMAP, Ecole Polytechnique (2003)
27. Trashorras, J.: Large deviations for a triangular array of exchangeable random variables. *Ann. Inst. H. Poincaré Probab. Stat.* **38**, 649–680 (2002)

Lévy Copulas: Review of Recent Results

Peter Tankov

Abstract We review and extend the now considerable literature on Lévy copulas. First, we focus on Monte Carlo methods and present a new robust algorithm for the simulation of multidimensional Lévy processes with dependence given by a Lévy copula. Next, we review statistical estimation techniques in a parametric and a non-parametric setting. Finally, we discuss the interplay between Lévy copulas and multivariate regular variation and briefly review the applications of Lévy copulas in risk management. In particular, we provide a new easy-to-use sufficient condition for multivariate regular variation of Lévy measures in terms of their Lévy copulas.

Keywords Lévy processes · Lévy copulas · Monte Carlo simulation · Statistical estimation · Risk management · Regular variation

1 Introduction

Introduced in [13, 31, 42], the concept of Lévy copula allows to characterize in a time-independent fashion the dependence structure of the pure jump part of a Lévy process. During the past ten years, several authors have proposed extensions of Lévy copulas, developed simulation and estimation techniques for these and related objects, and studied the applications of these tools to financial risk management. In this paper we review the early developments and the subsequent literature on Lévy copulas, present new simulation algorithms for Lévy processes with dependence given by a Lévy copula, discuss the link between Lévy copulas and multivariate regular variation and mention some risk management applications. The aim is to provide a summary of available tools and an entry point to the now considerable literature on Lévy copulas and more generally dependence models for multidimensional Lévy processes. We

P. Tankov (✉)

Laboratoire de Probabilités Et Modèles Aléatoires, Université Paris Diderot, Paris, France
e-mail: tankov@math.univ-paris-diderot.fr; peter.tankov@polytechnique.org

P. Tankov

International Laboratory of Quantitative Finance,
National Research University Higher School of Economics, Moscow, Russia

focus on practical aspects such as statistical estimation and Monte Carlo simulation rather than theoretical properties of Lévy copulas.

This chapter is structured as follows. In Sect. 2 we recall the main definitions and results from the theory of Lévy copulas and review some alternative constructions and dependence models proposed in the literature. Section 3 presents new algorithms for simulating Lévy processes with a given Lévy copula, via a series representation. Section 4 reviews the statistical procedures proposed in the literature for estimating Lévy copulas in the parametric or non-parametric setting. In Sect. 5 we discuss the interplay between these objects and multivariate regular variation. In particular, we present a new easy-to-use sufficient condition for multivariate regular variation of Lévy measures in terms of their Lévy copulas. In Sect. 6 we review the applications of Lévy copulas in risk management. Section 7 concludes the paper and discusses some directions for further research.

Remarks on notation In this chapter, the components of a vector are denoted by the same letter with superscripts: $X = (X^1, \dots, X^n)$. The scalar product of two vectors is written with angle brackets: $\langle X, Y \rangle = \sum_{i=1}^n X^i Y^i$, and the Euclidean norm of the vector X is denoted by $|X|$. The extended real line is denoted by $\mathbb{R} := (-\infty, \infty]$.

2 A Primer on Lévy Copulas

This section contains a brief review of the theory of Lévy copulas as exposed in [13, 31, 42]. We invite the readers to consult these references for additional details.

Recall that a Lévy process is a stochastic process with stationary and independent increments, which is continuous in probability. The law of a Lévy process $(X_t)_{t \geq 0}$ is completely determined by the law of X_t at any given time $t > 0$. The characteristic function of this law is given explicitly by the Lévy-Khintchine formula:

$$\begin{aligned} \mathbb{E}[e^{i\langle u, X_t \rangle}] &= e^{t\psi(u)}, \quad u \in \mathbb{R}^n, \\ \psi(u) &= -\frac{\langle Au, u \rangle}{2} + i\langle \gamma, u \rangle + \int_{\mathbb{R}^n} (e^{i\langle u, x \rangle} - 1 - i\langle u, x \rangle \mathbf{1}_{|x| \leq 1}) \nu(dx), \end{aligned}$$

where $\gamma \in \mathbb{R}^n$, A is a positive semi-definite $n \times n$ matrix and ν is a positive measure on $\mathbb{R}^n$ with $\nu(\{0\}) = 0$ such that $\int_{\mathbb{R}^n} (|x|^2 \wedge 1) \nu(dx) < \infty$. The triple (A, ν, γ) is called the *characteristic triple* of the Lévy process X .

The Lévy-Itô decomposition in turn gives a representation of the paths of X in terms of a Brownian motion and a Poisson random measure:

$$X_t = \gamma t + B_t + \int_0^t \int_{|x| \leq 1} x \tilde{J}(ds \times dx) + \int_0^t \int_{|x| > 1} x J(ds \times dx), \quad (1)$$

where B is a Brownian motion (centered Gaussian process with independent increments) with covariance matrix A at time $t = 1$, J is a Poisson random measure with

intensity measure $dt \times \nu(dx)$ and $\tilde{J}$ is the compensated version of J . $(B_t)_{t \geq 0}$ is thus the continuous martingale part of the process X , and the remaining terms

$$\gamma t + \int_0^t \int_{|x| \leq 1} x \tilde{J}(ds \times dx) + \int_0^t \int_{|x| > 1} x J(ds \times dx)$$

may be called the pure jump part of X . Since γ corresponds to a deterministic shift of every component, the law of the pure jump part of a Lévy process is determined essentially by the Lévy measure ν .

Lévy copulas provide a representation of the Lévy measure of a multidimensional Lévy process, which allows to specify separately the Lévy measures of the components and the information about the dependence between the components.¹ Similarly to copulas for probability measures, this gives a flexible approach for building multidimensional dynamic models based on Lévy processes.

The main ideas of Lévy copulas are simpler to explain in the context of Lévy measures on $[0, \infty)^n$, which correspond to Lévy processes with only positive jumps in every component. Formally, the definitions of Lévy copula, tail integrals etc. are different for Lévy measures on $[0, \infty)^n$ and on the full space, and we shall speak of Lévy copulas on $[0, \infty]^n$ and of Lévy copulas on $(-\infty, \infty]^n$, respectively. However, when there is no ambiguity, the explicit mention of the domain will be dropped. Moreover, by comparing the two definitions below it is easy to see that from a Lévy copula on $[0, \infty]^n$ one can always construct a Lévy copula on $(-\infty, \infty]^n$ by setting it to zero outside its original domain.

Lévy copulas on $[0, \infty]^n$ Similarly to probability measures, which can be represented through their distribution functions, Lévy measures can be represented by tail integrals.

Definition 1 (*Tail integral*) Let ν be a Lévy measure on $[0, \infty)^n$. The tail integral U of ν is a function $[0, \infty)^n \rightarrow [0, \infty]$ such that

1. $U(0, \dots, 0) = \infty$.
2. For $(x_1, \dots, x_n) \in [0, \infty)^n \setminus \{0\}$,

$$U(x_1, \dots, x_n) = \nu([x_1, \infty) \times \dots \times [x_n, \infty)).$$

The i -th one-dimensional marginal tail integral U_i of a $\mathbb{R}^n$ -valued Lévy process $X = (X^1, \dots, X^n)$ is the tail integral of the process X^i and can be computed as

$$U_i(z) = (U(x_1, \dots, x_n) | x_i = z; x_j = 0 \text{ for } j \neq i), \quad z \geq 0.$$

We recall that a function $F: \text{Dom } F \subseteq \bar{\mathbb{R}}^n \rightarrow \bar{\mathbb{R}}$ is called n -increasing if for all $a \in \text{Dom } F$ and $b \in \text{Dom } F$ with $a_i \leq b_i$ for all i we have

¹By “dependence” we mean the information on the law of a random vector which remains to be determined once the marginal laws of its components have been specified.

$$V_F((a, b]) := \sum_{c \in \text{Dom } F: c_j = a_i \text{ or } b_i, i=1 \dots n} \text{sgn}(c) F(c) \geq 0,$$

$$\text{sgn}(c) = \begin{cases} 1, & \text{if } c_k = a_k \text{ for an even number of indices,} \\ -1, & \text{if } c_k = a_k \text{ for an odd number of indices.} \end{cases}$$

Definition 2 (*Lévy copula*) A function $F : [0, \infty]^n \rightarrow [0, \infty]$ is a *Lévy copula* on $[0, \infty]^n$ if

1. $F(u_1, \dots, u_n) < \infty$ for $(u_1, \dots, u_n) \neq (\infty, \dots, \infty)$,
2. $F(u_1, \dots, u_n) = 0$ whenever $u_i = 0$ for at least one $i \in \{1, \dots, n\}$,
3. F is n -increasing,
4. $F_i(u) = u$ for any $i \in \{1, \dots, n\}$, $u \in [0, \infty]$, where

$$F_i(u) = (F(v_1, \dots, v_n) | v_i = u; v_j = 0 \text{ for } j \neq i).$$

The following theorem gives a representation of the tail integral of a Lévy measure (and thus of the Lévy measure itself) in terms of its marginal tail integrals and a Lévy copula. It may be called *Sklar's theorem* for Lévy copulas on $[0, \infty]^n$.

Theorem 1 Let ν be a Lévy measure on $[0, \infty)^n$ with tail integral U and marginal Lévy measures $\nu_1, \dots, \nu_n$. Then there exists a Lévy copula F on $[0, \infty]^n$ such that

$$U(x_1, \dots, x_n) = F(U_1(x_1), \dots, U_n(x_n)), \quad (x_1, \dots, x_n) \in [0, \infty)^n, \quad (2)$$

where $U_1, \dots, U_n$ are tail integrals of $\nu_1, \dots, \nu_n$. This Lévy copula is unique on $\prod_{i=1}^n \text{Ran } U_i$.

Conversely, if F is a Lévy copula on $[0, \infty]^n$ and $\nu_1, \dots, \nu_n$ are Lévy measures on $[0, \infty)$ with tail integrals $U_1, \dots, U_n$ then (2) defines a tail integral of a Lévy measure on $[0, \infty)^n$ with marginal Lévy measures $\nu_1, \dots, \nu_n$.

A basic example of a one-parameter family of Lévy copulas on $[0, \infty]^n$ is the Clayton family, given by

$$F_\theta(u_1, \dots, u_n) = (u_1^{-\theta} + \dots + u_n^{-\theta})^{-1/\theta}, \quad \theta > 0. \quad (3)$$

This family has as limiting cases the independence Lévy copula (when $\theta \rightarrow 0$)

$$F_\perp(u_1, \dots, u_n) = \sum_{i=1}^n u_i \prod_{j \neq i} 1_{\{\infty\}}(u_j)$$

and the complete dependence Lévy copula (when $\theta \rightarrow \infty$)

$$F_\parallel(u_1, \dots, u_n) = \min(u_1, \dots, u_n).$$

Since Lévy copulas are closely related to distribution copulas, many of the classical copula constructions can be modified to build Lévy copulas. This allows to define *Archimedean Lévy copulas* (see Propositions 5.6 and 5.7 in [13] for the case of Lévy copulas on $[0, \infty]^n$). Another example is the vine construction of Lévy copulas [26], where a Lévy copula on $[0, \infty]^n$ is constructed from $n(n-1)/2$ bivariate dependence functions ($n-1$ Lévy copulas and $(n-2)(n-1)/2$ distributional copulas).

Lévy copulas on $(-\infty, \infty]^n$ For Lévy measures on the full space $\mathbb{R}^n$, the definition of a Lévy copula is more complex because the singularity of the Lévy measure is located in the interior of the domain and not at the boundary. The tail integral is defined so as to avoid this singularity.

Definition 3 Let ν be a Lévy measure on $\mathbb{R}^n$. The *tail integral* of ν is the function $U : (\mathbb{R} \setminus \{0\})^n \rightarrow \mathbb{R}$ defined by

$$U(x_1, \dots, x_n) := \nu \left(\prod_{j=1}^n \mathcal{J}(x_j) \right) \prod_{i=1}^d \operatorname{sgn}(x_i), \quad (4)$$

where

$$\mathcal{J}(x) := \begin{cases} [x, \infty), & x \geq 0, \\ (-\infty, x), & x < 0. \end{cases}$$

In the above definition, the signs and the intervals are chosen in such way that the tail integral becomes a left-continuous n -increasing function on each orthant.

Given a set of indices $I \subset \{1, \dots, n\}$, the I -marginal tail integral U^I of the Lévy process $X = (X^1, \dots, X^n)$ is the tail integral of the process $(X^i)_{i \in I}$ containing the components of X with indices in I . The one-dimensional tail integrals are, as before, denoted by $U_i \equiv U^{\{i\}}$. Given an $\mathbb{R}^n$ -valued Lévy process X , its marginal tail integrals $\{U^I : I \subset \{1, \dots, n\} \text{ nonempty}\}$ are, of course, uniquely determined by its Lévy measure ν .

The tail integral for a Lévy measure on $\mathbb{R}^n$, as well as the marginal tail integrals, are only defined for nonzero arguments. This leads to a symmetric definition, which results in a simple statement for Sklar's theorem below. On the other hand, this means that when the Lévy measure is not absolutely continuous, it is not uniquely determined by its tail integral. However, it is always uniquely determined by the set $\{U^I, I \subseteq \{1, \dots, n\}, I \neq \emptyset\}$ containing its tail integral as well as all its marginal tail integrals.

Definition 4 A function $F : (-\infty, \infty]^n \rightarrow (-\infty, \infty]$ is a *Lévy copula* on $(-\infty, \infty]^n$ if

1. $F(u_1, \dots, u_n) < \infty$ for $(u_1, \dots, u_n) \neq (\infty, \dots, \infty)$,
2. $F(u_1, \dots, u_n) = 0$ if $u_i = 0$ for at least one $i \in \{1, \dots, n\}$,
3. F is n -increasing,
4. $F_i(u) = u$ for any $i \in \{1, \dots, n\}$, $u \in (-\infty, \infty]$, where F_i is defined below.

The margins of a Lévy copula on $(-\infty, \infty]^n$ are defined by

$$F^I((x_i)_{i \in I}) := \lim_{c \rightarrow \infty} \sum_{(x_j)_{j \in I^c} \in \{-c, \infty\}^{I^c}} F(x_1, \dots, x_n) \prod_{j \in I^c} \operatorname{sgn} x_j \quad (5)$$

with the convention $F_i = F^{(i)}$ for one-dimensional margins.

The following theorem is the analogue of Sklar's theorem for Lévy copulas on $(-\infty, \infty]^n$.

Theorem 2 *Let ν be a Lévy measure on $\mathbb{R}^n$. Then there exists a Lévy copula F such that the tail integrals of ν satisfy:*

$$U^I((x_i)_{i \in I}) = F^I((U_i(x_i))_{i \in I}) \quad (6)$$

for any non-empty $I \subseteq \{1, \dots, n\}$ and any $(x_i)_{i \in I} \in (\mathbb{R} \setminus \{0\})^I$. The Lévy copula F is unique on $\prod_{i=1}^n \operatorname{Ran} U_i$.

Conversely, if F is a n -dimensional Lévy copula and $\nu_1, \dots, \nu_n$ are Lévy measures on $\mathbb{R}$ with tail integrals U_i , $i = 1, \dots, n$ then there exists a unique Lévy measure on $\mathbb{R}^n$ with one-dimensional marginal tail integrals $U_1, \dots, U_n$ and whose marginal tail integrals satisfy (6) for any non-empty $I \subseteq \{1, \dots, n\}$ and any $(x_i)_{i \in I} \in (\mathbb{R} \setminus \{0\})^I$.

A basic example of a Lévy copula on $(-\infty, \infty]^n$ is the two-parameter Clayton family which has the form

$$F(u_1, \dots, u_n) = 2^{2-n} \left(\sum_{i=1}^n |u_i|^{-\theta} \right)^{-1/\theta} (\eta \mathbf{1}_{u_1 \dots u_n \geq 0} - (1 - \eta) \mathbf{1}_{u_1 \dots u_n < 0}), \quad (7)$$

for $\theta > 0$ and $\eta \in [0, 1]$. Here the parameter η determines the dependence of the sign of the jumps: for example, when $n = 2$ and $\eta = 1$, the two components always jump in the same direction. The parameter θ , as before, determines the dependence of the amplitude of the jumps. When $\eta = 1$ and $\theta \rightarrow 0$, the Clayton Lévy copula converges to the independence Lévy copula

$$F_{\perp}(u_1, \dots, u_n) = \sum_{i=1}^n u_i \prod_{j \neq i} \mathbf{1}_{\{0\}}(u_j),$$

and when $\eta = 1$ and $\theta \rightarrow \infty$, one recovers the complete dependence Lévy copula

$$F_{\perp}(u_1, \dots, u_n) = \min(|u_1|, \dots, |u_n|) \mathbf{1}_K(u_1, \dots, u_n) \prod_{i=1}^n \operatorname{sgn} u_i,$$

where $K = \{u \in \bar{\mathbb{R}}^n : \operatorname{sgn} u_1 = \dots = \operatorname{sgn} u_n\}$.

Other families of Lévy copulas on $(-\infty, \infty]^n$ may be obtained using the Archimedean Lévy copula construction (Theorem 5.1 in [31]) or, when a precise

description of dependence in each orthant is needed, from 2^n Lévy copulas on $[0, \infty]^n$ (Theorem 5.3 in [31]).

Alternative marginal transformations When F is a Lévy copula on $[0, \infty]^n$, the mapping

$$\chi_F((0, b_1] \times \cdots \times (0, b_n]) := F(b_1, \dots, b_n), \quad 0 \leq b_1, \dots, b_n \leq \infty$$

can be extended to a unique positive measure χ_F on the Borel sets of $[0, \infty]^n$, whose margins (projections on the coordinate axes) are uniform (standard Lebesgue) measures on $[0, \infty)$, and which has no atom at $\{\infty, \dots, \infty\}$. Similarly, a Lévy copula F on $(-\infty, \infty]^n$ can be associated to a positive measure χ_F whose margins are uniform measures on $(-\infty, \infty)$, and which satisfies

$$\chi_F((a_1, b_1] \times \cdots \times (a_n, b_n]) = V_F((a, b]), \quad -\infty < a_i \leq b_i \leq \infty, \quad i = 1, \dots, n. \quad (8)$$

The transformation to uniform margins offers a direct analogy with the distributional copulas, but other marginal transformations may be more convenient in specific contexts.

In particular, several authors [2, 18, 33] have considered the transformation to 1-stable margins. Following [2], introduce the “inversion map”

$$Q : [0, \infty]^n \rightarrow [0, \infty]^n, \quad (x_1, \dots, x_n) \mapsto (x_1^{-1}, \dots, x_n^{-1}),$$

where $1/0$ has to be interpreted as ∞ and $1/\infty$ as 0 . For a Lévy copula F on $[0, \infty]^n$, we define the measure ν_F as the image of the measure χ_F defined above under the mapping Q , that is,

$$\nu_F(B) = \chi_F(Q^{-1}(B)) \quad \forall B \text{ Borel set in } [0, \infty]^n.$$

Then, ν_F is a Lévy measure on $[0, \infty)^n$ with marginal tail integrals $U_k(x) = x^{-1}$, that is, ν_F has 1-stable margins. It is clear that the dependence structure of a Lévy measure with Lévy copula F can be alternatively characterized in terms of the Lévy measure ν_F . This construction has been extended to Lévy measures on $\mathbb{R}^n$ [18] (in this reference, the Lévy measure ν_F has been called *Pareto Lévy measure* in and its tail integral has been called *Pareto Lévy copula*).

A brief review of alternative dependence models Lévy copulas offer a very flexible approach to build multidimensional Lévy processes with a precise control over the dependence of joint jumps and joint extremes. However, this degree of precision is not needed for all applications and comes at the price of reduced analytical tractability. For this reason, several authors have proposed alternative approaches to generate dependency among Lévy processes, which may be less flexible, but lead to simpler models. Among the possible alternative approaches one can mention

- Brownian subordination (time change) with a one-dimensional subordinator [15, 35, 38] or a multi-dimensional subordinator [3, 16].
- Factor models based on linear combinations of independent Poisson shocks [34] or more generally independent Lévy processes [32, 36, 37].

Lévy copulas allow to construct a multidimensional Lévy process with given marginals and given dependence structure. The same question has been addressed for other classes of stochastic processes such as Markov processes [7] and semimartingales [46]. We refer the reader to [6] for a comprehensive review of copula-related concepts for these processes.

3 Monte Carlo Simulation of Lévy Processes with a Specified Lévy Copula

In models based on Lévy copulas, explicit computations are rarely possible (for instance, the characteristic function of the Lévy process is usually not known in explicit form), and to compute quantities such as option prices, one has to resort to numerical methods which can be either deterministic (partial integro-differential equations) or stochastic (Monte Carlo). Deterministic numerical methods for PIDEs arising in Lévy copula models have been developed in [24, 27, 39]. In this section, we propose a new algorithm for simulating multidimensional Lévy processes defined through a Lévy copula, which can be used for Monte Carlo simulation.

Let F be a Lévy copula such that for every $I \subseteq \{1, \dots, n\}$ nonempty,

$$\lim_{(x_i)_{i \in I} \rightarrow \infty} F(x_1, \dots, x_n) = F(x_1, \dots, x_n)|_{(x_i)_{i \in I} = \infty}. \quad (9)$$

As mentioned above, this Lévy copula is associated to a positive measure χ_F on $\bar{\mathbb{R}}^n$ with Lebesgue margins, and condition (9) guarantees that this measure is supported by $\mathbb{R}^n$.

The following technical lemma, whose proof can be found in the preprint [43], establishes the relation between χ_F and the Lévy measures of processes having F as their Lévy copula. For a one-dimensional tail integral U , the (generalized) inverse tail integral $U^{(-1)}$ is defined by

$$U^{(-1)}(u) := \begin{cases} \sup\{x > 0 : U(x) \geq u\} \vee 0, & u \geq 0 \\ \sup\{x < 0 : U(x) \geq u\}, & u < 0. \end{cases} \quad (10)$$

Lemma 1 *Let ν be a Lévy measure on $\mathbb{R}^n$ with marginal tail integrals U_i , $i = 1, \dots, n$, and Lévy copula F satisfying the condition (9), let χ_F be defined by (8) and let*

$$f : (u_1, \dots, u_n) \mapsto (U_1^{(-1)}(u_1), \dots, U_n^{(-1)}(u_n)).$$

Then ν is the image measure of χ_F by f .

In Theorems 3 and 4 below, to simulate the jumps of a multidimensional Lévy process (more precisely, of the corresponding Poisson random measure), we will choose one component of the Lévy process, simulate its jumps, and then simulate the jumps in the other components conditionally on the jumps in the chosen one. We therefore proceed by analyzing the conditional distributions of χ_F . To fix the ideas, we suppose that we have chosen the first component of the Lévy process, but the conditional distribution with respect to any other component is obtained in the same way.

By Theorem 2.28 in [1], there exists a family, indexed by $\xi \in \mathbb{R}$, of positive Radon measures $K_1(\xi, dx_2 \cdots dx_n)$ on $\mathbb{R}^{n-1}$, such that

$$\xi \mapsto K_1(\xi, dx_2 \cdots dx_n)$$

is Borel measurable and

$$\chi_F(dx_1 \dots dx_n) = dx_1 \times K_1(x_1, dx_2 \cdots dx_n). \quad (11)$$

In addition, $K_1(\xi, \mathbb{R}^{n-1}) = 1$ almost everywhere, that is, $K_1(\xi, \cdot)$ is, almost everywhere, a probability distribution. In the sequel we will call $\{K_1(\xi, \cdot)\}_{\xi \in \mathbb{R}}$ the family of conditional probability distributions with respect to the first component associated with the Lévy copula F . Similarly, the conditional distributions with respect to other components will be denoted by $K_2, \dots, K_n$.

Let F_1^ξ be the distribution function of the measure $K_1(\xi, \cdot)$:

$$F_1^\xi(x_2, \dots, x_n) := K_1(\xi, (-\infty, x_2] \times \cdots \times (-\infty, x_n]). \quad (12)$$

The following lemma, whose proof can also be found in [43], shows that it can be computed in a simple manner from the Lévy copula F . We recall that the law of a random variable is completely determined by the values of its distribution function at the continuity points of the latter.

Lemma 2 *Let F be a Lévy copula satisfying (9), and F_1^ξ be the corresponding conditional distribution function, defined by (12). Then, there exists a set $N \subset \mathbb{R}$ of zero Lebesgue measure such that for every fixed $\xi \in \mathbb{R} \setminus N$, $F_1^\xi(\cdot)$ is a probability distribution function, satisfying*

$$\begin{aligned} & F_1^\xi(x_2, \dots, x_n) \\ &= \operatorname{sgn}(\xi) \frac{\partial}{\partial \xi} V_F((\xi \wedge 0, \xi \vee 0] \times (-\infty, x_2] \times \cdots \times (-\infty, x_n]) \end{aligned} \quad (13)$$

in every point $(x_2, \dots, x_n)$, where F_1^ξ is continuous.

In the following two theorems we show how Lévy copulas may be used to simulate multidimensional Lévy processes with a specified dependence structure. Our results can be seen as an extension to Lévy processes, represented by Lévy copulas, of

the series representation results, developed by Rosinski and others (see [41] and references therein). The first result concerns the simpler case when the Lévy process has finite variation on compacts.

Theorem 3 (Simulation of multidimensional Lévy processes, finite variation case) *Let ν be a Lévy measure on $\mathbb{R}^n$, satisfying $\int (|x| \wedge 1) \nu(dx) < \infty$, with marginal tail integrals U_i , $i = 1, \dots, n$ and Lévy copula $F(x_1, \dots, x_n)$, such that (9) is satisfied, and let $K_1, \dots, K_n$ be the corresponding conditional probability distributions.*

Fix a truncation level τ . Let (V_k) and (W_k^i) for $1 \leq i \leq n$ and $k \geq 1$ be independent sequences of independent random variables, uniformly distributed on $[0, 1]$. Introduce n^2 random sequences $(\Gamma_k^{ij})_{k \geq 1}^{1 \leq i, j \leq n}$, independent from (V_i) and (W_i) such that

- For $i = 1, \dots, n$, $\sum_{k=1}^{\infty} \delta_{\{\Gamma_k^{ii}\}}$ are independent Poisson random measures on $\mathbb{R}$ with Lebesgue intensity measures.
- Conditionally on Γ_k^{ii} , the random vector $(\Gamma_k^{i1}, \dots, \Gamma_k^{i, i-1}, \Gamma_k^{i, i+1}, \dots, \Gamma_k^{in})$ is independent from Γ_l^{pq} with $1 \leq p, q \leq n$ and $l \neq k$ and from Γ_k^{pq} with $p \neq i$ and $1 \leq q \leq n$ and is distributed on $\mathbb{R}^{n-1}$ with law $K_i(\Gamma_k^{ii}, dx_1 \cdots dx_{n-1})$.

For each $k \geq 1$ and each $i = 1, \dots, n$, let $n_k^i = \#\{j = 1, \dots, n : |\Gamma_k^{ij}| \leq \tau\}$. Then the process $(Z_t^r)_{0 \leq t \leq 1}$ with components

$$Z_t^{\tau, j} = \sum_{k=1}^{\infty} \sum_{i=1}^n U_j^{(-1)}(\Gamma_k^{ij}) \mathbf{1}_{n_k^i W_k^i \leq 1} \mathbf{1}_{|\Gamma_k^{ii}| \leq \tau} \mathbf{1}_{[0, t]}(V_k), \quad j = 1, \dots, n, \quad (14)$$

is a Lévy process on $[0, 1]$ with characteristic function

$$E[e^{i\langle u, Z_t^r \rangle}] = \exp\left(t \int_{\mathbb{R}^n \setminus S_\tau} (e^{i\langle u, z \rangle} - 1) \nu(dz)\right), \quad (15)$$

where

$$S_\tau = (U_1^{(-1)}(-\tau), U_1^{(-1)}(\tau)) \times \cdots \times (U_n^{(-1)}(-\tau), U_n^{(-1)}(\tau)). \quad (16)$$

Moreover, there exists a Lévy process $(Z_t)_{0 \leq t \leq 1}$ with characteristic function

$$E[e^{i\langle u, Z_t \rangle}] = \exp\left(t \int_{\mathbb{R}^n} (e^{i\langle u, z \rangle} - 1) \nu(dz)\right) \quad (17)$$

such that

$$\mathbb{E}[\sup_{0 \leq t \leq 1} |Z_t^r - Z_t|] \leq \sum_{i=1}^n \int_{U_i^{(-1)}(-\tau)}^{U_i^{(-1)}(\tau)} |z| v_i(dz), \quad (18)$$

where v_i is the i -th margin of the Lévy measure.

Remark 1 Since the Lévy measure satisfies $\int(|x| \wedge 1)\nu(dx) < \infty$, the error (18) converges to 0 as $\tau \rightarrow +\infty$; in addition, the upper bound on the error does not depend on the Lévy copula F .

Remark 2 For the numerical computation of the sum in (14), we need to simulate only the variables Γ_k^{ii} for which $|\Gamma_k^{ii}| \leq \tau$. The number of such variables is a.s. finite and follows the Poisson distribution with intensity 2τ . They can therefore be simulated with the following two-step algorithm:

- Simulate a Poisson random variable N_i with intensity 2τ .
- Simulate N_i independent random variables $U_1, \dots, U_{N_i}$ with uniform distribution on $[-\tau, \tau]$ and let $\Gamma_k^{ii} = U_k$ for $k = 1, \dots, N_i$.

Remark 3 In [43] we proposed a simpler algorithm for simulating a Lévy process with a given Lévy copula, where all the components were simulated conditionally on the first one. As it turns out, this algorithm suffers from convergence problems when the components are weakly dependent. By contrast, the above algorithm treats all components in a symmetric way, which leads to a uniform bound (18) ensuring fast convergence even in the case of weak dependence, at the price of performing additional simulations and rejections. From Fig. 1, one can see that even in the case of very weak dependence, no single component appears to dominate the other one.

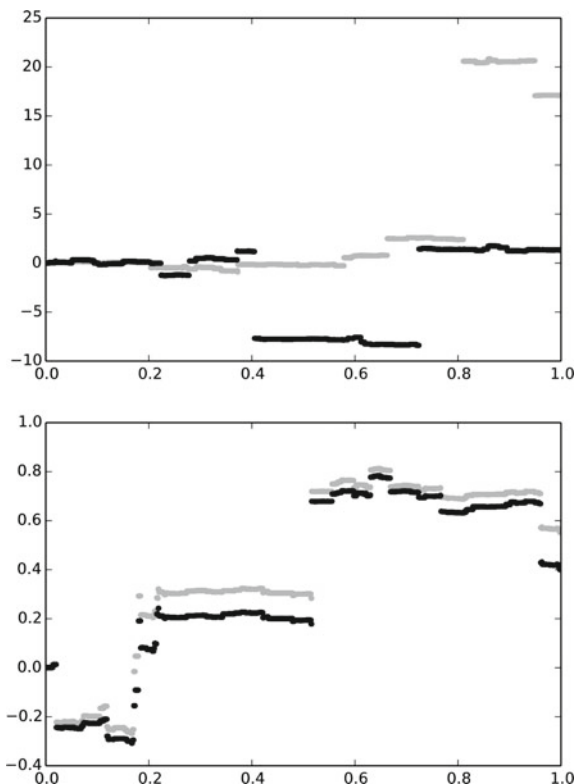
Example 1 Let $d = 2$ and F be the two-parameter Clayton Lévy copula (7). A straightforward computation yields:

$$F_1^\xi(x) = F_2^\xi(x) := F_\xi(x) = \left\{ (1 - \eta) + \left(1 + \left| \frac{\xi}{x} \right|^\theta \right)^{-1-1/\theta} (\eta - 1_{x < 0}) \right\} 1_{\xi \geq 0} \\ + \left\{ \eta + \left(1 + \left| \frac{\xi}{x} \right|^\theta \right)^{-1-1/\theta} (1_{x \geq 0} - \eta) \right\} 1_{\xi < 0}. \quad (19)$$

This conditional distribution function can be inverted analytically:

$$F_\xi^{-1}(u) = B(\xi, u) |\xi| \left\{ C(\xi, u)^{-\frac{\theta}{\theta+1}} - 1 \right\}^{-1/\theta} \\ \text{with } B(\xi, u) = \text{sgn}(u - 1 + \eta) 1_{\xi \geq 0} + \text{sgn}(u - \eta) 1_{\xi < 0} \\ \text{and } C(\xi, u) = \left\{ \frac{u - 1 + \eta}{\eta} 1_{u \geq 1 - \eta} + \frac{1 - \eta - u}{1 - \eta} 1_{u < 1 - \eta} \right\} 1_{\xi \geq 0} \\ + \left\{ \frac{u - \eta}{1 - \eta} 1_{u \geq \eta} + \frac{\eta - u}{\eta} 1_{u < \eta} \right\} 1_{\xi < 0}.$$

Fig. 1 Simulated trajectories of a two-dimensional α -stable process with dependence given by the Clayton Lévy copula (7). In both graphs, the marginal parameters are: $\alpha = 0.6$ and $c = 1$ and the truncation threshold is $\tau = 2000$. The dependence parameters are: $\theta = 0.1$ and $\eta = 0.99$ (top graph) and $\theta = 5$ and $\eta = 0.99$ (bottom graph). The different scale of the two graphs is due to the random nature of trajectories



Therefore, the sequences $(\Gamma_k^{i,j})_{k \geq 1}^{1 \leq i,j \leq 2}$, introduced in Theorem 3, can be constructed as follows:

- Simulate the variables N_1 and N_2 and the sequences $(\Gamma_k^{11})_{1 \leq k \leq N_1}$ and $(\Gamma_k^{22})_{1 \leq k \leq N_2}$ as described in Remark 2.
- For all $1 \leq k \leq N_1$, let $\Gamma_k^{12} = F_{\Gamma_k^{11}}^{-1}(Y_k^1)$ and for all $1 \leq k \leq N_2$ let $\Gamma_k^{21} = F_{\Gamma_k^{22}}^{-1}(Y_k^2)$, where $(Y_k^1)_{k \geq 0}$ and $(Y_k^2)_{k \geq 0}$ are independent sequences of independent random variables with uniform distribution on $[0, 1]$.

For the purpose of numerical illustration, we have simulated the trajectories of a two-dimensional α -stable process with marginal Lévy density $\nu(x) = \frac{\alpha c}{|x|^{1+\alpha}}$ for both components and dependence given by the Clayton Lévy copula (7) with weak dependence (Fig. 1, left graph) and strong dependence (Fig. 1, right graph). We used the truncation parameter $\tau = 2000$, which corresponds to keeping about 4000 jumps in every component. For this value, the upper bound on the error (18) is equal to $\frac{4c\alpha}{1-\alpha} \left(\frac{c}{\tau}\right)^{\frac{1-\alpha}{\alpha}} \approx 0.037$.

Proof (of Theorem 3) First note that (Γ_k^{ij}) are well defined since by Lemma 2, $K_i(\xi, \cdot)$ is a probability distribution for almost all ξ . Let

$$Z_t^{\tau, ij} = \sum_{k: |\Gamma_k^{ij}| \leq \tau} U_j^{(-1)}(\Gamma_k^{ij}) \mathbf{1}_{n_k^i w_k^i \leq 1} \mathbf{1}_{[0, t]}(V_k)$$

By Proposition 3.8 in [40],

$$Z_t^{\tau, ij} = \int_{[0, t] \times \mathbb{R}^n} U_j^{(-1)}(x_j) M_i(ds \times dx_1 \cdots dx_n),$$

where M_i is a Poisson random measure on $[0, 1] \times \mathbb{R}^n$ with intensity measure $dt \times \eta_i(x_1, \dots, x_n) \chi_F(dx_1 \cdots dx_n)$, and the function η_i is defined by

$$\eta_i(x_1, \dots, x_n) = \frac{\mathbf{1}_{|x_i| \leq \tau}}{1 + \#\{j = 1, \dots, n : j \neq i, |x_j| \leq \tau\}}.$$

Therefore,

$$Z_t^{\tau, j} = \sum_{i=1}^n Z_t^{\tau, ij} = \int_{[0, t] \times \mathbb{R}^n} U_j^{(-1)}(x_j) M(ds \times dx_1 \cdots dx_n),$$

where M_i is a Poisson random measure on $[0, 1] \times \mathbb{R}^n$ with intensity measure $dt \times \eta_i(x_1, \dots, x_n) \chi_F(dx_1 \cdots dx_n)$ with

$$\eta(x_1, \dots, x_n) = \sum_{i=1}^n \eta_i(x_1, \dots, x_n) = 1 - \mathbf{1}_{|x_i| > \tau, i=1, \dots, n}.$$

By Lemma 1 and Proposition 3.7 in [40],

$$Z_t^{\tau, j} = \int_{[0, t] \times \mathbb{R}^n} x_j N_\tau(ds \times dx_1 \cdots dx_n) \quad (20)$$

for a Poisson random measure N_τ on $[0, 1] \times \mathbb{R}^n$ with intensity measure $ds \times \nu_\tau(dx_1 \cdots dx_n)$, where

$$\nu_\tau := (1 - \mathbf{1}_{S_\tau}(x_1, \dots, x_n)) \nu(dx_1 \cdots dx_n). \quad (21)$$

The Lévy-Itô decomposition (1) then implies that Z_t^τ is a Lévy process on $[0, 1]$ with characteristic function

$$E \left[e^{i \langle u, Z_t^\tau \rangle} \right] = \exp \left(t \int_{\mathbb{R}^n} (e^{i \langle u, z \rangle} - 1) \nu_\tau(dz) \right) = \exp \left(t \int_{\mathbb{R}^n \setminus S_\tau} (e^{i \langle u, z \rangle} - 1) \nu(dz) \right).$$

Now let $R^\tau \in \mathbb{R}^n$ be defined by

$$R_t^{\tau,j} = \int_{[0,t] \times \mathbb{R}^n} x_j \bar{N}_\tau(ds \times dx_1 \cdots dx_n);$$

where $\bar{N}_\tau$ is a Poisson random measure on $[0, 1] \times \mathbb{R}^n$ with intensity measure $ds \times \mathbf{1}_{S_\tau}(x_1, \dots, x_n) \nu(dx_1 \cdots dx_n)$, independent from N_τ . It is clear that $Z = Z^\tau + R^\tau$ is a Lévy process with characteristic function (17). Finally,

$$\begin{aligned} \mathbb{E}[\sup_{0 \leq t \leq 1} |Z_t - Z_t^\tau|] &\leq \mathbb{E}\left[\sum_{0 \leq t \leq 1} |\Delta R_t^\tau|\right] \leq \sum_{i=1}^n \mathbb{E}\left[\sum_{0 \leq t \leq 1} |\Delta R_t^{\tau,i}|\right] \\ &= \sum_{i=1}^n \int_{\mathbb{R}^n} |x| \mathbf{1}_{S_\tau} \nu(dx) \leq \sum_{i=1}^n \int_{U_i^{(-1)}(-\tau)}^{U_i^{(-1)}(\tau)} |z| \nu_i(dz). \end{aligned}$$

which proves the bound (18).

If the Lévy process has paths of infinite variation on compact sets, it can no longer be represented as the sum of its jumps and we have to introduce a centering term into the series (14).

Theorem 4 (Simulation of multidimensional Lévy processes, infinite variation case) *Let ν be a Lévy measure on $\mathbb{R}^n$ with marginal tail integrals U_i , $i = 1, \dots, n$ and Lévy copula $F(x_1, \dots, x_n)$, such that the condition (9) is satisfied. Let $(V_k)_{k \geq 1}$, $(W_k^i)_{k \geq 1}$ for $1 \leq i \leq n$ and $(\Gamma_k^{ij})_{k \geq 1}$ for $1 \leq i, j \leq n$ be as in Theorem 3. Let*

$$A_k(\tau) = \int_{\mathbb{R}^n \setminus S_\tau} \mathbf{1}_{|x| \leq 1} x_k \nu(dx_1 \cdots dx_n), \quad k = 1 \dots n,$$

where S_τ is defined in (16). Then the process

$$(Z_t^\tau)_{0 \leq t \leq 1}, \quad \text{where} \quad Z_t^{\tau,j} = \sum_{k=1}^{\infty} \sum_{i=1}^n U_j^{(-1)}(\Gamma_k^{ij}) \mathbf{1}_{n_k^i} \mathbf{1}_{W_k^i \leq 1} \mathbf{1}_{|\Gamma_k^{ii}| \leq \tau} \mathbf{1}_{[0,t]}(V_k) - t A_j(\tau),$$

for $1 \leq j \leq n$ is a Lévy process on $[0, 1]$ with characteristic function

$$\mathbb{E}[e^{i\langle u, Z_t^\tau \rangle}] = \exp\left(t \int_{\mathbb{R}^n \setminus S_\tau} (e^{i\langle u, x \rangle} - 1 - i\langle u, x \rangle) \mathbf{1}_{|x| \leq 1} \nu(dx)\right). \quad (22)$$

Moreover, there exists a Lévy process $(Z_t)_{0 \leq t \leq 1}$ with characteristic function

$$\mathbb{E}[e^{i\langle u, Z_t \rangle}] = \exp\left(t \int_{\mathbb{R}^n} (e^{i\langle u, x \rangle} - 1 - i\langle u, x \rangle) \mathbf{1}_{|x| \leq 1} \nu(dx)\right)$$

such that

$$\mathbb{E} \left[\sup_{0 \leq t \leq 1} (Z_t - Z_t^\tau)^2 \right] \leq 4 \sum_{i=1}^n \int_{U_i^{-1}(-\tau)}^{U_i^{-1}(\tau)} z^2 v_i(dz).$$

Proof The proof is essentially the same as in Theorem 3. In this theorem, due to the presence of the compensating term $tA(\tau)$, the process Z^τ is a martingale and the error bound follows from Doob's martingale inequality.

4 Statistical Estimation Techniques

Lévy copulas give access to the distribution of sizes of simultaneous jumps of the different components of a Lévy process. Therefore, Lévy copula-based models are easy to estimate when jumps are observable. This is the case, for instance, in insurance models where jumps represent claims whose dates and amounts are known.² Esmaeli and Klüppelberg [21] use maximum likelihood for the statistical estimation of the parameters of a two-dimensional compound Poisson process with dependence given by a Lévy copula, assuming that all jumps in both components are perfectly observable. The method is then applied to the Danish fire insurance data (see also [19, paragraph 6.5.2] for a discussion of this data in a one-dimensional setting).

In the case of infinite jump intensity it is clearly not realistic to assume that all jumps are perfectly observable. For this reason, in the case of a bivariate stable subordinator with dependence given by the Clayton Lévy copula, Esmaeli and Klüppelberg [22] assume that one observes all simultaneous jumps whose value is larger than a certain small parameter ε for both components. Once again, the exact knowledge of jump times and sizes allows to write down the maximum likelihood estimator, which is shown to be consistent and asymptotically normal as ε tends to zero and/or the observation horizon tends to infinity.

In Esmaeli and Klüppelberg [23], the authors consider a slightly different observation scheme, assuming that jumps larger than ε in each single component are recorded (this sampling scheme is also adopted in [26]). A two-step parameter estimation scheme is developed, where one first estimates the parameters of the one-dimensional marginal Lévy processes, and next those of the Lévy copula. This reduces the computational cost compared to full likelihood (simultaneous estimation of all parameters), at the price of a somewhat lower efficiency. The example of a bivariate stable subordinator with Clayton dependence is worked out in detail. The two-step estimator is shown to be consistent and asymptotically normal and the efficiency loss as compared to the full likelihood estimator is shown to be quite small (the root mean square error increases by less than ten percentage points in most cases).

²The underlying assumption in Lévy insurance models is that dependence may only be present among simultaneous losses in different business lines. This is of course an over-simplification of reality since in practice, losses resulting from the same event may become known at different times, introducing dependence between non-simultaneous jumps.

In financial applications it is not always possible to assume that all jumps of sufficient size are perfectly observable: usually one only observes the Lévy process at a discrete time grid. If the monitoring frequency is sufficiently high though, the jumps remain “almost” observable and the Lévy copula can still be recovered.

Bücher and Vetter [11] develop a fully nonparametric approach for estimating the tail integrals and the Pareto Lévy copula (see Sect. 2) in the context of bivariate Lévy processes with only positive jumps. Their estimate for the bivariate tail integral becomes

$$U_n(x_1, x_2) = \frac{1}{k_n} \sum_{j=1}^n \mathbf{1}_{\{\Delta_j^n X^{(1)} \geq x_1, \Delta_j^n X^{(2)} \geq x_2\}},$$

where $\Delta_j^n X^{(i)} \geq x_i = X_{j\Delta_n}^{(i)} - X_{(j-1)\Delta_n}^{(i)}$ for $i = 1, 2$, $k_n = n\Delta_n$, Δ_n is the observation interval and n is the sample size. The method of using the increments of the process larger than a given size as a proxy for its jumps is rather common in the high frequency financial econometrics literature. The one-dimensional tail integrals are approximated by

$$U_{i,n}(x) = \frac{1}{k_n} \sum_{j=1}^n \mathbf{1}_{\{\Delta_j^n X^{(i)} \geq x\}}, \quad i = 1, \dots, 2,$$

and the estimator of the Pareto Lévy copula is given, up to some technical adjustments, by

$$\hat{\Gamma}_n(u_1, u_2) = U_n \left(U_{n,1}^{(-1)}(1/u_1), U_{n,2}^{(-1)}(1/u_2) \right).$$

These estimates for the tail integrals and the Pareto Lévy copula are shown to converge at the rate $\frac{1}{\sqrt{k_n}}$ as $n \rightarrow \infty$ provided that $\Delta_n \rightarrow 0$ (high frequency observation), $k_n \rightarrow \infty$ (infinite time horizon) and in addition $\sqrt{k_n} \Delta_n \rightarrow 0$.

We refer the reader to [11] for more details on the estimation procedure as well as for extensions to irregular and asynchronous sampling schemes. See also [9] for an alternative method of estimating the joint dependence of jumps using the extreme value theory. Another relevant reference is [25] where the so called jump tail dependence parameter (the analogue of the tail dependence index defined at the level of the Lévy copula) is estimated from discrete observations of the Lévy process.

5 Lévy Copulas and Multivariate Regular Variation

Multivariate regular variation is a widely accepted framework for risk analysis in a multidimensional setting (see e.g., [20] for examples), which turns out to be well suited for Lévy copula modeling.

For introduction to multivariate regular variation, see [4, 14, 30, 40]. Following [30], we let $\mathbf{M}_0$ denote the class of Borel measures on $\mathbb{R}^n$, whose restriction to $\mathbb{R}^n \setminus B_{0,r}$ is finite for each $r > 0$, where $B_{0,r}$ is the ball of radius r centered at the origin. Further, let $\mathcal{C}_0$ be the class of real-valued bounded and continuous functions on $\mathbb{R}^n$, vanishing in a neighborhood of zero. We say that a sequence of measures $(\mu_n)_{n \geq 1} \subset \mathbf{M}_0$ converges in $\mathbf{M}_0$ to a measure $\mu \in \mathbf{M}_0$ whenever

$$\int f(x) \mu_n(dx) \rightarrow \int f(x) \mu(dx), \quad n \rightarrow \infty,$$

for each $f \in \mathcal{C}_0$. This convergence is known as vague convergence of measures and it is, of course, similar to weak convergence, the only difference being that the test functions must vanish in the neighborhood of zero.

A measure $\nu \in \mathbf{M}_0$ is said to be regularly varying if there exists a norming sequence $(c_n)_{n \geq 1}$ of positive numbers with $c_n \uparrow \infty$ as $n \rightarrow \infty$ and a nonzero $\mu \in \mathbf{M}_0$ such that $n\nu(c_n \cdot) \rightarrow \mu$ in $\mathbf{M}_0$ as $n \rightarrow \infty$. Then necessarily the limit measure μ is homogeneous of some degree α and we write $\nu \in RV(\alpha, (c_n), \mu)$. A random vector $X \in \mathbb{R}^n$ is said to be regularly varying if its probability distribution is regularly varying.

For infinitely divisible random vectors, multivariate regular variation is characterized in terms of the Lévy measure: if X is infinitely divisible with Lévy measure ν then $X \in RV(\alpha, (c_n), \mu)$ if and only if $\nu \in RV(\alpha, (c_n), \mu)$ [29]. Under this condition, the Lévy process $(X_t)_{0 \leq t \leq 1}$ with Lévy measure ν is regularly varying on the space $D([0, 1], \mathbb{R}^d)$ of right-continuous functions with left limits [28]. The limiting measure is concentrated on the trajectories of the form $Z\mathbf{1}_{t \leq V}$ with $Z \in \mathbb{R}^d$ and $V \in [0, 1]$, so that intuitively the extremal behavior of such a process is determined by one large jump.

The link between multivariate regular variation and Lévy copulas is explored in detail in [18] using the closely related notion of Pareto Lévy copulas (see Sect. 2 above). Loosely speaking, if the Lévy measures of the components are regularly varying with the same index α and the Pareto Lévy measure is multivariate regularly varying with index 1 then the Lévy measure ν is multivariate regularly varying with index α (see Theorem 3.1 in [18] for a precise statement). Here, we shall summarize the main ideas in terms of standard Lévy copulas.

Definition 5 Let F be a Lévy copula on $(-\infty, \infty]^n$. We say that F is regularly varying if there exists a Lévy copula G on $(-\infty, \infty]^n$ such that for any nonempty $I \subseteq \{1, \dots, n\}$ and any $x \in (\mathbb{R} \setminus \{0\})^{|I|}$,

$$nF^I(n^{-1}x) \rightarrow G^I(x), \quad n \rightarrow \infty. \quad (23)$$

It is clear that in this case, the limiting Lévy copula G is homogeneous of order 1. Examples of Lévy copulas satisfying Assumption (23) include the independence Lévy copula, the complete dependence Lévy copula, or the Clayton family.

For regularly varying Lévy copulas, tail properties of the infinitely divisible random vector may be deduced from the corresponding properties of the Lévy copula. For example, the following result gives a sufficient condition for a Lévy measure (and thus for an infinitely divisible distribution) to be regularly varying.

Theorem 5 *Let ν be a Lévy measure on $\mathbb{R}^n$ with Lévy copula F and one-dimensional marginal tail integrals $U_1, \dots, U_n$. Assume that*

- *F is regularly varying with limiting Lévy copula G ;*
- *There exist a sequence of positive numbers $(c_n)_{n \geq 1}$, a constant $\alpha \in (0, 2)$ and nonnegative numbers $p_1^+, \dots, p_n^+, p_1^-, \dots, p_n^-$ with $p_i^+ + p_i^- > 0$ for at least one i , such that for all $i = 1, \dots, n$ and all $x > 0$,*

$$\lim_{n \rightarrow \infty} nU_i(c_n x) = p_i^+ |x|^{-\alpha} \quad \text{and} \quad \lim_{n \rightarrow \infty} nU_i(-c_n x) = -p_i^- |x|^{-\alpha}.$$

Then, $\nu \in RV(\alpha, (c_n), \hat{\nu})$, where $\hat{\nu}$ is a Lévy measure with Lévy copula G whose one-dimensional marginal tail integrals are given by

$$\hat{U}_i(x) = \begin{cases} p_i^+ |x|^{-\alpha}, & x > 0, \\ p_i^- |x|^{-\alpha}, & x < 0. \end{cases}$$

Remark 4 Since the Lévy copula G is homogeneous of order 1, the limiting measure $\hat{\nu}$ is the Lévy measure of an α -stable process (see [31, Theorem 4.8]).

Proof First observe that sets of the form

$$\{(y_1, \dots, y_n) \in \mathbb{R}^n : y_i \in \mathcal{J}(x_i), i \in I\}$$

for all $I \subseteq \{1, \dots, n\}$ nonempty and all $(x_i)_{i \in I} \in (\mathbb{R} \setminus \{0\})^{|I|}$, form a convergence-determining class for the $\mathbf{M}_0$ -convergence of measures on $\mathbb{R}^n$. Second, due to continuity of G with respect to each of its arguments and the continuity of $\hat{U}_1$, the measure $\hat{\nu}$ does not charge the boundaries of such sets. Therefore, it remains to prove that for all $I \subseteq \{1, \dots, n\}$ nonempty and all $(x_i)_{i \in I} \in (\mathbb{R} \setminus \{0\})^{|I|}$, the I -marginal tail integral of ν , denoted by U^I , satisfies

$$nU^I(c_n(x_i)_{i \in I}) \rightarrow \hat{U}^I((x_i)_{i \in I}),$$

as $n \rightarrow \infty$, where $\hat{U}^I$ is the I -marginal tail integral of $\hat{\nu}$. By Sklar's theorem, this is equivalent to

$$nF^I((U_i(c_n x_i))_{i \in I}) \rightarrow G^I((\hat{U}_i(x_i))_{i \in I}),$$

which follows from our assumptions and from the continuity of the limiting Lévy copula G .

As an application of Theorem 5, we can compute the tail dependence coefficients of an infinitely divisible random vector (see also [18] for a related discussion). Recall that for a two-dimensional random vector (X, Y) with continuous marginal distributions F_X and F_Y , the upper tail dependence coefficient is defined by

$$\lambda_U = \lim_{u \rightarrow 1-} \mathbb{P}[Y > F_Y^{-1}(u) | X > F_X^{-1}(u)]$$

and the lower tail dependence coefficient is defined by

$$\lambda_L = \lim_{u \rightarrow 0+} \mathbb{P}[Y \leq F_Y^{-1}(u) | X \leq F_X^{-1}(u)].$$

These coefficients depend only on the copula of (X, Y) .

Proposition 1 *Let (X_1, X_2) be an infinitely divisible random vector whose Lévy measure ν satisfies the assumptions of Theorem 5, with Lévy copula F and limiting Lévy copula G , and the coefficients p_1^+ , p_1^- , p_2^+ and p_2^- which are all strictly positive. Then,*

$$\lambda_U = \lim_{u \rightarrow 0+} u^{-1} F(u, u) = G(1, 1), \quad \lambda_L = \lim_{u \rightarrow 0+} u^{-1} F(-u, -u) = G(-1, -1).$$

Proof Denote the distribution functions of X_1 and X_2 by F_1 and F_2 , respectively. By Theorem 5, $\nu \in RV(\alpha, (c_n), \hat{\nu})$, and so $(X_1, X_2) \in RV(\alpha, (c_n), \hat{\nu})$. Under the assumptions of Theorem 5, it is easy to show that as $x \rightarrow +\infty$,

$$F_2^{-1}(F_1(x)) \sim x \left(\frac{p_1^+}{p_2^+} \right)^{-1/\alpha}.$$

Therefore,

$$\begin{aligned} \lambda_U &= \lim_{z \rightarrow +\infty} \frac{\mathbb{P}[X_2 > F_2^{-1}(F_1(z)), X_1 > z]}{\mathbb{P}[X_1 > z]} = \lim_{z \rightarrow +\infty} \frac{\mathbb{P}\left[X_2 > z \left(\frac{p_1^+}{p_2^+} \right)^{-1/\alpha}, X_1 > z\right]}{\mathbb{P}[X_1 > z]} \\ &= \lim_{n \rightarrow +\infty} \frac{n\mathbb{P}\left[X_2 > c_n \left(\frac{p_1^+}{p_2^+} \right)^{-1/\alpha}, X_1 > c_n\right]}{n\mathbb{P}[X_1 > c_n]} = \frac{\hat{\nu}\left((1, \infty) \times \left(\left(\frac{p_1^+}{p_2^+} \right)^{-1/\alpha}, \infty \right)\right)}{\hat{\nu}((1, \infty) \times \mathbb{R})} \\ &= \frac{G\left(\hat{U}_1(1), \hat{U}_2\left(\left(\frac{p_1^+}{p_2^+} \right)^{-1/\alpha}\right)\right)}{\hat{U}_1(1)} = G(1, 1) \end{aligned}$$

by the homogeneity of G . The coefficient λ_L can be computed along the same lines.

For example, for an infinitely divisible vector with regularly varying margins and Clayton Lévy copula (7), the tail dependence coefficients are given by

$$\lambda_U = \lambda_L = 2^{-1/\theta} \eta$$

6 Risk Management Applications

Lévy copulas allow for a high degree of precision in modeling joint jump dependence, and as such are well suited for risk management problems where joint extreme moves in different assets play a major role. One such application is to operational risk, which is defined by the Basel II capital accord as “the risk of losses resulting from inadequate or failed internal processes, people and systems, or from external events”. According to the Basel II framework, banks should allocate operational losses to one of eight business lines and to one of seven different loss event types. Therefore, a natural approach here is to model the different loss type/business line cells by a multidimensional compound Poisson process. Since a single loss event may affect several cells at the same time, it is essential to model the dependence of joint jumps, which is done in [8] through a Lévy copula approach. In this paper, the authors obtain explicit asymptotic formulas for the quantile of the total loss distribution (known as OpVaR) under various dependence assumptions: one dominating cell, completely dependent cells, independent cells and finally multivariate regular variation. Some of these results are extended to the Expected Shortfall risk measure in [5].

Another natural application of Lévy copulas is in insurance models. Here again, different business lines of an insurance company may be represented by compound Poisson or general Lévy processes with joint jump dependence due to the fact that certain events such as natural catastrophes may affect several business lines at the same time. We call such a representation a Lévy insurance model with Lévy copula dependence. Bregman and Klüppelberg [10] derive ruin estimates for a specific example of this model. Eder and Klüppelberg [17] compute the joint law of various quantities associated with the first passage over a fixed barrier of the sum of the components of a multidimensional Lévy process with dependence given by a Lévy copula. These computations are then used to study the ruin of an insurance company extending the results of [10]. Bäuerle and Blatter study the optimal investment and reinsurance policies for an insurance company in the Lévy insurance model with Lévy copula dependence.

Several option pricing applications also require precise modeling of joint jumps. An archetypical example is provided by a niche OTC product known as gap risk swap (sometimes also called gap note or crash note). A multiline version of this product may be structured as follows:

- At inception, the protection seller pays the notional amount N to the protection buyer and receives Libor plus spread monthly until maturity. If no gap event occurs, the protection seller receives the full notional amount at the maturity of the contract.

- A gap event is defined as a downside move of over 20 % during one business day in any underlying from a basket of 10 names.
- If a gap event occurs, the protection seller receives at maturity a reduced notional amount kN , where the reduction factor k is determined from the number M of gap events using the following table:

$$\begin{array}{ccccccc} M & 0 & 1 & 2 & 3 & \geq 4 \\ k & 1 & 1 & 1 & 0.5 & 0 \end{array}$$

This option is designed to capture large downside moves occurring simultaneously (or almost simultaneously) within a basket containing several stocks. Lévy copulas therefore provide a convenient tool for pricing and risk managing this product, assuming that the parameters can be reliably estimated. We refer the reader to [44] for explicit formulas for prices and hedge ratios of single-name and multiname gap risk swaps.

Risk analysis via multivariate regular variation Consider a multidimensional exponential Lévy model:

$$S_i = e^{X_i}, i = 1, \dots, n,$$

where $X = (X_1, \dots, X_n)$ is infinitely divisible with Lévy measure ν , let $\nu \in RV(\alpha, (c_n), \mu)$, and assume that one is interested in evaluating the potential loss of a long-only portfolio containing the stocks $S_1, \dots, S_n$. In other words, we are interested in the behavior of

$$\mathbb{P} \left[\sum_{i=1}^n e^{X_i} \leq z \right]$$

for small values of z . This probability admits the following bounds.

$$\mathbb{P}[X_i \leq \log z - \log n, i = 1, \dots, n] \leq \mathbb{P} \left[\sum_{i=1}^n e^{X_i} \leq z \right] \leq \mathbb{P}[X_i \leq \log z, i = 1, \dots, n]$$

The regular variation assumption implies that both bounds have the same asymptotic behavior as $z \rightarrow 0$. Indeed, since the measure μ is homogeneous, it does not charge the set $A := (-\infty, -1] \times \dots \times (-\infty, -1]$ and by Theorem 2.4 in [30],

$$\mathbb{P}[X_i \leq \log z - \log n, i = 1, \dots, n] \sim \mathbb{P}[X_i \leq \log z, i = 1, \dots, n] \sim \frac{\mu(A)}{n(z)}$$

as $z \rightarrow 0$, where $n(z) := \{n : c_n = \lfloor \log \frac{1}{z} \rfloor\}$, so that

$$\mathbb{P} \left[\sum_{i=1}^n e^{X_i} \leq z \right] \sim \frac{\mu(A)}{n(z)} \quad \text{as } z \rightarrow 0.$$

To be more specific, assume now that each asset follows the finite moment log stable model of Carr and Wu [12]. Recall that an α -stable random variable with $\alpha \neq 1$ has characteristic function

$$\phi(z) = \exp\left(-\sigma^\alpha |z|^\alpha \left(1 - i\beta \operatorname{sgn} z \tan \frac{\pi\alpha}{2}\right) + i\mu z\right),$$

where $\beta \in [-1, 1]$ is the asymmetry parameter, $\sigma > 0$ is the scale parameter and $\mu \in \mathbb{R}$ is the shift parameter. An α -stable law with these parameters will be denoted by $S_\alpha(\sigma, \beta, \mu)$. The finite moment log stable model assumes that $X_i \sim S_\alpha(\sigma_i, -1, \mu_i)$ with $\alpha \in (1, 2)$. With this choice of β , the Lévy measure of X_i is supported by $(-\infty, 0)$ and all moments of $S_i = e^{X_i}$ are finite.

To describe the joint behavior of the assets, assume that the Lévy copula of $X_1, \dots, X_n$ is the Clayton Lévy copula F given by (3) with parameter θ (note that since the Lévy measures of all components are supported on the negative half-axis, the dependence may be described by a Lévy copula on $[0, \infty]^n$).

This model satisfies the assumptions of Theorem 5 with the limiting Lévy copula $G = F$ and parameters $p_i^+ = 0$ and $p_i^- = \frac{\sigma_i^\alpha}{\Gamma(1-\alpha) \cos \frac{\pi\alpha}{2}}$ for $i = 1, \dots, n$ and $c_n = n^{1/\alpha}$. Therefore, as $z \downarrow 0$,

$$\mathbb{P}\left[\sum_{i=1}^n e^{X_i} \leq z\right] \sim \left(\log \frac{1}{z}\right)^{-\alpha} F_\theta(p_1^-, \dots, p_n^-).$$

7 Conclusion

In this paper we have reviewed the recent literature on Lévy copulas, including the numerical and statistical methods and some applications in risk management. We have also presented a new simulation algorithm and discussed the role of Lévy copulas in the context of multivariate regular variation.

Lévy copulas offer a very precise control over the joint jumps of a multidimensional Lévy process. For this reason, they are relevant for applications where one is interested in extremes and especially joint extremes of infinitely divisible random vectors or multidimensional stochastic processes. In other words, Lévy copulas provide a flexible modeling approach in the context of multivariate extreme value theory, the full potential of which is yet to be exploited. In this context it is important to note that multivariate regularly varying Lévy processes based on Lévy copulas have strong dependence in the tails (meaning that the extremes remain dependent), whereas other approaches, for example those based on subordination, may not lead to strong dependence [45], and are therefore not suitable for modeling joint extremes.

As we have seen in this chapter, multivariate regular variation allows to relate the tail properties of an infinitely divisible random vector to the properties of the Lévy copula in a very explicit way. This connection should certainly be developed further,

but another relevant question concerns the joint tail behavior of a multidimensional Lévy process outside the framework of multivariate regular variation. For instance, the components may exhibit faster than power law tail decay, or may be strongly heterogeneous.

From the point of view of applications, several authors have developed Lévy copula-based models for insurance, market risk and operational risk. While these domains are certainly very relevant, another important potential application appears to be in renewable energy production and distribution. Renewable energy production (for example, from wind), and electricity consumption are intermittent by nature, and spatially distributed, which naturally leads to models based on stochastic processes in large dimension. These processes exhibit jumps, spikes and non-Gaussian behavior, and understanding their joint extremes is crucial for the management of electrical distribution networks. Therefore, multidimensional Lévy processes based on Lévy copulas are natural building blocks for models in this important domain.

Acknowledgments I would like to thank the editor Robert Stelzer and the anonymous reviewer for their constructive comments on the first version of the manuscript. This research was partially supported by the grant of the Government of Russian Federation 14.12.31.0007.

References

1. Ambrosio, L., Fusco, N., Pallara, D.: Functions of Bounded Variation and Free Discontinuity Problems. Oxford University Press (2000)
2. Barndorff-Nielsen, O.E., Lindner, A.M.: Lévy copulas: dynamics and transforms of Upsilon type. *Scand. J. Stat.* **34**, 298–316 (2007)
3. Barndorff-Nielsen, O.E., Pedersen, J., Sato, K.-I.: Multivariate subordination, self-decomposability and stability. *Adv. Appl. Probab.* **33**, 160–187 (2001)
4. Basrak, B., Davis, R.A., Mikosch, T.: A characterization of multivariate regular variation. *Ann. Appl. Probab.* 908–920 (2002)
5. Biagini, F., Ulmer, S.: Asymptotics for operational risk quantified with expected shortfall. *Astin Bull.* **39**, 735–752 (2009)
6. Bielecki, T.R., Jakubowski, J., Niewegłowski, M.: Dynamic modeling of dependence in finance via copulae between stochastic processes, in *Copula theory and its applications, lecture notes in statistics*, vol.198, Part 1, pp. 33–76. Springer (2010)
7. Bielecki, T.R., Jakubowski, J., Vidozzi, A., Vidozzi, L.: Study of dependence for some stochastic processes. *Stoch. Anal. Appl.* **26**, 903–924 (2008)
8. Böcker, K., Klüppelberg, C.: Multivariate models for operational risk. *Quant. Finance* **10**, 855–869 (2010)
9. Bollerslev, T., Todorov, V., Li, S.Z.: Jump tails, extreme dependencies, and the distribution of stock returns. *J. Econom.* **172**, 307–324 (2013)
10. Bregman, Y., Klüppelberg, C.: Ruin estimation in multivariate models with clayton dependence structure. *Scand. Actuar. J.* **2005**, 462–480 (2005)
11. Bücher, A., Vetter, M.: Nonparametric inference on Lévy measures and copulas. *Ann. Stat.* **41**, 1485–1515 (2013)
12. Carr, P., Wu, L.: The finite moment logstable process and option pricing. *J. Finance* **58**, 753–778 (2003)
13. Cont, R., Tankov, P.: Financial Modelling with Jump Processes. Chapman and Hall / CRC Press (2004)

14. De Haan L., Ferreira, A.: *Extreme value theory: an introduction*. Springer (2007)
15. Eberlein, E.: Applications of generalized hyperbolic Lévy motion to finance. In: Barndorff-Nielsen, O., Mikosch, T., Resnick, S. (eds.) *Lévy Processes—Theory and Applications*, pp. 319–336. Birkhäuser, Boston (2001)
16. Eberlein, E., Madan, D.B.: On correlating Lévy processes. *J. Risk* **13**, 3–16 (2010)
17. Eder, I., Klüppelberg, C.: The first passage event for sums of dependent Lévy processes with applications to insurance risk. *Ann. Appl. Probab.* **19**, 2047–2079 (2009)
18. Eder, I., Klüppelberg, C.: Pareto Lévy measures and multivariate regular variation. *Adv. Appl. Probab.* **44**, 117–138 (2012)
19. Embrechts, P., Klüppelberg, C., Mikosch, T.: *Modelling extremal events for insurance and finance*. In: *Applications of Mathematics*, vol. 33. Springer, Berlin (1997)
20. Embrechts, P., Lambrigger, D.D., Wüthrich, M.V.: Multivariate extremes and the aggregation of dependent risks: examples and counter-examples. *Extremes* **12**, 107–127 (2009)
21. Esmaeili, H., Klüppelberg, C.: Parameter estimation of a bivariate compound Poisson process. *Insur.: Math. Econ.* **47**, 224–233 (2010)
22. Esmaeili, H., Klüppelberg, C.: Parametric estimation of a bivariate stable Lévy process. *J. Multivar. Anal.* **102**, 918–930 (2011)
23. Esmaeili, H., Klüppelberg, C.: Two-step estimation of a multi-variate Lévy process. *J. Time Ser. Anal.* **34**, 668–690 (2013)
24. Farkas, W., Reich, N., Schwab, C.: Anisotropic stable Lévy copula processes-analytical and numerical aspects. *Math. Models Methods Appl. Sci.* **17**, 1405–1443 (2007)
25. Grothe, O.: Jump tail dependence in Lévy copula models. *Extremes* **16**, 303–324 (2013)
26. Grothe, O., Nicklas, S.: Vine constructions of Lévy copulas. *J. Multivar. Anal.* **119**, 1–15 (2013)
27. Hilber, N., Reich, N., Schwab, C., Winter, C.: Numerical methods for Lévy processes. *Finance Stochast.* **13**, 471–500 (2009)
28. Hult, H., Lindskog, F.: Extremal behavior of regularly varying stochastic processes. *Stoch. Process. Appl.* **115**, 249–274 (2005)
29. Hult, H., Lindskog, F.: On regular variation for infinitely divisible random vectors and additive processes. *Adv. Appl. Probab.*, pp. 134–148 (2006)
30. Hult, H., Lindskog, F.: Regular variation for measures on metric spaces. *Publications de l'Institut Mathématique (Beograd)* **80**, 121–140 (2006)
31. Kallsen, J., Tankov, P.: Characterization of dependence of multidimensional Lévy processes using Lévy copulas. *J. Multivar. Anal.* **97**, 1551–1572 (2006)
32. Kawai, R.: A multivariate Lévy process model with linear correlation. *Quant. Finance* **9**, 597–606 (2009)
33. Klüppelberg, C., Resnick, S.I.: The Pareto copula, aggregation of risks, and the emperor's socks. *J. Appl. Probab.* **45**, 67–84 (2008)
34. Lindskog, F., McNeil, A.J.: Common Poisson shock models: applications to insurance and credit risk modelling. *Astin Bull.* **33**, 209–238 (2003)
35. Luciano, E., Schoutens, W.: A multivariate jump-driven financial asset model. *Quant. Finance* **6**, 385–402 (2006)
36. Madan, D., Yen, J.-Y.: Asset allocation with multivariate non-Gaussian returns. In: Birge, J., Linetsky, V. (eds.) *Financial Engineering Handbooks in Operations Research and Management Science*, vol. 15, pp. 949–969. North Holland, Amsterdam (2007)
37. Madan, D.B.: Equilibrium asset pricing: with non-Gaussian factors and exponential utilities. *Quant. Finance* **6**, 455–463 (2006)
38. Prause, K.: *The generalized hyperbolic model: estimation, financial derivatives, and risk measures*, PhD thesis, University of Freiburg (1999)
39. Reich, N., Schwab, C., Winter, C.: On Kolmogorov equations for anisotropic multivariate Lévy processes. *Finance Stochast.* **14**, 527–567 (2010)
40. Resnick, S.: *Extreme Values, Regular Variation, and Point Processes*. Springer (1987)
41. Rosiński, J.: Series representations of Lévy processes from the perspective of point processes. In: Barndorff-Nielsen, O., Mikosch, T., Resnick, S. (eds.) *Lévy Processes—Theory and Applications*, pp. 401–415. Birkhäuser, Boston (2001)

- 42. Tankov, P.: Dependence structure of spectrally positive multidimensional Lévy processes. <http://www.proba.jussieu.fr/pageperso/tankov/> (2003) [Unpublished manuscript]
- 43. Tankov, P.: Simulation and option pricing in Lévy copula models. <http://www.proba.jussieu.fr/pageperso/tankov/> (2004) [Unpublished manuscript]
- 44. Tankov, P.: Pricing and hedging gap risk. *J. Comput. Finan.* **13**, 1–27 (2010)
- 45. Tankov, P.: Left tail of the sum of dependent positive random variables. [arxiv:1402.4683](https://arxiv.org/abs/1402.4683) (2014) [preprint]
- 46. Vidozzi, L.: Two essays on multivariate stochastic processes and applications to credit risk modeling. PhD thesis, Illinois Institute of Technology, Chicago, IL (2009)

Weak Stationarity of Ornstein-Uhlenbeck Processes with Stochastic Speed of Mean Reversion

Fred Espen Benth and Asma Khedher

Abstract When modeling energy prices with the Ornstein-Uhlenbeck process, it was shown in Barlow et al. [2] and Zapranis and Alexandris [21] that there is a large uncertainty attached to the estimation of the speed of mean-reversion and that it is not constant but may vary considerably over time. In this paper we generalise the Ornstein-Uhlenbeck process to allow for the speed of mean reversion to be stochastic. We suppose that the mean-reversion is a Brownian stationary process. We apply Malliavin calculus in our computations and we show that this generalised Ornstein-Uhlenbeck process is stationary in the weak sense. Moreover we compute the instantaneous rate of change in the mean and in the squared fluctuations of the generalised Ornstein-Uhlenbeck process given its initial position. Finally, we derive the chaos expansion of this generalised Ornstein-Uhlenbeck process.

Keywords Ornstein-Uhlenbeck processes · Mean reversion · Stationarity

1 Introduction

An Ornstein-Uhlenbeck (OU) process X is defined as the solution of the stochastic differential equation

$$dX(t) = -\alpha X(t) dt + \sigma dW(t), \quad (1)$$

where W is a Brownian motion and σ and α are two positive constants. Such processes have applications, for example in the areas of physics and finance. The process X has a drift which will push it towards its long-term mean level at the origin, while the

F.E. Benth (✉)

Department of Mathematics, University of Oslo,
1053, N-0316 Blindern, Oslo, Norway
e-mail: fredb@math.uio.no

A. Khedher

Chair of Mathematical Finance, Technische Universität München,
Parking 11, 85748 Garching-Hochbruck, Germany
e-mail: asma.khedher@cma.uio.no

Brownian component introduces random fluctuations. It is well-known that X has a Gaussian stationary distribution.

The parameter α is sometimes referred to as the speed of mean-reversion. In practical applications, it may be hard to identify precisely the speed of mean-reversion α . For instance, in modelling energy prices based on OU processes, Barlow et al. [2] studied the problem of estimating parameters based on historical data. They found out that there is a large uncertainty attached to the estimation of the speed of mean-reversion. In their study of Paris daily temperature data, Zapranis and Alexandridis [21] showed by means of wavelet analysis that the mean reversion rate is not constant but may vary considerably over time.

In this paper we generalize the OU dynamics in (1) to allow for a speed of mean-reversion α being a stochastic process. We study the weak stationarity of this *generalised OU process*. In other words, we analyze the stationarity properties of the mean, the variance, and the covariance of the generalised OU process for simple specifications of α . Specifically, in our analysis we suppose α is a Brownian stationary process. Brownian stationary processes are themselves an extension of OU-processes, and have been intensively studied in the area of turbulence and finance (see Barndorff-Nielsen and Schmiegel [4], or Barndorff-Nielsen et al. [5] for the case of energy markets). Despite the fact that α may attain negative values, we are able to show the stationarity of the mean, the variance, and the covariance of X when the average speed of mean-reversion is sufficiently larger than its variance. Explicit conditions for these results to hold are derived. In our analysis, some aspects of Malliavin Calculus are applied.

We present the generalised OU (GOU) process as an alternative to Markov switching diffusions. In these latter models one can consider different states of α to model the changing in the speed of mean-reversion. However when the number of the states exceeds three, the parameter estimation to the model becomes cumbersome as one has also to estimate the probabilities of the change to the different states (see for example Carvalho and Lopez [11] and Perlin [18]).

In order to describe the behavior of the first and second order moment of increments of the generalised Ornstein-Uhlenbeck process X , we derive the instantaneous rate of change in the mean of X given the initial position of the process. We show that the latter is given in terms of the mean of α . Moreover we compute the instantaneous rate of change in the squared fluctuations of X given its initial position and we show that this is given in terms of the volatility σ . Hence, locally our process behaves like a classical Ornstein-Uhlenbeck process.

We further compute the chaos expansion of the generalised OU process. We show that for a specific choice of the process α , the chaos of order 1 converges pointwise to a function in $L^2(\mathbb{R})$. However, it does not converge in $L^2(\mathbb{R})$.

A generalised Ornstein-Uhlenbeck (OU) process is sometimes in the literature defined by

$$V(t) = \int_{-\infty}^t e^{-L(s-)} dU(s), \quad (2)$$

where $\{L, U\}$ is a bivariate Lévy process. See Carmona et al. [10] and Lindner and Sato [15] for basic properties of such processes. These processes have been applied in many areas, in particular in option pricing (see e.g. Yor [20]) or in insurance (see Dufresne [12]). The explicit dynamics V in (2) does not solve an OU-type stochastic differential equation and L is not immediately interpretable as a speed of mean-reversion. Another path of study is the so-called *quasi* Ornstein-Uhlenbeck processes, which are defined as processes X solving a stochastic differential equation of the type (1), however, with W being a general noise process with stationary increments (see Barndorff-Nielsen and Basse O'Connor [3]).

The paper is organized as follows. In Sect. 2, we describe the generalised Ornstein-Uhlenbeck process we consider in our analysis. In Sect. 3, we show that the mean and variance of this model class is stationary under some mild conditions on the model parameters. In Sect. 3 we compute the instantaneous rate of change of the first and the second order moment of increments of the generalised Ornstein-Uhlenbeck process X . In Sect. 4 we derive the chaos expansion of the model.

2 Generalised Ornstein-Uhlenbeck Processes

Let $(\Omega, \mathcal{F}, P)$ be a complete probability space equipped with a filtration $(\mathcal{F}_t)_{t \geq 0}$. We denote by $W = \{W(t)\}_{t \geq 0}$ a standard Brownian motion. We write D_t for the Malliavin derivative at time $t \geq 0$, defined on the subspace $\mathbb{D}^{1,2}$ of $L^2(P) := L^2(\Omega, \mathcal{F}, P)$. We define the Skorohod operator to be the adjoint of the Malliavin operator D . A stochastic process Y is said to be Skorohod integrable on $[0, T]$, for some $T < \infty$, if

$$\left| \mathbb{E} \left[\int_0^T (D_t Z) Y(t) dt \right] \right| \leq c \|Z\|_2,$$

for all $Z \in \mathbb{D}^{1,2}$, where c is a constant depending on Y and $\|\cdot\|_2$ is the $L^2(P)$ -norm. We denote the Skorohod integral of Y over $[0, T]$ by

$$\int_0^T Y(t) \delta W(t).$$

If Y is $(\mathcal{F}_t)$ -adapted, the Skorohod integral of Y coincides with the Itô integral, that is

$$\int_0^T Y(t) \delta W(t) = \int_0^T Y(t) dW(t).$$

For an introduction to Malliavin Calculus, we refer the reader to Nualart [17].

Introduce the *generalised Ornstein-Uhlenbeck process* (GOU) X as the solution to the stochastic differential equation

$$dX(t) = -\alpha(t)X(t) dt + \sigma dW(t), \quad (3)$$

where σ is a positive constant and α is an $(\mathcal{F}_t)$ -adapted stochastic process. Notice that α is *not* restricted to be positive, but may attain negative values as well. The initial condition $X(0) = X_0$ is assumed to be a constant. In the next proposition we derive the explicit solution to (3).

Proposition 2.1 *Assume that α is integrable on $[0, T]$ for a given $T < \infty$ and*

$$\mathbb{E} \left[\exp \left(2 \int_0^T \alpha(u) du \right) \right] < \infty .$$

Then, for $t \leq T$, the $(\mathcal{F}_t)$ -adapted process

$$\begin{aligned} X(t) = & X_0 \exp \left(- \int_0^t \alpha(u) du \right) \\ & + \exp \left(- \int_0^t \alpha(u) du \right) \int_0^t \sigma \exp \left(\int_0^s \alpha(u) du \right) dW(s) , \end{aligned}$$

is the solution to (3).

Proof We apply the Itô formula to obtain

$$\begin{aligned} d \left(X(t) \exp \left(\int_0^t \alpha(u) du \right) \right) \\ = \alpha(t) X(t) \exp \left(\int_0^t \alpha(u) du \right) dt + \exp \left(\int_0^t \alpha(u) du \right) dX(t) \\ = \sigma \exp \left(\int_0^t \alpha(u) du \right) dW(t) . \end{aligned}$$

Integrating yields the representation of X . Note that the stochastic integral is well-defined in Itô sense due to the $(\mathcal{F}_t)$ -adaptedness of α and the integrability condition.

Let the explicit dynamics X in Proposition 2.1 be our GOU process. Notice that in the case α is deterministic, it is usual to write

$$X(t) = X_0 \exp \left(- \int_0^t \alpha(u) du \right) + \int_0^t \sigma \exp \left(- \int_s^t \alpha(u) du \right) dW(s) .$$

That is, moving the exponential inside the Itô integral. In the general case, where α is stochastic, this is no longer a valid representation as $\exp(\int_0^t \alpha(u) du)$ is anticipating. However, the next proposition shows that we can move the exponential inside the integral when we interpret the stochastic integral in the sense of Skorohod. Moreover, we get an additional drift term involving the Malliavin derivative of the exponential.

Proposition 2.2 Assume that for every $0 \leq t \leq T$, $\alpha(t) \in \mathbb{D}^{1,2}$ and

$$\mathbb{E} \left[\int_0^t \exp \left(-2 \int_s^t \alpha(u) du \right) ds \right] < \infty. \quad (4)$$

Then, for every $0 \leq t \leq T$ we have a.s.

$$\begin{aligned} X(t) = & X_0 \exp \left(- \int_0^t \alpha(u) du \right) + \int_0^t \sigma \exp \left(- \int_s^t \alpha(u) du \right) \delta W(s) \\ & - \int_0^t \int_s^t \sigma (D_s \alpha(v)) \exp \left(- \int_s^t \alpha(u) du \right) dv ds. \end{aligned}$$

Proof Applying Proposition 1.3.3 in Nualart [17], we find from the assumptions in the proposition that

$$\begin{aligned} & \exp \left(- \int_0^t \alpha(u) du \right) \int_0^t \exp \left(\int_0^s \alpha(u) du \right) dW(s) \\ &= \int_0^t \exp \left(- \int_s^t \alpha(u) du \right) \delta W(s) \\ & \quad + \int_0^t \left(D_s \exp \left(- \int_0^t \alpha(u) du \right) \right) \exp \left(\int_0^s \alpha(u) du \right) ds. \end{aligned}$$

By the chain rule of the Malliavin derivative, we find

$$\begin{aligned} D_s \exp \left(- \int_0^t \alpha(u) du \right) &= - \left(D_s \int_0^t \alpha(v) dv \right) \exp \left(- \int_0^t \alpha(u) du \right) \\ &= - \int_0^t D_s \alpha(v) dv \exp \left(- \int_0^t \alpha(u) du \right). \end{aligned}$$

Since $\alpha(v)$ is $\mathcal{F}_v$ -measurable, we find that $D_s \alpha(v) = 0$ a.s. for $s > v$. Thus, a.s.,

$$\int_0^t D_s \alpha(v) dv = \int_s^t D_s \alpha(v) dv.$$

This proves the proposition.

Notice that condition (4) for the existence of the solution to the GOU process seems to be a strong condition. However later on we give an example of a broad class of processes α satisfying this condition (see Eq. (6)).

As the Skorohod integral has mean zero, we find from the Fubini-Tonelli theorem that

$$\begin{aligned} \mathbb{E} [X(t)] &= X_0 \mathbb{E} \left[\exp \left(- \int_0^t \alpha(u) du \right) \right] \\ & \quad - \sigma \int_0^t \int_s^t \mathbb{E} \left[(D_s \alpha(v)) \exp \left(- \int_s^t \alpha(u) du \right) \right] dv ds. \end{aligned}$$

The second order moment can be expressed by using the “isometry” for Skorohod integrals and integration by parts. This is the purpose of the next proposition.

Proposition 2.3 *Assume that $\alpha \in \mathbb{D}^{1,2}$ and that for every $0 \leq t \leq T$,*

$$D_w \alpha(v) e^{-\int_w^t \alpha(u) du} \in \mathbb{D}^{1,2}, \quad e^{-\int_0^t \alpha(u) du} \in \mathbb{D}^{1,2},$$

$e^{-\int_s^t \alpha(u) du}$ is Skorohod integrable, and $e^{-\int_0^t \alpha(u) du} e^{-\int_s^t \alpha(u) du} \in L^2(\Omega, \mathcal{F}, \mathbb{P})$. Then,

$$\begin{aligned} \mathbb{E}[X^2(t)] &= X_0^2 \mathbb{E}\left[e^{-2\int_0^t \alpha(u) du}\right] + \sigma^2 \int_0^t \mathbb{E}\left[e^{-2\int_s^t \alpha(u) du}\right] ds \\ &\quad + \sigma^2 \int_0^t \int_0^t \mathbb{E}\left[\int_s^t D_v \alpha(w) dw e^{-\int_s^t \alpha(u) du} \int_v^t D_s \alpha(w) dw e^{-\int_v^t \alpha(u) du}\right] dv ds \\ &\quad + \sigma^2 \mathbb{E}\left[\left(\int_0^t \int_s^t (D_s \alpha(v)) e^{-\int_s^t \alpha(u) du} dv ds\right)^2\right] \\ &\quad - 4X_0 \sigma \int_0^t \int_s^t \mathbb{E}\left[(D_s \alpha(v)) e^{-\int_0^t \alpha(u) du - \int_s^t \alpha(u) du}\right] dv ds \\ &\quad - 2\sigma^2 \int_0^t \int_0^t \int_w^t \mathbb{E}\left[D_s \left((D_w \alpha(v)) e^{-\int_w^t \alpha(u) du}\right) dv dw e^{-\int_s^t \alpha(u) du}\right] ds. \end{aligned}$$

Proof We have

$$\begin{aligned} X^2(t) &= X_0^2 e^{-2\int_0^t \alpha(u) du} + \left(\int_0^t \sigma e^{-\int_s^t \alpha(u) du} \delta W(s)\right)^2 \\ &\quad + \left(\int_0^t \int_s^t \sigma (D_s \alpha(v)) e^{-\int_s^t \alpha(u) du} dv ds\right)^2 \\ &\quad + 2X_0 e^{-\int_0^t \alpha(u) du} \int_0^t \sigma e^{-\int_s^t \alpha(u) du} \delta W(s) \\ &\quad - 2X_0 e^{-\int_0^t \alpha(u) du} \int_0^t \int_s^t \sigma (D_s \alpha(v)) e^{-\int_s^t \alpha(u) du} dv ds \\ &\quad - 2 \int_0^t \int_s^t \sigma (D_s \alpha(v)) e^{-\int_s^t \alpha(u) du} dv ds \int_0^t \sigma e^{-\int_s^t \alpha(u) du} \delta W(s). \end{aligned}$$

Now, the integration by parts in Eq. (1.48) of Nualart [17] and under the assumptions mentioned in the proposition, we get

$$\begin{aligned}
e^{-\int_0^t \alpha(u) du} \int_0^t e^{-\int_s^t \alpha(u) du} \delta W(s) &= \int_0^t e^{-\int_0^t \alpha(u) du - \int_s^t \alpha(u) du} \delta W(s) \\
&\quad + \int_0^t (D_s e^{-\int_0^t \alpha(u) du}) e^{-\int_s^t \alpha(u) du} ds \\
&= \int_0^t e^{-\int_0^t \alpha(u) du - \int_s^t \alpha(u) du} \delta W(s) \\
&\quad - \int_0^t \int_s^t (D_s \alpha(v)) e^{-\int_0^t \alpha(u) du - \int_s^t \alpha(u) du} dv ds.
\end{aligned}$$

Taking expectations, gives

$$\begin{aligned}
&\mathbb{E} \left[e^{-\int_0^t \alpha(u) du} \int_0^t e^{-\int_s^t \alpha(u) du} \delta W(s) \right] \\
&= - \int_0^t \int_s^t \mathbb{E} \left[(D_s \alpha(v)) e^{-\int_0^t \alpha(u) du - \int_s^t \alpha(u) du} \right] dv ds.
\end{aligned}$$

A similar argument shows that

$$\begin{aligned}
&\mathbb{E} \left[\int_0^t \int_w^t (D_w \alpha(v)) e^{-\int_w^t \alpha(u) du} dv dw \int_0^t e^{-\int_s^t \alpha(u) du} \delta W(s) \right] \\
&= \int_0^t \int_0^t \int_w^t \mathbb{E} \left[D_s \left((D_w \alpha(v)) e^{-\int_w^t \alpha(u) du} \right) dv dw e^{-\int_s^t \alpha(u) du} \right] ds.
\end{aligned}$$

Applying the covariance formula for Skorohod integrals in Eq. (1.48) of Nualart [17], it holds that

$$\begin{aligned}
&\mathbb{E} \left[\left(\int_0^t e^{-\int_s^t \alpha(u) du} \delta W(s) \right)^2 \right] \\
&= \int_0^t \mathbb{E} \left[e^{-2 \int_s^t \alpha(u) du} \right] ds \\
&\quad + \int_0^t \int_0^t \mathbb{E} \left[(D_v e^{-\int_s^t \alpha(u) du}) (D_s e^{-\int_v^t \alpha(u) du}) \right] dv ds \\
&= \int_0^t \mathbb{E} \left[e^{-2 \int_s^t \alpha(u) du} \right] ds \\
&\quad + \int_0^t \int_0^t \mathbb{E} \left[\int_s^t D_v \alpha(w) dw e^{-\int_s^t \alpha(u) du} \int_v^t D_s \alpha(w) dw e^{-\int_v^t \alpha(u) du} \right] dv ds.
\end{aligned}$$

After collecting terms the statement is proved.

In the next section we specify α to be a Brownian stationary process and analyse the weak stationarity of X .

3 Weak Stationarity

First, we recall the definition of weak stationarity (see for example Kloeden and Platen [14]).

Definition 3.1 Let X be a stochastic process, μ be a constant, and $c : \mathbb{R} \rightarrow \mathbb{R}$ be a function. Then X is called weakly stationary if its mean, its variance, and its covariance satisfy

$$\mathbb{E}[X(t)] = \mu, \quad \text{Var}[X(t)] = c(0), \quad \text{Cov}[X(t), X(t - \delta)] = c(\delta),$$

for all $t \in [0, T]$ and $\delta > 0$.

We weaken Definition 3.1 to processes with limiting mean and variance, that is processes for which the mean and the variance have a limit when time tends to infinity. Then we prove that when t goes to infinity, the first and the second order moments are not affected by a shift of the time variable. Let $g : \mathbb{R}_+ \mapsto \mathbb{R}$ be a measurable function such that

$$\int_0^\infty g^2(u) du < \infty. \quad (5)$$

For a constant $\mu \in \mathbb{R}$, define α to be the Brownian stationary (BS) process

$$\alpha(t) = \mu + \int_{-\infty}^t g(t-s) dW(s), \quad (6)$$

where W is considered to be a two-sided Wiener process on $\mathbb{R}$. That is a Wiener process W_1 and an independent copy of it, W_2 , are defined on $\mathbb{R}_+$. Then W is defined as $W(t) = W_1(t)$ for $t \geq 0$ and $W(t) = -W_2(-t)$, for $t < 0$, where W_1 is the same Brownian motion as in (3). For more about these processes, we refer to Sect. 6.2 in Benth and Šaltytė Benth [7].

Let $\mathcal{F}_0$ be the σ -algebra generated by $\{W_2(t), t \geq 0\}$. We define $(\mathcal{F}_t)_{t \geq 0}$ to be the filtration generated by $\mathcal{F}_0$ and $\sigma\{W_1(s), 0 \leq s \leq t\}$. By the condition on g , we immediately see by the Itô isometry and properties of Wiener integration that $\alpha(t)$ is a Gaussian $(\mathcal{F}_t)$ -adapted process with mean μ and variance $\int_0^\infty g^2(u) du$ independent of time t , thus being stationary. For more about Brownian stationary processes, see for example Brockwell and Davis [9] and Marquardt and Stelzer [16]. A typical example is to consider $g(u) = \eta \exp(-\beta u)$, for positive constants η and β . Thus α is the stationary solution of the OU process solving the stochastic differential equation

$$d\alpha(t) = \beta(\mu - \alpha(t)) dt + \eta dW(t). \quad (7)$$

In this specific case, the speed of mean reversion α is itself an OU process. If W has a positive increment, then so does α , and thus increasing the speed of mean reversion of X . The process α will decrease exponentially fast back at a rate β towards its

long-term mean level μ , being the average speed of mean reversion for X . A negative increment of W will push α downwards, yielding a slower mean reversion of X . Thus, positive increments of W implies faster mean-reversion, whereas negative increments means slower mean reversion. We can turn this relationship around by supposing η to be negative, meaning that α depends on W opposite to X . Note that α can become negative, as the Wiener integral in (6) is normally distributed and therefore takes values on the whole real line. However, for relatively large and positive values of μ , the probability of negative speeds of mean reversion will become small. As we shall see below, the mean, the variance, and the covariance of X are stationary when a specific relationship in the size between μ and g holds.

Observe that for any $0 \leq s < t < \infty$, Fubini-Tonelli theorem and Cauchy-Schwarz's inequality along with the Itô isometry yield

$$\begin{aligned} \mathbb{E} \left[\int_s^t \left| \int_{-\infty}^u g(u-v) dW(v) \right| du \right] &\leq \int_s^t \mathbb{E} \left[\left(\int_{-\infty}^u g(u-v) dW(v) \right)^2 \right]^{1/2} du \\ &= \sqrt{\int_0^\infty g^2(u) du} \times (t-s) < \infty. \end{aligned}$$

Thus, $\alpha(u)$ is integrable on any interval $[s, t]$, $0 \leq s < t < \infty$. Remark that the process α is in general not a semimartingale. In fact, by a suitable choice of g (see e.g. Alos et al. [1]) we can allow for α to be a fractional Brownian motion. It is possible to show that α is a semimartingale when $g(0)$ is well-defined and g is absolutely continuous and has a square integrable derivative defined almost everywhere (see Basse and Pedersen [6] and Benth and Eyjolfsson [8]). For example, the interesting case of continuous-time autoregressive moving average processes satisfy these properties (see Benth and Šaltytė Benth [7] for an application of these processes to weather modelling).

Introduce the function $h(x, y)$ for $0 \leq x \leq y$ by

$$h(x, y) = \int_x^y g(u) du. \quad (8)$$

We have the following useful lemma.

Lemma 3.2 *Assume that*

$$\int_0^\infty h^2(u, x+u) du < \infty,$$

for any $x > 0$. Then for every $0 \leq s \leq t$

$$\int_s^t \alpha(u) du = \mu(t-s) + \int_{-\infty}^s h(s-u, t-u) dW(u) + \int_s^t h(0, t-u) dW(u). \quad (9)$$

Proof First, observe that the condition on the function h ensures that the first Wiener integral on the right-hand side of (9) is well-defined since,

$$\int_{-\infty}^s h^2(s-u, t-u) du = \int_0^{\infty} h^2(u, u+t-s) du .$$

After appealing to the Cauchy-Schwartz inequality we find,

$$\int_s^t h^2(0, t-u) du = \int_0^{t-s} h^2(0, u) du \leq \int_0^{\infty} g^2(u) du \int_0^{t-s} u du < \infty ,$$

and therefore the second Wiener integral is also well-defined.

By definition of α , we find

$$\int_s^t \alpha(u) du = \mu(t-s) + \int_s^t \int_{-\infty}^u g(u-v) dW(v) du .$$

But

$$\int_{-\infty}^u g(u-v) dW(v) = \int_{-\infty}^s g(u-v) dW(v) + \int_s^u g(u-v) dW(v) .$$

By the stochastic Fubini-Tonelli theorem, we find

$$\int_s^t \int_s^u g(u-v) dW(v) du = \int_s^t \int_v^t g(u-v) du dW(v) = \int_s^t \int_0^{t-v} g(u) du dW(v) ,$$

and hence,

$$\int_s^t \int_s^u g(u-v) dW(v) du = \int_s^t h(0, t-v) dW(v) .$$

By the assumption on h , we again apply the stochastic Fubini-Tonelli theorem to find

$$\int_s^t \int_{-\infty}^s g(u-v) dW(v) du = \int_{-\infty}^s \int_s^t g(u-v) du dW(v) = \int_{-\infty}^s \int_{s-v}^{t-v} g(u) du dW(v) .$$

Hence, the Lemma follows after using the definition of h in the last integral.

We observe that $\int_s^t \alpha(u) du$ is represented as a sum of two independent Wiener integrals. This will enable us to compute exponential moments of $\int_s^t \alpha(u) du$ easily. Remark that from the Cauchy-Schwartz inequality

$$\left(\int_u^{u+t-s} g(v) dv \right)^2 = \left(\int_u^{u+t-s} 1 \cdot g(v) dv \right)^2 \leq x \int_u^{u+x} g^2(v) dv ,$$

and therefore a sufficient condition for $\int_0^\infty h^2(u, u+x) du$ to be finite is

$$\int_0^\infty \int_u^{u+x} g^2(v) dv du < \infty. \quad (10)$$

As an example, consider α to be an OU process, with $g(v) = \exp(-\beta v)$, for a constant $\beta > 0$. Then

$$\int_u^{u+x} g^2(v) dv = \frac{1}{2\beta} (1 - e^{-2\beta x}) e^{-2\beta u},$$

and hence (10) holds.

The Malliavin derivative of $\alpha(u)$ is simple to compute. It holds that

$$D_s \alpha(v) = g(v-s), \quad (11)$$

for $s < v$. In the case $s > v$, $D_s \alpha(v) = 0$. There is a potential problem at $v = s$ since g might not be defined there. However, as we are going to integrate expressions like $D_s \alpha(v)$ with respect to the Lebesgue measure, we leave the Malliavin derivative undefined for this singular point. If $g(0)$ is well-defined, there is no problem.

Using (11), we find

$$\begin{aligned} \int_0^t \int_s^t D_s \alpha(v) e^{-\int_s^t \alpha(u) du} dv ds &= \int_0^t \int_s^t g(v-s) dv e^{-\int_s^t \alpha(u) du} ds \\ &= \int_0^t \int_0^{t-s} g(v) dv e^{-\int_s^t \alpha(u) du} ds \\ &= \int_0^t h(0, t-s) e^{-\int_s^t \alpha(u) du} ds. \end{aligned}$$

We conclude from Proposition 2.2 that

$$X(t) = X_0 e^{-\int_0^t \alpha(u) du} + \int_0^t \sigma e^{-\int_s^t \alpha(u) du} \delta W(s) - \int_0^t \sigma h(0, t-s) e^{-\int_s^t \alpha(u) du} ds, \quad (12)$$

where $\int_s^t \alpha(u) du$ is expressed in Lemma 3.2. Note that the exponential integrability (4) in Proposition 2.2 which is a condition on $\int_s^t \alpha(u) du$ is satisfied as this is a normal random variable which has finite exponential moments of all orders. The condition of Malliavin differentiability is also readily verified in this explicit case. Recalling (9), since h is a deterministic function, it follows from Theorem 2.2.1 in Nualart [17] that $\int_s^t \alpha(u) du$ is Malliavin differentiable. By the chain rule, so is $\exp(-\int_s^t \alpha(u) du)$. We know from Proposition 1.3.1 of Nualart [17] that the space of Malliavin differentiable random variables is included in the domain of the Skorohod integral. This ensures the Skorohod integrability of $\exp(-\int_s^t \alpha(u) du)$.

From now on we assume that the function h defined in (8) satisfies the condition of Lemma 3.2 and in light of the above discussion, the model verifies Propositions 2.1 and 2.2. In the next subsection we study the stationarity of the mean of the GOU process.

3.1 Stationarity of the Mean

We compute the expectation of X and show that it has a limit when t goes to ∞ .

Proposition 3.3 *The expected value of X is*

$$\begin{aligned} \mathbb{E}[X(t)] = X_0 \exp \left(-\mu t + \frac{1}{2} \int_0^\infty h^2(u, t+u) du + \frac{1}{2} \int_0^t h^2(0, u) du \right) \\ - \int_0^t \sigma h(0, v) \exp \left(-\mu v + \frac{1}{2} \int_0^\infty h^2(u, v+u) du + \frac{1}{2} \int_0^v h^2(0, u) du \right) dv. \end{aligned}$$

If,

$$\lim_{t \rightarrow \infty} \exp \left(-\mu t + \frac{1}{2} \int_0^\infty h^2(u, t+u) du + \frac{1}{2} \int_0^t h^2(0, u) du \right) = 0, \quad (13)$$

and

$$\int_0^\infty \sigma |h(0, v)| \exp \left(-\mu v + \frac{1}{2} \int_0^\infty h^2(u, v+u) du + \frac{1}{2} \int_0^v h^2(0, u) du \right) dv < \infty, \quad (14)$$

then $\mathbb{E}[X(t)]$ has a limit when $t \rightarrow \infty$ given by

$$\begin{aligned} \lim_{t \rightarrow \infty} \mathbb{E}[X(t)] \\ = - \int_0^\infty \sigma h(0, v) \exp \left(-\mu v + \frac{1}{2} \int_0^\infty h^2(u, v+u) du + \frac{1}{2} \int_0^v h^2(0, u) du \right) dv. \end{aligned}$$

Proof As the expectation of the Skorohod integral is zero, we have by Fubini-Tonelli theorem

$$\mathbb{E}[X(t)] = X_0 \mathbb{E} \left[e^{-\int_0^t \alpha(u) du} \right] - \int_0^t \sigma h(0, t-s) \mathbb{E} \left[e^{-\int_s^t \alpha(u) du} \right] ds.$$

From Lemma 3.2, we find by independence of the Wiener integrals, that

$$\begin{aligned}\mathbb{E}\left[e^{-\int_0^t \alpha(u) du}\right] &= e^{-\mu t} \mathbb{E}\left[e^{-\int_{-\infty}^0 h(-u, t-u) dW(u)}\right] \mathbb{E}\left[e^{-\int_0^t h(0, t-u) dW(u)}\right] \\ &= e^{-\mu t} \exp\left(\frac{1}{2} \int_{-\infty}^0 h^2(-u, t-u) du + \frac{1}{2} \int_0^t h^2(0, t-u) du\right) \\ &= e^{-\mu t} \exp\left(\frac{1}{2} \int_0^\infty h^2(u, t+u) du + \frac{1}{2} \int_0^t h^2(0, u) du\right).\end{aligned}$$

Similarly, we find

$$\mathbb{E}\left[e^{-\int_s^t \alpha(u) du}\right] = e^{-\mu(t-s)} \exp\left(\frac{1}{2} \int_0^\infty h^2(u, t-s+u) du + \frac{1}{2} \int_0^{t-s} h^2(0, u) du\right). \quad (15)$$

Hence, the expression for $\mathbb{E}[X(t)]$ follows. It is simple to see that the integrability conditions imply the limit as claimed.

It is easily seen from the conditions in Proposition 3.3 that $\mu > 0$ is a necessary condition in order to have a limiting expectation. Notice that the limiting expectation of X may become negative. For example, if g in the definition of the process α is positive, then trivially $h(0, v)$ is positive and we have a negative limiting expectation of X . This is in sharp contrast to the classical OU process with a constant (and positive) speed of mean reversion α , as this has zero expectation in stationarity. We attribute the negative expected limiting value of X to the probability (however small) that α itself can be negative. However, as μ is positive, the mean of α will be positive as well. A negative α gives a non-stationary behavior, which will locally occur for X during times when α crosses zero from above

Let us consider the case where α is an OU process. In the following lemma we compute the stationary mean of the GOU process for this specific choice of α .

Lemma 3.4 *Let α be an OU process as in (7). If*

$$\mu\beta > \frac{\eta^2}{2\beta}, \quad (16)$$

then

$$\begin{aligned}\lim_{t \rightarrow \infty} \mathbb{E}[X(t)] \\ = - \int_0^\infty \frac{\sigma\eta}{\beta} (1 - e^{-\beta v}) e^{-\left(\mu - \frac{\eta^2}{2\beta^2}\right)v - \frac{\eta^2}{\beta^3}(1 - e^{-\beta v}) + \frac{\eta^2}{4\beta^3}(1 - e^{-2\beta v}) + \frac{\eta^2}{4\beta^3}(1 - e^{-\beta v})^2} dv.\end{aligned}$$

Proof Let α be an OU process, with $g(u) = \eta \exp(-\beta u)$, for constants η and β with $\beta > 0$. Then

$$h(x, y) = \int_x^y \eta e^{-\beta u} du = \frac{\eta}{\beta} (e^{-\beta x} - e^{-\beta y}) .$$

Hence,

$$\int_0^\infty h^2(u, t+u) du = \frac{\eta^2}{\beta^2} \int_0^\infty (e^{-\beta u} - e^{-\beta(u+t)})^2 du = \frac{\eta^2}{2\beta^3} (1 - e^{-\beta t})^2 .$$

Next,

$$\begin{aligned} \int_0^t h^2(0, u) du &= \frac{\eta^2}{\beta^2} \int_0^t (1 - e^{-\beta u})^2 du \\ &= \frac{\eta^2}{\beta^2} t - \frac{2\eta^2}{\beta^3} (1 - e^{-\beta t}) + \frac{\eta^2}{2\beta^3} (1 - e^{-2\beta t}) . \end{aligned}$$

Therefore, it holds that

$$\lim_{t \rightarrow \infty} \exp \left(-\mu t + \frac{1}{2} \int_0^\infty h^2(u, t+u) du + \frac{1}{2} \int_0^t h^2(0, u) du \right) = 0 ,$$

if and only if (16) holds true. Moreover,

$$\begin{aligned} -\mu v + \frac{1}{2} \int_0^\infty h^2(u, v+u) du + \frac{1}{2} \int_0^v h^2(0, u) du \\ = -\left(\mu - \frac{\eta^2}{2\beta^2} \right) v - \frac{\eta^2}{\beta^3} (1 - e^{-\beta v}) \\ + \frac{\eta^2}{4\beta^3} (1 - e^{-2\beta v}) + \frac{\eta^2}{4\beta^3} (1 - e^{-\beta v})^2 . \end{aligned}$$

This implies,

$$\begin{aligned} \int_0^\infty h(0, v) \exp \left(-\mu v + \frac{1}{2} \int_0^\infty h^2(u, v+u) du + \frac{1}{2} \int_0^v h^2(0, u) du \right) dv \\ \leq c \int_0^\infty (1 - e^{-\beta v}) \exp \left(-\left(\mu - \frac{\eta^2}{2\beta^2} \right) v \right) dv , \end{aligned}$$

for some constant $c > 0$. Under condition (16) the integral is finite and we prove the statement.

Note that by the definition of g , the stationary variance of α is $\eta^2/2\beta$. Hence for X to have a stationary mean value, the stationary mean of α times its speed of mean reversion must be larger than the stationary variance of α . In the case $\eta > 0$, this stationary mean value becomes negative, whereas $\eta < 0$ gives a positive stationary

mean since g and therefore $h(0, v)$ are negative. If $\eta < 0$ an increase in X due to a positive increment of W occurs in parallel to a decrease in α . This would mean that X is pushed away from its mean and reverts slower giving the rational for a positive stationary mean of the process X .

3.2 Stationarity of the Variance

In this section we analyze the second order moment of X and its limiting behavior. To reduce the number of terms, we suppose that $X_0 = 0$, which gives

$$X(t) = \int_0^t \sigma e^{-\int_s^t \alpha(u) du} \delta W(s) - \int_0^t \sigma h(0, t-s) e^{-\int_s^t \alpha(u) du} ds.$$

We find $\mathbb{E}[X^2(t)] = I_1 - 2I_2 + I_3$ where

$$I_1 = \mathbb{E} \left[\left(\int_0^t \sigma e^{-\int_s^t \alpha(u) du} \delta W(s) \right)^2 \right], \quad (17)$$

$$I_2 = \mathbb{E} \left[\int_0^t \sigma^2 e^{-\int_s^t \alpha(u) du} \delta W(s) \int_0^t h(0, t-s) e^{-\int_s^t \alpha(u) du} ds \right], \quad (18)$$

$$I_3 = \mathbb{E} \left[\left(\int_0^t \sigma h(0, t-s) e^{-\int_s^t \alpha(u) du} ds \right)^2 \right]. \quad (19)$$

We want to compute the three expectations (17)–(19). The approach is based on the same ideas as when we calculated the expectation of X .

Notice that when considering X_0 different from 0, there will be additive expectations of the type we studied in Proposition 3.3 and also of the type we will study in Lemmas 3.7 and 3.8. Using similar steps as in these lemmas we can derive expressions for these expectations depending on X_0 . Then imposing conditions similar to the one in (13), the terms involving X_0 will vanish for t going to infinity. We choose $X_0 = 0$ to shorten the paper and avoid similar proofs.

A key element in our derivations of the variance is the expectation of terms like $\exp(-\int_s^t \alpha(u) du - \int_v^t \alpha(u) du)$ for $s, v \in [0, t]$. This is the content of the next lemma.

Lemma 3.5 *It holds for $v \leq s \leq t$,*

$$\ln \mathbb{E} \left[\exp \left(-\int_s^t \alpha(u) du - \int_v^t \alpha(u) du \right) \right] = -2\mu(t-s) - \mu(s-v) + H(s-v, t-s), \quad (20)$$

where

$$H(x, y) = \frac{1}{2} \int_0^\infty (h(x+u, x+y+u) + h(u, x+y+u))^2 du \\ + \frac{1}{2} \int_0^x (h(u, y+u) + h(0, y+u))^2 du + 2 \int_0^y h^2(0, u) du ,$$

for $x, y \geq 0$.

Proof As

$$- \int_s^t \alpha(u) du = -\mu(t-s) - \int_{-\infty}^s h(s-u, t-u) dW(u) - \int_s^t h(0, t-u) dW(u) ,$$

we have for $v \leq s$

$$\begin{aligned} & - \int_s^t \alpha(u) du - \int_v^t \alpha(u) du \\ &= -\mu(t-s) - \mu(t-v) \\ & \quad - \left(\int_{-\infty}^s h(s-u, t-u) dW(u) + \int_{-\infty}^v h(v-u, t-u) dW(u) \right) \\ & \quad - \left(\int_s^t h(0, t-u) dW(u) + \int_v^t h(0, t-u) dW(u) \right) \\ &= -2\mu(t-s) - \mu(s-v) \\ & \quad - \left(\int_{-\infty}^v h(s-u, t-u) + h(v-u, t-u) dW(u) \right. \\ & \quad \left. + \int_v^s h(s-u, t-u) dW(u) \right) \\ & \quad - \left(2 \int_s^t h(0, t-u) dW(u) + \int_v^s h(0, t-u) dW(u) \right) \\ &= -2\mu(t-s) - \mu(s-v) \\ & \quad - \int_{-\infty}^v h(s-u, t-u) + h(v-u, t-u) dW(u) \\ & \quad - \int_v^s h(s-u, t-u) + h(0, t-u) dW(u) - 2 \int_s^t h(0, t-u) dW(u) . \end{aligned}$$

The three Wiener integrals will be independent by the properties of Brownian motion. Hence, the results follow after using the fact that the exponential of a Wiener integral is lognormally distributed.

We remark that the case $v > s$ is covered by the above result by simply interchanging the roles of v and s .

Let us now compute the three expectations (17)–(19). We start with the expectation in (17), which is computed in the following lemma.

Lemma 3.6 *It holds that*

$$\begin{aligned} & \mathbb{E} \left[\left(\int_0^t e^{-\int_s^t \alpha(u) du} \delta W(s) \right)^2 \right] \\ &= \int_0^t e^{-2\mu u + H(0,u)} du + \int_0^t \int_0^{t-v} h(0, v) h(u, u+v) e^{-2\mu v - \mu u + H(u,v)} du dv \\ & \quad + \int_0^t \int_0^v h(u, v) h(0, v-u) e^{-2\mu(v-u) - \mu u + H(u,v-u)} du dv. \end{aligned}$$

Proof By the formula for the variance of a Skorohod integral (Proposition 1.3.1 in Nualart [17]), we have

$$\begin{aligned} & \mathbb{E} \left[\left(\int_0^t e^{-\int_s^t \alpha(u) du} \delta W(s) \right)^2 \right] \\ &= \mathbb{E} \left[\int_0^t e^{-2\int_s^t \alpha(u) du} ds \right] + \mathbb{E} \left[\int_0^t \int_0^t \left(D_v e^{-\int_s^t \alpha(u) du} \right) \left(D_w e^{-\int_v^t \alpha(u) du} \right) dv dw \right] \\ &= \mathbb{E} \left[\int_0^t e^{-2\int_s^t \alpha(u) du} ds \right] \\ & \quad + \int_0^t \int_0^t \mathbb{E} \left[e^{-\int_s^t \alpha(u) du - \int_v^t \alpha(u) du} \right] \left(\int_s^t D_v \alpha(u) du \right) \left(\int_v^t D_w \alpha(u) du \right) dv dw. \end{aligned}$$

Here we have applied the chain rule for Malliavin differentiation together with the Fubini-tonelli theorem. Since $D_v \alpha(u) = g(u-v) \mathbf{1}(u > v)$, it holds

$$\begin{aligned} \int_s^t D_v \alpha(u) du &= \int_s^t g(u-v) \mathbf{1}(u > v) du = \int_{\max(v,s)}^t g(u-v) du \\ &= \int_{\max(v,s)-v}^{t-v} g(w) dw = h(\max(v, s) - v, t - v). \end{aligned} \quad (21)$$

Similarly,

$$\int_v^t D_w \alpha(u) du = h(\max(v, w) - w, t - w).$$

Therefore,

$$\begin{aligned}
& \mathbb{E} \left[\left(\int_0^t e^{-\int_s^t \alpha(u) du} \delta W(s) \right)^2 \right] \\
&= \mathbb{E} \left[\int_0^t e^{-2 \int_s^t \alpha(u) du} ds \right] \\
&\quad + \int_0^t \int_0^t \mathbb{E} \left[e^{-\int_s^t \alpha(u) du - \int_v^t \alpha(u) du} \right. \\
&\quad \quad \left. h(\max(s, v) - s, t - s) h(\max(s, v) - v, t - v) dv ds \right] \\
&= \int_0^t e^{-2\mu(t-s) + H(0, t-s)} ds \\
&\quad + \int_0^t \int_0^s h(0, t-s) h(s-v, t-v) e^{-2\mu(t-s) - \mu(s-v) + H(s-v, t-s)} dv ds \\
&\quad + \int_0^t \int_s^t h(v-s, t-s) h(0, t-v) e^{-2\mu(t-v) - \mu(v-s) + H(v-s, t-v)} dv ds,
\end{aligned}$$

where in the latter we applied Lemma 3.5. Hence, the statement of the lemma follows after a change of variables in the integrals.

We derive the second expectation (18) in the next lemma.

Lemma 3.7 *It holds that*

$$\begin{aligned}
& \mathbb{E} \left[\int_0^t e^{-\int_s^t \alpha(u) du} \delta W(s) \int_0^t h(0, t-s) e^{-\int_s^t \alpha(u) du} ds \right] \\
&= - \int_0^t \int_0^{t-v} h(0, u+v) h(0, v) e^{-2\mu v - \mu u + H(u, v)} du dv \\
&\quad - \int_0^t \int_0^v h(0, v-u) h(u, v) e^{-2\mu(v-u) - \mu u + H(u, v-u)} du dv.
\end{aligned}$$

Proof Using integration by parts for Skorohod integrals together with the Fubini-Tonelli theorem, we find

$$\begin{aligned}
& \mathbb{E} \left[\int_0^t e^{-\int_s^t \alpha(u) du} \delta W(s) \int_0^t h(0, t-s) e^{-\int_s^t \alpha(u) du} ds \right] \\
&= \int_0^t h(0, t-v) \mathbb{E} \left[e^{-\int_v^t \alpha(u) du} \int_0^t e^{-\int_s^t \alpha(u) du} \delta W(s) \right] dv \\
&= \int_0^t h(0, t-v) \mathbb{E} \left[\int_0^t e^{-\int_s^t \alpha(u) du - \int_v^t \alpha(u) du} \delta W(s) \right] dv \\
&\quad + \int_0^t h(0, t-v) \mathbb{E} \left[\int_0^t \left(D_s e^{-\int_v^t \alpha(u) du} \right) e^{-\int_s^t \alpha(u) du} ds \right] dv.
\end{aligned}$$

The Skorohod integral has zero expectation. Hence,

$$\begin{aligned} \mathbb{E} \left[\int_0^t e^{-\int_s^t \alpha(u) du} \delta W(s) \int_0^t h(0, t-s) e^{-\int_s^t \alpha(u) du} ds \right] \\ = - \int_0^t \int_0^t h(0, t-v) h(\max(v, s) - s, t-s) \mathbb{E} \left[e^{-\int_v^t \alpha(u) du - \int_s^t \alpha(u) du} \right] ds dv, \end{aligned}$$

after using the chain rule for the Malliavin derivative and the fact that $D_s \alpha(u) = g(u-s) \mathbf{1}(u > s)$. Invoking Lemma 3.5 proves the result.

Finally, we derive the expectation in (19).

Lemma 3.8 *It holds*

$$\begin{aligned} \mathbb{E} \left[\left(\int_0^t h(0, t-s) e^{-\int_s^t \alpha(u) du} ds \right)^2 \right] \\ = \int_0^t \int_0^{t-v} h(0, v) h(0, u+v) e^{-2\mu v - \mu u + H(u, v)} du dv \\ + \int_0^t \int_0^v h(0, v) h(0, v-u) e^{-2\mu(v-u) - \mu u + H(u, v-u)} du dv. \end{aligned}$$

Proof We find,

$$\begin{aligned} \mathbb{E} \left[\left(\int_0^t h(0, t-s) e^{-\int_s^t \alpha(u) du} ds \right)^2 \right] \\ = \mathbb{E} \left[\int_0^t h(0, t-s) e^{-\int_s^t \alpha(u) du} ds \int_0^t h(0, t-v) e^{-\int_v^t \alpha(u) du} dv \right] \\ = \int_0^t \int_0^t h(0, t-s) h(0, t-v) \mathbb{E} \left[e^{-\int_s^t \alpha(u) du - \int_v^t \alpha(u) du} \right] dv ds. \end{aligned}$$

Appealing to Lemma 3.5 and doing a change of variables prove the statement of the lemma.

Collecting together the results for the expectations in (17)–(19) calculated in Lemmas 3.6–3.8, we get

Proposition 3.9 *The second order moment of X is*

$$\begin{aligned} \mathbb{E} [X^2(t)] &= \sigma^2 \int_0^t e^{-2\mu u + H(0, u)} du \\ &+ \sigma^2 \int_0^t \int_0^v e^{-2\mu u - \mu(v-u) + H(v-u, u)} h(0, u) (h(v-u, v) + 3h(0, v)) du dv \\ &+ \sigma^2 \int_0^t \int_0^v e^{-2\mu(v-u) - \mu u + H(u, v-u)} h(0, v-u) (3h(u, v) + h(0, v)) du dv. \end{aligned}$$

If

$$\begin{aligned} \int_0^\infty e^{-2\mu u + H(0,u)} du &< \infty \\ \int_0^\infty \int_0^v |h(0,u)| |h(v-u,v) + 3h(0,v)| e^{-2\mu u - \mu(u+v) + H(v-u,u)} du dv &< \infty \\ \int_0^\infty \int_0^v |h(0,v-u)| |3h(u,v) + h(0,v)| e^{-2\mu(v-u) - \mu u + H(u,v-u)} du dv &< \infty, \end{aligned}$$

then the second order moment of X has a limit given by

$$\begin{aligned} \lim_{t \rightarrow \infty} \mathbb{E}[X^2(t)] \\ = \sigma^2 \int_0^\infty e^{-2\mu u + H(0,u)} du \\ + \sigma^2 \int_0^\infty \int_0^v h(0,u) (h(v-u,v) + 3h(0,v)) e^{-2\mu u - \mu(u+v) + H(v-u,u)} du dv \\ + \sigma^2 \int_0^\infty \int_0^v h(0,v-u) (3h(u,v) + h(0,v)) e^{-2\mu(v-u) - \mu u + H(u,v-u)} du dv. \end{aligned}$$

Proof The results of the Lemmas 3.6–3.8 give

$$\begin{aligned} \mathbb{E}[X^2(t)] \\ = \sigma^2 \int_0^t e^{-2\mu u + H(0,u)} du \\ + \sigma^2 \int_0^t \int_0^{t-v} h(0,v) (h(u,u+v) + 3h(0,u+v)) e^{-2\mu v - \mu u + H(u,v)} du dv \\ + \sigma^2 \int_0^t \int_0^v h(0,v-u) (h(0,v) + 3h(u,v)) e^{-2\mu(v-u) - \mu u + H(u,v-u)} du dv. \end{aligned}$$

Note from calculus that for sufficiently “nice” functions $q(t, v)$, it holds

$$\int_0^t q(t, v) dv = \int_0^t q(v, v) dv + \int_0^t \int_0^v \frac{\partial q}{\partial v}(v, u) du dv. \quad (22)$$

Hence, letting

$$xq(t, v) = \int_0^{t-v} h(0, v) (h(u, u+v) + 3h(0, u+v)) e^{-2\mu v - \mu u + H(u,v)} du,$$

we have $q(v, v) = 0$ and

$$\frac{\partial q}{\partial t}(t, v) = h(0, v) (h(t-v, t) + 3h(0, t-v)) e^{-2\mu v - \mu(t-v) + H(t-v,v)}.$$

Using (22), the proposition follows.

In the latter proposition we showed the stationarity of the second order moment of the GOU process. The stationarity of the variance follows immediately from Proposition 3.3, where besides of the integrability conditions of Proposition 3.9, we have to assume that (13) and (14) hold.

In the next lemma we investigate the conditions of stationarity of the second order moment of X , where we restrict our attention to the OU specification.

Lemma 3.10 *Let α be an OU process as in (7). If*

$$\mu > \frac{\eta^2}{\beta^2}, \quad (23)$$

then

$$\begin{aligned} \lim_{t \rightarrow \infty} \mathbb{E}[X^2(t)] &= \sigma^2 \int_0^\infty e^{-2\mu u + H(0,u)} du \\ &+ \int_0^\infty \int_0^v \frac{\sigma^2 \eta^2}{\beta^2} (e^{-\beta(v-u)} - 4e^{-\beta v} + 3)(1 - e^{-\beta u}) e^{-2\mu u - \mu(v-u) + H(v-u,u)} du dv \\ &+ \int_0^\infty \int_0^v \frac{\sigma^2 \eta^2}{\beta^2} (1 - e^{-\beta(v-u)})(3e^{-\beta u} - 4e^{-\beta v} + 1) e^{-2\mu(v-u) - \mu u + H(u,v-u)} du dv, \end{aligned}$$

where

$$\begin{aligned} H(x, y) &= \frac{\eta^2}{4\beta^3} (1 - 2e^{-\beta(x+y)} + e^{-\beta x})^2 + \frac{\eta^2}{2\beta^2} x \\ &+ \frac{\eta^2}{4\beta^3} (1 - 4e^{-\beta y} + 4e^{-2\beta y})(1 - e^{-2\beta x}) + \frac{\eta^2}{\beta^3} (1 - 2e^{-\beta y})(1 - e^{-\beta x}) \\ &+ \frac{2\eta^2}{\beta^2} y - \frac{4\eta^2}{\beta^3} (1 - e^{-\beta y}) + \frac{\eta^2}{\beta^3} (1 - e^{-2\beta y}). \end{aligned} \quad (24)$$

Proof We consider we have an OU process. Thus $g(u) = \eta \exp(-\beta u)$, for η and β constants and $\beta > 0$. As we have seen earlier,

$$h(x, y) = \frac{\eta}{\beta} (e^{-\beta x} - e^{-\beta y}).$$

We use this to compute $H(x, y)$. First, we find

$$\int_0^y h^2(0, u) du = \frac{\eta^2}{\beta^2} y - \frac{2\eta^2}{\beta^3} (1 - e^{-\beta y}) + \frac{\eta^2}{2\beta^3} (1 - e^{-2\beta y}).$$

Next,

$$\begin{aligned} \int_0^x (h(u, y+u) + h(0, y+u))^2 du &= \frac{\eta^2}{\beta^2} x + \frac{\eta^2}{2\beta^3} (1 - 4e^{-\beta y} + 4e^{-2\beta y})(1 - e^{-2\beta x}) \\ &\quad + \frac{2\eta^2}{\beta^3} (1 - 2e^{-\beta y})(1 - e^{-\beta x}). \end{aligned}$$

Finally,

$$\int_0^\infty (h(x+u, x+y+u) + h(u, x+y+u))^2 du = \frac{\eta^2}{2\beta^3} (1 - 2e^{-\beta(x+y)} + e^{-\beta x})^2.$$

Therefore, we find the expression (24) for $H(x, y)$. Consider now the first term in the second order moment of X . We find

$$\int_0^\infty e^{-2\mu u + H(0,u)} du \leq c \int_0^t \exp\left(-2\left(\mu - \frac{\eta^2}{\beta^2}\right)u\right) du < \infty,$$

as long as $\mu > \eta^2/\beta^2$. Hence, the limit of the first term exists when $t \rightarrow \infty$ under this restriction on the parameters. Next we consider the second term in the second order moment of X . It holds,

$$\begin{aligned} &\int_0^\infty \int_0^v \frac{\sigma^2 \eta^2}{\beta^2} |e^{-\beta(v-u)} - 4e^{-\beta v} + 3| |1 - e^{-\beta u}| e^{-2\mu u - \mu(v-u) + H(v-u,u)} du dv \\ &\leq c \int_0^\infty \int_0^v |e^{-u(\mu - 3\frac{\eta^2}{2\beta^2})} e^{-v(\mu - \frac{\eta^2}{2\beta^2})}| du dv \\ &\leq c \int_0^\infty \left| \frac{e^{-2v(\mu - \frac{\eta^2}{\beta^2})} - e^{-v(\mu - \frac{\eta^2}{2\beta^2})}}{-\mu + \frac{3\eta^2}{2\beta^2}} \right| dv < \infty, \end{aligned}$$

as long as $\mu > \eta^2/\beta^2$. Finally we consider the last term in the second order moment of X . We find

$$\begin{aligned} &\int_0^\infty \int_0^v \frac{\sigma^2 \eta^2}{\beta^2} |1 - e^{-\beta(v-u)}| |3e^{-\beta u} - 4e^{-\beta v} + 1| e^{-2\mu(v-u) - \mu u + H(u,v-u)} du dv \\ &\leq c \int_0^\infty \int_0^v e^{-2v(\mu - \frac{\eta^2}{\beta^2})} e^{-u(-\mu + \frac{3\eta^2}{2\beta^2})} du dv \\ &\leq c \int_0^\infty e^{-v(\mu - \frac{\eta^2}{2\beta^2})} - e^{-2v(\mu + \frac{\eta^2}{\beta^2})} dv < \infty, \end{aligned}$$

as long as $\mu > \eta^2/\beta^2$. With this condition on the parameters, the third term in the expression of the second order moment of X has a limit as time goes to infinity.

Therefore, the second order moment of X has a stationary limit as long as $\mu\beta$ is greater than twice the stationary variance of α , which we recall as $\eta^2/2\beta$. This condition is more restrictive than the one ensuring the stationarity of the mean. Thus, it obviously implies the stationarity of the expectation of X . In conclusion, when we specify α to be the OU process given by (7) then X has a stationary limit for the expectation and variance if

$$\mu\beta > \frac{\eta^2}{\beta}.$$

Observe that for fixed η , the stationary variance of α is decreasing with increasing speed of mean reversion β . On the other hand, for fixed μ , the expression $\mu\beta$ is obviously increasing with β . Hence, the condition for stationarity is less restrictive for models with high speed of mean reversion in the α than those of slow. Note that the slower the mean reversion in α becomes, the closer α gets to a non-stationary drifted Brownian motion. Hence, to obtain stationarity for small β 's, one must have sufficiently high mean levels μ , and/or, sufficiently small noise level η in the dynamics of α . Notice that our choice of X_0 being equal to zero is only for simplification in the calculations. However, it does not change the result on the stationarity of the variance.

3.3 Stationarity of the Covariance

In this section we compute the covariance of X and prove its stationarity. We do not present the proofs since they follow the same lines as the proofs of the results presented in Sect. 3.2. In the following proposition, we compute $\mathbb{E}[X(t)X(t-\delta)]$, for $\delta > 0$.

Proposition 3.11 *It holds for $\delta > 0$,*

$$\begin{aligned} & \mathbb{E}[X(t)X(t-\delta)] \\ &= \sigma^2 \int_{\delta}^t \exp(-\mu v - \mu(v-\delta) + \tilde{H}(0, v, v-\delta)) dv \\ &+ \sigma^2 \int_{\delta}^t \int_0^s \exp(-\mu v - \mu(s-\delta) + \tilde{H}(s-v, v, v-\delta)) \left\{ h(0, v-\delta) \right. \\ &\quad \left. [h(s-v, s) + h(0, s-\delta)] + h(0, v)[h(0, s) + h(0, s-\delta)] \right\} dv ds \\ &+ \sigma^2 \int_{\delta}^t \int_0^{s-\delta} \exp(-\mu s - \mu(s-v-\delta) + \tilde{H}(v, s-v, s-v-\delta)) \left\{ h(0, s-v) \right. \\ &\quad \left. [h(v, s-\delta) + h(v, s)] + h(0, s-v-\delta)[h(v, s-\delta) + h(0, s)] \right\} dv ds \\ &+ \sigma^2 \left(\int_0^{\delta} h(0, \delta-v) G(\delta-v) dv \right) \left(\int_0^{t-\delta} h(0, v) G(v) dv \right), \end{aligned}$$

where

$$\begin{aligned}\tilde{H}(x, y, z) = & \frac{1}{2} \int_0^\infty \{h(x+u, x+y+u) + h(u, z+x+u)\}^2 du \\ & + \frac{1}{2} \int_0^x \{h(u, u+y) + h(0, u+z)\}^2 du \\ & + \frac{1}{2} \int_0^z \{h(0, y-u) + h(0, z-u)\}^2 du + \frac{1}{2} \int_0^{y-z} h^2(0, y-z-u) du,\end{aligned}$$

for $x, y, z \geq 0$. If

$$\begin{aligned}& \int_\delta^\infty \exp(-\mu v - \mu(v-\delta) + \tilde{H}(0, v, v-\delta)) dv < \infty, \\ & \int_\delta^\infty \int_0^s \exp(-\mu v - \mu(s-\delta) + \tilde{H}(s-v, v, v-\delta)) \left| h(0, v-\delta) \right. \\ & \quad \left. [h(s-v, s) + h(0, s-\delta)] + h(0, v)[h(0, s) + h(0, s-\delta)] \right| dv ds < \infty, \\ & \int_\delta^\infty \int_0^{s-\delta} \exp(-\mu s - \mu(s-v-\delta) + \tilde{H}(v, s-v, s-v-\delta)) \\ & \quad \left| h(0, s-v)[h(v, s-\delta) + h(v, s)] + h(0, s-v-\delta)[h(v, s-\delta) + h(0, s)] \right| dv ds < \infty, \\ & \int_0^\infty |h(0, v)G(v)| dv < \infty,\end{aligned}$$

then

$$\begin{aligned}& \lim_{t \rightarrow \infty} \mathbb{E}[X(t)X(t-\delta)] \\ & = \sigma^2 \int_\delta^\infty \exp(-\mu v - \mu(v-\delta) + \tilde{H}(0, v, v-\delta)) dv \\ & \quad + \sigma^2 \int_\delta^\infty \int_0^s \exp(-\mu v - \mu(s-\delta) + \tilde{H}(s-v, v, v-\delta)) \left\{ h(0, v-\delta) \right. \\ & \quad \left. [h(s-v, s) + h(0, s-\delta)] + h(0, v)[h(0, s) + h(0, s-\delta)] \right\} dv ds \\ & \quad + \sigma^2 \int_\delta^\infty \int_0^{s-\delta} \exp(-\mu s - \mu(s-v-\delta) + \tilde{H}(v, s-v, s-v-\delta)) \left\{ h(0, s-v) \right. \\ & \quad \left. [h(v, s-\delta) + h(v, s)] + h(0, s-v-\delta)[h(v, s-\delta) + h(0, s)] \right\} dv ds \\ & \quad + \sigma^2 \left(\int_0^\delta h(0, \delta-v)G(\delta-v)dv \right) \left(\int_0^\infty h(0, v)G(v)dv \right).\end{aligned}$$

Considering the integrability conditions of the latter proposition together with the integrability conditions of Proposition 3.9, we prove the stationarity of the covariance and thus the weak stationarity of the GOU process.

Notice that when we consider $\delta = 0$ in Proposition 3.11, we recover the expression for the second order moment studied in Proposition 3.9. We chose to present the results for the second order moment in Sect. 3.2 to be more didactic.

In the next lemma we investigate the conditions of stationarity of the covariance where we restrict our attention to the case where α is an OU process.

Lemma 3.12 *Let α be an OU process as in (7). If*

$$\mu > \frac{\eta^2}{\beta^2},$$

then

$$\begin{aligned} & \lim_{t \rightarrow \infty} \mathbb{E}[X(t)X(t - \delta)] \\ &= \sigma^2 \int_{\delta}^{\infty} \exp(-\mu v - \mu(v - \delta) + \tilde{H}(0, v, v - \delta)) dv \\ &+ \sigma^2 \int_{\delta}^{\infty} \int_0^s \exp(-\mu v - \mu(s - \delta) + \tilde{H}(s - v, v, v - \delta)) \left\{ h(0, v - \delta) \right. \\ &\quad \left. [h(s - v, s) + h(0, s - \delta)] + h(0, v)[h(0, s) + h(0, s - \delta)] \right\} dv ds \\ &+ \sigma^2 \int_{\delta}^{\infty} \int_0^{s-\delta} \exp(-\mu s - \mu(s - v - \delta) + \tilde{H}(v, s - v, s - v - \delta)) \left\{ h(0, s - v) \right. \\ &\quad \left. [h(v, s - \delta) + h(v, s)] + h(0, s - v - \delta)[h(v, s - \delta) + h(0, s)] \right\} dv ds \\ &+ \sigma^2 \left(\int_0^{\delta} h(0, \delta - v) G(\delta - v) dv \right) \left(\int_0^{\infty} h(0, v) G(v) dv \right). \end{aligned}$$

where

$$\begin{aligned} & \tilde{H}(x, y, z) \\ &= \frac{\eta^2}{4\beta^3} \left(1 - e^{\beta(x+y)} - e^{-\beta(x+z)} + e^{-\beta x} \right)^2 + \frac{\eta^2}{\beta^3} (1 - e^{-\beta x})(1 - e^{-\beta z} - e^{-\beta y}) \\ &+ \frac{\eta^2}{2\beta^2} x + \frac{\eta^2}{4\beta^3} (1 - e^{-2\beta x}) \left(1 + e^{-2\beta y} - 2e^{-\beta y} + e^{-2\beta z} - 2e^{-\beta z} + 2e^{-\beta(y+z)} \right) \\ &+ \frac{2\eta^2}{\beta^2} z + \frac{\eta^2}{4\beta^3} (1 - e^{-2\beta z}) (1 + e^{-2\beta(y-z)} + 2e^{-\beta(y-z)}) \\ &+ \frac{2\eta^2}{\beta^3} e^{-\beta y} (1 - e^{\beta z}) + \frac{2\eta^2}{\beta^3} (e^{-\beta z} - 1) + \frac{\eta^2}{2\beta^2} (y - z) - \frac{3\eta^2}{4\beta^3} (1 - e^{-2\beta(y-z)}). \end{aligned}$$

and

$$G(x) = \exp \left(-\mu x + \frac{\eta^2}{4\beta^3} (1 - e^{-\beta x})^2 + \frac{\eta^2}{2\beta^2} x - \frac{\eta^2}{\beta^3} (1 - e^{-\beta x}) + \frac{\eta^2}{4\beta^3} (1 - e^{-2\beta x}) \right).$$

As a conclusion, when we consider α to be an OU process, then the GOU process is weakly stationary if $\mu\beta$ is greater than twice the stationary variance of α . Thus, it is weakly stationary if we impose the condition

$$\mu\beta > \frac{\eta^2}{\beta},$$

on the parameters of the model.

Remark 3.13 In this section we consider α to be a BS process. However we can also choose α to be a Lévy-driven OU process. That is

$$d\alpha(t) = -\beta\alpha(t) dt + \eta dL(t), \quad (25)$$

where L is a subordinator. In this case we assume that the filtration $\mathcal{F}_{t \geq 0}$, is generated by the Brownian motion W and the Lévy process L . Thus, considering the Malliavin derivative in the direction of W , we get $D_s\alpha(v) = 0$, for $s, v \in [0, T]$ since W and L are independent. Thus applying Proposition 2.2 we get the following expression for the process X

$$X(t) = X_0 \exp\left(-\int_0^t \alpha(u) du\right) + \int_0^t \sigma \exp\left(-\int_s^t \alpha(u) du\right) \delta W(s).$$

Moreover, we have

$$\int_0^t \alpha(u) du = \beta^{-1}(1 - e^{-\beta t}) + \int_0^t \beta^{-1}(1 - e^{-\beta(t-u)}) dL(u).$$

Therefore

$$\mathbb{E}[X(t)] = X_0 \exp(-\beta^{-1}(1 - e^{-\beta t})) \mathbb{E}\left[-\exp\left\{\int_0^t \beta^{-1}(1 - e^{-\beta(t-u)}) dL(u)\right\}\right].$$

Then we can write the last expression in terms of the cumulant function of the Lévy subordinator L and we compute the limit. Analytical expressions can also be computed for the variance and the correlation.

Another case would be to choose L to be a Lévy process driven by a pure jump part and a Brownian motion. Then we exploit again the independence structure between the jumps and the Brownian motion and we get the same expression for the Malliavin derivative of α as in the pure continuous case, where the Malliavin derivative is considered in the direction of W (see for example Solé et al. [19] for a Malliavin calculus for Lévy processes where a Malliavin operator in the direction of the Brownian motion and a Malliavin operator in the direction of the pure jump process were considered). We come back to our case as the challenging part in the analysis with the difference that the expression for $\int_0^t \alpha(u) du$ will include the jump part.

4 Local Behavior

In this section we study the instantaneous rate of change in the mean and in the squared fluctuations of the process X given that $X(s) = x$, where $x \in \mathbb{R}$. We do not present the proofs of the results studied in this section since they follow the same lines as the proofs of the propositions in Sect. 3.

We compute in the next lemma the mean of the increment $X(t) - X(s)$, $t \geq s$, given that $X(s) = x$.

Lemma 4.1 *We have for $s \leq t$,*

$$\mathbb{E}[X(t) - X(s) | X(s) = x] = x(f(t - s) - 1) + \tilde{f}(t - s),$$

where

$$f(v) = \exp \left\{ -\mu v + \frac{1}{2} \int_0^\infty h^2(u, v + u) du + \frac{1}{2} \int_0^v h^2(0, u) du \right\} \quad (26)$$

and

$$\tilde{f}(v) = - \int_0^v \sigma h(0, u) f(u) du.$$

From this lemma we can compute the instantaneous rate of change in the mean of the process X , given that $X(s) = x$. This is the content of the next proposition.

Proposition 4.2 *Assume that for all $t \in [0, T]$,*

$$h^2(u, t - s + u) \leq \zeta(s, u) \quad (27)$$

and

$$|g(t - s + u)| \leq \psi(s, u), \quad (28)$$

where $\int_0^\infty \zeta(s, u) du < \infty$ and $\int_0^\infty \psi(s, u) du < \infty$. Moreover assume

$$(\mathcal{A}) \quad \int_0^\infty |h(0, u) f(u)| < \infty, \quad \int_0^\infty h^2(0, u) < \infty.$$

Then we have for $s \leq t$,

$$a(x) = \lim_{t \rightarrow s} \frac{\mathbb{E}[X(t) - X(s) | X(s) = x]}{t - s} = -\mu x. \quad (29)$$

As we can see, locally the drift of X is behaving like the drift of an Ornstein-Uhlenbeck process. The speed of mean reversion becomes equal to the mean of α , namely μ .

In the next lemma we compute the squared fluctuations of $X(t) - X(s)$, $t \geq s$ given that $X(s) = x$.

Lemma 4.3 *We have for $s \leq t$,*

$$\mathbb{E}[(X(t) - X(s))^2 | X(s) = x] = \mathcal{I}_1 - 2\mathcal{I}_2 + \mathcal{I}_3 - 2x\mathcal{I}_4 + 2x\mathcal{I}_5 + x^2\mathcal{I}_6,$$

where

$$\begin{aligned} \mathcal{I}_1 &= \mathbb{E}\left[\left\{\int_s^t \sigma \exp\left(-\int_u^t \alpha(v)dv\right)\delta W(u)\right\}^2\right], \\ \mathcal{I}_2 &= \mathbb{E}\left[\left\{\int_s^t \sigma^2 \exp\left(-\int_u^t \alpha(v)dv\right)\delta W(u)\right\}\right. \\ &\quad \left.\left\{\int_s^t h(0, t-u) \exp\left(-\int_u^t \alpha(v)dv\right)du\right\}\right], \\ \mathcal{I}_3 &= \mathbb{E}\left[\left(\int_s^t \sigma h(0, t-u) \exp\left(-\int_u^t \alpha(v)dv\right)du\right)^2\right], \\ \mathcal{I}_4 &= \mathbb{E}\left[\int_s^t \sigma h(0, t-u) \exp\left(-\int_u^t \alpha(v)dv\right)du \left\{\exp\left(-\int_s^t \alpha(v)dv\right) - 1\right\}\right], \\ \mathcal{I}_5 &= \mathbb{E}\left[\exp\left(-\int_s^t \alpha(v)dv\right) \int_s^t \sigma \exp\left(-\int_u^t \alpha(v)dv\right)\delta W(v)\right], \\ \mathcal{I}_6 &= \mathbb{E}\left[\left(\exp\left(-\int_s^t \alpha(v)dv\right) - 1\right)^2\right]. \end{aligned}$$

With this lemma at hand, we can consider the instantaneous rate of change of the squared fluctuations of the process X given that $X(s) = x$.

Proposition 4.4 *Assume that (27), (28), and ($\mathcal{A}$) holds true. If moreover,*

$$\begin{aligned} (\mathcal{B}) \quad & \int_0^\infty \exp(-2\mu v + H(0, v))dv < \infty, \\ & \int_0^\infty \int_0^v |h(0, u)|e^{-2\mu(u)-\mu(v-u)+H(v-u, u)}|3h(v-u, v) + h(0, v)|dudv < \infty, \\ & \int_0^\infty \int_0^v |h(0, v-u)|e^{-2\mu(v-u)-\mu(u)+H(u, v-u)}|3h(0, v) + h(u, v)|dudv < \infty. \end{aligned}$$

Moreover, if

$$\begin{aligned} (\mathcal{C}) \quad & \int_0^\infty |h(0, v)|e^{-2\mu v + H(0, v)}dv < \infty, \\ & \int_0^\infty \int_0^v \left|h(0, u) \frac{\partial}{\partial v} H(v-u, u)\right|e^{-2\mu u - \mu(v-u) + H(v-u, u)}du dv < \infty, \\ & \int_0^\infty \left|h(0, v) \frac{\partial}{\partial v} H(0, v)\right|e^{-2\mu v + H(0, v)}dv < \infty, \end{aligned}$$

$$\int_0^\infty \int_0^v h(0, u) \left\{ e^{-2\mu u - \mu(v-u) + H(v-u, u)} \left(\frac{\partial}{\partial v} H(v-u, u) \right)^2 + \frac{\partial^2}{\partial v^2} H(v-u, u) \right\} du dv < \infty.$$

Then we have for $s \leq t$

$$b^2(x) = \lim_{t \rightarrow s} \frac{\mathbb{E}[(X(t) - X(s))^2 | X(s) = x]}{t - s} = \sigma^2.$$

The variance of the process X will locally behave as a constant, and therefore we may view the process X as an Ornstein-Uhlenbeck process for small increments. Notice that for the so-called diffusion processes (see for example Kloeden and Platen [14] for more about such processes), the quantity $a(x)$ is called the drift of the diffusion and $b(x)$ its diffusion coefficient at position x . These two terms describe the behavior of the first and second order moment of increments of the process X over an infinitesimal time interval $[s, s + ds]$.

Now we consider the case where α is the OU process given in (7). Under the assumption

$$\mu\beta > \frac{\eta^2}{\beta^2},$$

the results of Propositions 4.2 and 4.4 hold true. In fact when α is an OU process then assumptions (27), (28), ($\mathcal{A}$), and ($\mathcal{B}$) follow immediately from Lemmas 3.4 and 3.10. To prove ($\mathcal{C}$) we observe that $\frac{\partial}{\partial v} H(v-u, u)$ and $\frac{\partial^2}{\partial v^2} H(v-u, u)$ are bounded uniformly in u and v and thus ($\mathcal{C}$) becomes equivalent to ($\mathcal{B}$).

5 Chaos Expansion

In this section we develop the chaos expansion of X . We denote by $I_n(\phi_n)$, for a natural number n , the n th chaos with kernel function $\phi_n \in L^2(\mathbb{R}^n)$ being symmetric. That is,

$$\begin{aligned} I_n(\phi_n) &= \int_{\mathbb{R}^n} \phi_n(u_1, \dots, u_n) dW^{\otimes n}(u_1, \dots, u_n) \\ &= n! \int_{-\infty}^\infty \int_{-\infty}^{u_{n-1}} \cdots \int_{-\infty}^{u_2} \phi_n(u_1, \dots, u_n) dW(u_1) \cdots dW(u_n). \end{aligned}$$

Note that in particular, $I_0(\phi_0) = \phi_0$ is simply a constant and

$$I_1(\phi_1) = \int_{\mathbb{R}} \phi_1(u) dW(u),$$

a Gaussian random variable with zero mean and variance being the $L^2(\mathbb{R})$ -norm of ϕ_1 . Throughout this section we assume that the integrability condition on h in Lemma 3.2 holds.

Our first result concerns the chaos expansion of $-\int_s^t \alpha(u) du$.

Lemma 5.1 *It holds that*

$$-\int_s^t \alpha(u) du = -\mu(t-s) + I_1(\phi_{s,t}),$$

for

$$\phi_{s,t}(u) = -h(s-u, t-u)\mathbf{1}(u \leq s) - h(0, t-u)\mathbf{1}(s \leq u \leq t). \quad (30)$$

Proof From Lemma 3.2, we find

$$\begin{aligned} -\int_s^t \alpha(u) du &= -\mu(t-s) - \int_{-\infty}^s h(s-u, t-u) dW(u) - \int_s^t h(0, t-u) dW(u) \\ &= -\mu(t-s) + \int_{\mathbb{R}} \phi_{s,t}(u) dW(u). \end{aligned}$$

Hence, the result follows.

Next, we state explicitly the chaos expansion of $\exp\left(-\int_s^t \alpha(u) du\right)$.

Lemma 5.2 *It holds*

$$\exp\left(-\int_s^t \alpha(u) du\right) = \sum_{n=0}^{\infty} I_n\left(\frac{1}{n!} f(t-s) \phi_{s,t}^{\otimes n}\right),$$

where f is given by (26).

Proof From the theory of chaos expansions (see for example Exercise 1.1.1 in Nualart [17]), it holds

$$\exp\left(I_1(\phi_{s,t}) - \frac{1}{2}|\phi_{s,t}|_2^2\right) = \sum_{n=0}^{\infty} \frac{1}{n!} I_n(\phi_{s,t}^{\otimes n}),$$

where $|\cdot|_2$ denotes the $L^2(\mathbb{R})$ -norm. But,

$$\begin{aligned} |\phi_{s,t}|_2^2 &= \int_{\mathbb{R}} (h(s-u, t-u)\mathbf{1}(u \leq s) + h(0, t-u)\mathbf{1}(s \leq u \leq t))^2 du \\ &= \int_0^{\infty} h^2(u, t-s+u) du + \int_0^{t-s} h^2(0, u) du. \end{aligned}$$

Hence, the result follows from Lemma 5.1.

Using this, we can compute the chaos of the “drift-term” of X .

Lemma 5.3 *Assume that the function*

$$v \mapsto h(0, v) \exp \left(-\mu v + \int_0^\infty h^2(u, v + u) du + \int_0^v h^2(0, u) du \right)$$

is integrable on $[0, t]$. Then the integral $\int_0^t h(0, t-s) \exp(-\int_s^t \alpha(u) du) ds$ is well-defined as a Bochner integral in $L^2(\Omega, \mathcal{F}, P)$. Moreover, it holds

$$\int_0^t h(0, t-s) e^{-\int_s^t \alpha(u) du} ds = \sum_{n=0}^{\infty} I_n \left(\frac{1}{n!} \int_0^t h(0, t-s) f(t-s) \phi_{s,t}^{\otimes n} ds \right),$$

where f is defined in Lemma 5.2 and $\phi_{s,t}$ in (30).

Proof Recall that

$$\exp \left(-\int_s^t \alpha(u) du \right) = \exp \left(-\mu(t-s) + I_1(\phi_{s,t}) \right),$$

and hence

$$\begin{aligned} & \mathbb{E} \left[\exp \left(-2 \int_s^t \alpha(u) du \right) \right]^{1/2} \\ &= \exp \left(-\mu(t-s) + |\phi_{s,t}|_2^2 \right) \\ &= \exp \left(-\mu(t-s) + \int_0^\infty h^2(u, t-s+u) du + \int_0^{t-s} h^2(0, u) du \right), \end{aligned}$$

from the proof of Lemma 5.2. But then the condition implies that $s \mapsto h(0, t-s) \exp \left(-\int_s^t \alpha(u) du \right)$ is Bochner integrable on $[0, t]$. Invoking Proposition 8.1 and the following discussion on p. 281 in Hida et al. [13], the chaos expansion of a Bochner integral is given by integrating the chaos functions. Hence, the result follows.

Finally, we compute the chaos expansion of the Skorohod integral in the next lemma.

Lemma 5.4 *Assume that $e^{-\int_s^t \alpha(u) du}$ is Skorohod integrable. It holds that*

$$\int_0^t e^{-\int_s^t \alpha(u) du} \delta W(s) = \sum_{n=0}^{\infty} I_{n+1} \left(\frac{1}{n!} \psi_{n,t} \right),$$

where $\psi_{n,t} \in L^2(\mathbb{R}^{n+1})$ is the symmetrization of the function

$$\tilde{\psi}_{n,t}(u_1, \dots, u_n, u_{n+1}) = \mathbf{1}_{[0,t]}(u_{n+1}) f(t - u_{n+1}) \phi_{u_{n+1},t}^{\otimes n}(u_1, \dots, u_n).$$

Proof This follows from Proposition 1.3.3 in Nualart [17].

Collecting together the results in the lemmas above, we find the chaos expansion of $X(t)$ to be

$$\begin{aligned} X(t) = & X_0 f(t) + \int_0^t \sigma h(0, u) f(u) du \\ & + \sum_{n=1}^{\infty} I_n \left(\frac{1}{n!} \left(X_0 f(t) \phi_{0,t}^{\otimes n} + \int_0^t \sigma h(0, t-s) f(t-s) \phi_{s,t}^{\otimes n} ds + \sigma n \psi_{n-1,t} \right) \right). \end{aligned} \quad (31)$$

Note that the integral in the chaos functions can be alternatively written as

$$\int_0^t h(0, t-s) f(t-s) \phi_{s,t}^{\otimes n} ds = \int_0^t h(0, u) f(u) \phi_{t-u,t}^{\otimes n} du$$

after a change of variables.

We consider the case where α is the OU process given by (7). In this case, $g(u) = \eta e^{-\beta u}$ and $\exp(-\int_s^t \alpha(u) du)$ is Skorohod integrable as it was already discussed in Sect. 3. Since we are working in the setting of $L^2(\Omega, \mathcal{F}, P)$ -random variables, we suppose that $\mu\beta > \eta^2/\beta$, which we recall from the previous section that it is a sufficient condition for the second order moment of $X(t)$ to exist for all $t \geq 0$. We compute a chaos expansion for a random variable X in this particular case of specification of g and investigate the kernel functions in the chaos expansion.

Let us start with considering the zero order chaos, which is the mean of X . We recall the expression of $h(x, y)$ and $f(x)$ in (8) and (26), resp. We know the limit of the zero order chaos (the mean) from Proposition 3.3, giving

$$\lim_{t \rightarrow \infty} X_0 f(t) + \int_0^t h(0, u) f(u) du = \int_0^\infty h(0, u) f(u) du.$$

Next, we focus on chaos 1, that is the case $n = 1$ for simplicity. In the following lemma we compute the point wise limit of the kernel of chaos 1.

Lemma 5.5 *Assume $\mu\beta > \eta^2/2\beta$. The limit of the first order chaos is given by*

$$\lim_{t \rightarrow \infty} X_0 f(t) \phi_{0,t}(u) + \int_0^t \sigma h(0, t-s) f(t-s) \phi_{s,t}(u) ds = k e^{\beta u} \mathbf{1}_{(u \leq 0)}, \quad (32)$$

where $k = -\int_0^\infty \tilde{h}(u) du + \beta \int_0^\infty \int_0^\infty \tilde{h}(s) e^{-\beta(v-s)} ds dv$.

Proof As we have from (30)

$$\begin{aligned}\phi_{s,t}(u) &= -h(s-u, t-u)\mathbf{1}(u \leq s) - h(0, t-u)\mathbf{1}(s \leq u \leq t) \\ &= -\int_{s-u}^{t-u} g(v) dv \mathbf{1}(u \leq s) - \int_0^{t-u} g(v) dv \mathbf{1}(s \leq u \leq t) \\ &= \frac{\eta}{\beta} e^{\beta u} (e^{-\beta t} - e^{-\beta s}) \mathbf{1}(u \leq s) + \frac{\eta}{\beta} (e^{-\beta(t-u)} - 1) \mathbf{1}(s \leq u \leq t).\end{aligned}$$

for $u \in \mathbb{R}$. Let $\tilde{h}(v) = \frac{\eta}{\beta} h(0, v) f(v)$. Note that we have

$$\begin{aligned}h(0, u) f(u) &= \frac{\eta}{\beta} (1 - e^{-\beta u}) \exp\left(-\mu u + \frac{1}{2} \int_0^\infty h^2(v, u+v) dv + \frac{1}{2} \int_0^u h^2(0, v) dv\right) \\ &= \frac{\eta}{\beta} (1 - e^{-\beta u}) \exp\left(-\mu u - \frac{\eta^2}{2\beta^3} + \frac{\eta^2}{2\beta^3} e^{-\beta u} + \frac{\eta^2}{2\beta^2} u\right).\end{aligned}$$

Then

$$\begin{aligned}\int_0^t h(0, t-s) f(t-s) \phi_{s,t}(u) ds &= \int_0^t \tilde{h}(t-s) e^{\beta u} (e^{-\beta t} - e^{-\beta s}) \mathbf{1}(u \leq s) ds + \int_0^t \tilde{h}(t-s) (e^{-\beta(t-u)} - 1) \mathbf{1}(s \leq u \leq t) ds \\ &= \begin{cases} \int_0^t \tilde{h}(t-s) e^{\beta u} (e^{-\beta t} - e^{-\beta s}) ds, & u \leq 0, \\ \int_u^t \tilde{h}(t-s) e^{\beta u} (e^{-\beta t} - e^{-\beta s}) ds + \int_0^u \tilde{h}(t-s) (e^{-\beta(t-u)} - 1) ds, & u \in [0, t], \\ 0, & u > t. \end{cases}\end{aligned}$$

Consider first $u < 0$. After a simple change of variables, it holds,

$$\int_0^t \tilde{h}(t-s) e^{\beta u} (e^{-\beta t} - e^{-\beta s}) ds = e^{\beta u} \left(-\int_0^t \tilde{h}(s) e^{-\beta(t-s)} ds + e^{-\beta t} \int_0^t \tilde{h}(s) ds \right).$$

Since under the assumed condition, $\lim_{t \rightarrow \infty} \int_0^t \tilde{h}(s) ds = \frac{\eta}{\beta} \int_0^\infty h(0, s) f(s) ds$, then

$$\lim_{t \rightarrow \infty} e^{-\beta t} \int_0^t \tilde{h}(s) ds = 0.$$

From the “key formula” in (22), we find

$$\begin{aligned}\int_0^t \tilde{h}(s) e^{-\beta(t-s)} ds &= \int_0^t \tilde{h}(s) ds + \int_0^t \int_0^v \tilde{h}(u) (-\beta) e^{-\beta(v-u)} du dv \\ &= \int_0^t \tilde{h}(s) ds - \beta \int_0^t e^{-\beta v} \int_0^v \tilde{h}(u) e^{\beta u} du dv.\end{aligned}$$

Notice that the first term on the right-hand side has a limit as time goes to infinity. To have a limit for the second term, we need that the function

$$v \mapsto \exp(-\beta v) \int_0^v \tilde{h}(u) \exp(\beta u) du$$

is integrable on $\mathbb{R}_+$. But this is indeed the case, since

$$\begin{aligned} e^{-\beta v} \int_0^v \tilde{h}(u) e^{\beta u} du &\leq c e^{-\beta v} \int_0^v e^{\beta u - \mu u + \eta^2 u / (2\beta^2)} du \\ &\leq c \frac{e^{-\beta v} - e^{-(\mu - \frac{\eta^2}{2\beta^2})v}}{\mu - \frac{\eta^2}{2\beta^2} - \beta}, \end{aligned}$$

which is integrable on $\mathbb{R}_+$ by assumptions on the parameters. We can therefore conclude that for every $u < 0$, there exists a limit as $t \rightarrow \infty$ given by

$$\lim_{t \rightarrow \infty} \int_0^t \tilde{h}(t-s) \phi_{s,t}(u) ds = \left(- \int_0^\infty \tilde{h}(u) du + \beta \int_0^\infty e^{-\beta v} \int_0^v \tilde{h}(s) e^{\beta s} ds dv \right) e^{\beta u}. \quad (33)$$

Next, let us consider a fixed $u \in \mathbb{R}_+$. Note that for $t > u$,

$$\tilde{h}(t-s) \leq c e^{-(\mu - \eta^2 / (2\beta^2))(t-s)},$$

and hence for fixed u we will find that $\int_0^u \tilde{h}(t-s) ds \rightarrow 0$, when $t \rightarrow \infty$. But this yields that pointwise in $u \in \mathbb{R}_+$, $\int_0^u \tilde{h}(t-s)(e^{-\beta(t-u)} - 1) ds \rightarrow 0$, when $t \rightarrow \infty$. Next, after a change of variables and some straightforward manipulations,

$$\int_u^t \tilde{h}(t-s) e^{\beta u} (e^{-\beta t} - e^{-\beta s}) ds = - \int_0^{t-u} e^{\beta u} e^{-\beta t} \tilde{h}(v) (e^{\beta v} - 1) dv. \quad (34)$$

But for $v \leq t - u$, we have

$$\begin{aligned} |e^{\beta u} e^{-\beta t} \tilde{h}(v) (e^{\beta v} - 1)| &\leq c e^{\beta u} e^{-\beta t} e^{\beta v} e^{-(\mu - \eta^2 / (2\beta^2))v} \\ &\leq c e^{\beta u} e^{-\beta t} e^{\beta(t-u)} e^{-(\mu - \eta^2 / (2\beta^2))v}, \end{aligned}$$

which is integrable in $\mathbb{R}_+$. Thus we can take the limit inside the integral in (34) and we prove that the limit of $\int_u^t \tilde{h}(t-s) e^{\beta u} (e^{-\beta t} - e^{-\beta s}) ds$ goes to 0 when t goes to infinity and the statement is proved.

We note further that the pointwise limit in (32) is indeed a function in $L^2(\mathbb{R})$. One could hope that the kernel of chaos 1 converges not only pointwise to the $L^2(\mathbb{R})$ -function $\exp(\beta u)\mathbf{1}(u < 0)$, but also in $L^2(\mathbb{R})$. We demonstrate that this is *not* the case. Indeed, we have

$$\begin{aligned}
 & \left| \int_0^t \tilde{h}(t-s) \phi_{s,t}(\cdot) ds - k e^{\beta \cdot} \mathbf{1}(\cdot < 0) \right|_{L^2(\mathbb{R})} \\
 &= \left| \int_0^t \tilde{h}(t-s) (e^{-\beta t} - e^{-\beta s}) ds e^{\beta \cdot} \mathbf{1}(\cdot < 0) - k e^{\beta \cdot} \mathbf{1}(\cdot < 0) \right. \\
 &\quad + \int_0^t \tilde{h}(t-s) (e^{-\beta t} - e^{-\beta s}) ds e^{\beta \cdot} \mathbf{1}(0 \leq \cdot \leq t) \\
 &\quad \left. + \int_0^t \tilde{h}(t-s) (e^{-\beta(t-\cdot)} - 1) ds \mathbf{1}(0 \leq \cdot \leq t) \right|_{L^2(\mathbb{R})}^2 \\
 &= \frac{1}{2\beta} \left| \int_0^t \tilde{h}(t-s) (e^{-\beta t} - e^{-\beta s}) ds - k \right|^2 \\
 &\quad + \int_0^t \left(\int_u^t \tilde{h}(t-s) (e^{-\beta t} - e^{-\beta s}) ds e^{\beta u} + \int_0^u \tilde{h}(t-s) ds (e^{-\beta(t-u)} - 1) \right)^2 du
 \end{aligned}$$

The first term goes to zero by definition of k (using the arguments as above). We consider the second term. We have

$$\begin{aligned}
 & \int_0^t \left(\int_u^t \tilde{h}(t-s) (e^{-\beta t} - e^{-\beta s}) ds e^{\beta u} + \int_0^u \tilde{h}(t-s) ds (e^{-\beta(t-u)} - 1) \right)^2 du \\
 &= \int_0^t \left(\int_u^t \tilde{h}(t-s) (e^{-\beta t} - e^{-\beta s}) ds e^{\beta u} \right)^2 du \\
 &\quad + \int_0^t \left(\int_0^u \tilde{h}(t-s) ds (e^{-\beta(t-u)} - 1) \right)^2 du \\
 &\quad + 2 \int_0^t \left(\int_u^t \tilde{h}(t-s) (e^{-\beta t} - e^{-\beta s}) ds e^{\beta u} \right) \left(\int_0^u \tilde{h}(t-s) ds (e^{-\beta(t-u)} - 1) \right) du.
 \end{aligned} \tag{35}$$

However, for $s \leq t$, we have

$$\tilde{h}(t-s) \geq c(1 - e^{-\beta(t-s)})e^{-t(\mu - \frac{\eta^2}{2\beta})}e^{s(\mu - \frac{\eta^2}{2\beta})},$$

where c is a strictly positive constant. Thus

$$\left(\int_0^u \tilde{h}(t-s) ds (e^{-\beta(t-u)} - 1) \right)^2 \geq c(1 - e^{-\beta t})^4 e^{-2t(\mu - \frac{\eta^2}{2\beta})} (e^{2u(\mu - \frac{\eta^2}{2\beta})} - 1).$$

Integrating and taking the limit when t goes to infinity we get

$$\begin{aligned}
 & \lim_{t \rightarrow \infty} \int_0^t \left(\int_0^u \tilde{h}(t-s) ds (e^{-\beta(t-u)} - 1) \right)^2 du \\
 & \geq \lim_{t \rightarrow \infty} \int_0^t c(1 - e^{-\beta t})^4 e^{-2t(\mu - \frac{\eta^2}{2\beta})} (e^{2u(\mu - \frac{\eta^2}{2\beta})} - 1) du \\
 & \geq \lim_{t \rightarrow \infty} c(1 - e^{-\beta t})^4 e^{-2t(\mu - \frac{\eta^2}{2\beta})} (e^{2t(\mu - \frac{\eta^2}{2\beta})} - 1 - t) \\
 & = c,
 \end{aligned}$$

where c is a strictly positive constant. Since the first and third terms in (35) are positive then we have

$$\lim_{t \rightarrow \infty} \int_0^t \left(\int_u^t \tilde{h}(t-s) (e^{-\beta t} - e^{-\beta s}) ds e^{\beta u} + \int_0^u \tilde{h}(t-s) ds (e^{-\beta(t-u)} - 1) \right)^2 du > 0$$

Therefore the kernel function of chaos 1 converges pointwise to a square-integrable function, but not in L^2 .

Acknowledgments Asma Khedher thanks the KPMG Center of Excellence in Risk Management for the financial support.

References

1. Alos, E., Mazet, O., Nualart, D.: Stochastic calculus with respect to Gaussian processes. *Ann. Probab.* **29**(2), 766–801 (2001)
2. Barlow, M., Gusev, Y., Lai, M.: Calibration of multifactor models in electricity markets. *Int. J. Theor. Appl. Finan.* **7**(2), 101–120 (2004)
3. Barndorff-Nielsen, O.E., Basse O'Connor, A.: Quasi Ornstein-Uhlenbeck processes. *Bernoulli* **17**(3), 916–941 (2011)
4. Barndorff-Nielsen, O.E., Schmiegel, J.: Brownian semistationary processes and volatility/intermittency. In: Albrecher, H., Runggaldier, W., Schachermayer, W. (eds.) *Advanced Financial Modeling. Radon Series on Computational and Applied Mathematics*, vol. 8, pp. 1–26. W. de Gruyter, Berlin (2009)
5. Barndorff-Nielsen, O.E., Benth, F.E., Veraart, A.E.D.: Modelling energy spot prices by volatility modulated Lévy-driven Volterra processes. *Bernoulli* **19**(3), 803–845 (2013)
6. Basse, A., Pedersen, J.: Lévy driven moving averages and semimartingales. *Stoch. Proc. Appl.* **119**, 2970–2991 (2009)
7. Benth, F.E., Šaltytė Benth, J.: *Modelling and Pricing in Financial Markets for Weather Derivatives*. World Scientific (2013)
8. Benth, F.E., Eyjolfsson, H.: Stochastic modelling of power markets using stationary processes. In: Dalang, R., Dozzi, M., Russo, F. (eds.) *Stochastic Analysis, Random Fields and Applications. Progress in Probability, Ascona Proceedings 2011*, vol. 67, pp. 261–284. Birkhäuser Verlag (2012)
9. Brokwell, P.J., Davis, R.A.: *Introduction to Time Series and Forecasting*. Springer, New York (1996)

10. Carmona, P., Petit, F., Yor, M.: On the distribution and asymptotic results for exponential functionals of Lévy processes. In *Exponential Functionals and Principal Values related to Brownian motion. Bibl. Rev. Mat. Iberoamericana*, pp. 73–130. Rev. Mat. Iberoamericana, Madrid (1997)
11. Carvalho, C.M., Lopez, H.F.: Simulation-based sequential analysis of Markov switching stochastic volatility models. *Comput. Stat. Data Anal.* **51**, 4526–4542 (2007)
12. Dufresne, D.: The distribution of a perpetuity with application to risk theory and pension funding. *Scand. Actuar. J.* **9**, 39–79 (1990)
13. Hida, T., Kuo, H.H., Potthoff, J., Streit, L.: *White Noise—An Infinite Dimensional Calculus*. Kluwer Academic Publishers (1993)
14. Kloeden, P.E., Platen, E.: *Numerical Solution of Stochastic Differential Equations*. Springer (1999)
15. Lindner, A., Sato, K.I.: Continuity properties and infinite divisibility of stationary distributions of some generalised Ornstein-Uhlenbeck processes. *Ann. Probab.* **37**(1), 250–274 (2009)
16. Marquardt, T., Stelzer, R.: Multivariate CARMA processes. *Stochast. Process. Appl.* **117**(1), 96–120 (2007)
17. Nualart, D.: *The Malliavin Calculus and Related Topics*. Springer, Berlin (2006)
18. Perlin, M.: MS-Regress—The MATLAB Package for Markov Regime Switching Models. Working paper. Available at SSRN: <http://ssrn.com/abstract/1714016> (2010)
19. Solé, J.L., Utzet, F., Vives, J.: Canonical Lévy Process and Malliavin calculus. *Stochast. Process. Appl.* **117**, 165–187 (2007)
20. Yor, M.: *Exponential Functionals of Brownian Motion and Related Processes*. Springer, New York (2001)
21. Zapranis, A., Alexandridis, A.: Modelling the temperature time-dependent speed of mean reversion in the context of weather derivatives pricing. *Appl. Math. Finan.* **15**(4), 355–386 (2008)

Continuity of Local Time: An Applied Perspective

Jorge M. Ramirez, Enirque A. Thomann and Edward C. Waymire

Abstract Continuity of local time for Brownian motion ranks among the most notable mathematical results in the theory of stochastic processes. This article addresses its implications from the point of view of applications. In particular an extension of previous results on an explicit role of continuity of (natural) local time is obtained for applications to recent classes of problems in physics, biology and finance involving discontinuities in a dispersion coefficient. The main theorem and its corollary provide physical principles that relate macro scale continuity of deterministic quantities to micro scale continuity of the (stochastic) local time.

Keywords Dispersion · Discontinuous diffusion · Skew brownian motion · Semi-martingale local time · Diffusion local time · Occupation time

1 Introduction

Quoting from the backcover of the intriguing recent monograph [3]:

Random change of time is key to understanding the nature of various stochastic processes and gives rise to interesting mathematical results and insights of importance for the modelling and interpretation of empirically observed dynamic processes.

This point could hardly have been more aptly made with regard to the perspective of the present paper. The focus is on identifying the way in which continuity/discontinuity properties of certain local times of a diffusive Markov process inform interfacial discontinuities in large scale concentration, diffusion coefficients,

J.M. Ramirez
Universidad Nacional de Colombia, Medellin, Colombia
e-mail: jmr Ramirez@unal.edu.co

E.A. Thomann · E.C. Waymire (✉)
Oregon State University, Corvallis, OR, USA
e-mail: waymire@math.oregonstate.edu

E.A. Thomann
e-mail: thomann@science.oregonstate.edu

and transmission rates. For these purposes one may ignore processes with drift, and focus on discontinuities in diffusion rates and/or specific rate coefficients. This builds on related work of the authors where the focus was on interfacial effects on other functionals such as occupation time and first passage times; see [27] for a recent survey. We also restrict our study to one-dimensional processes. Some extensions to higher dimensions can be found in [25].

Dispersion problems in the physical sciences are often described by a second order linear parabolic equation in divergence form that results from a conservation law applied to the concentration of some substance. A particular class of interest for this paper is that of the one-dimensional diffusion equation (1) with discontinuous parameters D and η and specified discontinuities in the concentration. More precisely, we consider the solution $u(t, y)$ to the following problem for $y \in \mathbb{R} \setminus I$, $t \geq 0$,

$$\eta \frac{\partial u}{\partial t} = \frac{\partial}{\partial y} \left(\frac{1}{2} D \frac{\partial u}{\partial y} \right), \quad \left[D \frac{\partial u}{\partial y} \right]_{x_j} = 0, \quad \beta_j^+ u(t, x_j^+) = \beta_j^- u(t, x_j^-), \quad j \in \mathbb{Z}. \quad (1)$$

with a prescribed initial condition $u(0, y) = u_0(y)$, $y \in \mathbb{R}$, and under the following assumptions.

Assumptions 1 We consider a discrete set of “interfaces”

$$I := \{x_j : j \in \mathbb{Z}\}, \quad x_0 := 0, \quad x_j < x_{j+1}, \quad j \in \mathbb{Z}, \quad (2)$$

with no accumulation points. The functions D and η exhibit jump discontinuities only at points in I :

$$[D]_{x_j} := D(x_j^+) - D(x_j^-) \neq 0, \quad [\eta]_{x_j} := \eta(x_j^+) - \eta(x_j^-) \neq 0, \quad j \in \mathbb{Z}. \quad (3)$$

We further assume that D and η are functions of bounded variation in $\mathbb{R}$ with η being continuous, and D differentiable in (x_j, x_{j+1}) for all $j \in \mathbb{Z}$. Finally, there exist constants $0 < k < K < \infty$ such that

$$k \leq D(x) \leq K, \quad k \leq \eta(x) \leq K, \quad x \in \mathbb{R}. \quad (4)$$

The constants $\{\beta_j^\pm\}$ are strictly positive and such that

$$\sum_{j \in \mathbb{Z}} \frac{\beta_j^+}{\beta_j^-} < \infty. \quad (5)$$

Equation (1) is often referred to as a *continuity equation* for the conserved quantity $\eta(y)u(t, y)$; Fourier’s flux law for heat conduction and the corresponding Fick’s law for diffusion being among the most notable. The right-side of the pde is the divergence of the diffusive flux $\frac{1}{2} D \frac{\partial u}{\partial y}$ of u , and the first bracket condition is *continuity of flux* at the interfaces.

In Jean Perrin's historic determination of Avogadro's number N_A from observations of individual particle paths submerged in a homogeneous medium, the procedure was clear [24]. Einstein had provided a twofold characterization of the diffusion coefficient D in Fick's law: first as a function of variables at the molecular scale, including N_A ; and second, as the rate of growth of the variance of the position of particles in time. This meant that D , and therefore N_A could be statistically estimated.

If one regards (1) as a given physical law that embodies certain interfacial discontinuities at points in I , then the question we want to address is *what corresponding features should be specified about the paths of the stochastic process?* Our basic goal is to show that the answer resides in suitably interpreted continuity properties of local time.

In an informal conversation where we posed this question to David Aldous, his reply was that he wouldn't use a pde to model the phenomena in the first place! Of course this perspective makes the question all the more relevant to probabilistic modeling. The mathematical (probabilistic) tools are clearly available to do this, and much of the objective of this paper is to identify the most suitable way in which to express the stochastic model in relation to the underlying phenomena. It will be shown that the interfacial conditions at the pde level can be characterized in terms of the continuity properties of a certain local time process of the associated diffusion. In this regard, the continuity of local time of standard Brownian motion will be seen to indirectly play a pivotal role.

The evolution problem (1) can be viewed as the equation $\frac{\partial u}{\partial t} = \mathcal{L}^* u$ for the operator $\mathcal{L}^* : \text{Dom}(\mathcal{L}^*) \rightarrow L^2(dy)$ given by

$$\mathcal{L}^* f := \frac{1}{\eta} \frac{d}{dy} \left(\frac{1}{2} D \frac{df}{dy} \right), \quad (6)$$

for functions $f \in \text{Dom}(\mathcal{L}^*)$ satisfying, besides other decay conditions, that

$$[Df']_{x_j} = 0, \quad \beta_j^+ f'(x_j^+) = \beta_j^- f'(x_j^-), \quad j \in \mathbb{Z}. \quad (7)$$

Due to the presence of the coefficient $\eta(y)$, taking the adjoint in $L^2(dy)$ of $\mathcal{L}^*$ does not generally yield an operator $\mathcal{L}$ that generates a positive contraction semi-group on $C_0(\mathbb{R})$. In fact, integration by parts yields,

$$\mathcal{L} g := \frac{d}{dx} \left(\frac{1}{2} D \frac{d}{dx} \left(\frac{g}{\eta} \right) \right) \quad (8)$$

and any $g \in \text{Dom}(\mathcal{L})$ will satisfy the following interfacial conditions:

$$\left[\frac{g}{\eta} \right]_{x_j} = 0, \quad \frac{D(x_j^+)}{\beta_j^+} \left(\frac{g}{\eta} \right)'(x_j^+) = \frac{D(x_j^-)}{\beta_j^-} \left(\frac{g}{\eta} \right)'(x_j^-). \quad (9)$$

We refer to the corresponding evolution problem $\frac{\partial v}{\partial t} = \mathcal{L} v$ as the *adjoint equation*.

While physical laws are often formulated on the basis of conservation principles, not all such models are apriori conceived in conservation form. In fact some may be explicitly formulated as a specification of coefficients via a Kolmogorov backward equation with an operator of the form (10) below, or directly in the form of a stochastic differential equation; for a variety of examples of both in the present context, see [5, 11, 12, 15] from hydrology, [7, 19, 22, 26, 30] biology and ecology, [20] finance, [10] astrophysics, and [18] from physical oceanography.

To accommodate the broad class of such possible examples, the present paper follows the following general setting. Let $\mathcal{L}$ be as in (8) and define the operator $\mathcal{A}$ by

$$\mathcal{A}g := \frac{1}{\eta} \mathcal{L}(\eta g) = \frac{1}{\eta} \frac{d}{dx} \left(\frac{1}{2} D \frac{dg}{dx} \right). \quad (10)$$

The interfacial conditions satisfied by a function $g \in \text{Dom}(\mathcal{A})$ follow from (9) and can generally be written in the form:

$$[g]_{x_j} = 0, \quad \lambda_j g'(x_j^+) = (1 - \lambda_j) g'(x_j^-), \quad j \in \mathbb{Z} \quad (11)$$

for some values $\Lambda := \{\lambda_j : j \in \mathbb{Z}\} \subset (0, 1)$ for which we further assume the following decay condition (equivalent to (5) under (4)):

$$\sum_{j \in \mathbb{Z}} \frac{1 - \lambda_j}{\lambda_j} < \infty. \quad (12)$$

In Sect. 2 we construct a Feller process $X = \{X(t) : t \geq 0\}$ with generator $(\mathcal{A}, \text{Dom}(\mathcal{A}))$. The significance of this association is the following: the fundamental solution $p(t, x, y)$ to the backwards evolution problem $\frac{\partial v}{\partial t} = \mathcal{A}v$ is precisely the transition probability density of X . Namely, for an initial condition $v(0, x) = v_0(x)$, the solution $v(t, x)$ can be written as

$$v(t, x) = \int_{\mathbb{R}} p(t, x, y) v_0(y) dy = \mathbb{E}[v_0(X(t)) | X(0) = x]. \quad (13)$$

It follows, in turn, that the fundamental solution to the original diffusion problem (1) is given by

$$q(t, x, y) = \frac{\eta(x)}{\eta(y)} p(t, x, y). \quad (14)$$

This defines the interpretation we will use with respect to the physics of (1) as a “Fokker-Plank equation”, and to relate it with the diffusion process X .

Remark 1 One might note that this is the form of Doob’s h-transform under the further constraint that η is harmonic with respect to $\mathcal{A}$. However this latter condition is generally not appropriate for the physical examples of interest in the present paper.

On the other hand, were it applicable it could provide an alternative approach to our problem via conditioning, e.g. see [23].

In order to identify how the interfacial conditions (11) affect the sample paths of the process X , we will look at the behavior of the *natural local time* of X at points in I . This notion of local time was introduced in [2] as the density of the occupation time operator with respect to Lebesgue measure. Namely for $A \in \mathcal{B}(\mathbb{R})$, $\ell^X(t, x)$ is a previsible process, increasing with respect to t , such that

$$\int_0^t \mathbb{1}_A(X(s)) ds = \int_A \ell^X(t, x) dx. \quad (15)$$

Our main result is the following:

Theorem 2 Suppose D, η satisfy Assumptions 1. Let X be the Feller process with infinitesimal generator $(\mathcal{A}, \text{Dom}(\mathcal{A}))$ defined by (10, 11) and Λ satisfying (12). Then

$$\frac{\ell^X(t, x_j^+)}{\ell^X(t, x_j^-)} = \frac{\eta(x_j^+)}{\eta(x_j^-)} \frac{D(x_j^-)}{D(x_j^+)} \frac{\lambda_j}{1 - \lambda_j}, \quad j \in \mathbb{Z}. \quad (16)$$

with probability one, for any t such that $\ell^X(t, x_j^+) > 0$.

Corollary 1 Suppose D, η and $\{\beta_j^\pm : j \in \mathbb{Z}\}$ satisfy Assumptions 1. Let u be the solution to the diffusion Eq.(1) and X its associated Feller process via Eq.(10). Then

$$\frac{\ell^X(t, x_j^+)}{\ell^X(t, x_j^-)} = \frac{\eta(x_j^+)}{\eta(x_j^-)} \frac{\beta_j^-}{\beta_j^+} = \frac{\eta(x_j^+)u(t, x_j^+)}{\eta(x_j^-)u(t, x_j^-)}, \quad j \in \mathbb{Z}. \quad (17)$$

with probability one, for any t such that $\ell^X(t, x_j^+) > 0$.

The main principles to be taken from these mathematical results are as follows. First, Theorem 2 and its corollary demonstrate how the continuity properties of local time are reflected in the specifications of (1) at interfaces. It is noteworthy that under the continuity of flux condition, the diffusion coefficient plays no role with regard to continuity of local time. In particular, its determination would continue to be by statistical considerations of (local) variances along the lines used by Perrin, while the jumps in the natural local time are a manifestation of other characteristics of the model. That is, the relative values of η for given values of β_j^+/β_j^- , or vice-versa, are reflected in the local time behaviors of sample paths at interfacial points. An example is furnished below in which the β and η parameters are relative manifestations of geometries of both the medium and the dispersing individuals.

Under continuity of flux, the continuity of local time is equivalent to continuity of the conserved quantity ηu . In particular if $\eta \equiv 1$, or more generally is continuous, the continuity of u is a manifestation of the continuity of local time. These connections between continuities at the macro and micro scale are dependent on the continuity of flux in defining the physical model.

A second principle arises for those contexts in which (10) is a prescribed backward equation with $\eta = \beta \equiv 1$, e.g. as in financial mathematics, [3]. From a physical perspective the interface condition is *not* a continuity of flux condition, however continuity of local time occurs if and only if $\lambda_j = \frac{D_j^+}{D_j^- + D_j^+}$, $j = 1, 2, \dots$.

1.1 Example: Piecewise Constant Coefficients

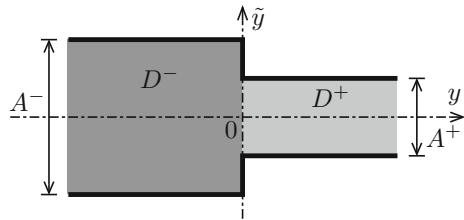
The scope and interest of our result may be illustrated with a single interface example motivated by applications to ecological dispersion in heterogeneous media. Consider a population of erratically moving individuals occupying an infinitely long, two-dimensional duct as depicted in Fig. 1. Let $A(y)$ be the cross-sectional area of the duct, and $1/\eta(y)$ the biomass of any individual occupying the cross-section at a distance y from the interface $y = 0$. Let $c(t, (y, \tilde{y}))$ denote the concentration of biomass, which is assumed continuous throughout, $\tilde{y}$ denoting the transversal spatial variable. Then $\eta(y)c(t, (y, \tilde{y}))$ is the concentration of individuals, for which we assume the following modification to Fick's law: the flux of individuals is given by $-D\mathbf{I}_{2 \times 2}\nabla c$, namely proportional to the gradient of the concentration of biomass. If D , A and η are taken to be piece-wise constant:

$$D(y) := \begin{cases} D^+ & \text{if } y > 0 \\ D^- & \text{if } y \leq 0, \end{cases}, \quad A(y) := \begin{cases} A^+ & \text{if } y > 0 \\ A^- & \text{if } y \leq 0, \end{cases}, \quad \eta(y) := \begin{cases} \eta^+ & \text{if } y > 0 \\ \eta^- & \text{if } y \leq 0, \end{cases} \quad (18)$$

then the concentration of biomass per unit length

$$u(t, y) = \int_{A(y)} c(t, (y, \tilde{y})) d\tilde{y} \quad (19)$$

Fig. 1 Schematics of an example of a problem leading to a one-dimensional diffusion process with one single interface at $y = 0$



satisfies the following one-dimensional diffusion equation

$$\frac{\partial u}{\partial t} = \frac{1}{\eta} \frac{\partial}{\partial y} \left(\frac{1}{2} D \frac{\partial u}{\partial y} \right), \quad \left[D \frac{\partial u}{\partial y} \right]_0 = 0, \quad \frac{1}{A^-} u(t, 0^-) = \frac{1}{A^+} u(t, 0^+), \quad (20)$$

which is of the form (1) with $\beta_0^\pm = 1/A^\pm$. The corresponding backwards operator $\mathcal{A}$ is given by (10) and (11) with $\lambda_0 = \frac{A^+ D^+}{A^+ D^+ + A^- D^-}$; let X denote the diffusion process generated by $\mathcal{A}$.

Given an initial biomass distribution u_0 , by virtue of (14), the solution to (20) can be written in terms of the transition probabilities of X as:

$$u(t, y) = \frac{1}{\eta(y)} \int_{\mathbb{R}} u_0(x) \eta(x) p(t, x, y) dx. \quad (21)$$

Namely, $p(t, x, y)$ is the density function of the *location* $X(t)$ of individuals that started at x , and the paths of X may be regarded as a model for the random movement of the individuals.

As in all cases of an operator $\mathcal{A}$ of the form (10) with *piece-wise constant* coefficients at a single interface, the associated process X may be expressed explicitly by re-scaling an appropriate skew Brownian motion (see [1, 2, 21, 27]). Specifically,

$$X(t) = \sqrt{\frac{D(B^{\alpha(D,A,\eta)}(t))}{\eta(B^{\alpha(D,A,\eta)}(t))}} B^{\alpha(D,A,\eta)}(t) \quad (22)$$

where $B^{\alpha(D,A,\eta)}$ is skew Brownian motion with transmission probability

$$\alpha(D, A, \eta) = \frac{A^+ \sqrt{\eta^+ D^+}}{A^+ \sqrt{\eta^+ D^+} + A^- \sqrt{\eta^- D^-}}. \quad (23)$$

The effect of the interface on the sample paths of the process X can be readily observed in the stochastic differential equation solved by X :

$$X(t) = \int_0^t \sqrt{\frac{D(X(r))}{\eta(X(r))}} dB(r) + \frac{A^+ D^+ - A^- D^-}{2A^+ D^+} L_+^X(t, x), \quad (24)$$

where B denotes standard Brownian motion, and $\{L_+^X(t, x) : t \geq 0\}$ is the right semi-martingale local time at x of the unknown process X . See [28] for details. For the current example, the right continuous version of natural local time is given by $\ell_+^X(t, x) := \frac{\eta(x)}{D(x)} L_+^X(t, x)$, and for every $t > 0$, $\ell_+^X(t, \cdot)$ is discontinuous at $x = 0$ with probability one, with a jump characterized by (16).

The nominal effect that a single interface exerts on the particle paths can now be elucidated by looking into further properties of skew Brownian motion, and natural local time. On one hand, (14) and the value of $\alpha(D, \lambda, \eta)$, inform us that for

an initial condition concentrated at $y = 0$, $u_0 = \delta_0(dy)$ in (1), the individuals will asymmetrically distribute on either side of the interface:

$$\int_0^\infty q(t, 0, y) dy = \mathbb{P}(X(t) > 0 | X(0) = 0) = \frac{\eta^-}{\eta^+} \alpha(D, \lambda, \eta), \quad \text{for all } t > 0. \quad (25)$$

On the other hand, natural local time $\ell^X(t, x)$ can be related to the time the process spends in a small vicinity of x . Of particular interest is the relative times particles spend at either side of zero. It follows from Corollary 1 that:

$$\lim_{\varepsilon \rightarrow 0} \frac{\int_0^t \mathbb{1}_{(0, \varepsilon)}(X(s)) ds}{\int_0^t \mathbb{1}_{(-\varepsilon, 0)}(X(s)) ds} = \frac{\eta^+ A^+}{\eta^- A^-}, \quad t > 0, \quad \text{a.s.} \quad (26)$$

1.2 Notation and Outline

The analytical treatment here revolves around functions $f : \mathbb{R} \rightarrow \mathbb{R}$ that are measurable with respect to the Borel σ -algebra $\mathcal{B}(\mathbb{R})$. The main function space is $C_b(\mathbb{R})$, the space of real valued continuous bounded functions. $B(\mathbb{R})$ denotes, in turn, the space of bounded measurable functions on $\mathbb{R}$. For a measure μ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, $L^2(\mu)$ denotes the Hilbert space $\{f : \mathbb{R} \rightarrow \mathbb{R}; \int_{\mathbb{R}} f^2(x) \mu(dx) < \infty\}$. The right and left limits of a function at x are denoted by $f(x^+)$ and $f(x^-)$, and their difference is the jump operator $[f]_x = f(x^+) - f(x^-)$. Whenever defined, the derivative of f is $\frac{df}{dx}$ or f' , while $f'_\pm(x) = f'(x^\pm)$ denote its left and right derivatives at x .

For the general theory of one-dimensional diffusion processes used here, we refer the reader to [28]. A diffusion process consists of a measurable space $(\Omega, \mathcal{F})$ and a family of probability measures $\{\mathbb{P}_x : x \in \mathbb{R}\}$. A sample path of the process is $X = \{X(t) : t \geq 0\} \in \Omega$. Under $\mathbb{P}_x$, the paths of X “start at x ”, namely, $\mathbb{P}_x(X(0) = x) = 1$ for all $x \in \mathbb{R}$. Only diffusions on $\mathbb{R}$, with infinite lifetime are considered here. Much of the analysis of such processes is undertaken in terms of their scale and speed measures, which we treat as follows. For $a \in \mathbb{R}$, the hitting time of a by X is $H_a^X = \inf\{t > 0 : X(t) = a\}$ and the scale measure $s(dx)$ is characterized by

$$\mathbb{P}_x(H_b^X < H_a^X) = \frac{s((a, x))}{s(a, b)}, \quad a < x < b. \quad (27)$$

Every scale measure in this paper will be absolutely continuous with respect to Lebesgue measure, $s(dx) = s'(x) dx$. We also define, without room for confusion, the scale function

$$s : \mathbb{R} \rightarrow \mathbb{R}, \quad s(x) := \int_0^x s'(y) dy. \quad (28)$$

The speed measure $m(dx)$ of X is the unique Radon measure such that

$$\mathbb{E}_x(H_a^X \wedge H_b^X) = \int_{\mathbb{R}} \frac{[s(x \wedge y) - s(a)][s(b) - s(x \vee y)]}{s(b) - s(a)} m(dy), \quad a < x < b. \quad (29)$$

The infinitesimal generator of the process can be written in terms of the speed and scale measures as follows

$$\mathcal{A}f = \frac{d}{dm} \frac{d}{ds} f, \quad f \in \text{Dom}(\mathcal{A}), \quad (30)$$

in the sense that,

$$\frac{df}{ds}(x_2) - \frac{df}{ds}(x_1) = \int_{x_1}^{x_2} \mathcal{A}f(x) m(dy). \quad (31)$$

Moreover,

$$f(X_t) - f(X_0) - \int_0^t \mathcal{A}f(X(s)) ds, \quad f \in \text{Dom}(\mathcal{A}), \quad x \in \mathbb{R} \quad (32)$$

is a martingale. For a given operator $(\mathcal{A}, \text{Dom}(\mathcal{A}))$ if (32) holds, we say that X solves the *martingale problem* for $\mathcal{A}$.

The rest of the paper is organized as follows. In the next section we provide a construction of the diffusion process X associated to the operator (10) and identify a stochastic differential equation for which X is the unique strong solution. In Sect. 3 we define three related but different notions of local time, including the *natural local time* and characterize its spatial continuity properties. The proof of the main Theorem 2 and Corollary 1 follows directly from such characterization.

2 On the Diffusion X

As illustrated in [27], the derivation of an associated process to an evolution operator (10) can be achieved in several ways, including the general theory of Dirichlet forms, or by martingale methods. Here, we “read off” the speed and scale measures from the backward operator written in the form (30) and construct the appropriate process via a stochastic differential equation. A similar approach was carried out in the case of piecewise constant coefficients by [25] and will be extended here to the present framework.

Recall Assumptions 1 on I , D and η , and let $(\mathcal{A}, \text{Dom}(\mathcal{A}))$ be as in (10), (11). Recursively define a sequence φ_j , $j \in \mathbb{Z}$ by

$$\frac{\varphi_j}{\varphi_{j-1}} = \frac{D(x_j^+)(1 - \lambda_j)}{D(x_j^-)\lambda_j} = \frac{\beta_j^+}{\beta_j^-}, \quad j \in \mathbb{Z}, \quad \varphi_0 := 1. \quad (33)$$

Then the generator $(\mathcal{A}, \text{Dom}(\mathcal{A}))$ given by (10) may be equivalently expressed by

$$\mathcal{A}f(x) = \frac{\varphi_j}{\eta(x)} \frac{d}{dx} \left(\frac{D(x)}{2\varphi_j} \frac{df}{dx} \right), \quad x \in (x_j, x_{j+1}), \quad j \in \mathbb{Z}, \quad (34)$$

acting on functions in $C_b(\mathbb{R})$ that are twice continuously differentiable within each (x_j, x_{j+1}) and such that

$$\frac{D(x_j^-)}{\varphi_{j-1}} f'(x_j^-) = \frac{D(x_j^+)}{\varphi_j} f'(x_j^+), \quad j \in \mathbb{Z}. \quad (35)$$

In view of (30) and (34), we take $m(dx) := m'(x) dx$ and $s(dx) = s'(x) dx$ with densities prescribed on $\mathbb{R} \setminus I$ by

$$s'(x) = \frac{2\varphi_j}{D(x)}, \quad m'(x) = \frac{\eta(x)}{\varphi_j}, \quad x \in (x_j, x_{j+1}), \quad j \in \mathbb{Z}. \quad (36)$$

The existence of an associated diffusion process is established in the following theorem.

Theorem 3 *Suppose D and η satisfy Assumptions 1, and Λ satisfies (12). Let m and s be measures with densities given by (36). Then there exist a Feller diffusion $X = (\Omega, \mathcal{F}, \{\mathbb{P}_x : x \in \mathbb{R}\})$ with speed and scale measures m and s , respectively, and whose transition probability function $p(t, x, dy)$ is the fundamental solution to the backwards equation $\frac{\partial v}{\partial t} = \mathcal{A}v$ with $(\mathcal{A}, \text{Dom}(\mathcal{A}))$ given by (34), (35). Moreover, q in (14) is the fundamental solution to the forward problem (1).*

Proof Note first that the boundedness assumptions on D and η , together with the fact that I has no accumulation points, make m and s Radon measures. Since m and s are assumed to have piecewise continuous densities, (30) takes the form $\mathcal{A}f(x) = \frac{1}{m'(x^+)} \frac{d}{dx} \left(\frac{f'(x)}{s'(x)} \right)$ with $\text{Dom}(\mathcal{A})$ being comprised of all functions $f \in C_b(\mathbb{R})$ such that $f'(x)/s'(x)$ is continuous on all of $\mathbb{R}$ and differentiable in $\mathbb{R} \setminus I$. Specializing to points in I , this specification is equivalent to the conditions in (35). The range $\text{Ran}(\mathcal{A})$ of $\mathcal{A}$ is contained in $B(\mathbb{R})$, and $\text{Dom}(\mathcal{A}) \times \text{Ran}(\mathcal{A})$ is a linear subset of $C_b(\mathbb{R}) \times B(\mathbb{R})$. The existence of a diffusion process X that solves the martingale problem for $(\mathcal{A}, \text{Dom}(\mathcal{A}))$ could now be established under very general conditions (see for example [31]). For our purposes however, Theorem 4 below explicitly constructs a diffusion X with speed and scale measures given by m and s , which therefore provides a solution to the martingale problem for $(\mathcal{A}, \text{Dom}(\mathcal{A}))$ for any $x \in \mathbb{R}$ ([28], Theorem 3.12, p. 308). Moreover, conditions (4) make the boundaries $\pm\infty$ inaccessible for the process X , and it follows from ([17], p. 38) that the transition probabilities $p(t, x, dy)$ of X make $T_t f(x) = \int_{\mathbb{R}} f(y) p(t, x, dy)$ a strongly continuous semigroup with the Feller property, namely $T_t : C_b(\mathbb{R}) \rightarrow C_b(\mathbb{R})$ for all $t \geq 0$. Let $(\mathcal{A}_0, \text{Dom}(\mathcal{A}_0))$ be the closure

of the infinitesimal generator of $\{T_t : t \geq 0\}$. By ([8], Theorem 4.1, p. 182) X is generated by $\mathcal{A}_0$ in $C_b(\mathbb{R})$ and is the unique solution to the martingale problem for $\mathcal{A}$. It follows that $\mathcal{A}_0 = \mathcal{A}$ in $\text{Dom}(\mathcal{A}_0) = \{f \in \text{Dom}(\mathcal{A}) : \mathcal{A}f \in C_b(\mathbb{R})\} \subset \text{Dom}(\mathcal{A})$. Moreover, from standard semigroup theory, $\frac{d}{dt}T_t f = \mathcal{A}f$ for all $f \in \text{Dom}(\mathcal{A}_0)$, namely, $p(t, x, dy)$ is the fundamental solution to the backwards equation $\frac{\partial v}{\partial t} = \mathcal{A}v$. The result now follows from the uniqueness of fundamental solutions for parabolic differential equations (see for example [9]).

We turn now to the construction of the diffusion X with speed and scale measures given by m and s in (36). The general procedure rests on the fact that the process $s(X)$ is on natural scale and can be written as an appropriate re-scaling of a time-changed Brownian motion (see [17] for the general theory, or [25] for the case of piecewise constant coefficients). Then, we derive the stochastic differential equation that the process X solves and verify that in fact, a strong solution exists.

We first establish a useful lemma regarding the processes X and $s(X)$.

Lemma 1 *Let X be a diffusion process with scale and speed measures s and m defined through (27)–(29), that are absolutely continuous with respect to Lebesgue measure, and have densities s' and m' respectively.*

1. *Denote $Y(t) = s(X(t))$, $t \geq 0$ where s is the scale function defined in (28). Then Y is a diffusion with scale function $s_Y(x) = x$ and speed measure m_Y with density satisfying:*

$$m'(x) = s'(x)m'_Y(s(x)) \quad a.e. \quad (37)$$

2. *The quadratic variation of X is given by:*

$$\langle X \rangle_t = \int_0^t \frac{2}{m'(X(s))s'(X(s))} ds. \quad (38)$$

Proof That $s_Y(x) = x$ follow because Y is on natural scale ([28], p. 302). Let $a < b$ and recall definition (29) of the speed measure. Denoting

$$g(a, b, x, y) := \frac{(x \wedge y - a)(b - x \vee y)}{b - a}, \quad x, y \in (a, b). \quad (39)$$

we can write

$$\mathbb{E}_x(H_a^X \wedge H_b^X) = \int_a^b g(s(a), s(b), s(x), s(y))m'(y) dy. \quad (40)$$

Since $x \rightarrow s(x)$ is an increasing function, $\mathbb{P}_x(X(t) \in (a, b)) = \mathbb{P}_{s(x)}(Y(t) \in (s(a), s(b)))$ and the expected exit time in (40) can also be written as

$$\mathbb{E}_{s(x)}(H_{s(a)}^Y \wedge H_{s(b)}^Y) = \int_{s(a)}^{s(b)} g(s(a), s(b), s(x), y) m'_Y(y) dy \quad (41)$$

$$= \int_a^b g(s(a), s(b), s(x), s(z)) s'(z) m'_Y(s(z)) dz. \quad (42)$$

The uniqueness of the measure m ([28], Theorem 3.6, p. 304) implies (37).

To prove the second assertion, let B be Brownian motion. It follows from ([6] Theorem 16.51) that a version of $Y = s(X)$ can be written as a time change of B as follows: let $\phi(t) := \int_0^t \frac{1}{2} m'_Y(B(r)) dr$ and $T(t) = \phi^{-1}(t)$, then $Y(t) = B(T(t))$, $t > 0$. In particular, $\langle Y \rangle_t = T(t)$. The quadratic variation of $X = s^{-1}(Y)$ is therefore

$$\langle X \rangle_t = \int_0^t [(s^{-1})'(Y(r))]^2 d\langle Y \rangle_r = \int_0^t \frac{1}{[s'(X(r))]^2} dT(r). \quad (43)$$

By (37), and performing a change of variables, we can also write T as

$$T(t) = \int_0^{T(t)} 2 \frac{1}{m'_Y(B(r))} d\phi(r) = \int_0^t 2 \frac{s'(X(r))}{m'(X(r))} dr. \quad (44)$$

Combining Eqs. (43) and (44) yields (38).

The following is an extension of results in [21] for the case of piecewise constant coefficient and a single interface, and concerns an equation of the general type considered by [16] and, more recently, [4].

Theorem 4 *Under Assumptions 1, the process X constructed in Theorem 3 is the pathwise unique strong solution to*

$$X(t) = \int_0^t \sqrt{\frac{D(X(s))}{\eta(X(s))}} dB(s) - \int_0^t \frac{D'(X(s))}{2\eta(X(s))} ds + \sum_{j \in \mathbb{Z}} \frac{2\lambda_j - 1}{2\lambda_j} L_+^X(t, x_j) \quad (45)$$

where $L_+^X(t, x)$ is right semimartingale local time of X and the functions D, η are taken to be left continuous at points in I .

Proof By Lemma 1 the continuous martingale $Y(t) = s(X(t))$ has absolutely continuous quadratic variation $\langle Y \rangle_t = \int_0^t Z^2(r) dr$ where $Z(r) := \sqrt{\frac{2s'(X(r))}{m'(X(r))}}$. It follows from ([14], Theorem 3.4.2) that there exists a filtered probability space with a Brownian motion B , such that $Y(t) = \int_0^t Z(r) dB(r)$. Since D is assumed to be of bounded variation, the function s^{-1} can be written as the difference of two convex functions and

$$(s^{-1})''(dx) = -\frac{s''(s^{-1}(x))}{[s'(s^{-1}(x))]^3} dx + \sum_{j \in \mathbb{Z}} \left[\frac{1}{s'(x_j^+)} - \frac{1}{s'(x_j^-)} \right] \delta_{s^{-1}(x_j)}(dx). \quad (46)$$

Applying the Ito-Tanaka and occupation times formulas ([28], Theorem 1.5 and Exercise 1.23, Chapter VI) on $X(t) = s^{-1}(Y(t))$ yields

$$X(t) = \int_0^t \sqrt{\frac{2}{s'_-(X(r))m'_-(X(r))}} dB(r) - \int_0^t \frac{s''(X(r))}{[s'(X(r))]^2 m'(X(r))} dr + \sum_{j \in \mathbb{Z}} \left[1 - \frac{s'(x_j^+)}{s'(x_j^-)} \right] L_+^X(t, x_j) \quad (47)$$

which coincides with (45). The pathwise uniqueness of strong solutions follows from [4, 16], by noting that under the current assumptions, $\sqrt{D/\eta}$ is a function of bounded variation, bounded away from zero, and the measure $(s^{-1})''(dx)$ in (47) is finite with $\frac{2\lambda_j - 1}{2\lambda_j} < \frac{1}{2}$ for all $j \in \mathbb{Z}$.

Having obtained the diffusion X corresponding to the conservation form of (1), in the next section we explore the role of continuity of flux in the structure of X and its local time.

3 Various Notions of Local Time

Local time has a striking mathematical role in the development of the modern theory of stochastic processes, from Brownian motion and diffusion, to continuous semimartingale calculus, e.g., see [13, 28, 29]. In the course of this development two particular variations on the notion of local time have occurred as follows:

Definition 1 Let X be a continuous semimartingale with quadratic variation $\langle X \rangle_t$. The *right, left, symmetric semimartingale local time (smlt)* of X is a stochastic process, respectively denoted $L_+^X(t, x)$, $L_-^X(t, x)$, $L_*^X(t, x)$, $x \in \mathbb{R}$, $t \geq 0$, continuous in t and determined by either being right-continuous in x , left continuous in x , or by averaging $L_*^X(t, x) = (L_+^X(t, x) + L_-^X(t, x))/2$, and such that in any case

$$\int_0^t \varphi(X(s)) d\langle X \rangle_s = \int_{\mathbb{R}} \varphi(x) L_{\pm,*}^X(t, x) dx, \quad (48)$$

almost surely for any positive Borel measurable function φ .

Remark 2 The notation here is slightly different from that of [28].

Observe that by choosing φ as an indicator function of an interval $[x, x + \varepsilon)$, $\varepsilon > 0$, one has by right-continuity, for example, that

$$L_+^X(t, x) = \lim_{\varepsilon \downarrow 0} \frac{1}{\varepsilon} \int_0^t \mathbb{1}_{[x, x+\varepsilon)}(X(s)) d\langle X \rangle_s. \quad (49)$$

Similarly, L_-^X can be obtained by using the indicator on the interval $(x - \varepsilon, x]$.

The next definition is that of diffusion local time (dlt) and requires Feller's notions of speed measure $m(dx)$ and scale function $s(x)$ (see (27) and (29)). It is customary to define dlt only for diffusions on natural scale; e.g., [13, 29]. However, since any diffusion in natural scale is a time change of Brownian motion, it follows that its local times will therefore be themselves time changes of the local time of Brownian motion, and therefore always (spatially) continuous. On the other hand, for a general Feller diffusion X with scale function s , the transformation $Y = s(X)$ produces a diffusion on natural scale. This transformation renders local time continuity as a *generic* property that does not further inform more specific structure of the diffusion X . We thus extend the definition to Feller diffusions with any scale.

Definition 2 Let Y be a diffusion with speed measure $m_Y(dy)$. Then the *diffusion local time (dlt)* of Y , denoted $\tilde{L}^Y(y, t)$ is specified by

$$\int_0^t \varphi(Y(r)) dr = \int_I \varphi(y) \tilde{L}^Y(y, t) m_Y(dy), \quad (50)$$

almost surely for any positive Borel measurable function φ .

By Lebesgue's differentiation theorem, it follows that

$$\tilde{L}_+^X(t, x) = \lim_{\varepsilon \downarrow 0} \frac{1}{m[x, x + \varepsilon)} \int_0^t \mathbb{1}_{[x, x + \varepsilon)}(X(s)) ds \quad (51)$$

with the corresponding formula for $\tilde{L}_-^X$, and $\tilde{L}_*^X$ as the average.

For the case of piecewise constant coefficients at a single interface, a particular *local time continuity* property at the interface was identified in [2]. It was useful there to consider a modification of the more standard notions of semimartingale and diffusion local time to one referred to as *natural local time*. This was achieved there by exploiting explicit connections with skew Brownian motion indicated above. An extension to piecewise continuous coefficients and multiple interfaces is obtained in the present note.

The following modification of the definition of local time will be seen as useful in precisely calibrating jump size relative to the interface parameters. Just as in the case of semimartingale local time, one may consider right, left, and symmetric versions.

Definition 3 Let X be a regular diffusion. The *natural local time (nlt)* of X , right, left, and symmetric, respectively, denoted $\ell_{\pm,*}^X(t, x)$ is specified by the occupation time formula

$$\int_0^t \varphi(X(s)) ds = \int_I \varphi(y) \ell_{\pm,*}^X(t, x) dx, \quad (52)$$

for any positive Borel measurable functions φ . The right and left versions are defined by the respective right-continuous, left-continuous versions, while the symmetric nlt is defined by the arithmetic average of these two.

Remark 3 In its simplest terms, the modification to natural local time is made physically natural by examination of its *units*, namely $[\frac{T}{L}]$, whereas those of smlt are those of (spatial) length $[L]$, while dlt is dimensionless. However, as previously noted, its essential feature resides in the implications of continuity properties relative to conservation laws. In particular, this puts a notion of stochastic local time on par with fundamentally important principles of *concentration flux* and *conservation of mass* for pdes.

The relationship between the three notions of local time above is summarized in the following proposition, the proof of which follows from a direct application of Definitions 1, 2 and 3.

Proposition 1 *Let X be a Feller diffusion with speed measure m and quadratic variation $\langle X \rangle_t = \int_0^t q(X(s)) ds$, then*

$$\ell^X(t, x) = \frac{L^X(t, x)}{q(x)} = m'(x) \tilde{L}^X(t, x), \quad \text{a.s.} \quad (53)$$

with right and left versions obtained by considering the right and left continuous versions of q and m' respectively.

The celebrated theorem of [32] on the (joint) continuity of local time for Brownian motion is well-known for the depth it provides to the analysis of Brownian motion paths. This result is also naturally at the heart of the following general characterization of continuity of natural local time for regular diffusions.

Theorem 5 *Let X be a Feller diffusion on $\mathbb{R}$ with absolutely continuous speed measure $m(dx) = m'(x) dx$ and scale function $s(x)$. Then the ratio $\ell^X(t, x)/m'(x)$ is continuous. Moreover, the natural local time of X is continuous at x if and only if m' is continuous at x .*

Proof Let $Y(t) = s(X(t)) = B(T(t))$ as in the proof of Lemma 1. Then $L^Y(t, x) = L^B(T(t), x)$, $t \geq 0$, $x \in \mathbb{R}$. On the other hand, (53) together with (38) and ([28] Exercise VI.1.23), give

$$\ell^X(t, x) = \frac{m'(x)s'(x)}{2} L^X(t, x) = \frac{m'(x)}{2} L^Y(t, s(x)) = \frac{m'(x)}{2} L^B(T(t), s(x)) \quad (54)$$

and the assertion follows from continuity of L^B and the scale function.

It is noteworthy that, in general, the semimartingale local time of X is *not* made continuous by division by m' as in Theorem 5, while, as has been previously noted, the diffusion local time of the process transformed to natural scale is *always* continuous. From the point of view of applications the theorem shows that natural local time furnishes a microscopic *probe* to detect interfacial parameters η , when $\beta = 1$, or β when $\eta = D$, respectively, in (1), through location and size of its discontinuities.

The desired Theorem 2 on the role of the continuity of flux for the process X defined by (34) now follows as a corollary.

Proof (Proof of Theorem 2) Let $x_j \in I$. By Theorem 5 and the definitions of m' , φ_j in (33), (36) give

$$\frac{\ell^X(t, x_j^+)}{\ell^X(t, x_j^-)} = \frac{m'(x_j^+)}{m'(x_j^-)} = \frac{\eta(x_j^+)\varphi_{j-1}}{\eta(x_j^-)\varphi_j} = \frac{\eta(x_j^+)D(x_j^-)\lambda_j}{\eta(x_j^-)D(x_j^+)(1 - \lambda_j)} \quad (55)$$

Acknowledgments This research was partially supported by a grant DMS-1408947 from the National Science Foundation.

References

1. Appuhamillage, T.A., Bokil, V.A., Thomann, E.A., Waymire, E.C., Wood, B.D.: Occupation and local times for skew brownian motion with applications to dispersion across an interface. *Ann. Appl. Probab.* **21**(1), 183–214 (2011)
2. Appuhamillage, T.A., Bokil, V.A., Thomann, E.A., Waymire, E.C., Wood B.D.: Skew dispersion and continuity of local time. *J. Stat. Phys.* 1–11 (2013)
3. Barndorff-Nielsen, O.E., Shiriaev, A.N.: Change of Time and Change of Measure, vol. 13. World Scientific (2010)
4. Bass, R.F., Chen, Z.-Q.: One-dimensional stochastic differential equations with singular and degenerate coefficients. *Sankhyā: Indian J. Stat.* 19–45 (2005)
5. Berkowitz, B., Cortis, A., Dror, I., Scher H.: Laboratory experiments on dispersive transport across interfaces: the role of flow direction. *Water Resour. Res.* **45**(2) (2009)
6. Breiman, L.: Probability, Volume 7 of Classics in Applied Mathematics. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA (1992)
7. Cantrell, R.S., Cosner, C.: Spatial Ecology via Reaction-diffusion Equations. John Wiley and Sons (2004)
8. Ethier, S.N., Kurtz, T.G.: Markov Processes: characterization and Convergence, vol. 282. Wiley (2009)
9. Friedman, A.: Partial Differential Equations of Parabolic Type. Courier Dover Publications (2013)
10. Guo, F., Jokipii, J., Kota, J.: Particle acceleration by collisionless shocks containing large-scale magnetic-field variations. *Astrophys. J.* **725**(1), 128 (2010)
11. Hill, A.E.: Leakage of barotropic slope currents onto the continental shelf. *J. Phys. Oceanogr.* **25**(7), 1617–1621 (1995)
12. Hoteit, H., Mose, R., Younes, A., Lehmann, F., Ackerer, P.: Three-dimensional modeling of mass transfer in porous media using the mixed hybrid finite elements and the random-walk methods. *Math. Geol.* **34**(4), 435–456 (2002)
13. Ito, K., MacKean, H.: Diffusion processes and their sample paths. *Class. Math.* (1974)
14. Karatzas, I., Shreve, S.: Brownian Motion and Stochastic Calculus, vol. 113, 2nd edn. Springer (1991)
15. Kuo, R.-K.H., Irwin, N., Greenkorn, R., Cushman, J.: Experimental investigation of mixing in aperiodic heterogeneous porous media: comparison with stochastic transport theory. *Transp. Porous Media* **37**(2), 169–182 (1999)
16. Le Gall, J.: One—dimensional stochastic differential equations involving the local times of the unknown process. In: *Stochastic Analysis and Applications*, pp. 51–82. Springer (1984)
17. Mandl, P.: Analytical Treatment of One-dimensional Markov Processes. Academia, Publishing House of the Czechoslovak Academy of Sciences (1968)
18. Matano, R.P., Palma, E.D.: On the upwelling of downwelling currents. *J. Phys. Oceanogr.* **38**(11), 2482–2500 (2008)

19. McKenzie, H.W., Lewis, M.A., Merrill, E.H.: First passage time analysis of animal movement and insights into the functional response. *Bull. Math. Biol.* **71**(1), 107–129 (2009)
20. Nilsen, W., Sayit, H.: No arbitrage in markets with bounces and sinks. *Int. Rev. Appl. Financ. Issues Econ.* **3**(4), 696–699 (2011)
21. Ouknine, Y.: “Skew-Brownian motion” and derived processes. *Theor. Probab. Appl.* **35**(1), 163–169 (1991)
22. Ovaskainen, O.: Habitat-specific movement parameters estimated using mark-recapture data and a diffusion model. *Ecology*. **85**(1), 242–257 (2004)
23. Perkowski, N., Ruf, J.: Conditioned martingales. *Electron. Commun. Probab.* **17**(48), 1–12 (2012)
24. Perrin, J.: Mouvement brownien et réalité moléculaire. *Annales de Chimie et de Physique* **18**, 5–104 (1909)
25. Ramirez, J.M.: Multi-skewed brownian motion and diffusion in layered media. *Proc. Am. Math. Soc.* **139**(10), 3739–3752 (2011)
26. Ramirez, J.M.: Population persistence under advection-diffusion in river networks. *J. Math. Biol.* **65**(5), 919–942 (2012)
27. Ramirez, J.M., Thomann, E.A., Waymire, E.C., et al.: Advection-dispersion across interfaces. *Stat. Sci.* **28**(4), 487–509 (2013)
28. Revuz, D., Yor, M.: *Continuous Martingales and Brownian Motion*, vol. 293. Springer (1999)
29. Rogers, L.C.G., Williams, D.: *Diffusions, Markov Processes, and Martingales: vol. 2*. Cambridge University Press, Itô Calculus (2000)
30. Schultz, C.B., Crone, E.E.: Patch size and connectivity thresholds for butterfly habitat restoration. *Conserv. Biol.* **19**(3), 887–896 (2005)
31. Stroock, D.W., Varadhan, S.S.: *Multidimensional Diffusion Processes*, vol. 233. Springer (1979)
32. Trotter, H., et al.: A property of brownian motion paths. *Ill. J. Math.* **2**(3), 425–433 (1958)

Simulation of Stochastic Volterra Equations Driven by Space–Time Lévy Noise

Bohan Chen, Carsten Chong and Claudia Klüppelberg

Abstract In this paper we investigate two numerical schemes for the simulation of stochastic Volterra equations driven by space–time Lévy noise of pure-jump type. The first one is based on truncating the small jumps of the noise, while the second one relies on series representation techniques for infinitely divisible random variables. Under reasonable assumptions, we prove for both methods L^p - and almost sure convergence of the approximations to the true solution of the Volterra equation. We give explicit convergence rates in terms of the Volterra kernel and the characteristics of the noise. A simulation study visualizes the most important path properties of the investigated processes.

Keywords Simulation of SPDEs · Simulation of stochastic Volterra equations · Space–time Lévy noise · Stochastic heat equation · Stochastic partial differential equation

1 Introduction

The aim of this paper is to investigate different simulation techniques for stochastic Volterra equations (SVEs) of the form

$$Y(t, x) = Y_0(t, x) + \int_0^t \int_{\mathbb{R}^d} G(t, x; s, y) \sigma(Y(s, y)) \Lambda(ds, dy), \quad (t, x) \in \mathbb{R}_+ \times \mathbb{R}^d, \quad (1)$$

B. Chen · C. Chong · C. Klüppelberg (✉)
Technische Universität München, Boltzmannstraße 3, 85748 Garching, Germany
e-mail: cklu@ma.tum.de

C. Chong
e-mail: carsten.chong@tum.de

B. Chen
e-mail: bohan.chen@mytum.de

where G is a deterministic kernel function, σ a Lipschitz coefficient and Λ a Lévy basis on $\mathbb{R}_+ \times \mathbb{R}^d$ of pure-jump type with no Gaussian part. In the purely temporal case where no space is involved and the kernel G is sufficiently regular on the diagonal $\{(t; s) \in \mathbb{R}_+ \times \mathbb{R}_+ : t = s\}$, the existence and uniqueness of the solution Y to (1) are established for general semimartingale integrators in [14]. The space–time case (1) is treated in [5] for quite general Lévy bases. In particular, G is allowed to be singular on the diagonal, which typically happens in the context of stochastic partial differential equations (SPDEs) where G is the Green’s function of the underlying differential operator. More details on the connection between SPDEs and the SVE (1) are presented in Sect. 2, or can be found in [2, 5, 20].

Since in most cases there exists no explicit solution formula for the SVE (1), it is a natural task to develop appropriate simulation algorithms. For SPDEs driven by Gaussian noise, research on this topic is rather far advanced, see e.g. [7, 9, 21]. However, for SPDEs driven by jump noises such as non-Gaussian Lévy bases, the related literature is considerably smaller, see [3] and the work of Hausenblas and coauthors [8, 10, 11]. The case $\sigma \equiv 1$ has been treated in [4]. The contribution of our paper can be summarized as follows:

- We propose and analyze two approximation schemes for (1), each of which replaces the original noise by a truncated noise that only has finitely many atoms on compact subsets of $\mathbb{R}_+ \times \mathbb{R}^d$. For the first scheme, we simply cut off all jumps whose size is smaller than a constant. For the second scheme, we use series representation techniques for the noise as in [17] such that the jumps to be dropped off are chosen randomly. Both methods have already been applied successfully to the simulation of Lévy processes, cf. [1, 18].
- In the case where G originates from an SPDE, the crucial difference of our numerical schemes to the Euler or finite element methods in the references mentioned before is that we do not simulate small space–time increments of the noise but successively the true jumps of the Lévy basis, which is an easier task given that one usually only knows the underlying Lévy measure. It is important to recognize that this is only possible because the noise Λ is of pure-jump type, and contains neither a Gaussian part nor a drift. We shall point out in Sect. 6 how to relax this assumption.

The remaining article is organized as follows: Sect. 2 gives the necessary background for the SVE (1). In particular, we present sufficient conditions for the existence and uniqueness of solutions, and address the connection between (1) and SPDEs. In Sect. 3 we construct approximations to the solution Y of (1) by truncating the small jumps of the Lévy basis. We prove in Theorem 1 their L^p -convergence, and in some cases also their almost sure (a.s.) convergence to the target process Y . In Sect. 4 we approximate the driving Lévy basis using series representation methods. This leads to an algorithm that produces approximations again converging in the L^p -sense, sometimes also almost surely, to Y , see Theorem 2. In both theorems, we find explicit L^p -convergence rates that only depend on the kernel G and the characteristics of Λ . Section 5 presents a simulation study for the stochastic heat equation

which highlights the typical path behaviour of stochastic Volterra equations. The final Sect. 6 compares the two simulation algorithms developed in this paper and discusses some further directions of the topic.

2 Preliminaries

We start with a summary of notations that will be employed in this paper.

$\mathbb{R}_+$	The set $[0, \infty)$ of <i>positive</i> real numbers
$\mathbb{N}$	The natural numbers $\{1, 2, \dots\}$
$\mathbb{B}$	A stochastic basis $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t \in \mathbb{R}_+}, \mathbb{P})$ satisfying the usual hypotheses of completeness and right-continuity
$\tilde{\Omega}, \tilde{\Omega}$	$\tilde{\Omega} := \Omega \times \mathbb{R}_+$ and $\tilde{\Omega} := \Omega \times \mathbb{R}_+ \times \mathbb{R}^d$ where $d \in \mathbb{N}$
$\mathcal{B}(\mathbb{R}^d)$	The Borel σ -field on $\mathbb{R}^d$
$\tilde{\mathcal{B}}_b$	The collection of all bounded Borel sets of $\mathbb{R}_+ \times \mathbb{R}^d$
$\mathcal{P}$	The predictable σ -field on $\mathbb{B}$ or the collection of all predictable processes $\tilde{\Omega} \rightarrow \mathbb{R}$
$\tilde{\mathcal{P}}$	The product $\mathcal{P} \otimes \mathcal{B}(\mathbb{R}^d)$ or the collection all $\mathcal{P} \otimes \mathcal{B}(\mathbb{R}^d)$ -measurable processes $\tilde{\Omega} \rightarrow \mathbb{R}$
$\tilde{\mathcal{P}}_b$	The collection of sets in $\tilde{\mathcal{P}}$ which are a subset of $\Omega \times [0, k] \times [-k, k]^d$ for some $k \in \mathbb{N}$
δ_a	The Dirac measure at a
p^*	$p \vee 1$
L^p	The space $L^p(\Omega, \mathcal{F}, \mathbb{P})$, $p \in (0, \infty)$ with $\ X\ _{L^p} := \mathbb{E}[X ^p]^{1/p^*}$
L^0	The space $L^0(\Omega, \mathcal{F}, \mathbb{P})$ of all random variables on $\mathbb{B}$ endowed with the topology of convergence in probability
B_{loc}^p	The set of all $Y \in \tilde{\mathcal{P}}$ for which $\ Y(t, x)\ _{L^p}$ is uniformly bounded on $[0, T] \times \mathbb{R}^d$ for all $T \in \mathbb{R}_+$ ($p \in (0, \infty)$)
A^c	The complement of A within the superset it belongs to (which will be clear from the context)
$A - B$	$\{x - y : x \in A, y \in B\}$
$-A$	$\{-x : x \in A\}$
Leb	The Lebesgue measure on $\mathbb{R}^d$ (d should be clear from the context)
$\ \cdot\ $	The Euclidean norm on $\mathbb{R}^d$
$C, C(T)$	Two generic constants in $\mathbb{R}_+$, one dependent and one independent of T , whose values are irrelevant in this paper and may therefore change from one place to the other

We suppose that the stochastic basis $\mathbb{B}$ supports a *Lévy basis*, that is, a mapping $\Lambda : \tilde{\mathcal{P}}_b \rightarrow L^0$ with the following properties:

- $\Lambda(\emptyset) = 0$ a.s.
- For all pairwise disjoint sets $(A_i)_{i \in \mathbb{N}} \subset \tilde{\mathcal{P}}_b$ with $\bigcup_{i=1}^{\infty} A_i \in \tilde{\mathcal{P}}_b$ we have

$$\Lambda\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \Lambda(A_i) \quad \text{in } L^0. \quad (2)$$

- $(\Lambda(\Omega \times B_i))_{i \in \mathbb{N}}$ is a sequence of independent random variables if $(B_i)_{i \in \mathbb{N}}$ are pairwise disjoint sets in $\mathcal{B}_b$.
- For every $B \in \mathcal{B}_b$, $\Lambda(\Omega \times B)$ has an infinitely divisible distribution.
- $\Lambda(A)$ is $\mathcal{F}_t$ -measurable when $A \in \tilde{\mathcal{P}}_b$ and $A \subseteq \Omega \times [0, t] \times \mathbb{R}^d$ for $t \in \mathbb{R}_+$.
- For every $t \in \mathbb{R}_+$, $A \in \tilde{\mathcal{P}}_b$ and $\Omega_0 \in \mathcal{F}_t$ we have a.s.

$$\Lambda(A \cap (\Omega_0 \times (t, \infty) \times \mathbb{R}^d)) = \mathbb{1}_{\Omega_0} \Lambda(A \cap (\Omega \times (t, \infty) \times \mathbb{R}^d)).$$

Just as Lévy processes are semimartingales and thus allow for an Itô integration theory, Lévy bases belong to the class of L^0 -valued σ -finite random measures. Therefore, it is possible to define the stochastic integral

$$\int_{\mathbb{R}_+ \times \mathbb{R}^d} H(s, y) \Lambda(ds, dy)$$

for $H \in \tilde{\mathcal{P}}$ that are *integrable* with respect to Λ , see [6] for the details.

Similarly to Lévy processes, there exist two notions of characteristics for Lévy bases: one going back to [15, Proposition 2.1] that is based on the Lévy-Khintchine formula and is independent of $\mathbb{F}$, and a filtration-based one that is useful for stochastic analysis [6, Theorem 3.2]. For the whole paper, we will assume that both notions coincide such that Λ has a canonical decomposition under the filtration $\mathbb{F}$ of the form

$$\begin{aligned} \Lambda(dt, dx) &= B(dt, dx) + \Lambda^c(dt, dx) + \int_{\mathbb{R}} z \mathbb{1}_{\{|z| \leq 1\}} (\mu - \nu)(dt, dx, dz) \\ &\quad + \int_{\mathbb{R}} z \mathbb{1}_{\{|z| > 1\}} \mu(dt, dx, dz), \end{aligned}$$

where B is a σ -finite signed Borel measure on $\mathbb{R}_+ \times \mathbb{R}^d$, Λ^c a Lévy basis such that $\Lambda(\Omega \times B)$ is normally distributed with mean 0 and variance $C(B)$ for all $B \in \tilde{\mathcal{B}}_b$, and μ a Poisson measure on $\mathbb{R}_+ \times \mathbb{R}^d$ relative to $\mathbb{F}$ with intensity measure ν (cf. [12, Definition II.1.20]). There exists also a σ -finite Borel measure λ on $\mathbb{R}_+ \times \mathbb{R}^d$ such that

$$\begin{aligned} B(dt, dx) &= b(t, x) \lambda(dt, dx), \quad C(dt, dx) = c(t, x) \lambda(dt, dx) \quad \text{and} \\ \nu(dt, dx, dz) &= \pi(t, x, dz) \lambda(dt, dx) \end{aligned} \quad (3)$$

with two functions $b: \mathbb{R}_+ \times \mathbb{R}^d \rightarrow \mathbb{R}$ and $c: \mathbb{R}_+ \times \mathbb{R}^d \rightarrow \mathbb{R}_+$ as well as a transition kernel π from $(\mathbb{R}_+ \times \mathbb{R}^d, \mathcal{B}(\mathbb{R}_+ \times \mathbb{R}^d))$ to $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ such that $\pi(t, x, \cdot)$ is a Lévy measure for each $(t, x) \in \mathbb{R}_+ \times \mathbb{R}^d$.

We have already mentioned in the introduction that we will assume

$$C = 0 \tag{4}$$

throughout the paper. For simplicity we will also make two further assumptions: first, that there exist $b \in \mathbb{R}$ and a Lévy measure π such that for all $(t, x) \in \mathbb{R}_+ \times \mathbb{R}^d$ we have

$$b(t, x) = b, \quad \pi(t, x, \cdot) = \pi \quad \text{and} \quad \lambda(dt, dx) = d(t, x); \tag{5}$$

second, that

$$\Lambda \in \mathcal{S} \cup \mathcal{V}_0, \tag{6}$$

where $\mathcal{S}$ is the collection of all *symmetric* Lévy bases and $\mathcal{V}_0$ is the class of Lévy bases with *locally finite variation and no drift*, defined by the property that

$$\int_{\mathbb{R}} |z| \mathbb{1}_{\{|z| \leq 1\}} \pi(dz) < \infty, \quad \text{and} \quad b_0 := b - \int_{\mathbb{R}} z \mathbb{1}_{\{|z| \leq 1\}} \pi(dz) = 0.$$

Furthermore, if Λ has a finite first moment, that is,

$$\int_{\mathbb{R}} |z| \mathbb{1}_{\{|z| > 1\}} \pi(dz) < \infty, \tag{7}$$

we define

$$\begin{aligned} B_1(dt, dx) &:= b_1 d(t, x), \quad b_1 := b + \int_{\mathbb{R}} z \mathbb{1}_{\{|z| > 1\}} \pi(dz), \\ M(dt, dx) &:= \Lambda(dt, dx) - B_1(dt, dx) = \int_{\mathbb{R}} z (\mu - \nu)(dt, dx, dz). \end{aligned}$$

Next, let us summarize the most important facts regarding the SVE (1). All details that are not explained can be found in [5]. First, many SPDEs of evolution type driven by Lévy noise can be written in terms of (1), where G is the Green's function of the corresponding differential operator. Most prominently, taking G being the heat kernel in $\mathbb{R}^d$, (1) is the so-called *mild formulation* of the stochastic heat equation (with constant coefficients and multiplicative noise). Typically for parabolic equations, the heat kernel is very smooth in general but explodes on the diagonal $t = s$ and $x = y$. In fact, it is only p -fold integrable on $[0, T] \times \mathbb{R}^d$ for $p < 1 + 2/d$. In particular, as soon as $d \geq 2$, it is not square-integrable, and as a consequence, no solution to the stochastic heat equation in the form (1) will exist for Lévy noises with non-zero Gaussian component. This is another reason for including assumption (4) in this paper.

Second, let us address the existence and uniqueness problem for (1). By a *solution* to this equation we mean a predictable process $Y \in \tilde{\mathcal{P}}$ such that for all $(t, x) \in \mathbb{R}_+ \times \mathbb{R}^d$, the stochastic integral on the right-hand side of (1) is well defined and the equation itself for each $(t, x) \in [0, T] \times \mathbb{R}^d$ holds a.s. We identify two solutions as soon as they are modifications of each other. Given a number $p \in (0, 2]$, the following conditions A1–A6 guarantee a unique solution to (1) in B_{loc}^p by [5, Theorem 3.1]:

- A1. $Y_0 \in B_{\text{loc}}^p$ is independent of Λ .
 A2. $\sigma: \mathbb{R} \rightarrow \mathbb{R}$ is Lipschitz continuous, that is, there exists $C \in \mathbb{R}_+$ such that

$$|\sigma(x) - \sigma(y)| \leq C|x - y|, \quad x, y \in \mathbb{R}. \quad (8)$$

- A3. $G: (\mathbb{R}_+ \times \mathbb{R}^d)^2 \rightarrow \mathbb{R}$ is a measurable function with $G(t, \cdot; s, \cdot) \equiv 0$ for $s > t$.

- A4. Λ satisfies (3)–(6) and

$$\int_{\mathbb{R}} |z|^p \pi(dz) < \infty. \quad (9)$$

- A5. If we define for $(t, x), (s, y) \in \mathbb{R}_+ \times \mathbb{R}^d$

$$\tilde{G}(t, x; s, y) := |G(t, x; s, y)| \mathbb{1}_{\{p > 1, \Lambda \notin \mathcal{S}\}} + |G(t, x; s, y)|^p, \quad (10)$$

then we have for all $T \in \mathbb{R}_+$

$$\sup_{(t,x) \in [0,T] \times \mathbb{R}^d} \int_0^T \int_{\mathbb{R}^d} \tilde{G}(t, x; s, y) d(s, y) < \infty. \quad (11)$$

- A6. For all $\varepsilon > 0$ and $T \in \mathbb{R}_+$ there exist $k \in \mathbb{N}$ and a partition $0 = t_0 < \dots < t_k = T$ such that

$$\sup_{(t,x) \in [0,T] \times \mathbb{R}^d} \sup_{i=1,\dots,k} \int_{t_{i-1}}^{t_i} \int_{\mathbb{R}^d} \tilde{G}(t, x; s, y) d(s, y) < \varepsilon. \quad (12)$$

Apart from A1–A6, we will add another assumption in this paper:

- A7. There exists a sequence $(U^N)_{N \in \mathbb{N}}$ of compact sets increasing to $\mathbb{R}^d$ such that for all $T \in \mathbb{R}_+$ and compact sets $K \subseteq \mathbb{R}^d$ we have, as $N \rightarrow \infty$,

$$\begin{aligned} r_1^N(T, K) := & \sup_{(t,x) \in [0,T] \times K} \left(\int_0^t \int_{(U^N)^c} |G(t, x; s, y)| \mathbb{1}_{\{p > 1, \Lambda \notin \mathcal{S}\}} d(s, y) \right. \\ & \left. + \left(\int_0^t \int_{(U^N)^c} |G(t, x; s, y)|^p d(s, y) \right)^{1/p^*} \right) \rightarrow 0. \end{aligned} \quad (13)$$

Conditions A6 and A7 are automatically satisfied if $|G(t, x; s, y)| \leq g(t - s, x - y)$ for some measurable function g and A5 holds with G replaced by g . For A6 see [5, Remark 3.3(3)]; for A7 choose $U^N := \{x \in \mathbb{R}^d: \|x\| \leq N\}$ such that for $p > 1$

$$\begin{aligned}
& \sup_{(t,x) \in [0,T] \times K} \int_0^t \int_{(U^N)^c} |G(t, x; s, y)|^p \, \mathrm{d}(s, y) \\
& \leq \sup_{(t,x) \in [0,T] \times K} \int_0^t \int_{(U^N)^c} g^p(t-s, x-y) \, \mathrm{d}(s, y) \\
& = \sup_{x \in K} \int_0^T \int_{x-(U^N)^c} g^p(s, y) \, \mathrm{d}(s, y) \leq \int_0^T \int_{K-(U^N)^c} g^p(s, y) \, \mathrm{d}(s, y) \\
& \rightarrow 0 \quad \text{as } N \rightarrow \infty
\end{aligned}$$

by the fact that $K - (U^N)^c \downarrow 0$. A similar calculation applies to the case $p \in (0, 1]$ and the first term in $r_1^N(T, K)$.

Example 1 We conclude this section with the stochastic heat equation in $\mathbb{R}^d$, whose mild formulation is given by the SVE (1) with

$$G(t, x; s, y) = g(t-s, x-y), \quad g(t, x) = \frac{\exp(-\|x\|^2/(4t))}{(4\pi t)^{d/2}} \mathbb{1}_{[0,t)}(s) \quad (14)$$

for $(t, x), (s, y) \in \mathbb{R}_+ \times \mathbb{R}^d$. We assume that Y_0 and σ satisfy conditions A1 and A2, respectively. Furthermore, we suppose that (3)–(6) are valid, and that (9) holds with some $p \in (0, 1 + 2/d)$. It is straightforward to show that then A3–A6 are satisfied with the same p . Let us estimate the rate $r_1^N(T, K)$ for $T \in \mathbb{R}_+$, $K := \{\|x\| \leq R\}$ with $R \in \mathbb{N}$, and $U^N := \{\|x\| \leq N\}$. We first consider the case $p \leq 1$ or $\Lambda \in \mathcal{S}$. Since $K - (U^N)^c = (U^{N-R})^c$ for $N \geq R$, the calculations after A7 yield ($\Gamma(\cdot, \cdot)$ denotes the upper incomplete gamma function and $p(d) := 1 + (1-p)d/2$)

$$\begin{aligned}
(r_1^N(T, K))^{p*} & \leq \int_0^T \int_{(U^{N-R})^c} g^p(t, x) \, \mathrm{d}(t, x) = \int_0^T \int_{N-R}^\infty \frac{\exp(-pr^2/(4t))}{(4\pi t)^{pd/2}} r^{d-1} \, \mathrm{d}r \, \mathrm{d}t \\
& = C \int_0^T t^{p(d)-1} \Gamma\left(\frac{d}{2}, \frac{p(N-R)^2}{4t}\right) \, \mathrm{d}t \\
& = C(T) \left((p(d))^{-1} \Gamma\left(\frac{d}{2}, \frac{p(N-R)^2}{4T}\right) - \left(\frac{p(N-R)^2}{4T}\right)^{p(d)} \right. \\
& \quad \times \left. \Gamma\left(\frac{d}{2} - p(d), \frac{p(N-R)^2}{4T}\right) \right) \\
& \leq C(T) \exp\left(-\frac{p(N-R)^2}{4T}\right) (N-R)^{d-2}, \quad (15)
\end{aligned}$$

which tends to 0 exponentially fast as $N \rightarrow \infty$. If $p > 1$ and $\Lambda \notin \mathcal{S}$, it follows from formula (13) that we need an extra summand for $r_1^N(T, K)$, namely (15) with $p = 1$.

3 Truncation of Small Jumps

In this section we approximate Eq. (1) by cutting off the small jumps of Λ . To this end, we first define for each $N \in \mathbb{N}$

$$G^N(t, x; s, y) := G(t, x; s, y) \mathbb{1}_{U^N}(y), \quad (t, x), (s, y) \in \mathbb{R}_+ \times \mathbb{R}^d, \quad (16)$$

where the meaning of the sets U^N is explained in A7. Furthermore, we introduce

$$r_2^N := \left(\int_{[-\varepsilon^N, \varepsilon^N]} |z|^p \pi(dz) \right)^{1/p^*}, \quad r_3^N := \left| \int_{[-\varepsilon^N, \varepsilon^N]} z \mathbb{1}_{\{p > 1, \Lambda \notin \mathcal{S}\}} \pi(dz) \right|, \quad (17)$$

where $(\varepsilon^N)_{N \in \mathbb{N}} \subseteq (0, 1)$ satisfies $\varepsilon^N \rightarrow 0$ as $N \rightarrow \infty$. Condition A4 implies that $r_2^N, r_3^N \rightarrow 0$ as $N \rightarrow \infty$. Next, defining truncations of the Lévy basis Λ by

$$\Lambda^N(dt, dx) := \int_{[-\varepsilon^N, \varepsilon^N]^c} z \mu(dt, dx, dz), \quad (18)$$

our approximation scheme for the solution Y to (1) is given as:

$$Y^N(t, x) := Y_0(t, x) + \int_0^t \int_{\mathbb{R}^d} G^N(t, x; s, y) \sigma(Y^N(s, y)) \Lambda^N(ds, dy) \quad (19)$$

for $(t, x) \in \mathbb{R}_+ \times \mathbb{R}^d$. Indeed, Y^N can be simulated exactly because for all $T \in \mathbb{R}_+$ the truncation Λ^N only has a finite (random) number $R^N(T)$ of jumps on $[0, T] \times U^N$, say at the space–time locations (τ_i^N, ξ_i^N) with sizes J_i^N . This implies that we have the following alternative representation of $Y^N(t, x)$ for $(t, x) \in [0, T] \times \mathbb{R}^d$:

$$Y^N(t, x) = Y_0(t, x) + \sum_{i=1}^{R^N(T)} G(t, x; \tau_i^N, \xi_i^N) \sigma(Y^N(\tau_i^N, \xi_i^N)) J_i^N \mathbb{1}_{\{\tau_i^N < t\}}. \quad (20)$$

What remains to do is to simulate $Y^N(\tau_i^N, \xi_i^N)$, $i = 1, \dots, R^N(T)$, iteratively, from which the values $Y(t, x)$ for all other $(t, x) \in [0, T] \times \mathbb{R}^d$ can be computed.

The following algorithm summarizes up the simulation procedure:

Algorithm 1 Consider a finite grid $\mathcal{G} \subseteq [0, T] \times \mathbb{R}^d$. For each N proceed as follows:

1. Draw a Poisson random variable $R^N(T)$ with parameter

$$R^N(T) := \int_0^T \int_{U^N} \int_{\mathbb{R}} \mathbb{1}_{\{z \in [-\varepsilon^N, \varepsilon^N]^c\}} \nu(dt, dx, dz) = T \text{Leb}(U^N) \pi([- \varepsilon^N, \varepsilon^N]^c).$$

2. For $i = 1, \dots, R^N(T)$:
 - a. Draw a pair (τ_i^N, ξ_i^N) with uniform distribution from $[0, T] \times U^N$.
 - b. Draw J_i^N from $[-\varepsilon^N, \varepsilon^N]^c$ with distribution $\pi/\pi([-\varepsilon^N, \varepsilon^N]^c)$.
 - c. Rename $(\tau_i^N, \xi_i^N, J_i^N: i = 1, \dots, R^N(T))$ by sorting $(\tau_i^N: i = 1, \dots, R^N(T))$ in increasing order.
3. For each $i = 1, \dots, R^N(T)$ and $(t, x) \in \mathcal{G}$ simulate $Y_0(\tau_i^N, \xi_i^N)$ and $Y_0(t, x)$.
4. For each $i = 1, \dots, R^N(T)$ set

$$Y^N(\tau_i^N, \xi_i^N) := Y_0(\tau_i^N, \xi_i^N) + \sum_{j=1}^{i-1} G(\tau_i^N, \xi_i^N; \tau_j^N, \xi_j^N) \sigma(Y^N(\tau_j^N, \xi_j^N)) J_j^N.$$

5. For each $(t, x) \in \mathcal{G}$ define $Y^N(t, x)$ via (20).

The next theorem determines the convergence behaviour of (19) to (1).

Theorem 1 *Grant assumptions A1–A7 under which the SVE (1) has a unique solution in B_{loc}^p . Then Y^N as defined in (19) belongs to B_{loc}^p for all $N \in \mathbb{N}$, and for all $T \in \mathbb{R}_+$ and compact sets $K \subseteq \mathbb{R}^d$ there exists a constant $C(T) \in \mathbb{R}_+$ independent of N and K such that*

$$\sup_{(t,x) \in [0,T] \times K} \|Y(t, x) - Y^N(t, x)\|_{L^p} \leq C(T)(r_1^N(T, K) + r_2^N + r_3^N). \quad (21)$$

Furthermore, if $\sum_{N=1}^{\infty} (r_1^N(T, K) + r_2^N + r_3^N)^{p^*} < \infty$ is fulfilled, then we also have for all $(t, x) \in [0, T] \times K$ that $Y^N(t, x) \rightarrow Y(t, x)$ a.s. as $N \rightarrow \infty$.

Proof It is obvious that $|G^N| \leq |G|$ pointwise and that we have $\nu^N \leq \nu$ for the third characteristic ν^N of Λ^N . Thus, A1–A6 are still satisfied when G and ν are replaced by G^N and ν^N (if $\Lambda \in \mathcal{S}$, also $\Lambda^N \in \mathcal{S}$). So Y^N as a solution to (1) with G^N and Λ^N instead of G and Λ belongs to B_{loc}^p as well. Moreover, for all $T \in \mathbb{R}_+$ there exists $C(T) \in \mathbb{R}_+$ independent of $N \in \mathbb{N}$ such that

$$\sup_{(t,x) \in [0,T] \times \mathbb{R}^d} \|Y^N(t, x)\|_{L^p} \leq C(T), \quad N \in \mathbb{N}. \quad (22)$$

We only sketch the proof for this statement. In fact, using [5, Lemma 6.1(1)] it can be shown that the left-hand side of (22) satisfies an inequality of the same type as in Lemma 6.4(3) of the same paper. In particular, it is bounded by a constant $C^N(T)$ that depends on N only through $|G^N|$ and ν^N , and that this constant is only increased if we replace $|G^N|$ and ν^N by the larger $|G|$ and ν . In this way, we obtain an upper bound $C(T)$ that does not depend on N .

Next, we prove the convergence of Y^N to Y as stated in (21). We have

$$\begin{aligned}
 Y(t, x) - Y^N(t, x) &= \int_0^t \int_{\mathbb{R}^d} [G(t, x; s, y) - G^N(t, x; s, y)] \sigma(Y(s, y)) \Lambda(ds, dy) \\
 &\quad + \int_0^t \int_{\mathbb{R}^d} G^N(t, x; s, y) [\sigma(Y(s, y)) - \sigma(Y^N(s, y))] \Lambda(ds, dy) \\
 &\quad + \int_0^t \int_{\mathbb{R}^d} G^N(t, x; s, y) \sigma(Y^N(s, y)) (\Lambda - \Lambda^N)(ds, dy) \\
 &=: I_1^N(t, x) + I_2^N(t, x) + I_3^N(t, x), \quad (t, x) \in \mathbb{R}_+ \times \mathbb{R}^d. \quad (23)
 \end{aligned}$$

If $p > 1$, we have by (8), Hölder's inequality and the Burkholder-Davis-Gundy-inequality

$$\begin{aligned}
 \|I_2^N(t, x)\|_{L^p} &\leq \left\| \int_0^t \int_{\mathbb{R}^d} G^N(t, x; s, y) [\sigma(Y(s, y)) - \sigma(Y^N(s, y))] B_1(ds, dy) \right\|_{L^p} \\
 &\quad + \left\| \int_0^t \int_{\mathbb{R}^d} G^N(t, x; s, y) [\sigma(Y(s, y)) - \sigma(Y^N(s, y))] M(ds, dy) \right\|_{L^p} \\
 &\leq C \left(\left(\int_0^t \int_{\mathbb{R}^d} |G(t, x; s, y)| |B_1|(ds, dy) \right)^{p-1} \right. \\
 &\quad \times \left. \int_0^t \int_{\mathbb{R}^d} |G(t, x; s, y)| \|Y(s, y) - Y^N(s, y)\|_{L^p}^p |B_1|(ds, dy) \right)^{1/p} \\
 &\quad + C \left(\int_0^t \int_{\mathbb{R}^d} |G(t, x; s, y)|^p \|Y(s, y) - Y^N(s, y)\|_{L^p}^p d(s, y) \right)^{1/p}. \quad (24)
 \end{aligned}$$

If $p \in (0, 1]$, we have $\Lambda \in \mathcal{V}_0$ by (6) and (9), and thus Jensen's inequality gives

$$\begin{aligned}
 \|I_2^N(t, x)\|_{L^p} &= \mathbb{E} \left[\left(\int_0^t \int_{\mathbb{R}^d} G^N(t, x; s, y) [\sigma(Y(s, y)) - \sigma(Y^N(s, y))] z \mu(ds, dy, dz) \right)^p \right] \\
 &\leq \mathbb{E} \left[\int_0^t \int_{\mathbb{R}^d} |G^N(t, x; s, y)|^p |\sigma(Y(s, y)) - \sigma(Y^N(s, y))|^p |z|^p \nu(ds, dy, dz) \right] \\
 &\leq C \int_0^t \int_{\mathbb{R}^d} |G(t, x; s, y)|^p \|Y(s, y) - Y^N(s, y)\|_{L^p}^p d(s, y). \quad (25)
 \end{aligned}$$

Inserting (24) and (25) back into (23), we have for $w^N(t, x) := \|Y(t, x) - Y^N(t, x)\|_{L^p}$

$$\begin{aligned}
 w^N(t, x) &\leq C(T) \left(\left(\int_0^t \int_{\mathbb{R}^d} |G(t, x; s, y)| \mathbb{1}_{\{p>1, \Lambda \notin \mathcal{S}\}} (w^N(s, y))^p d(s, y) \right)^{1/p} \right. \\
 &\quad + \left. \left(\int_0^t \int_{\mathbb{R}^d} |G(t, x; s, y)|^p (w^N(s, y))^{p^*} d(s, y) \right)^{1/p^*} \right) \\
 &\quad + \|I_1^N(t, x) + I_3^N(t, x)\|_{L^p}, \quad (t, x) \in [0, T] \times \mathbb{R}^d.
 \end{aligned}$$

By a Gronwall-type estimate, which is possible because of A5 (see the proof of [5, Theorem 4.7(3)] for an elaboration of an argument of this type), we conclude

$$\sup_{(t,x) \in [0,T] \times K} w^N(t, x) \leq C(T) \sup_{(t,x) \in [0,T] \times K} \|I_1^N(t, x) + I_3^N(t, x)\|_{L^p} .$$

where $C(T)$ does not depend on K because of (11). For $I_1^N(t, x)$ we have for $p > 1$

$$\begin{aligned} \|I_1^N(t, x)\|_{L^p} &\leq \left\| \int_0^t \int_{\mathbb{R}^d} [G(t, x; s, y) - G^N(t, x; s, y)] \sigma(Y(s, y)) B_1(ds, dy) \right\|_{L^p} \\ &\quad + \left\| \int_0^t \int_{\mathbb{R}^d} [G(t, x; s, y) - G^N(t, x; s, y)] \sigma(Y(s, y)) M(ds, dy) \right\|_{L^p} \\ &\leq C \left(1 + \sup_{(t,x) \in [0,T] \times \mathbb{R}^d} \|Y(t, x)\|_{L^p} \right) \left(\int_0^t \int_{(U^N)^c} |G(t, x; s, y)| |B_1|(ds, dy) \right. \\ &\quad \left. + \left(\int_0^t \int_{(U^N)^c} |G(t, x; s, y) z|^p \nu(ds, dy, dz) \right)^{1/p} \right) \\ &\leq C(T) r_1^N(T, K) , \end{aligned} \quad (26)$$

uniformly in $(t, x) \in [0, T] \times K$. In similar fashion one proves the estimate (26) for $p \in (0, 1]$, perhaps with a different $C(T)$. Next, when $p > 1$, (22) implies

$$\begin{aligned} \|I_3^N(t, x)\|_{L^p} &= \left\| \int_0^t \int_{\mathbb{R}^d} \int_{[-\varepsilon^N, \varepsilon^N]} G^N(t, x; s, y) \sigma(Y^N(s, y)) z (\mu - \nu)(ds, dy, dz) \right\|_{L^p} \\ &\quad + \left\| \int_0^t \int_{\mathbb{R}^d} \int_{[-\varepsilon^N, \varepsilon^N]} G^N(t, x; s, y) \sigma(Y^N(s, y)) z \mathbb{1}_{\{A \notin \mathcal{S}\}} \nu(ds, dy, dz) \right\|_{L^p} \\ &\leq C(T) \left(\left(\int_0^t \int_{\mathbb{R}^d} \int_{[-\varepsilon^N, \varepsilon^N]} |G(t, x; s, y) z|^p \pi(dz) d(s, y) \right)^{1/p} \right. \\ &\quad \left. + \left| \int_{[-\varepsilon^N, \varepsilon^N]} z \mathbb{1}_{\{A \notin \mathcal{S}\}} \pi(dz) \right| \int_0^t \int_{\mathbb{R}^d} |G(t, x; s, y)| \mathbb{1}_{\{A \notin \mathcal{S}\}} d(s, y) \right) \\ &\leq C(T) (r_2^N + r_3^N) . \end{aligned}$$

The case $p \in (0, 1]$ can be treated similarly, cf. the estimation of $I_2^N(t, x)$ above.

It remains to prove that for each $(t, x) \in [0, T] \times K$ the convergence of $Y^N(t, x)$ to $Y(t, x)$ is almost sure when $r_1^N(T, K)$, r_2^N and r_3^N are p^* -summable. To this end, choose an arbitrary sequence $(a_N)_{N \in \mathbb{N}} \subseteq (0, 1)$ converging to 0 such that

$$\sum_{N=1}^{\infty} A_N < \infty \quad \text{with} \quad A_N := \frac{(r_1^N(T, K) + r_2^N + r_3^N)^{p^*}}{a_N^p} .$$

Such a sequence always exists, see [13, Theorem 175.4], for example. So by (21) and Chebyshev's inequality we derive

$$\mathbb{P} \left[|Y(t, x) - Y^N(t, x)| \geq a_N \right] \leq \frac{\|Y(t, x) - Y^N(t, x)\|_{L^p}^{p^*}}{a_N^p} \leq C(T) A_N .$$

Our assertion now follows from the Borel-Cantelli lemma. $\square$

Example 2 The rates r_2^N and r_3^N from (17) only depend on the underlying Lévy measure π . Let $p, q \in (0, 2]$ with $q < p$, and assume that $\int_{[-1,1]} |z|^q \pi(dz) < \infty$. If $\Lambda \in \mathcal{V}_0$, assume that $q < 1$. Then

$$\begin{aligned} r_2^N &= \left(\int_{[-\varepsilon^N, \varepsilon^N]} |z|^p \pi(dz) \right)^{1/p^*} \leq \left(\int_{[-1,1]} |z|^q \nu(dz) (\varepsilon^N)^{p-q} \right)^{1/p^*} \\ &= \mathcal{O} \left((\varepsilon^N)^{(p-q)/p^*} \right) , \\ r_3^N &= \left| \int_{[-\varepsilon^N, \varepsilon^N]} z \mathbb{1}_{\{p>1, \Lambda \notin \mathcal{S}\}} \pi(dz) \right| \leq \mathcal{O} \left((\varepsilon^N)^{1-q} \mathbb{1}_{\{p>1, \Lambda \notin \mathcal{S}\}} \right) . \end{aligned}$$

For instance, if $\varepsilon^N = 1/N^k$, then the sequence $(r_2^N)^{p^*} = \mathcal{O}(N^{-k(p-q)})$ is summable for all $k > (p-q)^{-1}$. So in order to obtain a.s. convergence of $Y^N(t, x) \rightarrow Y(t, x)$, a sufficient condition is to choose the truncation rates ε^N small enough. Similar conclusions are valid for the other two rates $r_1^N(T, K)$ and r_3^N .

4 Truncation via Series Representations

Another approach to the simulation of (1) uses series representations for the Lévy basis. The idea, going back to [16, 17] and already applied to the simulation of Lévy processes [18], is to choose the jumps to be simulated in a random order. Instead of selecting the big jumps first and the smaller jumps later as in Sect. 3, we only choose the big jumps first more *likely*. The details are as follows: we fix a finite time horizon $T \in \mathbb{R}_+$ and, recalling A7, a partition $(Q^i)_{i \in \mathbb{N}}$ of $\mathbb{R}^d$ into pairwise disjoint compact sets such that $U^N = \bigcup_{i=1}^N Q^i$. We now assume that the jump measure μ of Λ on the strip $[0, T] \times \mathbb{R}^d \times \mathbb{R}$ can be represented in the form

$$\begin{aligned} \mu(dt, dx, dz) &= \sum_{i=1}^{\infty} \mu_i(dt, dx, dz) , \\ \mu_i(dt, dx, dz) &= \sum_{j=1}^{\infty} \delta_{(\tau_j^i, \xi_j^i, H(\Gamma_j^i, V_j^i))} (dt, dx, dz) \quad \text{a.s.} , \end{aligned} \tag{27}$$

where $H: (0, \infty) \times \mathbb{R} \rightarrow \mathbb{R}$ is a measurable function, satisfying $H(\cdot, v) = -H(\cdot, -v)$ for all $v \in \mathbb{R}$ when $\Lambda \in \mathcal{S}$, and the random variables involved have the following properties for each $i \in \mathbb{N}$:

- $(\tau_j^i : j \in \mathbb{N})$ and $(\xi_j^i : j \in \mathbb{N})$ are i.i.d. sequences with uniform distribution on $[0, T]$ and Q^i , respectively.
- $(\Gamma_j^i : j \in \mathbb{N})$ is a random walk whose increments are exponentially distributed with mean $1/T$.
- $(V_j^i : j \in \mathbb{N})$ is an i.i.d. sequence with distribution F on $\mathbb{R}$, which we should be able to simulate from. We assume that F is symmetric when $\Lambda \in \mathcal{S}$.
- The sequences τ^i, ξ^i, Γ^i and V^i are independent from each other.
- $(\tau^i, \xi^i, \Gamma^i, V^i)$ is independent from $(\tau^k, \xi^k, \Gamma^k, V^k : k \neq i)$.

Because of (6), μ can always be written in the form (27) whenever the underlying stochastic basis is rich enough. We give three examples of such series representations.

Example 3 The proofs that the following choices are valid can be found in [18, Sect. 3], where also more examples are discussed.

1. *LePage's method*: we set $F := (\delta_{-1} + \delta_1)/2$ and $H(r, \pm 1) := \pm \rho^\leftarrow(r, \pm 1)$, where $\rho^\leftarrow(r, \pm 1) := \inf\{x \in (0, \infty) : \pi(\pm[x, \infty)) < r\}$ for $r \in (0, \infty)$.
2. *Bondesson's method*: we assume that $\pi(A) = \int_0^\infty F(A/g(t)) dt$ for $A \in \mathcal{B}(\mathbb{R}^d)$ with some non-increasing $g : \mathbb{R}_+ \rightarrow \mathbb{R}_+$. Then we define $H(r, v) := g(r)v$.
3. *Thinning method*: we choose F in such a way that Q is absolutely continuous with respect to F with density q , and define $H(r, v) := v \mathbb{1}_{\{q(v) \geq r\}}$.

Our approximation scheme is basically the same as in Sect. 3: we define G^N by (16) and Y^N by (19), with the difference that Λ^N on $[0, T] \times \mathbb{R}^d$ is now defined as

$$\begin{aligned} \Lambda^N(dt, dx) &:= \int_{\mathbb{R}} z \mu^N(dt, dx, dz), \\ \mu^N(dt, dx, dz) &:= \sum_{i=1}^{\infty} \sum_{j: \Gamma_j^i \leq N} \delta_{(\tau_j^i, \xi_j^i, H(\Gamma_j^i, V_j^i))}(dt, dx, dz). \end{aligned} \quad (28)$$

We can therefore rewrite $Y^N(t, x)$ for $(t, x) \in [0, T] \times \mathbb{R}^d$ as

$$Y^N(t, x) = Y_0(t, x) + \sum_{i=1}^N \sum_{j: \Gamma_j^i \leq N} G(t, x; \tau_j^i, \xi_j^i) \sigma(Y^N(\tau_j^i, \xi_j^i)) H(\Gamma_j^i, V_j^i) \mathbb{1}_{\{\tau_j^i < t\}}. \quad (29)$$

This yields the following simulation algorithm:

Algorithm 2 Let $\mathcal{G}$ be a finite grid in $[0, T] \times \mathbb{R}^d$ and $N \in \mathbb{N}$.

1. For each $i = 1, \dots, N$ set $j := 1$ and repeat the following:
 - a. Draw E_j^i from an exponential distribution with mean $1/T$.
 - b. Define $\Gamma_j^i := \Gamma_{j-1}^i + E_j^i$ ($\Gamma_0^i := 0$).
 - c. If $\Gamma_j^i > N$, set $J_i := j - 1$ and leave the loop; otherwise set $j := j + 1$.

2. For each $i = 1, \dots, N$ and $j = 1, \dots, J_i$ simulate independently
 - a. a pair (τ_j^i, ξ_j^i) with uniform distribution on $[0, T] \times Q^i$;
 - b. a random variable V_j^i with distribution F ;
 - c. the random variable $Y_0(\tau_j^i, \xi_j^i)$.
3. Sort the sequence $(\tau_j^i : i = 1, \dots, N, j = 1, \dots, J_i)$ in increasing order, yielding sequences $(\tau_i, \xi_i, \Gamma_i, V_i : i = 1, \dots, \sum_{j=1}^N J_j)$. Now define

$$Y^N(\tau_i, \xi_i) := Y_0(\tau_i, \xi_i) + \sum_{j=1}^{i-1} G(\tau_i, \xi_i; \tau_j, \xi_j) \sigma(Y^N(\tau_j, \xi_j)) H(\Gamma_j, V_j) .$$

4. For each $(t, x) \in \mathcal{G}$ simulate $Y_0(t, x)$ and define $Y(t, x)$ by (29).

We can now prove a convergence theorem for Y^N to Y , similar to Theorem 1. Define

$$\begin{aligned} r_2^N &:= \left(\int_N^\infty \int_{\mathbb{R}} |H(r, v)|^p F(dv) dr \right)^{1/p^*} , \\ r_3^N &:= \left| \int_N^\infty \int_{\mathbb{R}} H(r, v) \mathbb{1}_{\{p>1, \Lambda \notin \mathcal{S}\}} F(dv) dr \right| . \end{aligned} \quad (30)$$

Theorem 2 *Grant assumptions A1–A7 under which the SVE (1) has a unique solution in B_{loc}^p . Further suppose that the jump measure μ of Λ has a representation in form of (27). Then Y^N as defined in (29) belongs to B_{loc}^p for all $N \in \mathbb{N}$, and for all $T \in \mathbb{R}_+$ and compact sets $K \subseteq \mathbb{R}^d$ there exists a constant $C(T) \in \mathbb{R}_+$ independent of N and K such that*

$$\sup_{(t,x) \in [0,T] \times K} \|Y(t, x) - Y^N(t, x)\|_{L^p} \leq C(T)(r_1^N(T, K) + r_2^N + r_3^N) . \quad (31)$$

If $\sum_{N=1}^\infty (r_1^N(T, K) + r_2^N + r_3^N)^{p^*} < \infty$, then we also have for all $(t, x) \in [0, T] \times K$ that $Y^N(t, x) \rightarrow Y(t, x)$ a.s. as $N \rightarrow \infty$.

Proof We start with some preliminaries. It follows from (27) and [18, Proposition 2.1] that on $[0, T] \times \mathbb{R}^d \times \mathbb{R}$ we have $v = \bar{v} \circ h^{-1}$ where $\bar{v}(dt, dx, dr, dv) = dt dx dr F(dv)$ and $h(t, x, r, v) = (t, x, H(r, v))$. Therefore, conditions (6) and (9) imply that

$$\int_0^\infty \int_{\mathbb{R}} |H(r, v)|^p F(dv) dr = \int_{\mathbb{R}} |z|^p \pi(dz) < \infty ,$$

$$\text{and } \int_0^\infty \int_{\mathbb{R}} |H(r, v)| \mathbb{1}_{\{p>1, \Lambda \notin \mathcal{S}\}} F(dv) dr = \int_{\mathbb{R}} |z| \mathbb{1}_{\{p>1, \Lambda \notin \mathcal{S}\}} \pi(dz) < \infty .$$

Consequently, r_2^N and r_3^N are well defined and converge to 0 when $N \rightarrow \infty$. Similarly, the compensator ν_N of the measure $\mu - \mu^N$ is given by $\nu_N(dt, dx, dz) = dt dx \pi_N(dz)$, where $\pi_N = (\text{Leb} \otimes F) \circ H_N^{-1}$ and $H_N(r, \nu) = H(r, \nu) \mathbb{1}_{(N, \infty)}(r)$.

For the actual proof of Theorem 2 one can basically follow the proof of Theorem 1. Only the estimation of $I_3^N(t, x)$ as defined in (23) is different, which we shall carry out now. In the case of $p > 1$, we again use the Burkholder-Davis-Gundy inequality and obtain for $(t, x) \in [0, T] \times \mathbb{R}^d$

$$\begin{aligned} \|I_3^N(t, x)\|_{L^p} &= \left\| \int_0^t \int_{\mathbb{R}^d} \int_{\mathbb{R}} G^N(t, x; s, y) \sigma(Y^N(s, y)) z (\mu_N - \nu_N)(ds, dy, dz) \right\|_{L^p} \\ &\quad + \left\| \int_0^t \int_{\mathbb{R}^d} \int_{\mathbb{R}} G^N(t, x; s, y) \sigma(Y^N(s, y)) z \mathbb{1}_{\{\Lambda \notin \mathcal{S}\}} \nu_N(ds, dy, dz) \right\|_{L^p} \\ &\leq C(T) \left(\left(\int_0^t \int_{\mathbb{R}^d} \int_N^\infty \int_{\mathbb{R}} |G(t, x; s, y) H(r, v)|^p F(dv) dr d(s, y) \right)^{1/p} \right. \\ &\quad \left. + \int_0^t \int_{\mathbb{R}^d} |G(t, x; s, y)| \left| \int_N^\infty \int_{\mathbb{R}} H(r, v) \mathbb{1}_{\{\Lambda \notin \mathcal{S}\}} F(dv) dr \right| d(s, y) \right) \\ &\leq C(T)(r_2^N + r_3^N). \end{aligned}$$

The case $p \in (0, 1]$ is treated analogously. One only needs to replace $\mu_N - \nu_N$ by μ_N and estimate via Jensen's inequality. $\square$

Example 4 (Continuation of Example 3) We calculate the rates r_2^N and r_3^N from (30) for the series representations given in Example 3. We assume that $p, q \in (0, 2]$ with $q < p$ are chosen such that $\int_{[-1, 1]} |z|^q \pi(dz) < \infty$, and $q < 1$ if $\Lambda \in \mathcal{V}_0$. For all three examples we use the fact that $\pi = (\text{Leb} \otimes F) \circ H^{-1}$ and that $r > N$ implies $|H(r, v)| \leq |H(N, v)|$ for all $v \in \mathbb{R}$.

1. *LePage's method:* We have

$$\begin{aligned} (r_2^N)^{p*} &= \int_N^\infty \frac{|H(r, 1)|^p + |H(r, -1)|^p}{2} dr \leq \frac{1}{2} \int_{[H(N, -1), H(N, 1)]} |z|^p \pi(dz) \\ &\leq \frac{1}{2} \int_{[H(1, -1), H(1, 1)]} |z|^q \pi(dz) (|H(N, -1)| \vee |H(N, 1)|)^{p-q}, \end{aligned}$$

and therefore

$$\begin{aligned} r_2^N &= \mathcal{O}((\rho^{\leftarrow}(N, 1) \vee \rho^{\leftarrow}(N, -1))^{(p-q)/p*}), \\ r_3^N &= \mathcal{O}((\rho^{\leftarrow}(N, 1) \vee \rho^{\leftarrow}(N, -1))^{1-q} \mathbb{1}_{\{p>1, \Lambda \notin \mathcal{S}\}}). \end{aligned}$$

2. *Bondesson's method:* Since $H(r, v) = g(r)v$ and g is non-increasing, we obtain

$$(r_2^N)^{p*} = \int_N^\infty \int_{\mathbb{R}} |g(r)v|^p F(dv) dr \leq (g(N))^{p-q} \int_0^\infty g^q(r) dr \int_{\mathbb{R}} |v|^p F(dv),$$

and consequently

$$r_2^N = \mathcal{O} \left(g(N)^{(p-q)/p^*} \right), \quad r_3^N = \mathcal{O} \left(g(N)^{1-q} \mathbb{1}_{\{p>1, A \notin \mathcal{S}\}} \right).$$

3. *Thinning method*: Here we have

$$\begin{aligned} r_2^N &= \left(\int_{\mathbb{R}} \int_N^{q(v) \vee N} |v|^p \, dr \, F(dv) \right)^{1/p^*} = \left(\int_{\mathbb{R}} |v|^p \frac{(q(v) - N) \vee 0}{q(v)} \pi(dv) \right)^{1/p^*} \\ &\leq \left(\int_{\mathbb{R}} |z|^p \mathbb{1}_{\{q(v) \geq N\}} \pi(dz) \right)^{1/p^*}, \\ r_3^N &\leq \int_{\mathbb{R}} |z| \mathbb{1}_{\{q(v) \geq N\}} \mathbb{1}_{\{p>1, A \notin \mathcal{S}\}} \pi(dz). \end{aligned}$$

In most situations, there exist $(\varepsilon^N)_{N \in \mathbb{N}} \subseteq \mathbb{R}_+$ with $\varepsilon^N \rightarrow 0$ as $N \rightarrow \infty$ such that $\{q(v) \geq N\} \subseteq [-\varepsilon^N, \varepsilon^N]$. In this case, one can apply the estimates in Example 2.

5 Simulation Study

In this section we visualize the sample path behaviour of the stochastic heat equation from Example 2 via a simulation study, using MATLAB programs from [4]. We take Λ to be a Lévy basis without drift, whose Lévy measure π is that of a gamma process, i.e.

$$\pi(dz) = \gamma z^{-1} \exp(-\lambda z) \mathbb{1}_{\{z>0\}} dz$$

with two parameters $\gamma, \lambda > 0$. In the figures below their values are always $\gamma = 10$ and $\lambda = 0.1$. Furthermore, we set $Y_0 \equiv 0$ and $\sigma \equiv 1$. Especially the latter choice simplifies the subsequent discussion a lot, but none of the issues we address below relies on this assumption. Thus, the process we would like to simulate is

$$Y(t, x) = \int_0^t \int_{\mathbb{R}^d} g(t-s, x-y) \Lambda(ds, dy), \quad (t, x) \in \mathbb{R}_+ \times \mathbb{R}^d, \quad (32)$$

with g being the heat kernel given in (14). In order to understand the path properties of Y , it is important to notice that g is smooth on the whole $\mathbb{R}_+ \times \mathbb{R}^d$ except at the origin where it explodes. More precisely, for every $t \in (0, \infty)$ the function $x \mapsto g(t, x)$ is the Gaussian density with mean 0 and variance $2t$, which is smooth and assumes its maximum at 0. Also, for every $x \neq 0$, the function $t \mapsto g(t, x)$ is smooth (also at $t = 0$), with maximum at $t = \|x\|^2/(2d)$. However, if $x = 0$, then $g(t, 0) = (4\pi t)^{-d/2}$ has a singularity at $t = 0$.

These analytical properties have direct consequences on the sample paths of Y . When Λ is of compound Poisson type, that is, has only finitely many atoms on compact sets, it can be readily seen from (32) that the evolution of Y after a jump J

at (τ, ξ) follows the shape of the heat kernel until a next jump arrives. In particular, for $x = \xi$, $Y(t, x)$ jumps to infinity at τ , and decays in t like $J(4\pi(t - \tau))^{-d/2}$ afterwards. But for every $x \neq \xi$, the evolution $t \mapsto Y(t, x)$ is *smooth* at $t = \tau$. In fact, it first starts to increase until $t = \tau + \|x - \xi\|^2/(2d)$ and then decays again. As a consequence, in space dimension 1, the space–time plot of Y shows a basically smoothly evolving path, only interrupted with slim poles at the jump locations of Λ ; see the case $N = 50$ in Fig. 1. However, when Λ has infinite activity, that is, has infinitely many jumps on any non-empty open set, then it is known from [16, Theorem 4] that on any such set Y is unbounded, at least with positive probability. Therefore, the space–time plots of the approximations of Y with finitely many jumps must be treated with caution: in the limiting situation, no smooth area exists any more, but there will be a dense subset of singularities on the plane, which is in line with Fig. 1.

Another interesting observation, however, is the following: if we consider a countable number of x - or t -sections of Y (for $x \in \mathbb{R}^d$, the x -section of Y is given by the function $t \mapsto Y(t, x)$; for $t \in \mathbb{R}_+$, the t -section of Y is the function $x \mapsto Y(t, x)$), then it is shown in [19, Sect. 2] that these are continuous with probability one. Intuitively, this is possible because a.s. the sections never hit a jump (although they are arbitrarily close). For instance, Figs. 2 and 3 show t -sections of a realization of (32) in one, respectively two space dimensions. So as long as we only take countably many “measurements”, we do not observe the space–time singularities of Y but only its relatively regular sections. In theory, this also includes the x -sections of the process Y . But if we plot them for one space dimension as in Fig. 4, one would conjecture from the simulation that they exhibit jumps in time. However, this is *not* true: the jump-like appearance of the x -sections are due to the fact that $g(\cdot, x)$ resembles a

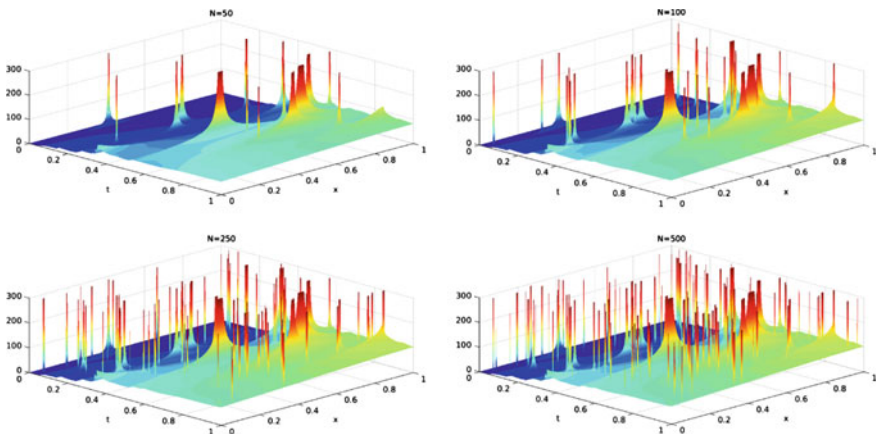


Fig. 1 Successive approximations of Y as given in (32) via Bondesson’s method in dimension 1 for $(t, x) \in [0, 1] \times [0, 1]$ with $N \in \{50, 100, 250, 500\}$ jumps in the region $[0, 1] \times [-1, 2]$

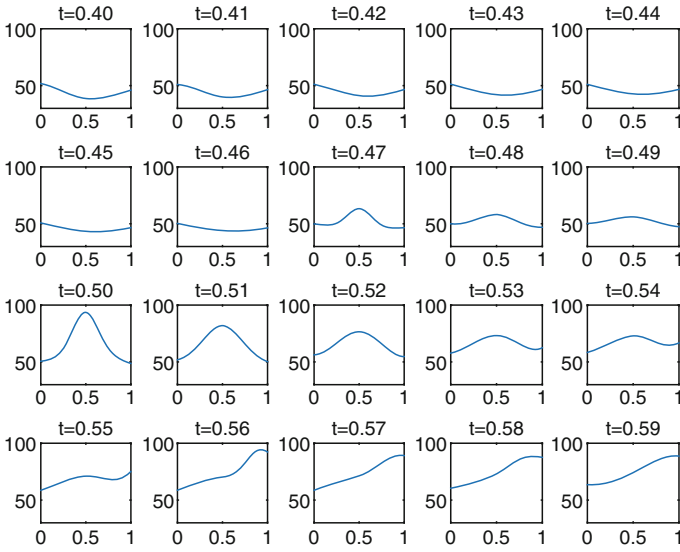


Fig. 2 Several t -sections of the realization of Y shown in Fig. 1 with $N = 500$

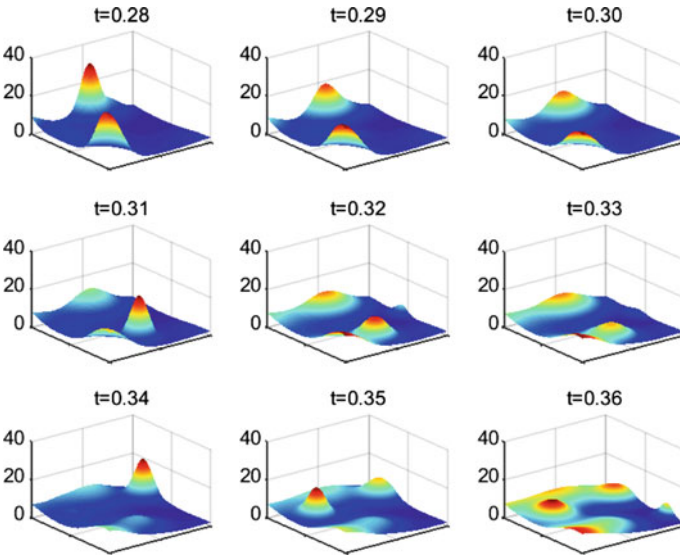


Fig. 3 Several t -sections in the region $[-1, 1]^2$ of a realization of Y in dimension 2 by Bondesson's method with $N = 500$ jumps within $[0, 1] \times [-2, 2]^2$

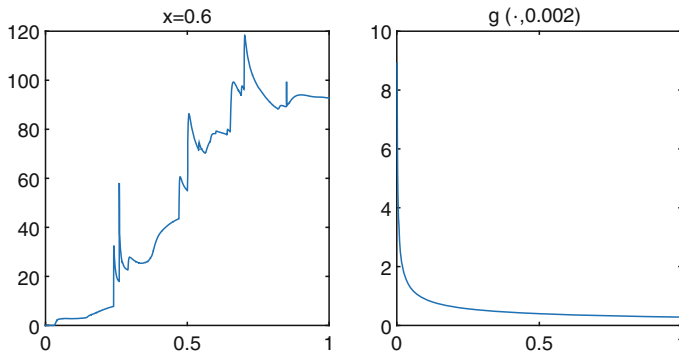


Fig. 4 The x -section of the realization of Y as in Fig. 1 with $N = 500$ at $x = 0.6$ and the heat kernel $g(\cdot, x)$ at $x = 0.002$

discontinuous function at $t = 0$ for small x . Of course, it follows right from the definition (14) that all x -sections of g are smooth everywhere.

Remark 1 Throughout this paper we have taken the stochastic heat equation as our key example for illustration purposes. Obviously, all theory developed in Sects. 3 and 4 holds under much more general assumptions, namely under A1–A7. However, the path behaviour of SVEs as examined through a simulation in this section heavily depends on the chosen kernel. While the path properties above are typical of kernels that originate from parabolic type SPDEs, or more generally, that are smooth everywhere except for an explosive singularity at the origin, a completely different picture arises when we consider bounded kernels. We refer to [2] and the references therein for examples of such kernels in different fields of application. In the case of a simple Ornstein-Uhlenbeck type kernel, a simulation study is carried out in [4, Chap. 5].

6 Conclusion and Outlook

In Sects. 3 and 4 we have presented two simulation algorithms for the SVE (1): Algorithms 1 and 2. In Theorems 1 and 2 we have determined the rate of convergence of the approximations Y^N to Y in the L^p -sense. If these rates are small enough, we have also proved a.s. convergence. Although the theoretical analysis of both schemes lead to quite similar results regarding their convergence behaviour, there are important differences which will decide on whether the one or the other method is preferable in concrete situations. For the first method of truncating the small jumps to work, one must be able to efficiently simulate from the truncated Lévy measure $\pi/\pi([- \varepsilon, \varepsilon]^c)$ for small ε . For the second method, which relies on series representations, the main challenge is to choose H and F in a way such that H is explicitly known and F can be easily simulated from. For instance, if one uses LePage’s method (see Example 3), then $F = (\delta_{-1} + \delta_1)/2$ is easily simulated, but for

H , which is given by the generalized inverse tails of the underlying Lévy measure, maybe no tractable expression exists.

Finally, let us comment on further generalizations of our results. Throughout this paper, we have assumed that the driving noise Λ is a homogeneous Lévy basis, i.e. satisfies (5). In fact, we have introduced this condition only for the sake of simplicity: with a straightforward adjustment, all results obtained in this paper also hold for time- and space-varying (but deterministic) characteristics. Another issue is the finite time perspective which we have taken up for our analysis. An interesting question would be under which conditions (1) has a stationary solution, and in this case, whether one can simulate from it. Sufficient conditions for the existence and uniqueness of stationary solutions to (1) are determined in [5, Theorem 4.8]. Under these conditions, the methods used to derive Theorems 1 and 2 can indeed be extended to the case of infinite time horizon. We leave the details to the reader at this point.

At last, also the hypothesis that Λ is of pure-jump type can be weakened. If Λ has an additional drift (including the case where Λ has locally infinite variation and is not symmetric) but still no Gaussian part, the approximations Y^N in (19) or (29) will contain a further term that is a Volterra integral with respect to the Lebesgue measure. So each time in between two simulated jumps, a deterministic Volterra equation has to be solved numerically, which boils down to a deterministic PDE in the case where G comes from an SPDE. For this subject, there exists a huge literature, which is, of course, also very different to the stochastic case as considered above. If Λ also contains a Gaussian part, then one has to apply techniques from the papers cited in Sect. 1 and ours simultaneously. We content ourselves with referring to [22], who numerically analyzes a Volterra equation driven by a drift plus a Brownian motion. Finally, let us remark that if $p = 2$ (in particular, G must be square-integrable), it is possible for some Lévy bases to improve the results of Sect. 3 if we do not neglect the small jumps completely but approximate them via a Gaussian noise with the same variance, cf. [1] in the case of Lévy processes.

Acknowledgments We take pleasure in thanking Jean Jacod for his valuable advice on this subject. The second author acknowledges support from the Studienstiftung des deutschen Volkes and the graduate programme TopMath at Technische Universität München.

References

1. Asmussen, S., Rosiński, J.: Approximations of small jumps of Lévy processes with a view towards simulation. *J. Appl. Probab.* **38**(2), 482–493 (2001)
2. Barndorff-Nielsen, O.E., Benth, F.E., Veraart, A.E.D.: Ambit processes and stochastic partial differential equations. In: Nunno, G.D., Øksendal, B. (eds.) *Advanced Mathematical Methods for Finance*, pp. 35–74. Springer, Berlin (2011)
3. Barth, A., Lang, A.: Simulation of stochastic partial differential equations using finite element methods. *Stochastics* **84**(2–3), 217–231 (2012)
4. Chen, B.: Simulation of stochastic partial differential equations. Master's thesis, Technische Universität München (2014). <https://mediatum.ub.tum.de/node?id=1225387>

5. Chong, C.: Lévy-driven Volterra equations in space–time (2014). Preprint under [arXiv:1407.8092](https://arxiv.org/abs/1407.8092) [math.PR]
6. Chong, C., Klüppelberg, C.: Integrability conditions for space–time stochastic integrals: theory and applications. *Bernoulli* **21**(4), 2190–2216 (2015)
7. Davie, A.M., Gaines, J.G.: Convergence of numerical schemes for the solution of parabolic stochastic partial differential equations. *Math. Comput.* **70**(233), 121–134 (2001)
8. Dunst, T., Hausenblas, E., Prohl, A.: Approximate Euler method for parabolic stochastic partial differential equations driven by space-time Lévy noise. *SIAM J. Numer. Anal.* **50**(6), 2873–2896 (2012)
9. Gyöngy, I.: Lattice approximations for stochastic quasi-linear parabolic partial differential equations driven by space-time white noise II. *Potential Anal.* **11**(1), 1–37 (1999)
10. Hausenblas, E.: Finite element approximation of stochastic partial differential equations driven by Poisson random measures of jump type. *SIAM J. Numer. Anal.* **46**(1), 437–471 (2008)
11. Hausenblas, E., Marchis, I.: A numerical approximation of parabolic stochastic differential equations driven by a Poisson random measure. *BIT Numer. Math.* **46**(4), 773–811 (2006)
12. Jacod, J., Shiryaev, A.N.: *Limit Theorems for Stochastic Processes*, 2nd edn. Springer, Berlin (2003)
13. Knopp, K.: *Theory and Application of Infinite Series*. Dover, New York (1990)
14. Protter, P.: Volterra equations driven by semimartingales. *Ann. Probab.* **13**(2), 519–530 (1985)
15. Rajput, B.S., Rosiński, J.: Spectral representations of infinitely divisible processes. *Probab. Theory Relat. Fields* **82**(3), 451–487 (1989)
16. Rosiński, J.: On path properties of certain infinitely divisible processes. *Stoch. Process. Appl.* **33**(1), 73–87 (1989)
17. Rosiński, J.: On series representations of infinitely divisible random vectors. *Ann. Probab.* **18**(1), 405–430 (1990)
18. Rosiński, J.: Series representations of Lévy processes from the perspective of point processes. In: Barndorff-Nielsen, O.E., Mikosch, T., Resnick, S.I. (eds.) *Lévy Processes*, pp. 401–415. Birkhäuser, Boston (2001)
19. Saint Loubert Bié, E.: Étude d’une EDPS conduite par un bruit poissonnien. *Probab. Theory Relat. Fields* **111**(2), 287–321 (1998)
20. Walsh, J.B.: An introduction to stochastic partial differential equations. In: Hennequin, P.L. (ed.) *École d’Été de Probabilités de Saint Flour XIV - 1984. Lecture Notes in Mathematics*, vol. 1180, pp. 265–439. Springer, Berlin (1986)
21. Walsh, J.B.: Finite element methods for parabolic stochastic PDE’s. *Potential Anal.* **23**(1), 1–43 (2005)
22. Zhang, X.: Euler schemes and large deviations for stochastic Volterra equations with singular kernels. *J. Differ. Equ.* **244**(9), 2226–2250 (2008)

On the Process of the Eigenvalues of a Hermitian Lévy process

Victor Pérez-Abreu and Alfonso Rocha-Arteaga

Abstract The dynamics of the eigenvalues (semimartingales) of a Lévy process X with values in Hermitian matrices is described in terms of Itô stochastic differential equations with jumps. This generalizes the well known Dyson-Brownian motion. The simultaneity of the jumps of the eigenvalues of X is also studied. If X has a jump at time t two different situations are considered, depending on the commutativity of $X(t)$ and $X(t-)$. In the commutative case all the eigenvalues jump at time t only when the jump of X is of full rank. In the noncommutative case, X jumps at time t if and only if all the eigenvalues jump at that time when the jump of X is of rank one.

Keywords Dyson–Brownian motion · Infinitely divisible random matrix · Bercovici–Pata bijection · Matrix semimartingale · Simultaneous jumps · Non-colliding process · Rank one perturbation · Stochastic differential equation with jumps

1 Introduction

This paper is on a topic that meets a number of subjects where Ole Barndorff-Nielsen has contributed. One is Ole's interests in the study of infinitely divisible random matrix models and their associated matrix-valued processes, like subordinators, Lévy and Ornstein–Uhlenbeck processes [3, 4, 6–8]. Another is the connection between classical and free infinite divisibility, where Ole's papers with Thorbjørnsen [9, 10] drove the study of the so-called Upsilon transformations of classical infinitely divisible distributions [5, 11, 12].

V. Pérez-Abreu (✉)

Departamento de Probabilidad y Estadística, CIMAT, Guanajuato, Mexico
e-mail: pabreu@cimat.mx; pabreuv@gmail.com

A. Rocha-Arteaga

Facultad de Ciencias Físico-Matemáticas,
Universidad Autónoma de Sinaloa, Culiacán, Mexico
e-mail: arteaga@uas.edu.mx

Matrix Lévy processes with jumps of rank one arise as a covariation process of two $\mathbb{C}^d$ -valued Lévy processes [23] and they provide natural examples of matrix subordinators and matrix Lévy processes of bounded variation, such as those considered in [3, 6].

In the context of free probability, Bercovici and Pata [14] introduced a bijection from the set of classical infinitely divisible distributions on $\mathbb{R}$ to the set of free infinitely divisible distributions on $\mathbb{R}$. This bijection was described in terms of ensembles of Hermitian random matrices $\{X_d : d \geq 1\}$ by Benaych-Georges [13] and Cabanal-Duvillard [20]. For fixed $d \geq 1$, the associated $d \times d$ matrix distribution of X_d is invariant under unitary conjugations and it is (matrix) infinitely divisible. The associated Hermitian Lévy process $\{X_d(t) : t \geq 0\}$ has been considered in [23] and it has the property that, in the non pure Gaussian case, the jumps $\Delta X_d(t) = X_d(t) - X_d(t-)$ are $d \times d$ matrices of rank one, see also [22, 29].

The simplest example of this connection is the well-known theorem of Wigner that gives the semicircle distribution (free infinitely divisible) as the asymptotic spectral distribution of $\{X_d : d \geq 1\}$ in the case of the Gaussian Unitary Ensemble (GUE), see [1]. In this case, for each fixed $d \geq 1$, the associated Lévy process of X_d is the $d \times d$ Hermitian Brownian motion $\{B(t) : t \geq 0\} = \{(B_{jk}(t)) : t \geq 0\}$ where $(B_{ij}(t))_{j=1}^d, (\operatorname{Re} B_{jk}(t))_{j < k}, (\operatorname{Im} B_{jk}(t))_{j < k}$ is a set of d^2 independent one-dimensional Brownian motions with parameter $\frac{t}{2}(1 + \delta_{jk})$.

Still in the GUE case, let $\lambda = \{(\lambda_1(t), \lambda_2(t), \dots, \lambda_d(t)) : t \geq 0\}$ be the d -dimensional stochastic process of eigenvalues of B . In a pioneering work, Dyson [24] showed that if the eigenvalues start at different positions $(\lambda_1(0) > \lambda_2(0) > \dots > \lambda_d(0))$ a.s.), then they never meet at any time $(\lambda_1(t) > \lambda_2(t) > \dots > \lambda_d(t) \quad \forall t > 0)$ a.s.) and furthermore they form a diffusion process (and a semimartingale) satisfying the Itô Stochastic Differential Equation (SDE)

$$d\lambda_i(t) = dW_i(t) + \sum_{j \neq i} \frac{dt}{\lambda_i(t) - \lambda_j(t)}, \quad t \geq 0, 1 \leq i \leq d, \quad (1)$$

where $W_1, \dots, W_d$ are independent one-dimensional standard Brownian motions (see [1, Sect. 4.3.1]). The stochastic process $\{\lambda(t) : t \geq 0\}$ is called the Dyson non-colliding Brownian motion corresponding to the GUE, or, in short, Dyson-Brownian motion. The study of the eigenvalue process of other continuous matrix-valued processes has received considerable attention, see [17–19, 25–28]. The eigenvalues of all these processes are strongly dependent and do not collide at any time almost surely.

The aim of this paper is to understand the behavior on time of the eigenvalue process $\lambda = \{(\lambda_1(t), \lambda_2(t), \dots, \lambda_d(t)) : t \geq 0\}$ of a $d \times d$ Hermitian Lévy process $X = \{X(t) : t \geq 0\}$. The first goal is to give conditions for the simultaneity of the jumps of the eigenvalues. In particular, when the jump of X is of rank one and $X(t)$ and $X(t-)$ do not commute, we show that a single eigenvalue jumps at time t if and only if all the eigenvalues jump at that time. This fact which is due to the strong dependence between the eigenvalue processes contrasts with the case of the no simultaneity of

the jumps at any time of d independent real Lévy processes; in Example 1 we give a Hermitian Lévy process with such eigenvalue process. We observe that in general, the eigenvalue process is not a Lévy process (as in the Dyson-Brownian case) but a semimartingale given in terms of the Hermitian Lévy process X .

The second goal is to describe the dynamics of the eigenvalue process, analogously to (1). This leads to first considering the appropriate Itô's formula for matrix-valued semimartingales with jumps, more general than those of bounded variation considered by Barndorff-Nielsen and Stelzer [6].

Our main results and the structure of the paper are as follows. In Sect. 2 we gather several results on Hermitian Lévy processes, including versions of Itô's formula which are useful for studying the corresponding eigenvalue processes. We also point out conditions under which the spectrum of X will be simple for every $t > 0$ and conditions for the differentiability of the eigenvalues. In Sect. 3 we consider the simultaneous jumps of the eigenvalue (semimartingale) process of a $d \times d$ Hermitian Lévy process X . When $\Delta X(t) = X(t) - X(t-)$ is not zero and $X(t)$ and $X(t-)$ commute, we show that the eigenvalues jump simultaneously if and only if the matrix jump $\Delta X(t)$ is of rank d . On the other hand, if $\Delta X(t) \neq 0$ is of rank one and $X(t)$ and $X(t-)$ do not commute, then the process X jumps at time $t > 0$ if and only if all the components of λ jump at that time. This leads to analyzing the relation between the eigenvalues of $X(t)$ and $X(t-)$ when the jump $\Delta X(t)$ is a rank one matrix. A key step in this direction is the recent paper [31] on the spectrum of a rank one perturbation of an unstructured matrix. Extensions to Hermitian semimartingales are also pointed out. Finally, Sect. 4 considers the dynamics of the eigenvalues of a Hermitian Lévy process, extension to the Dyson-Brownian motion, showing that a repulsion force appears in the bounded variation part of the eigenvalues only when there is a Gaussian component of X . One of the main problems to overcome is the fact that even when $X(t)$ is Hermitian and has a simple spectrum in an open subset almost surely, $X(t-)$ may not.

We shall use either $X(t)$ or X_t to denote a stochastic process in a convenient way when the dimension or the entries of a matrix have also to be specified.

2 Preliminaries on Hermitian Lévy Processes

In this section we gather several results on matrix-valued semimartingales and general Hermitian Lévy processes.

Let $\mathbb{M}_d = \mathbb{M}_d(\mathbb{C})$ denote the linear space of $d \times d$ matrices with complex entries with scalar product $\langle A, B \rangle = \text{tr}(B^*A)$ and the Frobenius norm $\|A\| = [\text{tr}(A^*A)]^{1/2}$ where tr denotes the (non-normalized) trace. The set of Hermitian random matrices in $\mathbb{M}_d$ is denoted by $\mathbb{H}_d$, $\mathbb{H}_d^0 = \mathbb{H}_d \setminus \{0\}$ and $\mathbb{H}_d^1$ is the set of rank one matrices in $\mathbb{H}_d$.

We will use the identification of a $d \times d$ Hermitian matrix $X = (x_{ij} + iy_{ij})_{1 \leq i, j \leq d}$ with an element in $\mathbb{R}^{d^2}$, that is,

$$X = (x_{ij} + iy_{ij})_{1 \leq i, j \leq d} \leftrightarrow \widehat{X} := \left(x_{11}, \dots, x_{dd}, (x_{ij}, y_{ij})_{1 \leq i < j \leq d} \right)^\top. \quad (2)$$

Thus, the set $\mathbb{H}_d$ is identified with $\mathbb{R}^{d^2}$ and one has that for $X, Y \in \mathbb{H}_d$, $\text{tr}(XY) = \widehat{Y}^\top \widehat{X}$ is real.

An $\mathbb{M}_d$ -valued process $X = \{(x_{ij})(t) : t \geq 0\}$ is a matrix semimartingale if $x_{ij}(t)$ is a complex semimartingale for each $i, j = 1, \dots, d$. Let $X = \{(x_{ij})(t) : t \geq 0\} \in \mathbb{H}_d$ and $Y = \{(y_{ij})(t) : t \geq 0\} \in \mathbb{H}_d$ be semimartingales. Similar to the case of matrices with real entries in [6], the matrix covariation of X and Y is defined as the $\mathbb{H}_d$ -valued process $[X, Y] := \{[X, Y](t) : t \geq 0\}$ with entries

$$[X, Y]_{ij}(t) = \sum_{k=1}^d [x_{ik}, y_{kj}](t), \quad (3)$$

where $[x_{ik}, y_{kj}](t)$ is the covariation of the $\mathbb{C}$ -valued semimartingales $\{x_{ik}(t) : t \geq 0\}$ and $\{y_{kj}(t) : t \geq 0\}$, see [30, pp. 83]. One has a decomposition into a continuous part and a pure jump part by

$$[X, Y](t) = [X^c, Y^c](t) + \sum_{s \leq t} \Delta X_s \Delta Y_s^*, \quad (4)$$

where $[X^c, Y^c]_{ij}(t) := \sum_{k=1}^d [x_{ik}^c, y_{kj}^c](t)$. We recall that for any semimartingale x , the process x^c is the a.s. unique continuous local martingale m such that $[x - m]$ is purely discontinuous.

Matrix-valued Lévy processes are examples of matrix-valued semimartingales. In particular, so are the Hermitian Lévy processes.

2.1 The Lévy–Khintchine Representation and Lévy–Itô Decomposition

A $d \times d$ matrix Hermitian Lévy process $X = \{X(t) : t \geq 0\}$ is characterized by the Lévy–Khintchine representation of its Fourier transform $\mathbb{E} e^{\text{itr}(\Theta X(t))} = \exp(t\psi(\Theta))$ with Laplace exponent

$$\psi(\Theta) = \text{itr}(\Theta \mathcal{A} \Theta) - \frac{1}{2} \text{tr}(\Theta \mathcal{A} \Theta) + \int_{\mathbb{H}_d} \left(e^{\text{itr}(\Theta \xi)} - 1 - i \frac{\text{tr}(\Theta \xi)}{1 + \|\xi\|^2} \right) \nu(d\xi), \quad \Theta \in \mathbb{H}_d,$$

where $\mathcal{A} : \mathbb{H}_d \rightarrow \mathbb{H}_d$ is a positive symmetric linear operator (i.e., $\text{tr}(\Phi \mathcal{A} \Phi) \geq 0$ for $\Phi \in \mathbb{H}_d$ and $\text{tr}(\Theta_2 \mathcal{A} \Theta_1) = \text{tr}(\Theta_1 \mathcal{A} \Theta_2)$ for $\Theta_1, \Theta_2 \in \mathbb{H}_d$), ν is a measure on

$\mathbb{H}_d$ (the Lévy measure) satisfying $\nu(\{0\}) = 0$ and $\int_{\mathbb{H}_d} (1 \wedge \|x\|^2) \nu(dx) < \infty$, and $\Psi \in \mathbb{H}_d$. The triplet $(\mathcal{A}, \nu, \Psi)$ uniquely determines the distribution of X .

Interesting matrix Gaussian distributions for some examples of $\mathcal{A}$ are the following (here $\Psi = 0$ and $\nu = 0$):

- (i) $\mathcal{A} = \sigma^2 \mathbf{I}_d$, $\sigma^2 > 0$, corresponds to the GUE case of parameter σ^2 ;
- (ii) More generally, $\mathcal{A}\Theta = \Sigma_1 \Theta \Sigma_2$, for Σ_1, Σ_2 nonnegative definite matrices in $\mathbb{H}_d$, corresponds to the matrix Gaussian distribution with Kronecker covariance $\Sigma_1 \otimes \Sigma_2$;
- (iii) $\mathcal{A}\Theta = \text{tr}(\Theta) \sigma^2 \mathbf{I}_d$, $\sigma^2 > 0$, is the covariance operator of the Gaussian random matrix $g\mathbf{I}_d$ where g is a one-dimensional random variable with zero mean and variance σ^2 .

Any $\mathbb{H}_d$ -valued Lévy process $X = \{X(t) : t \geq 0\}$ with triplet $(\mathcal{A}, \nu, \Psi)$ is a semi-martingale with the Lévy–Itô decomposition

$$X(t) = t\Psi + B_{\mathcal{A}}(t) + \int_{[0,t]} \int_{\|y\| \leq 1} y \tilde{J}_X(ds, dy) + \int_{[0,t]} \int_{\|y\| > 1} y J_X(ds, dy), \quad t \geq 0, \quad (5)$$

where $\{B_{\mathcal{A}}(t) : t \geq 0\}$ is a $\mathbb{H}_d$ -valued Brownian motion with covariance $\mathcal{A}$, i.e., it is a Lévy process with continuous sample paths (a.s.) and each $B_{\mathcal{A}}(t)$ is centered Gaussian with

$$\mathbb{E} \{ \text{tr}(\Theta_1 B_{\mathcal{A}}(s)) \text{tr}(\Theta_2 B_{\mathcal{A}}(t)) \} = \min(s, t) \text{tr}(\Theta_1 \mathcal{A} \Theta_2) \text{ for each } \Theta_1, \Theta_2 \in \mathbb{H}_d;$$

$J_X(\cdot, \cdot)$ is a Poisson random measure of jumps on $[0, \infty) \times \mathbb{H}_d^0$ with intensity measure $\text{Leb} \otimes \nu$, independent of $\{B_{\mathcal{A}}(t) : t \geq 0\}$ and the compensated measure is given by

$$\tilde{J}_X(dt, dy) = J_X(dt, dy) - dt\nu(dy).$$

For a systematic study of Lévy processes with values in Banach spaces, see [2].

2.2 Hermitian Lévy Processes with Simple Spectrum

Hermitian Lévy processes with simple spectrum (all eigenvalues distinct) are of special importance in this paper. For each $t > 0$, if $X(t)$ has an absolutely continuous distribution then $X(t)$ has a simple spectrum almost surely, since the set of Hermitian matrices whose spectrum is not simple has zero Lebesgue measure [33], see also [1]. The absolute continuity gives also that $X(t)$ is a nonsingular random matrix since for a Hermitian matrix A , $\det(A)$ is a real polynomial in d^2 variables and $\{A \in \mathbb{H}_d : \det(A) = 0\}$ has zero Lebesgue measure in $\mathbb{R}^{d^2}$.

If there is a Gaussian component, i.e., $\mathcal{A} \neq 0$, then $X(t)$ has an absolutely continuous distribution for each $t > 0$. Next we point out a sufficient condition on the

Lévy measure of a Hermitian Lévy process without Gaussian component to have an absolutely continuous distribution and hence a simple spectrum for each fixed time.

Let $\mathbb{S}_d$ be the unit sphere of $\mathbb{H}_d$. A polar decomposition of the Lévy measure ν is a family of pairs (π, ρ_ξ) where π , the spherical component, is a finite measure on $\mathbb{S}_d$ with $\pi(\mathbb{S}_d) \geq 0$ and ρ_ξ , the radial component, is a measure on $(0, \infty)$ for each $\xi \in \mathbb{S}_d$ with $\rho_\xi((0, \infty)) > 0$ such that

$$\nu(E) = \int_{\mathbb{S}_d} \pi(d\xi) \int_{(0, \infty)} 1_E(u\xi) \rho_\xi(du), \quad E \in \mathcal{B}(\mathbb{H}_d^0).$$

We say that ν satisfies the condition **D** if, for each $\xi \in \mathbb{S}_d$, there is a nonnegative function g_ξ on $(0, \infty)$ such that $\rho_\xi(B) = \int_B g_\xi(u) du$ for any $B \in \mathcal{B}(\mathbb{R} \setminus \{0\})$ and the following divergence condition holds

$$\int_{(0, \infty)} g_\xi(u) du = \infty \quad \pi\text{-a.e. } \xi.$$

As an example, recall that a random matrix M is called self-decomposable (and hence infinitely divisible) if for any $b \in (0, 1)$ there exists M_b independent of M such that M and $bM + M_b$ have the same matrix distribution. An infinitely divisible random matrix M in $\mathbb{H}_d$ with triplet $(\mathcal{A}, \nu, \Psi)$ is self-decomposable if and only if

$$\nu(E) = \int_{\mathbb{S}_d} \pi(d\xi) \int_{(0, \infty)} 1_E(u\xi) \frac{k_\xi(u)}{u} (du), \quad E \in \mathcal{B}(\mathbb{H}_d^0),$$

with a finite measure π on $\mathbb{S}_d$ and a nonnegative decreasing function k_ξ on $(0, \infty)$ for each $\xi \in \mathbb{S}_d$. In this case ν satisfies the condition **D** for $g_\xi(u) = k_\xi(u)/u$ where we can choose the measure π to be zero on the set $\{\xi \in \mathbb{S}_d : k_\xi(0+) = 0\}$.

The following proposition for Hermitian Lévy processes follows from the corresponding real vector case, Theorems 27.10 and 27.13 in [32], by using the identification (2) of $\mathbb{H}_d$ with $\mathbb{R}^{d^2}$. We recall that a Lévy process is nondegenerate if its distribution for any fixed time is nondegenerate.

Proposition 1 (a) *If $\{X(t) : t \geq 0\}$ is a nondegenerate Lévy process in $\mathbb{H}_d$ without Gaussian component satisfying condition **D**, then $X(t)$ has absolutely continuous distribution for each $t > 0$.*

(b) *If $\{X(t) : t \geq 0\}$ is a nondegenerate self-decomposable Lévy process in $\mathbb{H}_d$ without Gaussian component, then $X(t)$ has absolutely continuous distribution for each $t > 0$.*

2.3 Smoothness of the Spectrum of Hermitian Matrices

Next we consider several facts about eigenvalues as functions of the entries of a Hermitian matrix, similar to [28] for symmetric matrices, that will be useful in the sequel.

Following [1], let $\mathbb{U}_d^{VG}$ be the set of unitary matrices $U = (u_{ij})$, with $u_{ii} > 0$ for all i , $u_{ij} \neq 0$ for i, j , and all minors of U having non-zero determinants. Let us denote by $\mathbb{H}_d^{VG}$ the set of matrices X in $\mathbb{H}_d$ such that

$$X = UDU^*, \quad (6)$$

where D is a diagonal matrix with entries $\lambda_i = D_{ii}$ such that $\lambda_1 > \lambda_2 > \dots > \lambda_d$, $U \in \mathbb{U}_d^{VG}$. An element of $\mathbb{H}_d^{VG}$ is called a very good matrix. The set of very good matrices $\mathbb{H}_d^{VG}$ can be identified with an open subset of $\mathbb{R}^{d^2}$. It is known that the complement of $\mathbb{H}_d^{VG}$ has zero Lebesgue measure. If $T : \mathbb{U}_d^{VG} \rightarrow \mathbb{R}^{d(d-1)}$ is the mapping defined by

$$T(U) = \left(\frac{u_{12}}{u_{11}}, \dots, \frac{u_{1d}}{u_{11}}, \frac{u_{23}}{u_{22}}, \dots, \frac{u_{2d}}{u_{22}}, \dots, \frac{u_{(d-1)d}}{u_{(d-1)(d-1)}} \right),$$

then Lemma 2.5.6 in [1] shows that T is bijective, smooth, and the complement of $\mathbb{U}_d^{VG}$ is a closed subset of zero Lebesgue measure.

Let $\mathcal{S}_d$ be the open set

$$\mathcal{S}_d = \{(\lambda_1, \lambda_2, \dots, \lambda_d) \in \mathbb{R}^d : \lambda_1 > \lambda_2 > \dots > \lambda_d\} \quad (7)$$

and let D_λ be the diagonal matrix $D_{ii} = \lambda_i$ for any $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_d) \in \mathcal{S}_d$. Then the map $\tilde{T} : \mathcal{S}_d \times T(\mathbb{U}_d^{VG}) \rightarrow \mathbb{H}_d^{VG}$ defined by $\tilde{T}(\lambda, z) = T^{-1}(z)^* D_\lambda T^{-1}(z)$ is a smooth bijection and the inverse mapping of $\tilde{T}$ satisfies

$$\tilde{T}^{-1}(X) = (\lambda(X), \tilde{U}(X)),$$

where $\tilde{U}(X) = T(U)$ for $U \in \mathbb{U}_d^{VG}$ satisfying (6). As a consequence of these results, there exist an open subset $\tilde{G} \subset \mathbb{R}^{d^2}$ such that $\mathbb{R}^{d^2} \setminus \tilde{G}$ has zero Lebesgue measure and a smooth function

$$\Phi : \mathbb{R}^{d^2} \rightarrow \mathbb{R}^d \quad (8)$$

such that $\lambda(X) = \Phi(X)$ for $X \in \tilde{G}$.

2.4 The Itô Formula for Hermitian Lévy Processes

We shall first consider a convenient notation for the Fréchet derivative of a function f from $\mathbb{H}_d$ to $\mathbb{R}$ (see [15, X.4]), using the identification (2) of an element $X = (x_{ij} + iy_{ij})_{1 \leq i, j \leq d}$ in $\mathbb{H}_d$ as the element $(x_{11}, \dots, x_{dd}, (x_{ij}, y_{ij})_{1 \leq i < j \leq d})^\top$ in $\mathbb{R}^{d^2}$.

The derivative of f at the point u is denoted by $Df(u)$; it can be thought as a linear operator from $\mathbb{R}^{d^2}$ to $\mathbb{R}$. Its representation as a Hermitian matrix is

$$Df = \left(\frac{\partial}{\partial x_{ij}} f + i \frac{\partial}{\partial y_{ij}} f 1_{\{i \neq j\}} \right)_{1 \leq i \leq j \leq d}. \quad (9)$$

Let $L(\mathbb{R}^{d^2}, \mathbb{R})$ denote the space of linear operators from $\mathbb{R}^{d^2}$ to $\mathbb{R}$. If the map $Df : \mathbb{R}^{d^2} \rightarrow L(\mathbb{R}^{d^2}, \mathbb{R})$ defined by $u \rightarrow Df(u)$ is differentiable at a point u , we say that f is twice differentiable at u . We denote the derivative of the map Df at the point u by $D^2f(u)$. Recall that $D^2f(u)$ is a bilinear operator from $\mathbb{R}^{d^2} \times \mathbb{R}^{d^2}$ to $\mathbb{R}$ which is symmetric, that is, $D^2f(u)(x, y) = D^2f(u)(y, x)$. We use the following notation for the first and second partial derivative operators, where $z_{ij} = x_{ij}$ or $z_{ij} = y_{ij}$, with $y_{ii} = 0$, $i = 1, \dots, d$

$$\sum_{(i,j)} \frac{\partial}{\partial z_{ij}} := \sum_{1 \leq i \leq d} \frac{\partial}{\partial x_{ii}} + 2 \sum_{1 \leq i < j \leq d} \left(\frac{\partial}{\partial x_{ij}} + \frac{\partial}{\partial y_{ij}} \right), \quad (10)$$

$$\begin{aligned} \sum_{(i,j)(k,h)} \frac{\partial^2}{\partial z_{ij} z_{kh}} &:= \sum_{1 \leq i \leq j \leq d} \frac{\partial^2}{\partial x_{ii} x_{jj}} + \sum_{1 \leq i \leq j \leq d} \sum_{1 \leq k < h \leq d} \left(\frac{\partial^2}{\partial x_{ij} x_{kh}} + \frac{\partial^2}{\partial x_{ij} y_{kh}} \right) \\ &+ \sum_{1 \leq i < j \leq d} \sum_{1 \leq k < h \leq d} \frac{\partial^2}{\partial y_{ij} y_{kh}}. \end{aligned} \quad (11)$$

For a linear operator $\mathcal{A} : \mathbb{H}_d \rightarrow \mathbb{H}_d$, we denote its $d^2 \times d^2$ matrix representation by $\mathcal{A} = (A_{ij, kh})$.

Following Barndorff-Nielsen and Stelzer [6], we say that for each open subset $G \subset \mathbb{M}_d$, the process $X = \{X_t : t \geq 0\}$ is locally bounded within G if there exist a sequence of stopping times $(T_n)_{n \geq 1}$ increasing to infinity almost surely and a sequence of compact convex subsets $D_n \subset G$ with $D_n \subset D_{n+1}$ such that $X_t \in D_n$ for all $0 \leq t < T_n$. Observe that if the process X is locally bounded within G , it cannot get arbitrarily close to the boundary of G in finite time, hence X_t and X_{t-} are in D_n at all times $t \geq 0$.

The following result can be proved similarly to Proposition 3.10.10 in Bichteler [16], who considered the multivariate case. The bounded variation multivariate case can be seen in [6].

Theorem 1 (The Itô formula for matrix semimartingales) *Let $\{X_t : t \geq 0\}$ be a matrix semimartingale, $G \subset \mathbb{H}_d$ an open subset, and $f : G \rightarrow \mathbb{R}$ twice continuously differentiable. If $\{X_t : t \geq 0\}$ is locally bounded within G , then X_{t-} belongs to G for all $t > 0$, the process $f(X_t)$ is a semimartingale, and*

$$\begin{aligned} f(X_t) = f(X_0) &+ \operatorname{tr} \left(\int_{0+}^t Df(X_{s-}) dX_s \right) + \frac{1}{2} \int_{0+}^t \sum_{(i,j)(k,h)} \frac{\partial^2}{\partial z_{ij} z_{kh}} f(X_s) d[z_{ij}^c, z_{kh}^c]_s \\ &+ \int_{(0,t] \times \mathbb{H}_d^0} [f(X_s) - f(X_{s-}) - \operatorname{tr}(Df(X_{s-})y)] J_X(ds, dy). \end{aligned}$$

The following two propositions for Hermitian Lévy processes are special cases of the above result. The first one is the analog of the multivariate case in [21, Proposition 8.15].

Proposition 2 (The Itô formula for matrix Lévy processes) *Let $\{X_t : t \geq 0\}$ be a Hermitian Lévy process with triplet $(\mathcal{A}, \nu, \Psi)$, $G \subset \mathbb{H}_d$ an open subset, and $f : G \rightarrow \mathbb{R}$ twice continuously differentiable. If $\{X_t : t \geq 0\}$ is locally bounded within G , then X_{t-} belongs to G for all $t > 0$, the process $f(X_t)$ is a semimartingale, and*

$$\begin{aligned} f(X_t) = f(X_0) &+ \operatorname{tr} \left(\int_0^t Df(X_{s-}) dX_s \right) + \frac{1}{2} \int_0^t \sum_{(i,j)(k,h)} A_{ij,kh} \frac{\partial^2}{\partial z_{ij} z_{kh}} f(X_s) ds \quad (12) \\ &+ \int_{(0,t] \times \mathbb{H}_d^0} [f(X_{s-} + y) - f(X_{s-}) - \operatorname{tr}(Df(X_{s-})y)] J_X(ds, dy). \end{aligned}$$

The next martingale-drift decomposition of functions of a Hermitian Lévy process X can be obtained by replacing the Lévy-Itô decomposition (5) for X into the stochastic integral part of (12).

Proposition 3 *Let $\{X_t : t \geq 0\}$ be a Hermitian Lévy process with triplet $(\mathcal{A}, \nu, \Psi)$, $G \subset \mathbb{H}_d$ an open subset, and $f : G \rightarrow \mathbb{R}$ twice continuously differentiable. If $\{X_t : t \geq 0\}$ is locally bounded within G , then X_{t-} belongs to G for all $t > 0$, and the process $f(X_t)$ can be decomposed as $f(X_t) = M_t + V_t$ where M_t is a martingale and V_t is a bounded variation process such that*

$$\begin{aligned} M_t = f(X_0) &+ \operatorname{tr} \left(\int_0^t Df(X_{s-}) B_{\mathcal{A}}(ds) \right) \\ &+ \int_{(0,t] \times \mathbb{H}_d^0} [f(X_{s-} + y) - f(X_{s-})] \tilde{J}_X(ds, dy) \end{aligned}$$

and

$$V_t = \frac{1}{2} \int_0^t \sum_{(i,j)(k,h)} A_{ij,kh} \frac{\partial^2}{\partial z_{ij} \partial z_{kh}} f(X_s) ds + \text{tr} \left(\int_0^t \Psi Df(X_s) ds \right) \\ + \int_{(0,t] \times \mathbb{H}_d^0} [f(X_{s-} + y) - f(X_{s-}) - \text{tr}(Df(X_{s-})y) 1_{\{\|y\| \leq 1\}}] ds \nu(dy).$$

Remark 1 If a Hermitian Lévy process X is invariant under unitary conjugation with triplet $(\sigma^2 I_d \otimes I_d, \nu, \psi I_d)$, where ν has the polar decomposition (π, ρ) , and π is the uniform measure on $\mathbb{S}_d$, then the Itô formula for X in Proposition 3 becomes in

$$M_t = f(X_0) + \text{tr} \left(\int_0^t Df(X_{s-}) B_{\mathcal{A}}(ds) \right) \\ \int_{(0,t] \times \mathbb{S}_d \times (0,\infty)} [f(X_{s-} + ru) - f(X_{s-})] (J_X(ds, du, dr) - ds \pi(du) \rho(dr))$$

and

$$V_t = \frac{\sigma^2}{2} \int_0^t \sum_{(i,j)(k,h)} \frac{\partial^2}{\partial z_{ij} \partial z_{kh}} f(X_s) ds + \psi \text{tr} \left(\int_0^t Df(X_s) ds \right) \\ + \int_{(0,t] \times \mathbb{S}_d \times (0,\infty)} [f(X_{s-} + ru) - f(X_{s-}) - r \text{tr}(Df(X_{s-})u) 1_{\{r \leq 1\}}] ds \pi(du) \rho(dr).$$

3 Simultaneous Jumps of a Hermitian Lévy Process and Its Eigenvalues

Let $\{X(t) : t \geq 0\}$ be a $d \times d$ Hermitian Lévy process and let $\lambda(t) = (\lambda_1(t), \dots, \lambda_d(t))$ be the vector of eigenvalues of $X(t)$ where $\lambda_1(t) \geq \lambda_2(t) \geq \dots \geq \lambda_d(t)$ for each $t \geq 0$. If λ jumps at time t , it is clear that $\Delta X(t) \neq 0$, since the d -tuple of ordered eigenvalues is a continuous function on the space of Hermitian matrices. This follows from the Hoffman-Wielandt inequality (see [1, Remark 2.1.20]) where for each $t \geq 0$

$$\|\Delta \lambda(t)\|_2 \leq \|\Delta X(t)\|, \quad (13)$$

where $\|\cdot\|_2$ denotes the Euclidean norm in $\mathbb{R}^d$.

A natural question is whether the jump of the process X at time t implies the jump of the eigenvalues at that time. Our goal in this section is to provide conditions for the simultaneous jumps of the eigenvalue process λ whenever X jumps.

We analyze the non zero jumps of X in two situations, when $X(t)$ and $X(t-)$ commute and when they do not commute. In the commutative case we find that exactly k eigenvalues jump when X has a jump of rank k at time t , and that all the eigenvalues would jump simultaneously if and only if the jump of X at time t

is of rank d . In the noncommutative case, we prove that all the eigenvalues jump simultaneously if the jump of X at time t is of rank one.

We present below natural examples of both, commutative and noncommutative situations.

Example 1 Let $\{(\lambda_1(t), \dots, \lambda_d(t)) : t \geq 0\}$ be a d -dimensional Lévy process where $\{\lambda_i(t) : t \geq 0\}$, $i = 1, \dots, d$, are independent one-dimensional Lévy processes. For an arbitrary non random unitary $d \times d$ matrix U , define the $d \times d$ Hermitian Lévy process

$$X(t) = U \text{diag}(\lambda_1(t), \dots, \lambda_d(t)) U^*, \quad t \geq 0.$$

Since independent Lévy processes cannot jump at the same time almost surely, X is an Hermitian Lévy process whose jumps are of rank one almost surely and we observe that in that case $X(t)$ and $X(t-)$ are simultaneously diagonalizable.

Example 2 Let us consider a compound Poisson $d \times d$ matrix process

$$X(t) = \sum_{j=1}^{N(t)} u_j u_j^*, \quad t \geq 0,$$

where $u_j, j \geq 1$ are i.i.d. random vectors uniformly distributed on the unit sphere of $\mathbb{C}^d$ and $N(t), t \geq 0$, is a Poisson process independent of $u_j, j \geq 1$. Since $\Delta X(t) = u_{N(t)} u_{N(t)}^* 1_{\{\Delta X(t)=1\}}$ the process X has jumps of rank one almost surely and $X(t) = X(t-) + u_{N(t)} u_{N(t)}^*$ whenever there is a jump at time t . Hence we observe that $X(t)$ and $X(t-)$ do not commute since $u_{N(t)} u_{N(t)}^*$ and $X(t-) = u_1 u_1^* + \dots + u_{N(t)-1} u_{N(t)-1}^*$ do not commute.

Example 3 Let us consider $\{X(t) : t \geq 0\}$ a Hermitian Lévy process in $\mathbb{H}_d^+$ the set of nonnegative definite $d \times d$ Hermitian matrices. Assume that X has jumps of rank one. Then by Theorem 4.1 in [23] X is the quadratic variation of a $\mathbb{C}^d$ -valued Lévy process $\{Y(t) : t \geq 0\}$, that is

$$X(t) = [Y](t), \quad t \geq 0.$$

For each $t > 0$, $X(t) = \sum_{0 < u \leq t} \Delta Y_u (\Delta Y_u)^*$ and $X(t-) = \lim_{s \nearrow t} \sum_{0 < r \leq s} \Delta Y_r (\Delta Y_r)^*$ do not commute in general. In particular, $X(t)$ and $X(t-)$ commute if for each $u, r \in (0, t]$ the term $\Delta Y_u (\Delta Y_u)^* \Delta Y_r (\Delta Y_r)^*$ is Hermitian.

Example 4 Let $\{Z(t) : t \geq 0\}$ be a Hermitian Lévy process in $\mathbb{H}_d$ of bounded variation with jumps of rank one. By Theorem 4.2 in [23] Z is the difference of the quadratic variations of two $\mathbb{C}^d$ -valued Lévy processes $\{X(t) : t \geq 0\}$ and $\{Y(t) : t \geq 0\}$,

$$Z(t) = [X](t) - [Y](t), \quad t \geq 0,$$

where $[X](t), t \geq 0$ and $[Y](t), t \geq 0$ are independent Lévy processes in $\mathbb{H}_d^+$. It follows from Example 3 that $Z(t) = [X](t) - [Y](t)$ and $Z(t-) = [X](t-) -$

$[Y](t-)$ do not commute in general. In particular, $Z(t)$ and $Z(t-)$ commute if for each $u, r \in (0, t]$ the terms $\Delta X_u (\Delta X_u)^* \Delta X_r (X_r)^*$, $\Delta Y_u (\Delta Y_u)^* \Delta Y_r (\Delta Y_r)^*$ and $\Delta X_u (\Delta X_u)^* \Delta Y_r (\Delta Y_r)^*$ are Hermitian.

3.1 Commutative Case

Assume that $X(t)$ and $X(t-)$ commute. Then there exists a random unitary matrix $U = U_t$ such that $X(t) = U \text{diag}(\lambda_1(t), \dots, \lambda_d(t)) U^*$ and $X(t-) = U \text{diag}(\lambda_1(t-), \dots, \lambda_d(t-)) U^*$. Observe that the eigenvalues of $\Delta X(t) = X(t) - X(t-)$ are the difference between the eigenvalues of $X(t)$ and the eigenvalues of $X(t-)$.

If $\Delta X(t)$ is of rank one, there is exactly one eigenvalue λ_i such that $\Delta X(t) = U \text{diag}(0, \dots, \Delta \lambda_i(t), \dots, 0) U^*$ and all the other eigenvalues do not jump at that time t . Similarly if $\Delta X(t)$ is of rank k where $1 \leq k \leq d$ there are exactly k eigenvalues that jump at time t and all other $d - k$ eigenvalues remain without jumping at t . Observe that the only case when all eigenvalues jump simultaneously at t is when the jump $\Delta X(t) \neq 0$ is of rank d .

3.2 Noncommutative Case

When $X(t)$ and $X(t-)$ do not commute the situation is completely different from the commutative case, since the eigenvalues of $X(t) - X(t-)$ are not any more the difference between the eigenvalues of $X(t)$ and the eigenvalues of $X(t-)$. In this case we prove that if the jump $\Delta X(t)$ is of rank one, all eigenvalues jump simultaneously at this time t , as we will see in Theorem 2.

We first consider some facts about the eigenvalues of rank one matrix perturbations. Let A be a $d \times d$ complex matrix. A rank one perturbation of A is a $d \times d$ complex matrix $A + ruv^\top$ where $u, v \in \mathbb{C}^d$ and $r \in \mathbb{C}$. If S is any set of polynomials in $\mathbb{C}[x_1, \dots, x_d]$, we define $V(S) = \{u \in \mathbb{C}^d : f(u) = 0 \text{ for all } f \in S\}$. If S is a finite set of polynomials $f_1, \dots, f_n$ we write $V(f_1, \dots, f_n)$ for $V(S)$. Recall that a subset W of $\mathbb{C}^d$ is an algebraic set if $W = V(f_1, \dots, f_n)$ for some $f_1, \dots, f_n$. A non empty subset $Q \subset \mathbb{C}^d$ is generic if the complement $\mathbb{C}^d \setminus Q$ is contained in an algebraic set W which is not $\mathbb{C}^d$. In such a case, $\mathbb{C}^d \setminus Q$ is nowhere dense and of $2d$ dimensional Lebesgue measure zero.

When a Hermitian Lévy process X has simple spectrum for each $t > 0$, then X_t and X_{t-} have no common eigenvalues when the jump at time t is of rank one.

Proposition 4 *Let $\{X_t : t \geq 0\}$ be an absolutely continuous $d \times d$ Hermitian Lévy process. Let $\lambda(t) = (\lambda_1(t), \dots, \lambda_d(t))$ be the vector of eigenvalues of X_t for each $t \geq 0$. If X has a rank one jump at time $t > 0$, then*

$$\{\lambda_1(X_t), \dots, \lambda_d(X_t)\} \cap \{\lambda_1(X_{t-}), \dots, \lambda_d(X_{t-})\} = \emptyset \text{ a.s.}$$

Proof Observe that $Q = \{z \in \mathbb{C}^{2d} : \text{tr}(zz^*) \neq 0\}$ is a generic set since $\{z \in \mathbb{C}^{2d} : \text{tr}(zz^*) = 0\} = \{0\}$ is an algebraic set which is equal to $V(f_1, \dots, f_{2d})$ where $f_i(z) = z_i$ for any $z = (z_1, \dots, z_{2d})^\top \in \mathbb{C}^{2d}$ and $i = 1, \dots, 2d$.

Since X has a jump of rank one at time t , $\Delta X_t = ruu^*$ with $r = r_t \in \mathbb{R} \setminus \{0\}$ and $u = u_t \in \mathbb{C}^d \setminus \{0\}$. That is, we have the rank one perturbation $X_{t-} = X_t + ruu^*$. Since X_t is absolutely continuous and Hermitian, the eigenvalues $\lambda_1(X_t), \dots, \lambda_d(X_t)$ do not collide almost surely, hence the degree of the minimal polynomial of X_t is d . By Proposition 2.2 (iii) of [31] applied to $z = (u, \bar{u}) \in Q$, there are exactly d eigenvalues of X_{t-} that are not eigenvalues of X_t .

Remark 2 Proposition 2.2 in [31] also gives insight into the behavior of the non-Hermitian rank one jumps of a general complex matrix Lévy process for which X_t no longer has a simple spectrum. In this case, $\Delta X_t = ruv^\top$ with $r \in \mathbb{C} \setminus \{0\}$ and $u, v \in \mathbb{C}^d \setminus \{0\}$, the number $m_t \leq d$ of distinct eigenvalues of X_t might be random and there are exactly m_t eigenvalues of X_{t-} that are not eigenvalues of X_t .

We now present the main result of this paper, giving conditions for the simultaneous jumps of a Hermitian Lévy process and its eigenvalue process. Moreover, if one eigenvalue component of the eigenvalue process jumps, then all the other eigenvalue components jump.

Theorem 2 *Let $\{X(t) : t \geq 0\}$ be an absolutely continuous $d \times d$ Hermitian Lévy process and let $\lambda(t) = (\lambda_1(t), \dots, \lambda_d(t))$ be the vector of eigenvalues of $X(t)$ where $\lambda_1(t) \geq \lambda_2(t) \geq \dots \geq \lambda_d(t)$ for each $t \geq 0$.*

If X has a jump of rank one at $t > 0$ and $X(t)$ and $X(t-)$ do not commute then $\Delta\lambda(t) \neq 0$. Moreover $\Delta\lambda_i(t) \neq 0$ for some i if and only if $\Delta\lambda_j(t) \neq 0$ for all $j \neq i$.

Proof If the Lévy process X jumps at time t , then, by Proposition 4, $\lambda_i(X_t) - \lambda_j(X_{t-}) \neq 0$ for all $i, j = 1, 2, \dots, d$ and therefore $\Delta\lambda(t) \neq 0$.

Moreover, if λ jumps at time t , it is clear from (13) that $\Delta X(t) \neq 0$, and then by the first part of the proof all the eigenvalues jump, that is $\Delta\lambda_j(t) \neq 0$ for all $j \neq i$.

Remark 3 (a) We finally observe that Proposition 4 and Theorem 2 are true also for a Hermitian semimartingale X with simple spectrum for each $t > 0$. This is guaranteed whenever $X(t)$ has an absolutely continuous distribution for each $t > 0$.

(b) However, we have written the results of this section in the framework of Hermitian Lévy processes since the subject is one of Ole's interests.

4 The Dynamics of the Eigenvalues of Hermitian Lévy Processes

4.1 Eigenvalue Derivatives of Hermitian Lévy Processes

Consider a Hermitian matrix X with decomposition $X = UDU^*$ given by (6) where $U = (u_{ij})$ is a unitary matrix, with $u_{ii} > 0$ for all i , $u_{ij} \neq 0$ for i, j , and all minors of U having non-zero determinants and D is a diagonal matrix with entries $\lambda_i = D_{ii}$ such that $\lambda_1 > \lambda_2 > \dots > \lambda_d$.

Now if we assume that $X = (x_{ij} + iy_{ij})_{1 \leq i, j \leq d}$ is a smooth function of a real parameter θ , the Hadamard variation formulae hold: for each $m = 1, 2, \dots, d$,

$$\partial_\theta \lambda_m = (U^* \partial_\theta XU)_{mm},$$

$$\partial_\theta^2 \lambda_m = (U^* \partial_\theta^2 XU)_{mm} + 2 \sum_{j \neq m} \frac{|(U^* \partial_\theta XU)_{mj}|^2}{\lambda_m - \lambda_j}.$$

In particular, for $\theta = x_{kh}$ and y_{kh} with $k \leq h$, we observe that

$$\frac{\partial \lambda_m}{\partial x_{kh}} = 2 \operatorname{Re}(\bar{u}_{km} u_{hm}) 1_{\{k \neq h\}} + |u_{km}|^2 1_{\{k=h\}},$$

$$\frac{\partial \lambda_m}{\partial y_{kh}} = 2 \operatorname{Im}(\bar{u}_{km} u_{hm}) 1_{\{k \neq h\}},$$

$$\frac{\partial^2 \lambda_m}{\partial x_{kh}^2} = 2 \sum_{j \neq m} \frac{|\bar{u}_{km} u_{hj} + \bar{u}_{hm} u_{kj}|^2}{\lambda_m - \lambda_j} 1_{\{k \neq h\}} + 2 \sum_{j \neq m} \frac{|\bar{u}_{km} u_{kj}|^2}{\lambda_m - \lambda_j} 1_{\{k=h\}},$$

$$\frac{\partial^2 \lambda_m}{\partial y_{kh}^2} = 2 \sum_{j \neq m} \frac{|\bar{u}_{km} u_{hj} - \bar{u}_{hm} u_{kj}|^2}{\lambda_m - \lambda_j} 1_{\{k \neq h\}}.$$

We point out that the spectral dynamics of Hermitian matrices is studied in [34], giving a Hadamard variation formula of any order for each eigenvalue in terms of a gradient. It is noted there that an eigenvalue depends smoothly on X whenever that eigenvalue is simple (even if other eigenvalues are not).

Using the former derivatives of the eigenvalues and the smooth function given in (8) we have the following result for Hermitian Lévy processes with simple spectrum for each time $t > 0$.

Lemma 1 *Let $\{X_t : t \geq 0\}$ be a $d \times d$ Hermitian Lévy process with absolutely continuous distribution for each $t > 0$. Let $\lambda(t) = (\lambda_1(t), \dots, \lambda_d(t))$ with $\lambda_1(t) > \dots >$*

$\lambda_d(t)$ the eigenvalues of X_t for each $t \geq 0$. For each $m = 1, 2, \dots, d$, there exists $\Phi_m : \mathbb{R}^{d^2} \rightarrow \mathbb{R}$ which is C^∞ in an open subset $\tilde{G} \subset \mathbb{R}^{d^2}$, with $\text{Leb}(\mathbb{R}^{d^2} \setminus \tilde{G}) = 0$, such that for each $t > 0$, $\lambda_m(t) = \Phi_m(X_t)$ and for $k \leq h$

$$\frac{\partial \Phi_m}{\partial x_{kh}} = 2 \operatorname{Re}(\bar{u}_{km} u_{hm}) 1_{\{k \neq h\}} + |u_{km}|^2 1_{\{k=h\}}, \quad (14)$$

$$\frac{\partial \Phi_m}{\partial y_{kh}} = 2 \operatorname{Im}(\bar{u}_{km} u_{hm}) 1_{\{k \neq h\}}, \quad (15)$$

and

$$\frac{\partial^2 \Phi_m}{\partial x_{kh}^2} = 2 \sum_{j \neq m} \frac{|\bar{u}_{km} u_{hj} + \bar{u}_{hm} u_{kj}|^2}{\lambda_m - \lambda_j} 1_{\{k \neq h\}} + 2 \sum_{j \neq m} \frac{|\bar{u}_{km} u_{kj}|^2}{\lambda_m - \lambda_j} 1_{\{k=h\}}, \quad (16)$$

$$\frac{\partial^2 \Phi_m}{\partial y_{kh}^2} = 2 \sum_{j \neq m} \frac{|\bar{u}_{km} u_{hj} - \bar{u}_{hm} u_{kj}|^2}{\lambda_m - \lambda_j} 1_{\{k \neq h\}}, \quad (17)$$

where for simplicity, the dependence on t is omitted in $\lambda_m = \lambda_m(t)$ and $u_{kl} = u_{kl}(t)$, which are random.

4.2 The SDE for the Process of Eigenvalues

Using Lemma 1 and Proposition 2, we can describe the dynamics of the process of eigenvalues of a Hermitian Lévy process.

Theorem 3 Let $\{X_t : t \geq 0\}$ be a Hermitian Lévy process with triplet $(\mathcal{A}, \nu, \Psi)$ and locally bounded within the open set G in $\mathbb{H}_d$ identified with the open set $\tilde{G}$ in $\mathbb{R}^{d^2}$ of Lemma 1. Assume that X_t has an absolutely continuous distribution for each $t > 0$. Let $\lambda(t) = (\lambda_1(t), \dots, \lambda_d(t))$ be the eigenvalues of X_t for each $t \geq 0$.

Then, for each $m = 1, \dots, d$ the process $\lambda_m(X_t)$ is a semimartingale and

$$\begin{aligned} \lambda_m(X_t) = \lambda_m(X_0) + \operatorname{tr} \left(\int_0^t D\lambda_m(X_{s-}) dX_s \right) + \frac{1}{2} \int_0^t \sum_{(i,j) \in (k,h)} A_{ij,kh} \frac{\partial^2}{\partial z_{ij} \partial z_{kh}} \lambda_m(X_s) ds \\ + \int_{(0,t] \times \mathbb{H}_d^0} [\lambda_m(X_{s-} + y) - \lambda_m(X_{s-}) - \operatorname{tr}(D\lambda_m(X_{s-})y)] J_X(ds, dy), \end{aligned} \quad (18)$$

where (a)

$$D\lambda_m(X_{s-}) = \left(\frac{\partial \lambda_m(s-)}{\partial x_{kh}} + i \frac{\partial \lambda_m(s-)}{\partial y_{kh}} 1_{\{k \neq h\}} \right)_{1 \leq k \leq h \leq d},$$

$$\frac{\partial \lambda_m(s-)}{\partial x_{kh}} = 2 \operatorname{Re} (\bar{u}_{km}(s-) u_{hm}(s-)) 1_{\{k \neq h\}} + |u_{km}(s-)|^2 1_{\{k=h\}}, \quad (19)$$

$$\frac{\partial \lambda_m(s-)}{\partial y_{kh}} = 2 \operatorname{Im} (\bar{u}_{km}(s-) u_{hm}(s-)) 1_{\{k \neq h\}}. \quad (20)$$

(b)

$$\begin{aligned} \sum_{(i,j)(k,h)} A_{ij,kh} \frac{\partial^2}{\partial z_{ij} \partial z_{kh}} \lambda_m(X_s) &= 2 \sum_{k=1}^d A_{kk,kk} \sum_{j \neq m} \frac{|u_{km}(s) u_{kj}(s)|^2}{\lambda_m(s) - \lambda_j(s)} \\ &+ 4 \sum_{1 \leq k < h \leq d} A_{kh,kh} \sum_{j \neq m} \frac{|u_{km}(s) u_{hj}(s)|^2 + |u_{hm}(s) u_{kj}(s)|^2}{\lambda_m(s) - \lambda_j(s)}. \end{aligned} \quad (21)$$

Proof By Lemma 1, for each $m = 1, 2, \dots, d$ there exists $\Phi_m : \mathbb{R}^{d^2} \rightarrow \mathbb{R}$ which is C^∞ in an open subset $\tilde{G}$ of $\mathbb{R}^{d^2}$ such that $\lambda_m(t) = \Phi_m(X_t)$ for each $t > 0$, and the Φ_m satisfy (14)–(17).

Now, taking $f = \lambda_m$ in Proposition 2, where the open set G is identified with the open subset $\tilde{G}$, we have that $\lambda_m(X_t)$ is a semimartingale where formula (12) becomes (18). The derivative matrix $D\lambda_m(X_{s-})$ in (18), see (9), is given by $D\lambda_m(X_{s-}) = \left(\frac{\partial \Phi_m(s-)}{\partial x_{kh}} + i \frac{\partial \Phi_m(s-)}{\partial y_{kh}} 1_{\{k \neq h\}} \right)_{1 \leq k \leq h \leq d}$, where the expressions of $\frac{\partial \Phi_m(s-)}{\partial x_{kh}}$ and $\frac{\partial \Phi_m(s-)}{\partial y_{kh}}$ in (14) and (15) turn out to be in (19) and (20).

The second derivative term in (18) is given by $\sum_{(i,j)(k,h)} A_{ij,kh} \frac{\partial^2}{\partial z_{ij} \partial z_{kh}} \lambda_m(X_s)$
 $= \sum_{(i,j)(k,h)} A_{ij,kh} \frac{\partial^2}{\partial z_{ij} \partial z_{kh}} \Phi_m(s)$. It can be seen that $\frac{\partial^2}{\partial x_{ii} \partial x_{jj}} \Phi_m(s) = 0$ for $i \neq j$, $\frac{\partial^2}{\partial x_{ij} \partial x_{kh}} \Phi_m(s) = 0$ for $i < j, k < h$ with $ij \neq kh$, $\frac{\partial^2}{\partial x_{ij} \partial y_{kh}} \Phi_m(s) = 0$ for $i \leq j, k < h$, and $\frac{\partial^2}{\partial y_{ij} \partial y_{kh}} \Phi_m(s) = 0$ for $i < j, k < h$ with $ij \neq kh$.

Using the previous computations, the operator (11) applied to $\Phi_m(s)$ gives

$$\begin{aligned} \sum_{(i,j)(k,h)} A_{ij,kh} \frac{\partial^2}{\partial z_{ij} \partial z_{kh}} \Phi_m(s) &= \sum_{k=1}^d A_{kk,kk} \frac{\partial^2}{\partial x_{kk}^2} \Phi_m(X_s) \\ &+ \sum_{1 \leq k < h \leq d} A_{kh,kh} \left(\frac{\partial^2}{\partial x_{kh}^2} + \frac{\partial^2}{\partial y_{kh}^2} \right) \Phi_m(X_s), \end{aligned} \quad (22)$$

and now from (16) and (17) we get

$$\sum_{k=1}^d A_{kk,kk} \frac{\partial^2}{\partial x_{kk}^2} \Phi_m(X_s) = 2 \sum_{k=1}^d A_{kk,kk} \sum_{j \neq m} \frac{|\bar{u}_{km} u_{kj}|^2}{\lambda_m - \lambda_j}, \quad (23)$$

and

$$\begin{aligned}
& \sum_{1 \leq k < h \leq d} A_{kh, kh} \left(\frac{\partial^2}{\partial x_{kh}^2} + \frac{\partial^2}{\partial y_{kh}^2} \right) \Phi_m(X_s) \\
&= 2 \sum_{1 \leq k < h \leq d} A_{kh, kh} \sum_{j \neq m} \left\{ \frac{|\bar{u}_{km} u_{hj} + \bar{u}_{hm} u_{kj}|^2}{\lambda_m - \lambda_j} + \frac{|\bar{u}_{km} u_{hj} - \bar{u}_{hm} u_{kj}|^2}{\lambda_m - \lambda_j} \right\} \\
&= 2 \sum_{1 \leq k < h \leq d} A_{kh, kh} \sum_{j \neq m} 2 \frac{|\bar{u}_{km}(s) u_{hj}(s)|^2 + |\bar{u}_{hm}(s) u_{kj}(s)|^2}{\lambda_m(s) - \lambda_j(s)}.
\end{aligned} \tag{24}$$

Finally, (21) follows from (22)–(24).

Remark 4 (a) Theorem 3 is the analog of the Dyson–Brownian motion (1) for a Hermitian Lévy process. In this case, $X(t)$ has a simple spectrum almost surely for each $t > 0$. The problem of noncolliding eigenvalues for each $t > 0$ almost surely will be considered elsewhere.

(b) However, we observe that the repulsion term between each pair of eigenvalues $1/(\lambda_i(s) - \lambda_j(s))$, $i \neq j$, appears in (18) and (21) when there is a Gaussian component (i.e., $\mathcal{A} \neq 0$).

(c) The corresponding expression of Proposition 3 can be obtained for the dynamics of eigenvalues in terms of the local martingale and bounded variation components of the eigenvalues. Let $\{X_t : t \geq 0\}$ be a Hermitian Lévy process with triplet $(\mathcal{A}, \nu, \Psi)$ satisfying the same conditions as in Theorem 3. For each $m = 1, 2, \dots, d$ it holds the decomposition $\lambda_m(X_t) = M_t^m + V_t^m$ where M_t^m is a martingale and V_t^m is a process of bounded variation such that

$$\begin{aligned}
M_t^m &= \lambda_m(X_0) + \text{tr} \left(\int_0^t D\lambda_m(X_{s-}) B_{\mathcal{A}}(ds) \right) \\
&\quad + \int_{(0,t] \times \mathbb{H}_d^0} [\lambda_m(X_{s-} + y) - \lambda_m(X_{s-})] \tilde{J}_X(ds, dy)
\end{aligned}$$

and

$$\begin{aligned}
V_t^m &= \frac{1}{2} \int_0^t \sum_{(i,j) \in (k,h)} A_{ij, kh} \frac{\partial^2}{\partial z_{ij} \partial \bar{z}_{kh}} \lambda_m(X_s) ds + \text{tr} \left(\int_0^t \Psi D\lambda_m(X_s) ds \right) \\
&\quad + \int_{(0,t] \times \mathbb{H}_d^0} [\lambda_m(X_{s-} + y) - \lambda_m(X_{s-}) - \text{tr}(D\lambda_m(X_{s-})y) 1_{\{\|y\| \leq 1\}}] ds \nu(dy).
\end{aligned} \tag{25}$$

(d) Hermitian Lévy processes with rank one jumps and with their distributions invariant under unitary conjugation appear in the context of a connection between classical and free one-dimensional infinitely divisible distributions [13, 20, 22, 23]. Let us consider an example with the triplet $\left(\frac{\sigma^2}{2} (1 + 1_{\{k=h \leq d\}}) I_d \otimes I_d, \nu, \psi I_d \right)$. From (c), for each $m = 1, \dots, d$, the eigenvalue λ_m has the decomposition $\lambda_m(t) = M_t^m +$

V_t^m , where

$$M_t^m = \lambda_m(X_0) + \text{tr} \left(\int_0^t D\lambda_m(X_{s-}) B_{\mathcal{A}}(ds) \right) \\ + \int_{(0,t] \times \mathbb{H}_d^1} [\lambda_m(X_{s-} + u) - \lambda_m(X_{s-})] (J_X(ds, du) - ds\nu(du))$$

and

$$V_t^m = \sigma^2 \int_0^t \sum_{j \neq m} \frac{1}{\lambda_m(X_s) - \lambda_j(X_s)} ds + \psi \text{tr} \left(\int_0^t D\lambda_m(X_s) ds \right) \\ + \int_{(0,t] \times \mathbb{H}_d^1} [\lambda_m(X_{s-} + u) - \lambda_m(X_{s-}) - \text{tr}(D\lambda_m(X_{s-})u) 1_{\{\|u\| \leq 1\}}] ds\nu(du).$$

In particular, when the process is symmetric, the Lévy measure is symmetric and $\psi = 0$, and hence

$$V_t^m = \sigma^2 \int_0^t \sum_{j \neq m} \frac{1}{\lambda_m(X_s) - \lambda_j(X_s)} ds + \int_{(0,t] \times \mathbb{H}_d^1} [\lambda_m(X_{s-} + u) - \lambda_m(X_{s-})] ds\nu(du).$$

Acknowledgments Most of this work was done while Alfonso Rocha-Arteaga was visiting CIMAT in the summer of 2014 and Víctor Pérez-Abreu was visiting the Autonomous University of Sinaloa in the winter of 2015. The authors would like to thank Michał Wojtylak for useful comments on the use of Proposition 2.2 of [31], and Armando Domínguez, José-Luis Pérez and the referee for valuable comments.

References

1. Anderson, W., Guionnet, A., Zeitouni, O.: An Introduction to Random Matrices. Cambridge University Press (2009)
2. Applebaum, D.: Lévy processes and stochastic integrals in Banach spaces. *Probab. Math. Statist.* **27**, 75–88 (2007)
3. Barndorff-Nielsen, O.E., Pérez-Abreu, V.: Matrix subordinators and related Upsilon transformations. *Theory Probab. Appl.* **52**, 1–23 (2008)
4. Barndorff-Nielsen, O.E., Pérez-Abreu, V., Rocha-Arteaga, A.: MatG random matrices. *Stoch. Mod.* **22**, 723–734 (2006)
5. Barndorff-Nielsen, O.E., Pérez-Abreu, V., Thorbjørnsen, S.: Lévy mixings. *ALEA Lat. Am. J. Probab. Math. Stat.* **10**, 1013–1062 (2013)
6. Barndorff-Nielsen, O.E., Stelzer, R.: Positive-definite matrix processes of finite variation. *Probab. Math. Statist.* **27**, 3–43 (2007)
7. Barndorff-Nielsen, O.E., Stelzer, R.: The multivariate supOU stochastic volatility model. *Math. Finance* **23**, 275–296 (2011)
8. Barndorff-Nielsen, O.E., Stelzer, R.: Multivariate supOU processes. *Ann. Appl. Probab.* **21**, 140–182 (2011)
9. Barndorff-Nielsen, O.E., Thorbjørnsen, S.: A connection between free and classical infinite divisibility. *Inf. Dim. Anal. Quantum Probab.* **7**, 573–590 (2004)

10. Barndorff-Nielsen, O. E., Thorbjørnsen, S.: Classical and free infinite divisibility and Lévy processes. In: Schürmann, M., Franz, U. (eds) *Quantum Independent Increment Processes II. Lecture Notes in Mathematics*, vol. 1866. Springer, Berlin (2006)
11. Barndorff-Nielsen, O.E., Maejima, M., Sato, K.: Some classes of multivariate infinitely divisible distributions admitting stochastic integral representation. *Bernoulli* **12**, 1–33 (2004)
12. Barndorff-Nielsen, O.E., Rosiński, J., Thorbjørnsen, S.: General γ -transformations. *ALEA Lat. Am. J Probab. Math. Stat.* **4**, 131–165 (2008)
13. Benaych-Georges, F.: Classical and free infinitely divisible distributions and random matrices. *Ann. Probab.* **33**, 1134–1170 (2005)
14. Bercovici, H., Pata, V.: Stable laws and domains of attraction in free probability theory. *Ann. Math.* **149**, 1023–1060 with an appendix by P. Biane (1999)
15. Bhatia, R.: *Matrix Analysis*. Graduate Texts in Mathematics, vol. 169. Springer, Berlin (1997)
16. Bichteler, K.: *Stochastic Integration with Jumps*. Cambridge University Press (2002)
17. Bru, M.F.: Diffusions of perturbed principal component analysis. *J. Multivar. Anal.* **29**, 127–136 (1989)
18. Bru, M.F.: Wishart processes. *J. Theoret. Probab.* **9**, 725–751 (1991)
19. Chan, T.: The Wigner semicircle law and eigenvalues of matrix diffusions. *Probab. Theory Relat. Fields* **93**, 249–272 (1992)
20. Cabanal-Duvillard, T.: A matrix representation of the Bercovici-Pata bijection. *Electron. J. Probab.* **10**, 632–661 (2005)
21. Cont, R., Tankov, P.: *Financial Modelling with Jump Processes*. Chapman & Hall/CRC, Chichester (2004)
22. Domínguez-Molina, J.A., Rocha-Arteaga, A.: Random matrix models of stochastic integrals type for free infinitely divisible distributions. *Periodica Math. Hung.* **64**(2), 145–160 (2012)
23. Domínguez-Molina, J.A., Pérez-Abreu, V., Rocha-Arteaga, A.: Covariation representation for Hermitian Lévy process ensembles of free infinitely divisible distributions. *Electron. Comm. Probab.* **18**, 1–14 (2013)
24. Dyson, F.J.: A Brownian-motion model for the eigenvalues of a random matrix. *J. Math. Phys.* **3**, 1191–1198 (1962)
25. Katori, M., Tanemura, K.: Symmetry of matrix stochastic processes and noncolliding diffusion particle systems. *J. Math. Phys.* **45**, 3058–3085 (2004)
26. Katori, M., Tanemura, K.: Complex Brownian motion representation of the Dyson model. *Elect. Comm. Probab.* **18**, 1–16 (2013)
27. König, W., O Connell, N.: Eigenvalues of the Laguerre process as non-colliding squared Bessel processes. *Elect. Comm. Probab.* **6**, 107–114 (2001)
28. Nualart, D., Pérez-Abreu, V.: On the eigenvalue process of a matrix fractional Brownian motion. *Stoch. Process. Appl.* **124**, 4266–4282 (2014)
29. Pérez-Abreu, V., Sakuma, N.: Free generalized gamma convolutions. *Elect. Comm. Probab.* **13**, 526–539 (2008)
30. Protter, P.: *Stochastic integration and differential equations. Stochastic Modelling and Applied Probability*. vol. 21. Springer, Berlin (2004)
31. Ran, A.C.M., Wojtylak, M.: Eigenvalues of rank one perturbations of unstructured matrices. *Linear Algebra Appl.* **437**, 589–600 (2012)
32. Sato, K.: *Lévy Processes and Infinitely Divisible Distributions*. Cambridge University Press (1999)
33. Tao, T.: *Topics in Random Matrix Theory. Graduate Studies in Mathematics*, vol. 132. American Mathematical Society, Providence (2012)
34. Tao, T., Vu, V.: Random matrices: universality of local eigenvalue statistics. *Acta Math.* **206**, 127–204 (2011)

Likelihood Inference for Exponential-Trawl Processes

Neil Shephard and Justin J. Yang

Abstract Integer-valued trawl processes are a class of serially correlated, stationary and infinitely divisible processes that Ole E. Barndorff-Nielsen has been working on in recent years. In this chapter, we provide the first analysis of likelihood inference for trawl processes by focusing on the so-called exponential-trawl process, which is also a continuous time hidden Markov process with countable state space. The core ideas include prediction decomposition, filtering and smoothing, complete-data analysis and EM algorithm. These can be easily scaled up to adapt to more general trawl processes but with increasing computation efforts.

Keywords Integer-valued process · Trawl process · Poisson random measure · Hidden Markov process · Maximum likelihood estimate · Filtering and smoothing · EM algorithm

1 Introduction

In recent years, Ole E. Barndorff-Nielsen has been working on a class of stochastic models called integer-valued trawl processes. References include [2, 6, 7]. These are flexible models whose core randomness is driven by Poisson random measures. Trawl processes are related to the up-stairs processes of [25] and the random measure processes of [24]. Both of these processes are stationary. Barndorff-Nielsen et al. [7] also brings out the relationship between their processes and $M/G/\infty$ queues (e.g. [16, 18] and [8, Chap. 6.31]) and mixed moving average processes (e.g. [22]). Related discrete time count models include [9, 10, 12–15, 17, 23, 26]. Trawl processes also

N. Shephard (✉)
Department of Economics, Harvard University,
1802 Cambridge Street, Cambridge, MA 02138, USA
e-mail: shephard@fas.harvard.edu

J.J. Yang
Department of Statistics, Harvard University, 1 Oxford Street,
Cambridge, MA 02138, USA
e-mail: juchenjustinyang@fas.harvard.edu

fall within the wide class of the so-called ambit fields (e.g. [3, 4]). Recently, [21] models high frequency financial data by using a trawl process to allow for fleeting movements to prices in addition to an integer-valued Lévy process proposed by [5].

As far as we know, there is no existing literature that directly and completely addresses likelihood inference for these trawl processes—or equivalently the prediction based upon it. Even though there are a large number of papers that focus on likelihood inference for marked point processes (see [11] for a survey), it only indirectly and partially describes trawl processes in terms of their jumps. A thorough likelihood inference for trawl processes needs to include the information in the initial value of the process.

In this article, we provide a thorough analysis of likelihood inference for integer-valued trawl processes and demonstrate the core ideas—prediction decomposition, filtering, smoothing and EM algorithm—by focusing on the so-called exponential trawl. It is not only a simplification of the modelling framework but also an intellectually interesting special case of its own, as in this special case the resulting trawl process is a continuous time hidden Markov process with countable state space. The theoretical analysis for the filtering and smoothing problems for this type of process has been discussed in details by [19, 20], using the classical theory of Kolmogorov’s forward and backward differential equations. We particularly emphasize that the resulting EM algorithm in this special case is exact in the sense that there are no discretization errors in its computation.

The major goal of this article is to derive filtering and smoothing results in the framework of trawl processes, so the analysis adopted here can be easily scaled up to adapt to the discussions of other general trawls or even the inclusion of a non-stationary component proposed in [21]. These general discussions will be dealt with elsewhere, for they require a significantly more sophisticated particle filtering and smoothing device. We also discuss non-negative trawl processes, which are particularly easy to work with.

The structure of this article is as follows. In Sect. 2, we remind the reader how to construct trawl processes using the exponential trawl. Section 3 includes details of how to carry out filtering and smoothing for these models. In Sect. 4, we show likelihood inference for exponential-trawl processes based on these filters and smoothers. Section 5 discusses the important but analytically tractable special case of non-negative trawl processes. We finally conclude in Sect. 6. The Appendix contains the proofs and derivations of various results given in this article.

2 Exponential-Trawl Processes

In this section, we build our notation, definitions and key structures for the exponential-trawl process that will be focused on throughout this article. We also provide its log-likelihood function based on observed data.

2.1 Definition

Our model will be based on a homogeneous Lévy basis on $[0, 1] \times \mathbb{R} \rightarrow \mathbb{Z}$, which models the discretely scattered events of integer size (with direction) $y \in \mathbb{Z} \setminus \{0\}$ at each point in time $s \in \mathbb{R}$ and height $x \in [0, 1]$. It is defined by

$$L(dx, ds) \triangleq \int_{-\infty}^{\infty} y N(dy, dx, ds), \quad (x, s) \in [0, 1] \times \mathbb{R},$$

where N is a three-dimensional Poisson random measure with intensity measure

$$\mathbb{E}(N(dy, dx, ds)) = \nu(dy) dx ds.$$

Here ds means the arrival times are uniformly scattered (over $\mathbb{R}$), dx means the random heights are also uniformly scattered (over $[0, 1]$) and $\nu(dy)$ is a Lévy measure concentrated on the non-zero integers $\mathbb{Z} \setminus \{0\}$. Without any confusion, we will abuse the notation $\nu(y)$ to denote the mass of the Lévy measure centered at y . Throughout this article, we assume that

$$\int_{-\infty}^{\infty} \nu(dy) = \sum_{y \in \mathbb{Z} \setminus \{0\}} \nu(y) < \infty.$$

Following [7], we think of dragging a fixed Borel measurable set $A \subseteq [0, 1] \times (-\infty, 0]$ through time

$$A_t \triangleq A + (0, t), \quad t \geq 0,$$

so the trawl process is defined by

$$Y_t \triangleq L(A_t) = \int_{[0, 1] \times \mathbb{R}} 1_A(x, s - t) L(dx, ds).$$

Throughout the rest of this article, we will focus on the exponential trawl

$$A \triangleq \{(x, s) : s \leq 0, 0 \leq x < d(s) \triangleq \exp(s\phi)\}, \quad \phi > 0,$$

to simplify our exposition of the key ideas. We will leave results on more general trawls in another study.

Example 1 Suppose that

$$\nu(dy) = \nu^+ \delta_{\{1\}}(dy) + \nu^- \delta_{\{-1\}}(dy), \quad \nu^+, \nu^- > 0,$$

where $\delta_{\{\pm 1\}}(dy)$ is the Dirac point mass measure centered at ± 1 . The corresponding $L(dx, ds)$ is called a Skellam Lévy basis, while the special case of $\nu^- = 0$ is called

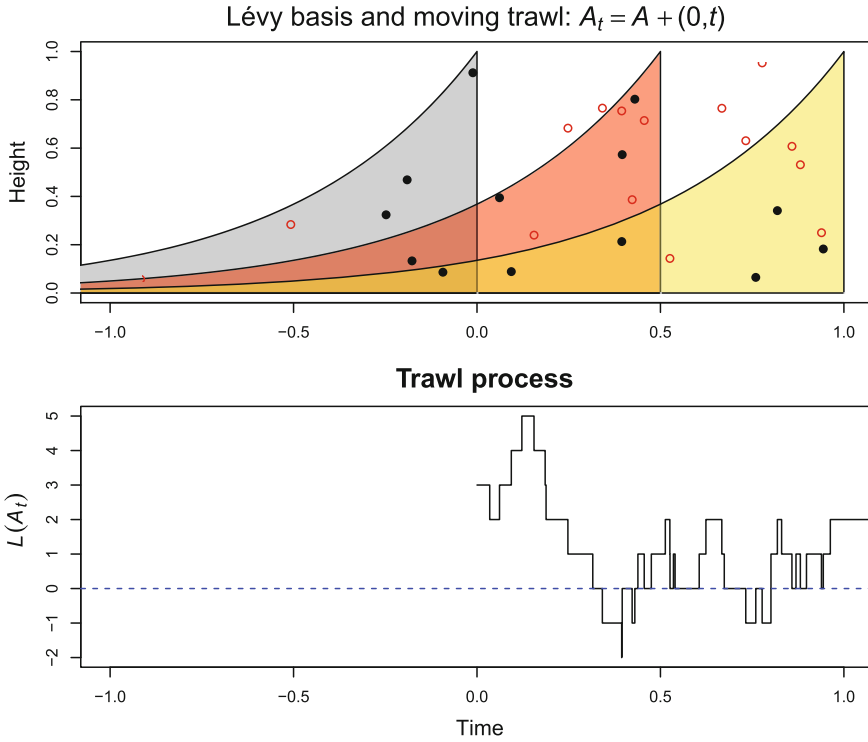


Fig. 1 A moving trawl A_t is joined by the Skellam Lévy basis $L(dx, ds)$, where the horizontal axis s is time and the vertical axis x is height. The shaded area is an example of the exponential trawl A , while we also show the outlines of A_t when $t = 1/2$ and $t = 1$. Also shown below is the implied trawl process $Y_t = L(A_t)$. Code: EPTprocess_illustration.R

Poisson. The upper panel of Fig. 1 shows events in L using $\nu^+ = \nu^- = 10$, taking sizes on $1, -1$ with black and white dots respectively and with equal probability. The lower panel of Fig. 1 then illustrates the resulting Skellam exponential-trawl process $Y_t = L(A_t)$ using $\phi = 2$, which sums up all the effects (both positive and negative) captured by the exponential trawl. Dynamically, $L(A_t)$ will move up by 1 if the moving trawl A_t either captures one positive event or releases a negative one; conversely, it will move down by 1 if vice versa. Notice that $Y_0 = L(A_0)$ might not be necessarily zero and the path of Y at negative time is not observed.

2.2 Markovian Counting Process

For $y \in \mathbb{Z} \setminus \{0\}$, let $C_t^{(y)} \in \{0, 1, 2, \dots\}$ be the total counts of surviving events of size y in the trawl at time t , which also includes the event that arrives *exactly* at time

t , so each $C_t^{(y)}$ must be càdlàg (right-continuous with left-limits). Then clearly the trawl process can be represented as

$$Y_t = \sum_{y \in \mathbb{Z} \setminus \{0\}} y C_t^{(y)}, \quad t \geq 0. \quad (1)$$

Note that each $C_t^{(y)}$ is not only a Poisson exponential-trawl process with (different) intensity of arrivals $\nu(y)$ (and sharing the same trawl) but also a M/M/ ∞ queue and hence a continuous time Markov process. Hence, for $\{\mathcal{C}_t^{(y)}\}_{t \geq 0}$ being the natural filtration generated by the counting process $C_t^{(y)}$, i.e., $\mathcal{C}_t^{(y)} \triangleq \sigma\left(\{C_s^{(y)}\}_{0 \leq s \leq t}\right)$, it has (infinitesimal) transition probabilities (or rates or intensities)

$$\lim_{dt \rightarrow 0} \frac{\mathbb{P}\left(C_t^{(y)} - C_{t-dt}^{(y)} = j \mid \mathcal{C}_{t-dt}^{(y)}\right)}{dt} = \begin{cases} \nu(y), & \text{if } j = 1 \\ \phi C_{t-}^{(y)}, & \text{if } j = -1 \\ 0, & \text{if } j \in \mathbb{Z} \setminus \{-1, 1\} \end{cases}. \quad (2)$$

The cases of $j = 1$ or -1 —which correspond to the arrival of a new event of size y and the departure of an old one—are the only two possible infinitesimal movements of $C_t^{(y)}$ due to the point process nature of the Lévy basis. Note that the arrival rate and departure rate are controlled by the Lévy measure ν and the trawl parameter ϕ respectively. Derivation of (2) can be found in many standard references for queue theory (e.g. [1]).

Remark 1 Let $\Delta X_t \triangleq X_t - X_{t-}$ denote the *instantaneous* jump of any process X at time t . Then the transition probability (2) can be conveniently written in a differential form

$$\mathbb{P}\left(\Delta C_t^{(y)} = j \mid \mathcal{C}_{t-}^{(y)}\right) = \begin{cases} \nu(y) dt, & \text{if } j = 1 \\ \phi C_{t-}^{(y)} dt, & \text{if } j = -1 \\ 0, & \text{if } j \in \mathbb{Z} \setminus \{-1, 1\} \end{cases}.$$

Throughout this article, our analysis will be majorly based on this infinitesimal point of view for the ease of demonstration. All of our arguments can be rephrased in a mathematically tighter way.

The independence property of the Lévy basis implies the independence between each $C_t^{(y)}$ for $y \in \mathbb{Z} \setminus \{0\}$, so the joint counting process

$$\mathbf{C}_t \triangleq \left(\dots, C_t^{(-2)}, C_t^{(-1)}, C_t^{(1)}, C_t^{(2)}, \dots \right)$$

is also Markovian, which serves as the unobserved state process for the observed *hidden Markov process* Y_t and will be the central target for the filter and smoother

we will discuss in a moment. Let $\mathcal{C}_t \triangleq \sigma(\{C_s\}_{0 \leq s \leq t}) = \bigvee_{y \in \mathbb{Z} \setminus \{0\}} \mathcal{C}_t^{(y)}$ be the joint filtration. Clearly, from (2), C_t has (infinitesimal) transition probabilities

$$\mathbb{P}(\Delta C_t = \mathbf{j} | \mathcal{C}_{t-}) = \begin{cases} \nu(y) dt, & \text{if } \mathbf{j} = \mathbf{1}^{(y)} \text{ for some } y \\ \phi C_{t-}^{(y)} dt, & \text{if } \mathbf{j} = -\mathbf{1}^{(y)} \text{ for some } y, \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

where $\mathbf{1}^{(y)} \in \mathbb{Z}^\infty$ is the vector that takes 1 at y -th component and 0 otherwise.

The trawl process Y_t can be also written as

$$Y_t = \sum_{y=1}^{\infty} y Y_t^{(y)}, \quad Y_t^{(y)} \triangleq C_t^{(y)} - C_t^{(-y)},$$

where each $Y_t^{(y)}$ is a Skellam exponential-trawl process. Each $Y_t^{(y)}$ is observed from the path of Y_t up to its initial value $Y_0^{(y)}$, for we can exactly observe all the jumps of Y_t and hence allocate them into the appropriate $Y_t^{(y)}$. In other words, we can regard the observed trawl process as (i) a *marked point process* $\Delta Y_t \in \mathbb{Z} \setminus \{0\}$, which consists of several independent (given all the $Y_0^{(y)}$) marked point processes $\Delta Y_t^{(y)} \in \{-1, 1\}$, plus (ii) the initial value Y_0 . The missing components $Y_0^{(y)}$'s will have some mild effects on $\Delta Y_t^{(y)}$. It is this initial value challenge that differentiates the likelihood analysis of trawl processes from that of marked point processes.

The special case where Y_t is always non-negative has further simpler structure, as we must have $C_t^{(-y)} = 0$ for all $y = 1, 2, \dots$ and hence $C_t^{(y)} = Y_t^{(y)}$ is directly observed up to its initial condition $C_0^{(y)}$, which can be well-approximated if the observation period T is large enough. We will go through these details in Sect. 5.

2.3 Conditional Intensities and Log-Likelihood

Let $\{\mathcal{F}_t\}_{t \geq 0}$ be the natural filtration generated by the observed trawl process Y_t , i.e. $\mathcal{F}_t \triangleq \sigma(\{Y_s\}_{0 \leq s \leq t})$. Define the càdlàg conditional intensity process of the trawl process Y as

$$\lambda_{t-}^{(y)} \triangleq \lim_{dt \rightarrow 0} \frac{\mathbb{P}(Y_t - Y_{t-dt} = y | \mathcal{F}_{t-dt})}{dt}, \quad y \in \mathbb{Z} \setminus \{0\}, \quad t > 0 \quad (4)$$

or conveniently in a differential form

$$\lambda_{t-}^{(y)} dt \triangleq \mathbb{P}(\Delta Y_t = y | \mathcal{F}_{t-}). \quad (5)$$

It means the (time-varying) predictive intensity of a size y move at time t of the trawl process, conditional on information instantaneously before time t .

Remark 2 To emphasize the $\mathcal{F}_t$ -predictability of $\lambda^{(y)}$, i.e., being adapted to the left natural filtration $\mathcal{F}_{t-}$, we will keep the subscript $t-$ throughout this article. This is particularly informative in the implementation of likelihood calculations, reminding us to take the *left-limit* of the intensity process whenever there is a jump.

For any two σ -fields $\mathcal{F}$ and $\mathcal{G}$, let the Radon-Nikodym derivative over $\mathcal{F}|\mathcal{G}$ between two probability measures $\mathbb{P}$ and $\mathbb{Q}$ be

$$\left(\frac{d\mathbb{P}}{d\mathbb{Q}}\right)_{\mathcal{F}|\mathcal{G}} \triangleq \frac{\left(\frac{d\mathbb{P}}{d\mathbb{Q}}\right)_{\mathcal{F} \vee \mathcal{G}}}{\left(\frac{d\mathbb{Q}}{d\mathbb{Q}}\right)_{\mathcal{G}}}.$$

In particular, when $\mathcal{G} = \sigma(X)$ for any random variable X , we will simply write the subscript as $\mathcal{F}|X$. The following classical result serves as the foundation for all likelihood inference for jump processes.

Theorem 1 *Let X_t be any integer-valued stochastic process and $\{\mathcal{F}_t^X\}_{t \geq 0}$ be its associated natural filtration. Assume that, under both $\mathbb{P}$ and $\mathbb{Q}$, (i) it has finite expected number of jumps during $(0, T]$, and (ii) the conditional intensities $\lambda_{t-}^{(y), \mathbb{P}}$ and $\lambda_{t-}^{(y), \mathbb{Q}}$ are well-defined using (4) and $\mathcal{F}_t^X$. Then $\mathbb{P} \ll \mathbb{Q}$ over $\mathcal{F}_T^X|X_0$ if and only if $\lambda_{t-}^{(y), \mathbb{Q}}$ is strictly positive. In this case, the logarithmic Radon-Nikodym derivative over $\mathcal{F}_T^X|X_0$ is*

$$\begin{aligned} \log \left(\frac{d\mathbb{P}}{d\mathbb{Q}}\right)_{\mathcal{F}_T^X|X_0} &= \sum_{0 < t \leq T} \sum_{y \in \mathbb{Z} \setminus \{0\}} \log \left(\frac{\lambda_{t-}^{(y), \mathbb{P}}}{\lambda_{t-}^{(y), \mathbb{Q}}}\right) 1_{\{\Delta X_t = y\}} \\ &\quad - \int_{t \in (0, T]} \sum_{y \in \mathbb{Z} \setminus \{0\}} \left(\lambda_{t-}^{(y), \mathbb{P}} - \lambda_{t-}^{(y), \mathbb{Q}}\right) dt. \end{aligned}$$

Proposition 14.4.I of [11] provides a complete and mathematically rigorous treatment for this Theorem. For completeness, we also provide an intuitive and heuristic derivation in the Appendix. A direct application of Theorem 1 gives the following Corollary.

Corollary 1 *The log-likelihood function of the (general) trawl process is (ignoring the constant)*

$$l_{\mathcal{F}_T}(\theta) = \sum_{0 < t \leq T} \sum_{y \in \mathbb{Z} \setminus \{0\}} \log \lambda_{t-}^{(y)} 1_{\{\Delta Y_t = y\}} - \int_{t \in (0, T]} \sum_{y \in \mathbb{Z} \setminus \{0\}} \lambda_{t-}^{(y)} dt + l_{Y_0}(\theta), \quad (6)$$

where the parameters of interest θ include the Lévy measure $\nu(dy)$ (i.e. $\nu(y)$'s) and the trawl parameter ϕ .

The study of likelihood inference for trawl processes then reduces to the calculations of conditional intensities $\lambda_{t-}^{(y)}$ for $y \in \mathbb{Z} \setminus \{0\}$. Now, by law of iterated expectations and the fact that $\mathcal{C}_t \supseteq \mathcal{F}_t$ for all t (because of (1)), we have

$$\begin{aligned} \lambda_{t-}^{(y)} dt &= \mathbb{E} (\mathbb{P} (\Delta Y_t = y | \mathcal{C}_{t-}) | \mathcal{F}_{t-}) \\ &= \mathbb{E} (\mathbb{P} (\Delta \mathbf{C}_t = \mathbf{1}^{(y)} | \mathcal{C}_{t-}) | \mathcal{F}_{t-}) + \mathbb{E} (\mathbb{P} (\Delta \mathbf{C}_t = -\mathbf{1}^{(-y)} | \mathcal{C}_{t-}) | \mathcal{F}_{t-}) \\ &= v(y) dt + \phi \mathbb{E} (C_{t-}^{(-y)} | \mathcal{F}_{t-}) dt, \end{aligned}$$

where the second line follows because the event $\Delta Y_t = y$ must come from either an arrival of a new size y event or a departure of an old size $-y$ event; the third line follows from (3). Thus,

$$\lambda_{t-}^{(y)} = v(y) + \phi \mathbb{E} (C_{t-}^{(-y)} | \mathcal{F}_{t-}), \quad y \in \mathbb{Z} \setminus \{0\}. \quad (7)$$

In the next section, we will study an exact filtering scheme to numerically calculate $\mathbb{E} (C_{t-}^{(-y)} | \mathcal{F}_{t-})$.

The non-negative exponential-trawl process, where we always have positive events, admits a further simplification

$$\lambda_{t-}^{(y)} = v(y), \quad \lambda_{t-}^{(-y)} = \phi \mathbb{E} (C_{t-}^{(y)} | \mathcal{F}_{t-}), \quad y = 1, 2, \dots, \quad (8)$$

so likelihood inference for such a case is easier. In the Poisson case, all the impacts are of size one, so in particular $C_0^{(1)} = Y_0$ is also observed (as $C_0^{(y)} = 0$ for all $y \neq 1$), which allows us to bypass the conditional expectation in (8) for $y = 1$.

3 Exact Filter and Smoother for Exponential-Trawl Processes

3.1 Filtering

In general we need to solve the filtering problems for $\mathbf{C}_t$ to implement (6) and (7). Denote the filtering probability mass function as

$$p_{t,s}(\mathbf{j}) \triangleq \mathbb{P}(\mathbf{C}_t = \mathbf{j} | \mathcal{F}_s), \quad \mathbf{j} = (\dots, j_{-2}, j_{-1}, j_1, j_2, \dots), \quad j_y = 0, 1, 2, \dots, \quad t, s \geq 0.$$

Also, let $\|\mathbf{j}\|_1 \triangleq \sum_{y \in \mathbb{Z} \setminus \{0\}} j_y$ and $D_t \triangleq \|\mathbf{C}_t\|_1 = \sum_{y \in \mathbb{Z} \setminus \{0\}} C_t^{(y)}$.

Our goal here is to sequentially update $p_{t-,t-}(\mathbf{j})$, where the initial distribution is derived from

$$C_0^{(y)} \stackrel{\text{indep.}}{\sim} \text{Poisson}\left(\frac{\nu(y)}{\phi}\right) \text{ subject to } \sum_{y \in \mathbb{Z} \setminus \{0\}} y C_0^{(y)} = Y_0,$$

so, by letting $\text{Poisson}(x|\lambda) \triangleq \lambda^x e^{-\lambda}/x!$, we have

$$p_{0,0}(\mathbf{j}) = \frac{\prod_{y \in \mathbb{Z} \setminus \{0\}} \text{Poisson}(j_y | \nu(y)/\phi)}{\mathbb{P}\left(\sum_{y \in \mathbb{Z} \setminus \{0\}} y C_0^{(y)} = Y_0\right)},$$

where the denominator can be numerically calculated using the inverse fast Fourier transform [21].

Notice that the filtering distribution not only updates at the times when the process jumps but also at the inactivity periods. We discuss these two cases separately.

Theorem 2 (Forward Filtering)

1. **[Update by inactivity]** Assume that the last jump time is τ (or $\tau = 0$) and the current time is $t-$, where $\Delta Y_s = 0$ for $\tau < s < t$ (and $\Delta Y_\tau \neq 0$ if $\tau > 0$). Then

$$p_{t-,t-}(\mathbf{j}) = \frac{e^{-\phi \|\mathbf{j}\|_1(t-\tau)} p_{\tau,\tau}(\mathbf{j})}{\sum_{\mathbf{k}} e^{-\phi \|\mathbf{k}\|_1(t-\tau)} p_{\tau,\tau}(\mathbf{k})}, \quad (9)$$

where $p_{\tau,\tau}$ is the filtering distribution we have already known at time τ .

2. **[Update by jump]** Assume that the current time is $\tau-$ and $\Delta Y_\tau = y$ for some $y \in \mathbb{Z} \setminus \{0\}$. Then

$$p_{\tau,\tau}(\mathbf{j}) = \frac{1}{\lambda_{\tau-}^{(y)}} \left(\nu(y) p_{\tau-, \tau-}(\mathbf{j} - \mathbf{1}^{(y)}) + \phi (j_{-y} + 1) p_{\tau-, \tau-}(\mathbf{j} + \mathbf{1}^{(-y)}) \right), \quad (10)$$

where $p_{\tau-, \tau-}$ is the filtering distribution we have already known at time $\tau-$.

Overall, the filtering procedures (9) and (10) imply that $p_{t-,t-}(\mathbf{j})$ can be updated in continuous time without discretization errors at any set of finite discrete time points, so we call it an exact filter.

Example 2 For the Skellam exponential-trawl process with Lévy intensities ν^+ and ν^- , we always have

$$Y_{t-} = C_{t-}^{(+)} - C_{t-}^{(-)}, \quad t > 0,$$

so knowing $p_{t-,t-}(j) \triangleq \mathbb{P}\left(C_{t-}^{(-)} = j \mid \mathcal{F}_{t-}\right)$ immediately gives us $p_{t-,t-}(j, k)$. Hence, the filtering updating scheme reduces to the following: starting from $\tau = 0$,

$$\begin{aligned}
p_{t-,t-}(j) &\propto e^{-\phi(2j+Y_{\tau})(t-\tau)} p_{\tau,\tau}(j) \text{ if } \Delta Y_s = 0 \text{ for } \tau < s < t, \\
p_{\tau,\tau}(j) &\propto v^+ p_{\tau-, \tau-}(j) + \phi(j+1) p_{\tau-, \tau-}(j+1) \text{ if } \Delta Y_{\tau} = 1, \\
p_{\tau,\tau}(j) &\propto v^- p_{\tau-, \tau-}(j-1) + \phi(j+Y_{\tau-}) p_{\tau-, \tau-}(j) \text{ if } \Delta Y_{\tau} = -1.
\end{aligned}$$

We then renormalize $p_{t-,t-}(j)$ such that $\sum_{j=0}^{\infty} p_{t-,t-}(j) = 1$ in each step of the updates. Knowing the filtering distributions $p_{t-,t-}(j)$ allows us to calculate

$$\mathbb{E}\left(C_{t-}^{(-)} \middle| \mathcal{F}_{t-}\right) = \sum_{j=0}^{\infty} j p_{t-,t-}(j), \quad \mathbb{E}\left(C_{t-}^{(+)} \middle| \mathcal{F}_{t-}\right) = \sum_{j=0}^{\infty} j p_{t-,t-}(j) + Y_{t-}.$$

Using the following settings, with time unit being second,

$$v^+ = 0.013, \quad v^- = 0.011, \quad \phi = 0.034, \quad T = 21 \times 60^2 = 75,600 \text{ (s)}, \quad (11)$$

Figure 2 shows a simulated path of the trawl process Y_t together with the filtering expectations of $C_t^{(+)}$, $C_t^{(-)}$ and $D_t = C_t^{(+)} + C_t^{(-)}$, the total number of surviving (both positive and negative) events in the trawl at time t .

3.2 Smoothing

We now consider the smoothing procedure for the exponential-trawl process Y_t , which is necessary for the likelihood inference based on the EM algorithm we will see in a moment.

Running the filtering procedure up to time T , we then start from $p_{T,T}$ to conduct the smoothing procedure.

Theorem 3 (Backward Smoothing)

1. **[Update by inactivity]** Assume that the (backward) last jump time is τ (or $\tau = T$) and the current time is t , where $\Delta Y_s = 0$ for $t \leq s < \tau$ (and $\Delta Y_{\tau} \neq 0$ if $\tau < T$). Then

$$p_{t,T}(\mathbf{j}) = p_{\tau-,T}(\mathbf{j}),$$

where $p_{\tau-,T}$ is the smoothing distribution we have already known at time $\tau-$.

2. **[Update by jump]** Assume that the current time is τ and $\Delta Y_{\tau} = y$ for some $y \in \mathbb{Z} \setminus \{0\}$. Then

$$p_{\tau-,T}(\mathbf{j}) = \frac{p_{\tau-, \tau-}(\mathbf{j})}{\lambda_{\tau-}^{(y)}} \left(v(y) \frac{p_{\tau,T}(\mathbf{j} + \mathbf{1}^{(y)})}{p_{\tau,\tau}(\mathbf{j} + \mathbf{1}^{(y)})} + \phi j_{-y} \frac{p_{\tau,T}(\mathbf{j} - \mathbf{1}^{(-y)})}{p_{\tau,\tau}(\mathbf{j} - \mathbf{1}^{(-y)})} \right), \quad (12)$$

where $p_{\tau-, \tau-}$ and $p_{\tau,\tau}$ are from the forward filtering procedure and $p_{\tau,T}$ is the smoothing distribution we have already known at time τ .

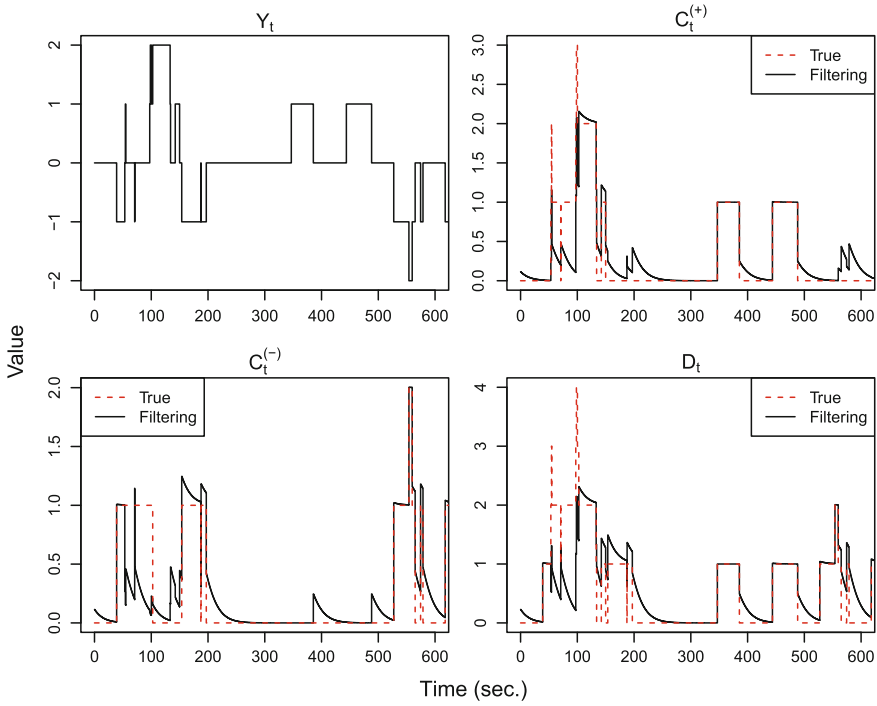


Fig. 2 *Top left* A simulated path for the Skellam exponential-trawl process Y_t . *Top right, Bottom left, Bottom right* Paths of the true hidden counting processes $C_t^{(+)}$, $C_t^{(-)}$ and $D_t = C_t^{(+)} + C_t^{(-)}$ of surviving events in the trawl along with their filtering estimations. Code: EPTprocess_FilteringSmoothing_Illustration.R

The two terms in (12) are

$$\mathbb{P}(C_{\tau-} = \mathbf{j}, C_{\tau} = \mathbf{j} + \mathbf{1}^{(y)} | \mathcal{F}_T) \text{ and } \mathbb{P}(C_{\tau-} = \mathbf{j}, C_{\tau} = \mathbf{j} - \mathbf{1}^{(y)} | \mathcal{F}_T)$$

respectively, so, in particular,

$$\mathbb{P}(\Delta C_{\tau}^{(y)} = 1 | \mathcal{F}_T) = \sum_{\mathbf{j}} \frac{p_{\tau-, \tau-}(\mathbf{j})}{\lambda_{\tau-}^{(y)}} \left(v(y) \frac{p_{\tau, T}(\mathbf{j} + \mathbf{1}^{(y)})}{p_{\tau, \tau}(\mathbf{j} + \mathbf{1}^{(y)})} \right), \quad (13)$$

$$\mathbb{P}(\Delta C_{\tau}^{(y)} = -1 | \mathcal{F}_T) = \sum_{\mathbf{j}} \frac{p_{\tau-, \tau-}(\mathbf{j})}{\lambda_{\tau-}^{(y)}} \left(\phi_{j-y} \frac{p_{\tau, T}(\mathbf{j} - \mathbf{1}^{(-y)})}{p_{\tau, \tau}(\mathbf{j} - \mathbf{1}^{(-y)})} \right). \quad (14)$$

These (total) weights in (12) will be recorded for every jump time τ as by-products of the smoothing procedure, for later they will play important roles in the EM algorithm introduced in Sect. 4.3.

Example 3 (Continued from Example 2) For the Skellam exponential-trawl process, the smoothing updating scheme reduces to the following: starting from $\tau = T$,

$$\begin{aligned}
 p_{t,T}(j) &= p_{\tau-,T}(j) \quad \text{if } \Delta Y_s = 0 \text{ for } t \leq s < \tau, \\
 p_{\tau-,T}(j) &\propto p_{\tau-, \tau-}(j) \left(v^+ \frac{p_{\tau,T}(j)}{p_{\tau,\tau}(j)} + \phi j \frac{p_{\tau,T}(j-1)}{p_{\tau,\tau}(j-1)} \right) \quad \text{if } \Delta Y_\tau = 1, \\
 p_{\tau-,T}(j) &\propto p_{\tau-, \tau-}(j) \left(v^- \frac{p_{\tau,T}(j+1)}{p_{\tau,\tau}(j+1)} + \phi (Y_{\tau-} + j) \frac{p_{\tau,T}(j)}{p_{\tau,\tau}(j)} \right) \quad \text{if } \Delta Y_\tau = -1.
 \end{aligned}$$

We also renormalize $p_{t,T}(j)$ in each step of the updates.

Using the same simulated path and the same setting (11) as in Example 2, we show the smoothing expectations of $C_t^{(+)}$, $C_t^{(-)}$ and D_t in Fig. 3. For most of the time, the smoothing expectations can match the truth quite well and will remove the peaks of filtering expectations resulting from departures (such as the one close to $t = 400$ in the plot for $C_t^{(-)}$).

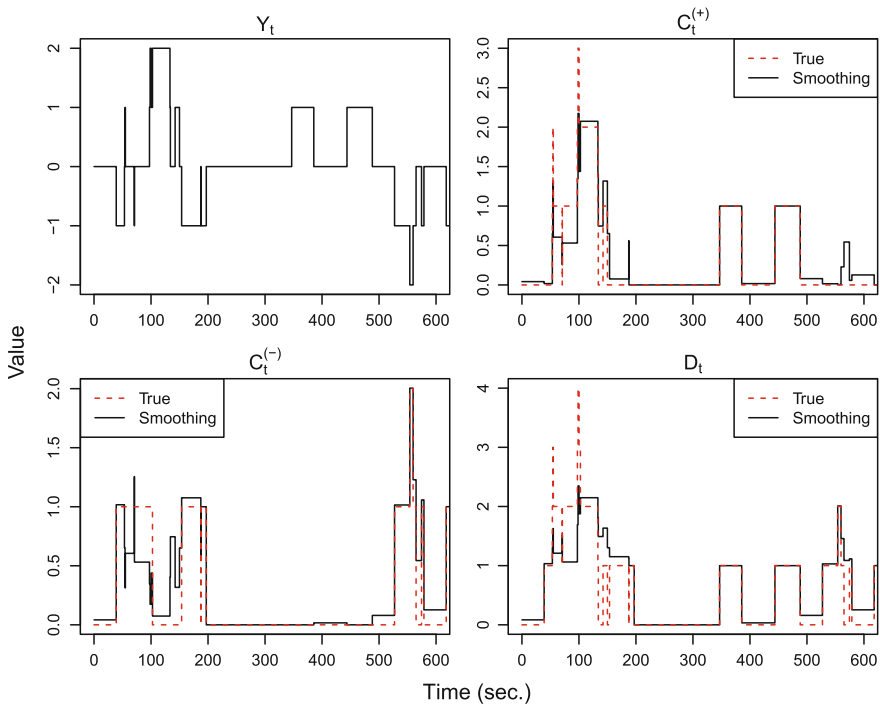


Fig. 3 Top left A simulated path for the Skellam exponential-trawl process Y_t . Top right, Bottom left, Bottom right Paths of the true hidden counting processes $C_t^{(+)}$, $C_t^{(-)}$ and $D_t = C_t^{(+)} + C_t^{(-)}$ of surviving events in the trawl along with their smoothing estimations. Code: EPTprocess_FilteringSmoothing_Illustration.R

Remark 3 In Figs. 2 and 3, we present the estimation of the hidden counting processes $C_t^{(+)}$ and $C_t^{(-)}$ in terms of their filtering and smoothing expectations, which are, of course, not necessarily integer-valued. To preserve such discrete natures in the estimate(s), one can consider the modes of the filtering and smoothing distributions instead.

Now we are capable of conducting likelihood inference for exponential-trawl processes as one of the most important applications of the filtering and smoothing procedures we have already built here.

4 Likelihood Inference for General Exponential-Trawl Processes

It has been reported by [7, 21] that the moment-based inference for the family of trawl processes could be easily performed, but such inference is arbitrarily dependent on its procedure design. In this section, we focus on the maximum likelihood estimate (MLE) calculation for exponential-trawl processes with general Lévy basis and demonstrate its correctness using several examples.

4.1 MLE Calculation Based on Filtering

Recall that the evaluation of the log-likelihood (6) requires the calculations of the conditional intensities $\lambda_{t-}^{(y)}$ and their integrals

$$\int_{t \in (0, T]} \lambda_{t-}^{(y)} dt = \nu(y) T + \phi \sum_{\mathbf{j}} j_{-y} \int_{t \in (0, T]} p_{t-, t-}(\mathbf{j}) dt, \quad (15)$$

which follows from (7) and $\mathbb{E} \left(C_{t-}^{(-y)} \middle| \mathcal{F}_{t-} \right) = \sum_{\mathbf{j}} j_{-y} p_{t-, t-}(\mathbf{j})$.

However, we do not know the integral $\int_{t \in (0, T]} p_{t-, t-}(\mathbf{j}) dt$ analytically, as the denominator in (9) also depends on t . Hence, we have to calculate (9) in a dense grid of time points—separated by a time gap $\delta_{\text{inactivity}}$ during the inactivity periods—and approximate (15) by linear interpolation. Clearly, the smaller the time gap $\delta_{\text{inactivity}}$, the smaller the numerical error in (15) but the larger the computational burden.

Example 4 Using the true parameters in (11) and simulating a 10-day-long data with $T = 756,000$ (s), Fig. 4 shows how an inappropriate choice of $\delta_{\text{inactivity}}$ will depict a wrong log-likelihood surface no matter how long the correct simulated data we supply, where the comparison is made with respect to the first day portion (75,600 (s)) of the 10-day-long simulated data. Using the same one-day-long data, Fig. 5 also shows the corresponding log-likelihood function over ν^+ or ν^- with other

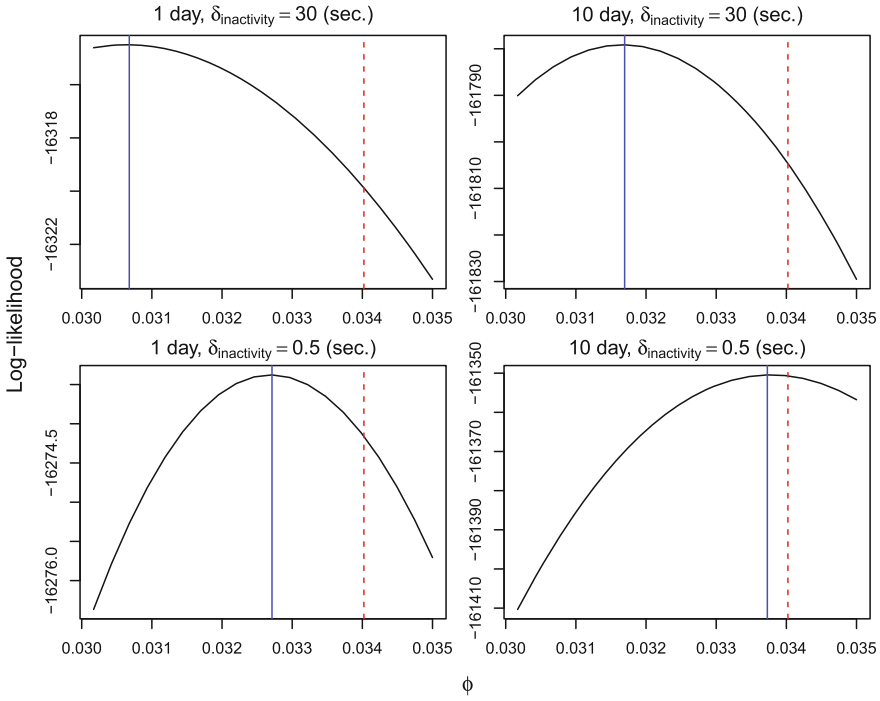


Fig. 4 Log-likelihood plots over ϕ (with ν^+ and ν^- fixed at the truth) using different $\delta_{\text{inactivity}}$ and a simulated 10-day-long ($T = 756,000$ (s)) Skellam exponential-trawl process. The one-day-long data is the first tenth of the simulated data. The *dashed lines* indicate the true value of ϕ , while the *solid lines* indicate the optimal value of ϕ in each plot. The p -values using the likelihood ratio test are 0.104 % (Top left), 21.0 % (Bottom left), 8.82×10^{-13} (Top right) and 46.1 % (Bottom right). Code: EPTprocess_MLE_Inference_Simulation_Small_vs_Large.R

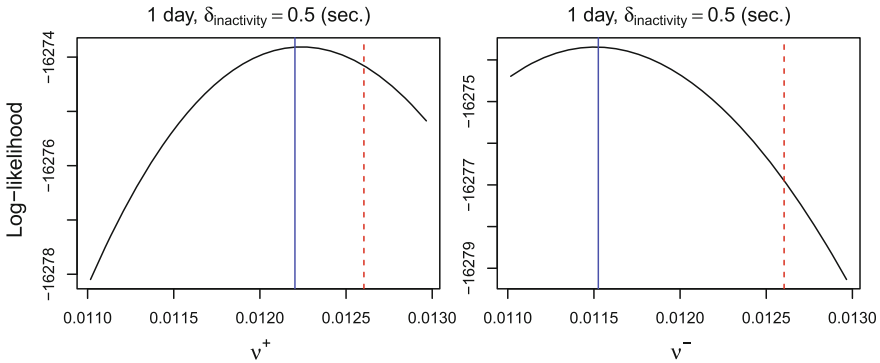


Fig. 5 Log-likelihood plots over either ν^+ or ν^- for one simulated Skellam exponential-trawl process. The *dashed lines* indicate the true value, while the *solid lines* indicate the optimal value of ν^+ or ν^- in the individual plot. The p -values using the likelihood ratio test are 40.5 % (Left) and 33.4 % (Right). Code: EPTprocess_MLE_Inference_Simulation_Small_vs_Large.R

parameters fixed at the truth when we specify a proper $\delta_{\text{inactivity}}$. All of the MLE's (solid lines) in the three plots (the bottom left one in Fig. 4 and the two in Fig. 5) are reasonably close to the true values (dashed lines), for the likelihood ratio test suggests that all the p -values in each plot are greater than 20 %.

4.2 Complete-Data Likelihood Inference

Even though in general it would be computationally expensive to calculate the MLE by direct filtering, the maximum complete-data likelihood estimate (MCLE) is much simpler. A comprehensive analysis of the complete-data likelihood inference is performed in the following.

Let $N_t^{(y),A}$ and $N_t^{(y),D}$ be the counting process of the arrival of size y events and the departure of size y events during the period $(0, T]$. Also let

$$N_t^{\text{type}} \triangleq \sum_{y \in \mathbb{Z} \setminus \{0\}} N_t^{(y), \text{type}}, \quad \text{type} = A, D.$$

Theorem 4 *The complete-data log-likelihood function of the exponential-trawl process is (ignoring the constant)*

$$\begin{aligned} l_{\mathcal{C}_T}(\theta) = & \sum_{y \in \mathbb{Z} \setminus \{0\}} \left(\log(v(y)) \left(N_T^{(y),A} + C_0^{(y)} \right) - v(y) (T + \phi^{-1}) \right) \\ & + \log(\phi) (N_T^D - D_0) - \phi \int_{t \in (0, T]} D_t - dt, \end{aligned} \quad (16)$$

so the corresponding MCLE's for the Lévy measure and the trawl parameter are

$$\begin{aligned} \hat{v}_{\text{MCLE}}(y) &= \frac{N_T^{(y),A} + C_0^{(y)}}{T + \hat{\phi}_{\text{MCLE}}^{-1}}, \quad y \in \mathbb{Z} \setminus \{0\}, \\ \hat{\phi}_{\text{MCLE}} &= \frac{\Xi_T + \sqrt{\Xi_T^2 + 4 \frac{N_T^A + N_T^D}{T} \int_{t \in (0, T]} D_t - dt}}{2 \int_{t \in (0, T]} D_t - dt}, \\ \Xi_T &\triangleq N_T^D - D_0 - \frac{1}{T} \int_{t \in (0, T]} D_t - dt. \end{aligned} \quad (17)$$

Furthermore, the MCLE's above are strongly consistent: with probability 1, as $T \rightarrow \infty$,

$$\hat{\phi}_{\text{MCLE}} \rightarrow \phi \text{ and } \hat{v}_{\text{MCLE}}(y) \rightarrow v(y), \quad y \in \mathbb{Z} \setminus \{0\}.$$

We note that $\hat{\phi}_{\text{MCLE}}$ depends on $\int_{t \in (0, T]} D_t - dt$, the total number of possible departures, weighted by the individual survival time during the period $(0, T]$.

4.3 MLE Calculation Based on EM Algorithm

In this subsection, we introduce an EM algorithm that is particularly suitable for exponential-trawl processes, as there are no discretization errors. The EM algorithm is also computationally efficient. Compared with generic optimization methods like *limited-memory BFGS* (*L-BFGS*), the updating scheme suggested by EM can converge to the MLE in fewer steps and with no error. Clearly, the use of EM needs some extra computations in each step for backward smoothing, but in aggregate EM performs much faster than L-BFGS as EM skips intermediate filtering calculations during the inactivity periods.

E-Step The linear form of the complete-data log-likelihood (16) allows us to easily take expectation on it with respect to $\mathbb{P}(\cdot|\mathcal{F}_T)$ (under a set of old estimated parameters $\hat{\theta}_{\text{old}}$), which then requires the calculations of the following quantities using the smoothing distribution $p_{t,T}$:

$$\begin{aligned}\mathbb{E}\left(N_T^{(y),A}\middle|\mathcal{F}_T\right) &= \sum_{0<t\leq T}\mathbb{P}\left(\Delta C_t^{(y)}=1\middle|\mathcal{F}_T\right), \\ \mathbb{E}\left(N_T^{(y),D}\middle|\mathcal{F}_T\right) &= \sum_{0<t\leq T}\mathbb{P}\left(\Delta C_t^{(y)}=-1\middle|\mathcal{F}_T\right), \\ \mathbb{E}\left(C_0^{(y)}\middle|\mathcal{F}_T\right) &= \sum_{\mathbf{j}}j_y p_{0,T}(\mathbf{j}), \quad \mathbb{E}\left(D_0\middle|\mathcal{F}_T\right) = \sum_{\mathbf{j}}\|\mathbf{j}\|_1 p_{0,T}(\mathbf{j}), \\ \mathbb{E}\left(D_{t-}\middle|\mathcal{F}_T\right) &= \sum_{\mathbf{j}}\|\mathbf{j}\|_1 p_{t-,T}(\mathbf{j}),\end{aligned}\tag{18}$$

where (13) and (14) will be extensively used. Note that $\mathbb{E}(D_{t-}|\mathcal{F}_T)$ will be a step function of t , so the calculation of $\int_{t\in(0,T]}\mathbb{E}(D_{t-}|\mathcal{F}_T)dt$ is trivially exact.

M-Step Since the *E-Step* generates a Q function that takes the same functional form of θ as (16), the solution to *M-Step* takes the same form as the MCLE in (17), where we just replace each of the hidden data related terms by their smoothing expectations in (18). This can be also viewed as a representation of the plug-in principle for (17), i.e., replacing unknown quantities (e.g. $1_{\{\Delta C_t^{(y)}=1\}}$) by known ones (e.g. $\mathbb{P}\left(\Delta C_t^{(y)}=1\middle|\mathcal{F}_T\right)$). We further use the solution of this *M-Step* for next iteration.

Table 1 The MLE calculations on one simulated Skellam exponential-trawl process using L-BFGS-B procedure in R (with default settings) and EM algorithm (with uniform tolerance 10^{-6} on the parameter space)

Estimation	Parameter			Log-likelihood	
	ν^+	ν^-	ϕ	$\delta_{\text{inactivity}} = 0.5$	$\delta_{\text{inactivity}} = 0.01$
Truth	0.01260	0.01111	0.03402	-15,974.98	-15,974.9543
L-BFGS-B	0.01201	0.01128	0.03362	-15,973.92	-15,973.8915
EM	0.01199	0.01126	0.03354	-15,973.91	-15,973.8881

The R elapsed time is 137.4 (s) for L-BFGS-B and 3.3 (s) for EM, which is about 40 times speed up. Code: EPTprocess_MLE_Inference_Simulation_LBFGS_vs_EM.R

Example 5 (Continued from Example 2)

Using the same simulated Skellam exponential-trawl process path, Table 1 compares the MLE derived from (i) the L-BFGS-B procedure in the `optim` function of the R language (using the default tolerance settings) with that from (ii) the EM algorithm (using the same initial parameter value), which stops until each parameter differs less than a uniform tolerance 10^{-6} .

As expected, using the EM algorithm gives estimation values that are very close to the direct optimization of the log-likelihood function (using $\delta_{\text{inactivity}} = 0.5$). An interesting feature here is that the MLE found by the EM algorithm has a slightly larger log-likelihood value (even for $\delta_{\text{inactivity}} = 0.01$) than by the L-BFGS-B, which might attribute to the numerical insufficiency of the default optimization tolerance setting of R.

The L-BFGS-B procedure uses 27 evaluations of the filtering procedure (9 of them for objective function evaluations and 18 of them for numerical gradients); as a comparison, the EM algorithm takes 12 evaluations of the filtering procedure plus 12 more of the smoothing procedure. In aggregate, the EM algorithm is over 40 times faster than the L-BFGS-B in terms of the computation time.

Starkly different from Example 4, the EM algorithm does not require the fine evaluation of the integrals of $\lambda_{t-}^{(y)}$, so the filtering procedure in each iteration of the EM is faster, as it skips the grid calculations of $\lambda_{t-}^{(y)}$ during the inactivity periods. Furthermore, the convergent result of EM will maximize the *numerically errorless* log-likelihood, as it has nothing to do with $\delta_{\text{inactivity}}$ to conduct EM. As a conclusion, using EM algorithm to search the MLE for exponential-trawl processes will dominate the direct optimization of log-likelihood both on the numerical quality and on the computation speed.

4.4 Likelihood Inference Without the Initial Information

If we consider the complete-data log-likelihood given the information $\mathbf{C}_0$, i.e. $l_{\mathcal{C}_T|\mathbf{C}_0}(\theta)$, then the MCLE's are even simpler:

$$\hat{\nu}_{\text{MCLE}}(y) = \frac{N_T^{(y),A}}{T}, \quad \hat{\phi}_{\text{MCLE}} = \frac{N_T^D}{\int_{t \in (0,T]} D_{t-} dt}.$$

Note that these estimates are the most natural frequency estimates providing that we know the hidden state process $\mathbf{C}_t$: $\nu(y)$ is estimated by the sample intensity of all the arrivals of size y events, while ϕ^{-1} is estimated by the average lifetime among all the departures of the temporary events, for the lifetime of any temporary event is exponentially distributed with mean $1/\phi$.

However, here is a subtle statistical inconsistency if one wants to build an EM algorithm based on $l_{\mathcal{C}_T|\mathbf{C}_0}(\theta)$. In practice, all the initial values $\mathbf{C}_0^{(y)}$'s are unknown, so the only way we can work on $l_{\mathcal{C}_T|\mathbf{C}_0}(\theta)$ is to treat them as nuisance parameters. Thus, the EM Q function is defined by

$$Q(\theta', \mathbf{C}'_0 | \theta, \mathbf{C}_0) = \mathbb{E}_\theta(l_{\mathcal{C}_T|\mathbf{C}'_0}(\theta') | \mathbf{C}_0, \mathcal{F}_T),$$

which not only requires the smoothing scheme based on $\mathbb{P}_\theta(\cdot | \mathcal{F}_T, \mathbf{C}_0)$ —not $\mathbb{P}_\theta(\cdot | \mathcal{F}_T)$ —but also finally gives us the MLE of the *joint* log-likelihood function $l_{\mathcal{F}_T}(\theta, \mathbf{C}_0)$ —not the MLE of $l_{\mathcal{F}_T}(\theta)$ nor of $l_{\mathcal{F}_T|Y_0}(\theta)$. On the other hand, one might also define the EM Q function as

$$Q(\theta' | \theta) = \mathbb{E}_\theta(l_{\mathcal{C}_T|\mathbf{C}_0}(\theta') | \mathcal{F}_T),$$

but in this case

$$Q(\theta | \theta) = l_{\mathcal{F}_T|Y_0}(\theta) - \mathbb{E}_\theta(l_{\mathbf{C}_0|Y_0}(\theta) | \mathcal{F}_T) \neq l_{\mathcal{F}_T|Y_0}(\theta),$$

which then *breaks* the fundamental monotonicity that guarantees the availability of EM:

$$l_{\mathcal{F}_T|Y_0}(\theta^*) \geq Q(\theta^* | \theta) = \max_{\theta'} Q(\theta' | \theta) \geq Q(\theta | \theta) = l_{\mathcal{F}_T|Y_0}(\theta).$$

Therefore, even though the direct filtering allows the calculations of the MLE whether we include the initial information Y_0 or not (i.e. to maximize $l_{\mathcal{F}_T}(\theta)$ or $l_{\mathcal{F}_T|Y_0}(\theta)$), a *correct* EM-based inference will automatically enforce the consideration of Y_0 (i.e. to maximize $l_{\mathcal{F}_T}(\theta)$ using EM). This is a bit different from likelihood inference for marked point processes, which usually ignores the effect of the initial value Y_0 . This mild difference will clearly disappear asymptotically as $T \rightarrow \infty$, but here we still prefer to present a complete likelihood analysis for trawl processes instead of treating them the same as marked point processes.

5 Likelihood Inference for Non-Negative Exponential-Trawl Processes

In this section, we focus on exponential-trawl processes that are always non-negative. Then all the negative movements of this type of processes must be attributed to the departures of the positive events in the trawl, so it is natural to split up Y into the counting process of size y jumps

$$N_t^{(y)} \triangleq \sum_{0 < s \leq t} 1_{\{\Delta Y_s = y\}}, \quad y \in \mathbb{Z} \setminus \{0\},$$

which relates to $C_t^{(y)}$ via

$$C_t^{(y)} = C_0^{(y)} + N_t^{(y)} - N_t^{(-y)}. \quad (19)$$

Then, as mentioned in the end of Sect. 2.2,

$$Y_t = \sum_{y=1}^{\infty} y C_t^{(y)} = \sum_{y=1}^{\infty} y C_0^{(y)} + \sum_{y=1}^{\infty} y (N_t^{(y)} - N_t^{(-y)}).$$

Clearly, the path of Y_t reveals the path of each of the individual $N_t^{(y)}$ for $y \in \mathbb{Z} \setminus \{0\}$, so $N_t^{(y)} \in \mathcal{F}_t$. Thus, the only unknown objects here are $C_0^{(y)}$'s, for we just see $Y_0 = \sum_{y=1}^{\infty} y C_0^{(y)}$ and all the departures resulting from $C_0^{(y)}$'s. If we can know $C_0^{(y)}$, then we will see the complete path of $C_t^{(y)}$ and hence likelihood inference will be particularly tractable.

5.1 Partial Likelihood Inference

We can specialize Corollary 1 using (8) and write down the log-likelihood for the non-negative case (ignoring the constant):

$$\begin{aligned} l_{\mathcal{F}_T}(\theta) &= \sum_{y=1}^{\infty} \left(\log(v(y)) N_T^{(y)} - v(y) T \right) \\ &\quad + \log(\phi) N_T^{(-)} - \phi \int_{t \in (0, T]} \mathbb{E}_{\theta} \left(C_{t-}^{(+)} \middle| \mathcal{F}_{t-} \right) dt \\ &\quad + \sum_{0 < t \leq T} \sum_{y=1}^{\infty} \log \mathbb{E}_{\theta} \left(C_{t-}^{(y)} \middle| \mathcal{F}_{t-} \right) 1_{\{\Delta Y_t = -y\}} + l_{Y_0}(\theta), \end{aligned}$$

where $N_T^{(-)} \triangleq \sum_{y=1}^{\infty} N_T^{(-y)}$ and $C_{t-}^{(+)} \triangleq \sum_{y=1}^{\infty} C_{t-}^{(y)}$.

Like the general case we studied in Sect. 4, there are no analytic expressions available for the filtering expectations $\mathbb{E}_{\theta} \left(C_{t-}^{(y)} \middle| \mathcal{F}_{t-} \right)$ and the initial likelihood $l_{Y_0}(\theta)$, so finding $\hat{\theta}_{\text{MLE}}$ also requires the EM techniques we introduced before. However, the first part of $l_{\mathcal{F}_T}(\theta)$ that involves $\nu(y)$'s is particularly analytically tractable, so this leads us to consider the following maximum partial likelihood estimate (MPLE) for the Lévy measure:

$$\hat{\nu}_{\text{MPLE}}(y) = \frac{N_T^{(y)}}{T}, \quad y = 1, 2, 3, \dots,$$

which is a non-parametric moment estimate that is apparent from the non-negative setting.

Even though $\hat{\nu}_{\text{MPLE}}$ is not $\hat{\nu}_{\text{MLE}}$, it has several advantages. First, it has strong consistency, i.e., with probability 1, $\hat{\nu}_{\text{MPLE}}(y) \rightarrow \nu(y)$ as $T \rightarrow \infty$. Second, it is asymptotically equivalent to the MCLE, because

$$\hat{\nu}_{\text{MCLE}}(y) = \frac{N_T^{(y)} + C_0^{(y)}}{T + \hat{\phi}_{\text{MCLE}}^{-1}} = \frac{\frac{N_T^{(y)}}{T} + \frac{C_0^{(y)}}{T}}{1 + \frac{\hat{\phi}_{\text{MCLE}}^{-1}}{T}} \approx \hat{\nu}_{\text{MPLE}},$$

where the MCLE of θ is simply given from (17) but we need to replace the D_{t-} by $C_{t-}^{(+)}$. Third, it allows to estimate each component of the Lévy measure separately from themselves and from ϕ , as given a long enough path of Y , including the initial value $C_0^{(y)}$ and $\hat{\phi}_{\text{MCLE}}^{-1}$ has no strong improvement on the estimation quality of $\hat{\nu}_{\text{MPLE}}$.

Alternatively, a parameterized common intensity function $\nu(y|\eta)$ can be used, where η is some finite dimensional parameter. Then the MPLE is found by solving

$$\hat{\eta}_{\text{MPLE}} \triangleq \underset{\eta}{\operatorname{argmax}} \sum_{y=1}^{\infty} \left(\log(\nu(y|\eta)) N_T^{(y)} - \nu(y|\eta) T \right)$$

and letting $\hat{\nu}_{\text{MPLE}}(y) = \nu(y|\hat{\eta}_{\text{MPLE}})$.

To infer on the trawl parameter ϕ , we can simply plug-in the $\hat{\nu}_{\text{MPLE}}$ (either parametric or non-parametric) and then do the filtering procedure to calculate $\mathbb{E}_{\theta=(\hat{\nu}_{\text{MPLE}}, \phi)} \left(C_{t-}^{(y)} \middle| \mathcal{F}_{t-} \right)$ for $y = 1, 2, \dots$ and $t \in (0, T]$. Combining this with an (one-dimensional) optimization procedure we can find

$$\hat{\phi}_{\text{MPLE}} \triangleq \underset{\phi}{\operatorname{argmax}} l_{\mathcal{F}_T}(\hat{\nu}_{\text{MPLE}}, \phi).$$

5.2 Estimate the Missing Initial Values

Except in the Poisson case ($Y_0 = C_0^{(1)}$, $C_0^{(y)} = 0$ for all $y > 1$ and hence in particular $\hat{\theta}_{\text{MCLE}} = \hat{\theta}_{\text{MLE}}$), every $C_0^{(y)}$'s are missing, so in principle we need to estimate these initial values in order to get (an approximation of) $\hat{\theta}_{\text{MCLE}}$. Indeed, the EM algorithm also does so through the smoothing expectations

$$\mathbb{E} \left(C_{t-}^{(y)} \middle| \mathcal{F}_T \right) = \mathbb{E} \left(C_0^{(y)} \middle| \mathcal{F}_T \right) + N_t^{(y)} - N_t^{(-y)},$$

but it just iterates (17) until converges. Nevertheless, there is another simpler estimation of $C_0^{(y)}$ thanks to the special non-negative feature.

The following Proposition only relies on the fact that Y_t is non-negative and in fact does not depend on the choice of the trawl.

Proposition 1 *Assume that the (general) trawl process Y_t is non-negative. If*

$$C_{0,T}^{(y),L} \triangleq \sup_{t \in [0,T]} \left(N_t^{(-y)} - N_t^{(y)} \right), \quad C_{0,T}^{(y),U} \triangleq \left\lfloor \frac{Y_0 - \sum_{y' \neq y} y' C_{0,T}^{(y'),L}}{y} \right\rfloor,$$

where $N_0^{(y)} \triangleq 0$ conventionally and $\lfloor x \rfloor$ means the integer part of x , then

$$C_{0,T}^{(y),U} \geq C_0^{(y)} \geq C_{0,T}^{(y),L}.$$

Furthermore,

$$\lim_{T \rightarrow \infty} C_{0,T}^{(y),U} = \lim_{T \rightarrow \infty} C_{0,T}^{(y),L} = C_0^{(y)}.$$

Thus, a straightforward and sharp estimation to $C_0^{(y)}$ can be given by, e.g.,

$$\hat{C}_0^{(y)} \triangleq \left\lfloor \frac{C_{0,T}^{(y),U} + C_{0,T}^{(y),L}}{2} \right\rfloor,$$

so using this estimation in (17) will give an estimate of θ that is almost as good as $\hat{\theta}_{\text{MCLE}}$.

Example 6 Figure 6 illustrates Proposition 1 with a non-negative geometric Lévy basis, where

$$\begin{aligned} \nu(y|\eta) &= \| \nu \| \eta (1 - \eta)^{y-1}, \quad y = 1, 2, \dots, \\ \| \nu \| &= 3, \quad \eta = 0.5, \quad \phi = 0.5, \quad T = 100. \end{aligned}$$

The paths of the upper bound $C_{0,t}^{(y),U}$ and the lower bound $C_{0,t}^{(y),L}$ are shown as step functions of time t in Fig. 6. We can observe a strong convergent pattern, as all the

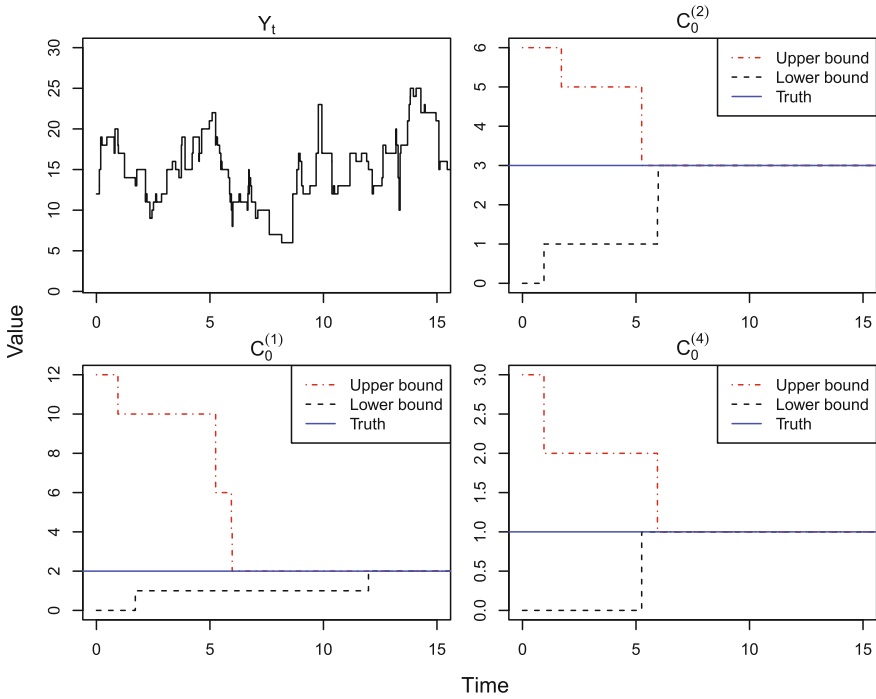


Fig. 6 *Top left* A simulated path for the exponential-trawl process Y_t using a non-negative geometric Lévy basis. *Top right, Bottom left, Bottom right* Paths of $C_{0,t}^{(y),U}$ and $C_{0,t}^{(y),L}$ along with the true $C_0^{(y)}$ for $y = 1, 2, 4$. Code: `EPTprocess _NonNegativeInitialEstimate.R`

bounds for different y converge after $t > 15$ —the perfect estimations of the initial values $C_0^{(y)}$'s. Furthermore, as $Y_0 = C_0^{(1)} + 2C_0^{(2)} + 4C_0^{(4)}$ in this case, all the other $C_0^{(y)}$'s for $y \neq 1, 2, 4$ must be zero. We then have discovered all the initial values and can use them to conduct MCLE by (17).

6 Conclusion

In this article, we studied likelihood-based inference of the trawl processes by explicitly working on the filtering and smoothing procedures inherited from this model. It is plausible and practically implementable under the exponential trawl. We used some simulation examples to justify the correctness of our procedures.

The major contribution of this article is to provide an easiest beginning step toward likelihood inference for all of the other more general trawl processes, which might even allow the inclusion of a non-stationary Lévy process component. Shephard and Yang [21] calls it a fleeting price process and extensively uses it for the study of high frequency financial econometrics.

The filters for the fleeting price process they proposed will allow an econometrically interesting decomposition of observed prices into equilibrium prices and market microstructure noises. More empirical analysis about these will be addressed in the future work.

Appendix: Proofs and Derivations

6.1 Heuristic Proof of Theorem 1

Our heuristic derivation starts from the following prediction decomposition of the Radon-Nikodym derivative:

$$\log \left(\frac{d\mathbb{P}}{d\mathbb{Q}} \right)_{\mathcal{F}_T^X | X_0} = \int_{t \in (0, T]} \log \left(\frac{d\mathbb{P}}{d\mathbb{Q}} \right)_{X_t | \mathcal{F}_{t-}^X}, \quad (20)$$

where the integral over $t \in (0, T]$ means a continuous sum of the integrand random variables. Thus,

$$\begin{aligned} \left(\frac{d\mathbb{P}}{d\mathbb{Q}} \right)_{X_t | \mathcal{F}_{t-}^X} &= \left(\frac{d\mathbb{P}}{d\mathbb{Q}} \right)_{\Delta X_t | \mathcal{F}_{t-}^X} \\ &= \sum_{y \in \mathbb{Z} \setminus \{0\}} \frac{\mathbb{P}(\Delta X_t = y | \mathcal{F}_{t-}^X)}{\mathbb{Q}(\Delta X_t = y | \mathcal{F}_{t-}^X)} 1_{\{\Delta X_t = y\}} + \frac{\mathbb{P}(\Delta X_t = 0 | \mathcal{F}_{t-}^X)}{\mathbb{Q}(\Delta X_t = 0 | \mathcal{F}_{t-}^X)} 1_{\{\Delta X_t = 0\}} \\ &= \sum_{y \in \mathbb{Z} \setminus \{0\}} \frac{\lambda_{t-}^{(y), \mathbb{P}} dt}{\lambda_{t-}^{(y), \mathbb{Q}} dt} 1_{\{\Delta X_t = y\}} + \frac{1 - \sum_{y \in \mathbb{Z} \setminus \{0\}} \lambda_{t-}^{(y), \mathbb{P}} dt}{1 - \sum_{y \in \mathbb{Z} \setminus \{0\}} \lambda_{t-}^{(y), \mathbb{Q}} dt} 1_{\{\Delta X_t = 0\}}, \end{aligned}$$

where the first equality follows because X_{t-} is known in $\mathcal{F}_{t-}^X$; the third equality follows from (5). Therefore, (20) can be rewritten as

$$\begin{aligned}
\int_{t \in (0, T]} \log \left(\frac{d\mathbb{P}}{d\mathbb{Q}} \right)_{X_t | \mathcal{F}_{t-}^X} &= \sum_{0 < t \leq T} \sum_{y \in \mathbb{Z} \setminus \{0\}} \log \left(\frac{\lambda_{t-}^{(y), \mathbb{P}} dt}{\lambda_{t-}^{(y), \mathbb{Q}} dt} \right) 1_{\{\Delta X_t = y\}} \\
&\quad + \int_{\{t \in (0, T] : \Delta X_t = 0\}} \log \left(\frac{1 - \sum_{y \in \mathbb{Z} \setminus \{0\}} \lambda_{t-}^{(y), \mathbb{P}} dt}{1 - \sum_{y \in \mathbb{Z} \setminus \{0\}} \lambda_{t-}^{(y), \mathbb{Q}} dt} \right) \\
&= \sum_{0 < t \leq T} \sum_{y \in \mathbb{Z} \setminus \{0\}} \log \left(\frac{\lambda_{t-}^{(y), \mathbb{P}}}{\lambda_{t-}^{(y), \mathbb{Q}}} \right) 1_{\{\Delta X_t = y\}} \\
&\quad - \int_{t \in (0, T]} \sum_{y \in \mathbb{Z} \setminus \{0\}} \left(\lambda_{t-}^{(y), \mathbb{P}} - \lambda_{t-}^{(y), \mathbb{Q}} \right) dt,
\end{aligned}$$

where the second equality follows from $\log(1 - x) \approx -x$ for small x and $\{t \in (0, T] : \Delta X_t \neq 0\}$ has Lebesgue measure 0.

6.2 Heuristic Proof of Theorem 2

6.2.1 Update by Inactivity

We want to update $p_{\tau, \tau}(\mathbf{j})$ by incorporating the information $\mathcal{F}_{(\tau, t)} \triangleq \sigma(\{\Delta Y_s = 0, \tau < s < t\})$ using Bayes' Theorem:

$$\begin{aligned}
\mathbb{P}(\mathbf{C}_{t-} = \mathbf{j} | \mathcal{F}_{t-}) &= \mathbb{P}(\mathbf{C}_\tau = \mathbf{j} | \mathcal{F}_{t-}) = \mathbb{P}(\mathbf{C}_\tau = \mathbf{j} | \mathcal{F}_\tau, \mathcal{F}_{(\tau, t)}) \\
&\propto \mathbb{P}(\mathcal{F}_{(\tau, t)} | \mathcal{F}_\tau, \mathbf{C}_\tau = \mathbf{j}) \mathbb{P}(\mathbf{C}_\tau = \mathbf{j} | \mathcal{F}_\tau),
\end{aligned}$$

where the first equality holds because there is no activity of Y_s for $s \in (\tau, t)$ and hence the hidden state $\mathbf{C}$ must stay the same.

Using the prediction decomposition, we have

$$\begin{aligned}
\log \mathbb{P}(\mathcal{F}_{(\tau, t)} | \mathcal{F}_\tau, \mathbf{C}_\tau = \mathbf{j}) &= \int_{s \in (\tau, t)} \log \mathbb{P}(\Delta Y_s = 0 | \mathcal{F}_\tau, \mathcal{F}_{(\tau, s)}, \mathbf{C}_\tau = \mathbf{j}) \\
&= \int_{s \in (\tau, t)} \log \left(1 - \sum_{y \in \mathbb{Z} \setminus \{0\}} \nu(y) ds - \sum_{y \in \mathbb{Z} \setminus \{0\}} \phi j_y ds \right) \\
&= - \sum_{y \in \mathbb{Z} \setminus \{0\}} \nu(y) (t - \tau) - \phi \|\mathbf{j}\|_1 (t - \tau),
\end{aligned}$$

where the second equality intuitively holds because we know the instantaneous departure probability of a size y event at time s is $\phi C_{s-}^{(y)} ds$ but $C_{s-}^{(y)} = C_\tau^{(y)} = j_y$ under $\mathcal{F}_{(\tau, s)}$; the third equality follows from $\log(1 - x) \approx -x$ for small x . Therefore,

$$\mathbb{P}(\mathbf{C}_{t-} = \mathbf{j} | \mathcal{F}_{t-}) \propto e^{-\phi \|\mathbf{j}\|_1 (t - \tau)} \mathbb{P}(\mathbf{C}_\tau = \mathbf{j} | \mathcal{F}_\tau),$$

where we throw out the term $\exp\left(-\sum_{y \in \mathbb{Z} \setminus \{0\}} \nu(y)(t - \tau)\right)$ because it doesn't depend on $\mathbf{j}$. Normalizing the equation above leads to the desired result.

6.2.2 Update by Jump

We want to update $p_{\tau-, \tau-}(\mathbf{j})$ by incorporating the piece of information, $\Delta Y_\tau = y$. First note that

$$\begin{aligned} \mathbb{P}(\mathbf{C}_\tau = \mathbf{j} | \mathcal{F}_\tau) &= \mathbb{P}(\mathbf{C}_\tau = \mathbf{j} | \mathcal{F}_{\tau-}, \Delta Y_\tau = y) \\ &= \mathbb{P}(\mathbf{C}_\tau = \mathbf{j}, \mathbf{C}_{\tau-} = \mathbf{j} - \mathbf{1}^{(y)} | \mathcal{F}_{\tau-}, \Delta Y_\tau = y) \\ &\quad + \mathbb{P}(\mathbf{C}_\tau = \mathbf{j}, \mathbf{C}_{\tau-} = \mathbf{j} + \mathbf{1}^{(-y)} | \mathcal{F}_{\tau-}, \Delta Y_\tau = y), \end{aligned}$$

which corresponds to the arrival of a new size y event and the departure of an old size $-y$ event.

For the first term,

$$\begin{aligned} &\mathbb{P}(\mathbf{C}_\tau = \mathbf{j}, \mathbf{C}_{\tau-} = \mathbf{j} - \mathbf{1}^{(y)} | \mathcal{F}_{\tau-}, \Delta Y_\tau = y) \\ &= \frac{\mathbb{P}(\mathbf{C}_\tau = \mathbf{j}, \mathbf{C}_{\tau-} = \mathbf{j} - \mathbf{1}^{(y)}, \Delta Y_\tau = y | \mathcal{F}_{\tau-})}{\mathbb{P}(\Delta Y_\tau = y | \mathcal{F}_{\tau-})} \\ &= \frac{\mathbb{P}(\mathbf{C}_\tau = \mathbf{j}, \Delta Y_\tau = y | \mathbf{C}_{\tau-} = \mathbf{j} - \mathbf{1}^{(y)}, \mathcal{F}_{\tau-}) \mathbb{P}(\mathbf{C}_{\tau-} = \mathbf{j} - \mathbf{1}^{(y)} | \mathcal{F}_{\tau-})}{\mathbb{P}(\Delta Y_\tau = y | \mathcal{F}_{\tau-})} \\ &= \frac{\mathbb{P}(\Delta \mathbf{C}_\tau = \mathbf{1}^{(y)} | \mathbf{C}_{\tau-} = \mathbf{j} - \mathbf{1}^{(y)}, \mathcal{F}_{\tau-}) \mathbb{P}(\mathbf{C}_{\tau-} = \mathbf{j} - \mathbf{1}^{(y)} | \mathcal{F}_{\tau-})}{\mathbb{P}(\Delta Y_\tau = y | \mathcal{F}_{\tau-})} \\ &= \frac{\nu(y)}{\lambda_{\tau-}^{(y)}} \mathbb{P}(\mathbf{C}_{\tau-} = \mathbf{j} - \mathbf{1}^{(y)} | \mathcal{F}_{\tau-}), \end{aligned}$$

where the fourth equality follows from (3) (using $\mathcal{C}_{\tau-} \supseteq \mathcal{F}_{\tau-}$) and (5).

Using similar arguments, the second term is

$$\mathbb{P}(\mathbf{C}_\tau = \mathbf{j}, \mathbf{C}_{\tau-} = \mathbf{j} + \mathbf{1}^{(-y)} | \mathcal{F}_{\tau-}, \Delta Y_\tau = y) = \frac{\phi(j_{-y} + 1)}{\lambda_{\tau-}^{(y)}} \mathbb{P}(\mathbf{C}_{\tau-} = \mathbf{j} + \mathbf{1}^{(-y)} | \mathcal{F}_{\tau-}).$$

Combining all of these gives us the required result.

6.3 Heuristic Proof of Theorem 3

The case of updating smoothing distribution $p_{\tau-, \tau}(\mathbf{j})$ due to inactivity is trivial because the hidden configuration $\mathbf{C}$ must stay unchanged because of the inactivity during the time period $[t, \tau)$.

6.3.1 Update by Jump

We now consider the case of (backward) updating the smoothing distribution $p_{\tau,T}(\mathbf{j})$ due to the jump $\Delta Y_\tau = y$. Then

$$\begin{aligned} \mathbb{P}(\mathbf{C}_{\tau-} = \mathbf{j} | \mathcal{F}_T) &= \mathbb{P}(\mathbf{C}_{\tau-} = \mathbf{j}, \mathbf{C}_\tau = \mathbf{j} + \mathbf{1}^{(y)} | \mathcal{F}_T) + \mathbb{P}(\mathbf{C}_{\tau-} = \mathbf{j}, \mathbf{C}_\tau = \mathbf{j} - \mathbf{1}^{(-y)} | \mathcal{F}_T) \\ &= \mathbb{P}(\mathbf{C}_{\tau-} = \mathbf{j} | \mathcal{F}_T, \mathbf{C}_\tau = \mathbf{j} + \mathbf{1}^{(y)}) \mathbb{P}(\mathbf{C}_\tau = \mathbf{j} + \mathbf{1}^{(y)} | \mathcal{F}_T) \\ &\quad + \mathbb{P}(\mathbf{C}_{\tau-} = \mathbf{j} | \mathcal{F}_T, \mathbf{C}_\tau = \mathbf{j} - \mathbf{1}^{(-y)}) \mathbb{P}(\mathbf{C}_\tau = \mathbf{j} - \mathbf{1}^{(-y)} | \mathcal{F}_T). \end{aligned}$$

Note that

$$\begin{aligned} \mathbb{P}(\mathbf{C}_{\tau-} = \mathbf{j} | \mathcal{F}_T, \mathbf{C}_\tau = \mathbf{k}) &= \mathbb{P}(\mathbf{C}_{\tau-} = \mathbf{j} | \mathcal{F}_\tau, \mathbf{C}_\tau = \mathbf{k}) \tag{21} \\ &= \frac{\mathbb{P}(\mathbf{C}_\tau = \mathbf{k} | \mathbf{C}_{\tau-} = \mathbf{j}, \mathcal{F}_\tau) \mathbb{P}(\mathbf{C}_{\tau-} = \mathbf{j} | \mathcal{F}_\tau)}{\mathbb{P}(\mathbf{C}_\tau = \mathbf{k} | \mathcal{F}_\tau)} \\ &\quad \frac{\mathbb{P}(\mathbf{C}_\tau = \mathbf{k} | \mathbf{C}_{\tau-} = \mathbf{j}, \mathcal{F}_\tau) \mathbb{P}(\Delta Y_\tau = y | \mathbf{C}_{\tau-} = \mathbf{j}, \mathcal{F}_{\tau-})}{\times \mathbb{P}(\mathbf{C}_{\tau-} = \mathbf{j} | \mathcal{F}_{\tau-})} \\ &= \frac{\mathbb{P}(\mathbf{C}_\tau = \mathbf{k} | \mathcal{F}_\tau) \mathbb{P}(\Delta Y_\tau = y | \mathcal{F}_{\tau-})}{\mathbb{P}(\mathbf{C}_\tau = \mathbf{k}, \Delta Y_\tau = y | \mathbf{C}_{\tau-} = \mathbf{j}, \mathcal{F}_{\tau-}) \mathbb{P}(\mathbf{C}_{\tau-} = \mathbf{j} | \mathcal{F}_{\tau-})}, \\ &= \frac{\lambda_{\tau-}^{(y)} dt}{\mathbb{P}(\mathbf{C}_\tau = \mathbf{k} | \mathcal{F}_\tau)}, \end{aligned}$$

where the first equality holds due to the Markov property of $\mathbf{C}_t$, a heuristic derivation is given later; the second and third equalities follow from the Bayes' Theorem. Since

$$\begin{aligned} \mathbb{P}(\mathbf{C}_\tau = \mathbf{j} + \mathbf{1}^{(y)}, \Delta Y_\tau = y | \mathbf{C}_{\tau-} = \mathbf{j}, \mathcal{F}_{\tau-}) &= \mathbb{P}(\Delta \mathbf{C}_\tau = \mathbf{1}^{(y)} | \mathbf{C}_{\tau-} = \mathbf{j}, \mathcal{F}_{\tau-}) \\ &= v(y) dt, \\ \mathbb{P}(\mathbf{C}_\tau = \mathbf{j} - \mathbf{1}^{(-y)}, \Delta Y_\tau = y | \mathbf{C}_{\tau-} = \mathbf{j}, \mathcal{F}_{\tau-}) &= \mathbb{P}(\Delta \mathbf{C}_\tau = -\mathbf{1}^{(-y)} | \mathbf{C}_{\tau-} = \mathbf{j}, \mathcal{F}_{\tau-}) \\ &= \phi_{j-y} dt, \end{aligned}$$

combining all of these gives us the required result.

6.3.2 Derivation of (21)

Let $\mathcal{F}_{(\tau,T]} \triangleq \sigma(\{Y_t\}_{\tau < t \leq T})$ and $\mathcal{C}_{(\tau,T]} \triangleq \sigma(\{\mathbf{C}_t\}_{\tau < t \leq T})$. Note that heuristically the Bayes' Theorem implies

$$\begin{aligned} \mathbb{P}(\mathbf{C}_{\tau-} = \mathbf{j} | \mathcal{F}_T, \mathbf{C}_\tau = \mathbf{k}) &= \mathbb{P}(\mathbf{C}_{\tau-} = \mathbf{j} | \mathcal{F}_\tau, \mathcal{F}_{(\tau, T]}, \mathbf{C}_\tau = \mathbf{k}) \\ &= \frac{\left(\frac{d\mathbb{P}}{d\mathbb{Q}}\right)_{\mathcal{F}_{(\tau, T]} | \mathcal{F}_\tau, \mathbf{C}_\tau = \mathbf{k}, \mathbf{C}_{\tau-} = \mathbf{j}}}{\left(\frac{d\mathbb{P}}{d\mathbb{Q}}\right)_{\mathcal{F}_{(\tau, T]} | \mathcal{F}_\tau, \mathbf{C}_\tau = \mathbf{k}}} \mathbb{P}(\mathbf{C}_{\tau-} = \mathbf{j} | \mathcal{F}_\tau, \mathbf{C}_\tau = \mathbf{k}). \end{aligned}$$

Since $\mathcal{F}_{(\tau, T]} \subseteq \mathcal{C}_{(\tau, T]}$ (each $Y_t = \sum_{y \in \mathbb{Z} \setminus \{0\}} C_t^{(y)}$), the Markov property of $\mathbf{C}_t$ implies

$$\left(\frac{d\mathbb{P}}{d\mathbb{Q}}\right)_{\mathcal{F}_{(\tau, T]} | \mathcal{F}_\tau, \mathbf{C}_\tau = \mathbf{k}, \mathbf{C}_{\tau-} = \mathbf{j}} = \left(\frac{d\mathbb{P}}{d\mathbb{Q}}\right)_{\mathcal{F}_{(\tau, T]} | \mathcal{F}_\tau, \mathbf{C}_\tau = \mathbf{k}},$$

because given the current information $\mathbf{C}_\tau$ the information in the past $\mathbf{C}_{\tau-}$ is irrelevant. This then proves that

$$\mathbb{P}(\mathbf{C}_{\tau-} = \mathbf{j} | \mathcal{F}_T, \mathbf{C}_\tau = \mathbf{k}) = \mathbb{P}(\mathbf{C}_{\tau-} = \mathbf{j} | \mathcal{F}_\tau, \mathbf{C}_\tau = \mathbf{k}).$$

6.4 Proof of Theorem 4

Since each $C_t^{(y)}$ is independent for different y , the complete-data log-likelihood can be written as

$$l_{\mathcal{C}_T}(\theta) = \sum_{y \in \mathbb{Z} \setminus \{0\}} l_{\mathcal{C}_T^{(y)} | C_0^{(y)}}(\theta) + \sum_{y \in \mathbb{Z} \setminus \{0\}} l_{C_0^{(y)}}(\theta),$$

where we recall that $\mathcal{C}_t^{(y)}$ is the natural filtration generated by $C_t^{(y)}$,

$$\begin{aligned} l_{\mathcal{C}_T^{(y)} | C_0^{(y)}}(\theta) &= \sum_{0 < t \leq T} \left(\log(\nu(y)) 1_{\{\Delta C_t^{(y)} = 1\}} + \log(\phi C_{t-}^{(y)}) 1_{\{\Delta C_t^{(y)} = -1\}} \right) \\ &\quad - \int_{t \in (0, T]} (\nu(y) + \phi C_{t-}^{(y)}) dt \\ &= \log(\nu(y)) N_T^{(y), A} - \nu(y) T + \log(\phi) N_T^{(y), D} - \phi \int_{t \in (0, T]} C_{t-}^{(y)} dt, \end{aligned}$$

where the first equality follows directly from Theorem 1 (ignoring the constant), and

$$l_{C_0^{(y)}}(\theta) = C_0^{(y)} (\log v(y) - \log \phi) - \frac{v(y)}{\phi}$$

because of $C_0^{(y)} \sim \text{Poisson}(v(y)/\phi)$. Thus, collecting terms will give us the required result (16). The derivations of the MCLE are elementary.

Let

$$\|v\| \triangleq \int v(dy) = \sum_{y=1}^{\infty} v(y).$$

The ergodicity of D_{t-} implies that as $T \rightarrow \infty$

$$\frac{1}{T} \int_{t \in (0, T]} D_{t-} dt \rightarrow \mathbb{E}(D_{t-}) = \frac{\|v\|}{\phi}.$$

Since $\frac{N_T^D}{T} \approx \frac{N_T^A}{T} \rightarrow \|v\|$, we have

$$\frac{\mathcal{E}_T}{T} = \frac{N_T^D}{T} - \frac{D_0 + T^{-1} \int_{t \in (0, T]} D_{t-} dt}{T} \rightarrow \|v\|, \text{ too.}$$

Thus,

$$\begin{aligned} \hat{\phi}_{\text{MCLE}} &= \frac{\frac{\mathcal{E}_T}{T} + \sqrt{\left(\frac{\mathcal{E}_T}{T}\right)^2 + 4T^{-1} \frac{N_T^A + N_T^D}{T} T^{-1} \int_{t \in (0, T]} D_{t-} dt}}{2T^{-1} \int_{t \in (0, T]} D_{t-} dt} \\ &\rightarrow \frac{\|v\| + \sqrt{\|v\|^2 + 0}}{2 \frac{\|v\|}{\phi}} = \phi. \end{aligned}$$

Finally, for any $y \in \mathbb{Z} \setminus \{0\}$, $\frac{N_T^{(y)}}{T} \rightarrow v(y)$ and $\hat{\phi}_{\text{MCLE}}^{-1} \rightarrow \phi^{-1} < \infty$, so we easily have

$$\hat{v}_{\text{MCLE}}(y) = \frac{\frac{N_T^{(y)}}{T} + \frac{C_0^{(y)}}{T}}{1 + \frac{\hat{\phi}_{\text{MCLE}}^{-1}}{T}} \rightarrow v(y) \text{ as well.}$$

6.5 Proof of Proposition 1

As $C_t^{(y)} \geq 0$, (19) implies that

$$C_0^{(y)} \geq C_{0,T}^{(y),L} = \sup_{t \in [0,T]} \left(N_t^{(-y)} - N_t^{(y)} \right), \quad y = 1, 2, \dots,$$

where we set $N_0^{(y)} \triangleq 0$ conventionally. Now

$$C_0^{(y)} = \frac{Y_0 - \sum_{y' \neq y} y' C_0^{(y')}}{y} \leq \left\lfloor \frac{Y_0 - \sum_{y' \neq y} y' C_{0,T}^{(y'),L}}{y} \right\rfloor = C_{0,T}^{(y),U},$$

so we have

$$C_{0,T}^{(y),U} \geq C_0^{(y)} \geq C_{0,T}^{(y),L}.$$

Let $N_t^{(-y),*}$ be the counting processes of $-y$ jumps resulting from the departures of the initial events of size y that constitute $C_0^{(y)}$. Let τ be the time when $N^{(-y),*}$ achieve $C_0^{(y)}$. Then we have

$$\begin{aligned} C_{0,T}^{(y),L} &= C_{0,\tau}^{(y),L} \vee \sup_{t \in (\tau, T]} \left(N_t^{(-y),*} - \left(N_t^{(y)} - \left(N_t^{(-y)} - N_t^{(-y),*} \right) \right) \right) \\ &= C_{0,\tau}^{(y),L} \vee \left(C_0^{(y)} - \inf_{t \in (\tau, T]} \left(N_t^{(y)} - \left(N_t^{(-y)} - N_t^{(-y),*} \right) \right) \right). \end{aligned}$$

Observe that $N_t^{(y)} - \left(N_t^{(-y)} - N_t^{(-y),*} \right)$ is a $M/G/\infty$ queue initiated at state 0, so by the ergodicity we must have with probability 1

$$\lim_{T \rightarrow \infty} \inf_{t \in (\tau, T]} \left(N_t^{(y)} - \left(N_t^{(-y)} - N_t^{(-y),*} \right) \right) = 0.$$

This then shows that actually

$$\lim_{T \rightarrow \infty} C_{0,T}^{(y),L} = C_{0,\tau}^{(y),L} \vee C_0^{(y)} = C_0^{(y)},$$

where the last equality follows because $C_{0,\tau}^{(y),L} \leq C_0^{(y)}$. Correspondingly,

$$\lim_{T \rightarrow \infty} C_{0,T}^{(y),U} = \left\lfloor \frac{Y_0 - \sum_{y' \neq y} y' C_0^{(y')}}{y} \right\rfloor = C_0^{(y)}.$$

References

1. Asmussen, S.: *Applied Probability and Queues*. Springer, New York (2003)
2. Barndorff-Nielsen, O.E.: Stationary infinitely divisible processes. *Braz. J. Probab. Stat.* **25**, 294–322 (2011)
3. Barndorff-Nielsen, O.E., Schmiegel, J.: Ambit processes; with applications to turbulence and tumour growth. In: Benth, F.E., Di Nunno, G., Lindstrøm, T., Øksendal, B., Zhang, T. (eds.) *Stochastic Analysis and Applications*, pp. 93–124. Springer, Berlin Heidelberg (2007)
4. Barndorff-Nielsen, O.E., Benth, F.E., Veraart, A.E.D.: Ambit processes and stochastic partial differential equations. In: Di Nunno, G., Øksendal, B. (eds.) *Advanced Mathematical Methods for Finance*, pp. 35–74. Springer, Berlin (2011)
5. Barndorff-Nielsen, O.E., Pollard, D.G., Shephard, N.: Integer-valued Lévy processes and low latency financial econometrics. *Quant. Finan.* **12**, 587–605 (2012)
6. Barndorff-Nielsen, O.E., Benth, F.E., Veraart, A.E.D.: Recent advances in ambit stochastics with a view towards tempo-spatial stochastic volatility, intermittency. ArXiv e-prints. Unpublished paper. Department of Mathematics, Imperial College London (2012)
7. Barndorff-Nielsen, O.E., Lunde, A., Shephard, N., Veraart, A.E.D.: Integer-valued trawl processes: a class of stationary infinitely divisible processes. *Scand. J. Stat.* **41**, 693–724 (2014)
8. Bartlett, M.S.: *An Introduction to Stochastic Processes, with Special Reference to Methods and Applications*, 3rd edn. Cambridge University Press, Cambridge (1978)
9. Cameron, C.A., Trivedi, P.K.: *Regression Analysis of Count Data*. Cambridge University Press, Cambridge (1998)
10. Cui, Y., Lund, R.: A new look at time series of counts. *Biometrika* **96**, 781–792 (2009)
11. Daley, D., Vere-Jones, D.: *Evolutionary processes and predictability. An Introduction to the Theory of Point Processes. Probability and Its Applications*, pp. 355–456. Springer, New York (2008)
12. Davis, R.A., Wu, R.: A negative binomial model for time series of counts. *Biometrika* **96**, 735–749 (2009)
13. Fokianos, K., Kedem, B.: Regression theory for categorical time series. *Stat. Sci.* **18**, 357–376 (2003)
14. Jacobs, P.A., Lewis, P.A.W.: Discrete time series generated by mixtures. I: correlational and runs properties. *J. Roy. Stat. Soc. Ser. B (Methodological)* **40**, 94–105 (1978)
15. Jung, R., Tremayne, A.: Useful models for time series of counts or simply wrong ones? *AStA Adv. Stat. Anal.* **95**, 59–91 (2011)
16. Lindley, D.V.: The estimation of velocity distributions from counts. In: *Proceedings of the International Congress of Mathematicians*, vol. 3, pp. 427–444. North-Holland, Amsterdam (1956)
17. McKenzie, D.J.: Measuring inequality with asset indicators. *J. Popul. Econ.* **18**, 229–260 (2005)
18. Reynolds, J.F.: On the autocorrelation and spectral functions of queues. *J. Appl. Probab.* **5**, 467–475 (1968)
19. Rudemo, M.: State estimation for partially observed Markov chains. *J. Math. Anal. Appl.* **44**, 581–611 (1973)
20. Rudemo, M.: Prediction and smoothing for partially observed Markov chains. *J. Math. Anal. Appl.* **49**, 1–23 (1975)
21. Shephard, N., Yang, J.J.: Continuous time analysis of fleeting discrete price moves. ArXiv e-prints. Unpublished paper. Department of Statistics, Harvard University (2014)
22. Surgailis, D., Rosinski, J., Mandrekar, V., Cambanis, S.: Stable mixed moving averages. *Probab. Theory Relat. Fields* **97**, 543–558 (1993)
23. Weiß, C.: Thinning operations for modeling time series of counts—a survey. *AStA Adv. Stat. Anal.* **92**, 319–341 (2008)
24. Wolpert, R.L., Brown, L.D.: Stationary infinitely-divisible Markov processes with non-negative integer values. Working paper, Department of Statistics, Duke University (2011)

25. Wolpert, R.L., Taqqu, M.S.: Fractional Ornstein-Uhlenbeck Lévy processes and the telecom process: upstairs and downstairs. *Sig. Process.* **85**, 1523–1545 (2005)
26. Zhu, R., Joe, H.: A new type of discrete self-decomposability and its application to continuous-time Markov processes for modeling count data time series. *Stoch. Models* **19**, 235–254 (2003)

The Different Asymptotic Regimes of Nearly Unstable Autoregressive Processes

Thibault Jaisson and Mathieu Rosenbaum

Abstract We extend the results of [14, 27, 29] about the convergence of nearly unstable $\text{AR}(p)$ processes to the infinite order case. To do so, we proceed as in [19, 20] by using limit theorems for some well chosen geometric sums. We prove that when the coefficients sequence has a light tail, nearly unstable $\text{AR}(\infty)$ processes behave as Ornstein-Uhlenbeck models. However, in the heavy tail case, we show that fractional diffusions arise as limiting laws for such processes.

Keywords Autoregressive processes · $\text{AR}(\infty)$ · Nearly unstable processes · Limit theorems · Ornstein-Uhlenbeck processes · Fractional diffusions · Volatility modeling

1 Introduction

In the field of time series analysis, autoregressive processes (AR processes for short) probably represent the most classical class of models. In this work, a discrete time process y is said to be autoregressive if it satisfies $y_0 = \varepsilon_0$ and for $k \geq 1$,

$$y_k = \varepsilon_k + \sum_{i=1}^k \phi_i y_{k-i},$$

where ϕ is a sequence of non-negative coefficients and the ε_k are iid centered random variables with finite second order moment. The process is called $\text{AR}(\infty)$ if an infinite number of coefficients are non-zero and $\text{AR}(p)$ if they are all equal to zero after rank p . A first reason for the popularity of AR processes is the fact that they are quite

T. Jaisson
CMAP, École Polytechnique, Palaiseau, France
e-mail: thibault.jaisson@polytechnique.edu

M. Rosenbaum (✉)
LPMA, Université Marie et Pierre Curie (Paris 6), Paris, France
e-mail: mathieu.rosenbaum@upmc.fr; mathieu.rosenbaum@polytechnique.edu

tractable. Furthermore, it is usually easy to give an interpretation for such dynamics in practice. That is why they are used in various fields such as population dynamics, see [14, 28], finance, see [17], or telecommunications, see [1].¹

For classical AR processes starting at $-\infty$ and with non-negative coefficients, it is well known that a necessary condition for the existence of a stationary solution is

$$\sum_{i=1}^{+\infty} \phi_i < 1. \quad (1)$$

Thus, (1) is called stability condition. However, in many applications, notably finance, the case where this stability condition is almost saturated:

$$0 \leq 1 - \sum_{i=1}^{+\infty} \phi_i \ll 1$$

seems to be relevant, see [14, 24].

To address this near instability situation, several authors have considered the framework of sequences of AR(1) processes of the form

$$y_k^n = \rho_n y_{k-1}^n + \varepsilon_k,$$

with the autoregression parameter ρ_n tending to one as n goes to infinity, see for example [14, 25, 27]. In these works, it is shown that after suitable renormalization, such processes asymptotically behave as Ornstein-Uhlenbeck models. The estimation of the AR coefficient has also been extensively studied in such an asymptotic, see [7, 9–12] for more recent related developments. The situation where the number of coefficients in the autoregressive process is larger than one is treated in [8, 21, 29]. However, to our knowledge, no result is available in the case of nearly unstable AR(∞) processes. The aim of this work is to fill this gap, using a methodology originally developed for Hawkes processes in [19, 20].

A Hawkes process (N_t) is a point process whose intensity at time t , denoted by λ_t , is of the form

$$\lambda_t = \mu + \int_{-\infty}^t \phi(t-s) dN_s,$$

with μ a constant and ϕ a non-negative measurable function, see [18]. The linearity in the previous equation implies that Hawkes processes have a common structure with autoregressive processes. In particular, under the stability condition that the L^1 norm of ϕ , denoted by $|\phi|_{L^1}$, is strictly smaller than one, Hawkes processes admit a stationary intensity. Let us also mention that the counterpart of the Yule-Walker equations linking the autocorrelations of an AR process to its coefficients is the Wiener-Hopf equation, see [4, 18].

¹Of course the references given here are just some examples among many others.

In the recent papers [19, 20], nearly unstable Hawkes processes are introduced and their asymptotic behavior is investigated. Such processes are defined as Hawkes processes for which the stability condition is almost saturated: $1 - |\phi|_{L^1} \ll 1$. The key idea to understand the limiting laws of nearly unstable Hawkes processes is to relate their dynamics with some geometric sums whose parameter tends to zero. In this work, we follow the same strategy in order to derive asymptotic results for nearly unstable $\text{AR}(\infty)$ processes.

We show that the limiting behavior of nearly unstable $\text{AR}(\infty)$ processes strongly depends on the properties of the sequence $\phi = (\phi_1, \phi_2, \dots)$. Two main situations have to be considered: the light tail case, where ϕ_i goes rapidly to zero (or is even equal to zero for i large enough) and the heavy tail case, where the decay of the sequence ϕ is slow. We prove that in the first situation, the cumulated process asymptotically behaves as an integrated Ornstein-Uhlenbeck model, in agreement with the results obtained in the literature for $\text{AR}(p)$ processes. However, in the heavy tail case, a fractional limit is obtained. More precisely, the renormalized cumulated $\text{AR}(\infty)$ process behaves as a kind of integrated fractional Brownian motion with Hurst parameter smaller than $1/2$, or as a process close to a fractional Brownian motion with Hurst parameter larger than $1/2$, depending on the decay rate of the sequence ϕ .

The paper is organized as follows. In Sect. 2, we introduce the geometric sums and derive their limiting behavior as their parameter tends to zero. We describe our asymptotic framework in Sect. 3. Section 4 contains our main results about the limiting laws of nearly unstable $\text{AR}(\infty)$ processes and their proofs. Finally, we interpret these results in terms of volatility modeling in Sect. 5.

2 Geometric Sums

In this section, we consider the asymptotic behavior of some geometric sums. Understanding these sums is actually the key point in order to be able to derive limiting laws for nearly unstable $\text{AR}(\infty)$ processes.

2.1 Geometric Sums and Convolution

Let $(a_n)_{n \geq 1}$ and $(\phi_i)_{i \geq 0}$ be two sequences of non-negative numbers such that $a_n < 1$, $\phi_0 = 0$ and

$$\sum_{i=1}^{+\infty} \phi_i = 1.$$

In the next sections, a_n will correspond to the sum of the autoregression coefficients and the ϕ_i to the (fixed) shape of these coefficients, see Sect. 3.

In this work, geometric sums, see [22], are random variables of the form

$$Y^n = \sum_{i=1}^{I^n} X^i,$$

where I^n is a geometric random variable with parameter $1 - a_n$ ² (with the convention $\sum_1^0 = 0$) and (X^i) is a sequence of strictly positive integer-valued iid random variables such that for any $i \geq 1$,

$$\mathbb{P}[X^1 = i] = \phi_i.$$

Now we define the sequence ψ^n by

$$\psi^n = \sum_{l=0}^{+\infty} (a_n \phi)^{*l},$$

with ϕ^{*0} the sequence indexed by $\mathbb{N}$ such that $\phi_0^{*0} = 1$ and for $i > 0$, $\phi_i^{*0} = 0$ and for $n \geq 0$, $\phi^{*(n+1)} = \phi^{*n} * \phi$, where $*$ denotes the convolution operator for sequences. Note that

$$\sum_{i=0}^{+\infty} \psi_i^n = \frac{1}{1 - a_n}.$$

Actually, the random variable Y^n and the sequence ψ^n are closely related since for any $i \geq 0$,

$$\mathbb{P}[Y^n = i] = \frac{\psi_i^n}{\sum_{j \geq 0} \psi_j^n}.$$

We will use the limiting behavior of Y^n to obtain asymptotic properties for the sequence ψ^n , see Corollaries 1 and 2. This will provide us key results for our study of nearly unstable AR(∞) processes at large time scales. Indeed, in the next sections, for n tending to infinity, we consider on time intervals of the form $[0, nt]$ autoregressive processes whose coefficients sequence can be written $a_n \phi$:

$$y_k^n = \varepsilon_k + \sum_{i=1}^k a_n \phi_i y_{k-i}^n,$$

with a_n tending to one (nearly unstable case). Then, it is easy to obtain the following moving average representation of the process:

² $\forall i \geq 0$, $\mathbb{P}[I^n = i] = a_n^i (1 - a_n)$.

$$y_k^n = \sum_{i=0}^k \psi_{k-i}^n \varepsilon_i$$

and in particular

$$y_{[nt]}^n = \sum_{i=0}^{[nt]} \psi_{[nt]-i}^n \varepsilon_i.$$

Thanks to the previous representation, we see that the long term behavior of y^n is linked with the asymptotic properties of the $\psi_{[nt]-i}^n$ and therefore with those of Y^n/n . Indeed, for example, for nt an integer,

$$\mathbb{P}[Y^n/n = t] = \frac{\psi_{nt}^n}{\sum_{j \geq 0} \psi_j^n}.$$

Remark 1 Note that our approach being based on geometric sums, the autoregression coefficients have to be non-negative (so that they can be seen as density weights). We thus cannot retrieve as general asymptotic behaviors as in [21] where the roots of the polynomial can be asymptotically close to any point of the unit circle.

We now recall some results from [19, 20] which explain the two possible types of asymptotic behaviors for our geometric sums. Exhibiting one behavior or the other depends on the decay rate of the coefficients of ϕ , that is on the tail of the distribution of X^1 .

2.2 Light Tail Case

The first asymptotic regime of interest is the one which occurs when the expectation of X^1 is finite. Thus we consider the following assumption.

Assumption 1 We have

$$\sum_{i=1}^{+\infty} i \phi_i = m < +\infty.$$

In that situation, the “appropriate” speed of convergence of a_n towards one for the nearly unstable case is given in the next assumption.

Assumption 2 We have $a_n < 1$ and there exists $\lambda > 0$ such that

$$n(1 - a_n) \xrightarrow{n \rightarrow +\infty} \lambda.$$

Indeed, recall Proposition 2.2 of [19].

Proposition 1 *Under Assumptions 1 and 2, the random variable Y^n/n converges in law towards an exponential random variable with parameter λ/m .*

Using Dini's theorem, this implies the next corollary.

Corollary 1 *Under Assumptions 1 and 2,*

$$\sup_{x \in [0,1]} \left| \sum_{i=0}^{\lfloor nx \rfloor} \psi_i^n(1 - a_n) - (1 - e^{-\frac{\lambda}{m}x}) \right| \xrightarrow{n \rightarrow +\infty} 0.$$

2.3 Heavy Tail Case

The second asymptotic regime of interest is the one which occurs when the tail of the distribution of X^1 is heavy. This can be formalized as follows.

Assumption 3 There exists $K > 0$ and $\alpha \in (0, 1)$ such that

$$\sum_{i=N}^{+\infty} \phi_i \underset{N \rightarrow +\infty}{\sim} \frac{K}{N^\alpha}.$$

Let us now write

$$\delta = K \Gamma(1 - \alpha) / \alpha,$$

with Γ the Gamma function. In that heavy tail case, the “appropriate” convergence speed of a_n towards one is given in the next assumption.

Assumption 4 We have $a_n < 1$ and there exists $\lambda > 0$ such that

$$n^\alpha (1 - a_n) \xrightarrow{n \rightarrow +\infty} \lambda \delta.$$

Indeed, recall Proposition 2.3 of [19].

Proposition 2 *Under Assumptions 3 and 4, the random variable Y^n/n converges in law towards the random variable whose Laplace transform is*

$$\frac{\lambda}{\lambda + z^\alpha}.$$

It is explained in [19] that this random variable has the density

$$f^{\alpha, \lambda}(x) = \lambda x^{\alpha-1} E_{\alpha, \alpha}(-\lambda x^\alpha),$$

where $E_{\alpha, \alpha}$ is the Mittag-Leffler function, see [26] for definition.

Writing

$$F^{\alpha,\lambda}(x) = \int_0^x f^{\alpha,\lambda}(s)ds$$

and using Dini's theorem, we have the following result.

Corollary 2 *Under Assumptions 3 and 4,*

$$\sup_{x \in [0,1]} \left| \sum_{i=0}^{\lfloor nx \rfloor} \psi_i^n (1 - a_n) - F^{\alpha,\lambda}(x) \right| \xrightarrow{n \rightarrow +\infty} 0.$$

3 The Asymptotic Setting

As previously explained, we consider a sequence of autoregressive processes of infinite order indexed by n . More precisely, $y_0^n = \varepsilon_0$ and for any $k \geq 1$,

$$y_k^n = \varepsilon_k + \sum_{i=1}^k \phi_i^n y_{k-i}^n,$$

where the ϕ_i^n are non-negative coefficients. We assume that the ε_k do not depend on n , are iid centered, and satisfy (for simplicity)

$$\mathbb{E}[(\varepsilon_k)^2] = 1.$$

We are interested in the long term behavior of the cumulative sums of such autoregressive processes when the stability condition (1) is almost saturated. To study this, somehow as in [19, 20], we consider that the “shape” of the coefficients sequence is fixed and that their L^1 norm a_n tends to one from below. More precisely, we take

$$\phi_i^n = a_n \phi_i,$$

where

$$\sum_{i=1}^{+\infty} \phi_i = 1$$

and $a_n \rightarrow 1$ as n tends to infinity, with $a_n < 1$.

Since we want to investigate the asymptotic properties of cumulated sums of such processes, we introduce the suitably renormalized Donsker line for $t \in [0, 1]$:

$$Z_t^n = \frac{1 - a_n}{\sqrt{n}} \left(\sum_{j=0}^{\lfloor nt \rfloor} y_j^n + (nt - \lfloor nt \rfloor) y_{\lfloor nt \rfloor + 1}^n \right).$$

4 Main Results

4.1 The Light Tail Case

We give here the behavior of the Donsker line when the autoregression coefficients converge rapidly to zero. Note that the next theorem covers in particular the case of $\text{AR}(p)$ processes already studied in the literature.

Theorem 1 *Under Assumptions 1 and 2, (Z^n) converges in law towards the process Z defined by*

$$Z_t = \int_0^t (1 - e^{-\frac{\lambda}{m}(t-s)}) dW_s = \int_0^t \int_0^s \frac{\lambda}{m} e^{-\frac{\lambda}{m}(s-u)} dW_u ds,$$

where W is a Brownian motion.

This theorem means that when observed on a time scale of order $1/(1 - a_n)$, a cumulated nearly unstable $\text{AR}(\infty)$ process whose coefficients sequence has a light tail behaves as an integrated Ornstein-Uhlenbeck process.

To show this result, the first step is to get the following proposition.

Proposition 3 *Under Assumptions 1 and 2, the sequence (Z^n) is tight.*

Proof of Proposition 3

In the sequel, c denotes a positive constant which may vary from line to line. Before proving Proposition 3, we need a technical lemma.

Lemma 1 *There exists $c > 0$ such that for any n, k, k' ,*

$$\mathbb{E}[y_k^n y_{k'}^n] \leq \frac{c}{1 - a_n}.$$

Proof For any k , using the moving average representation of the process, since ε is a white noise, we get

$$\mathbb{E}[(y_k^n)^2] = \sum_{j=0}^k (\psi_{k-j}^n)^2 \leq \sum_{j=0}^{+\infty} (\psi_j^n)^2.$$

Moreover, using Parseval's Theorem,³

$$\begin{aligned}\sum_{j=0}^{+\infty} (\psi_j^n)^2 &= 2 \int_0^{1/2} |\widehat{\psi}^n(z)|^2 dz \\ &\leq 2 \int_0^{1/2} \frac{1}{|1 - a_n|^2} dz,\end{aligned}$$

where we have used that

$$|\widehat{\psi}^n(z)| = \left| \sum_{k \geq 0} (a_n \widehat{\phi}(z))^k \right| = \frac{1}{|1 - a_n \widehat{\phi}(z)|} \leq |\widehat{\psi}^n(0)| = \frac{1}{|1 - a_n|}.$$

□

Let us now prove Proposition 3.

The result follows from an application of Kolmogorov's criterion, see [6]. We show here that there exists $c > 0$ such that for any n, s and t ,

$$\mathbb{E}[|Z_t^n - Z_s^n|^2] \leq c|t - s|^2.$$

We begin with the case where $\lfloor nt \rfloor = \lfloor ns \rfloor$. Using Lemma 1,

$$\begin{aligned}\mathbb{E}[|Z_t^n - Z_s^n|^2] &= \frac{(1 - a_n)^2}{n} n^2 |t - s|^2 \mathbb{E}[(y_{\lfloor nt \rfloor + 1}^n)^2] \\ &\leq cn(1 - a_n)|t - s|^2.\end{aligned}$$

Using Assumption 2, this ends the proof for $\lfloor nt \rfloor = \lfloor ns \rfloor$.

In the case $\lfloor nt \rfloor = \lfloor ns \rfloor + 1$,

$$\begin{aligned}\mathbb{E}[|Z_t^n - Z_s^n|^2] &\leq c(\mathbb{E}[|Z_t^n - Z_{\lfloor nt \rfloor/n}^n|^2] + \mathbb{E}[|Z_{(\lfloor ns \rfloor + 1)/n}^n - Z_s^n|^2]) \\ &\leq c|t - s|^2 + \frac{(1 - a_n)^2}{n} (1 - ns + \lfloor ns \rfloor)^2 \mathbb{E}[(y_{\lfloor ns \rfloor + 1}^n)^2] \\ &\leq c|t - s|^2 + \frac{(1 - a_n)}{n} (\lfloor nt \rfloor - ns)^2 \\ &\leq c|t - s|^2.\end{aligned}$$

³In this work, the Fourier transform of a sequence (f_n) is defined for $z \in [-1/2, 1/2]$ as $\widehat{f}(z) = \sum_{k \geq 0} f_k e^{-2\pi i k z}$.

We now treat the case where $t = k^t/n$ and $s = k^s/n$, where k^t and k^s are integers so that $k^t > k^s$. Using again Lemma 1 together with Assumption 2, we have

$$\begin{aligned} \mathbb{E}[|Z_t^n - Z_s^n|^2] &= \frac{(1 - a_n)^2}{n} \sum_{k_1=k^s+1}^{k^t} \sum_{k_2=k^s+1}^{k^t} \mathbb{E}[y_{k_1}^n y_{k_2}^n] \\ &\leq \frac{(1 - a_n)^2}{n} (k^t - k^s)^2 \frac{c}{1 - a_n} \\ &\leq c(1 - a_n)n \left(\frac{k^t - k^s}{n} \right)^2 \leq c|t - s|^2. \end{aligned}$$

Finally, for any $t > s$ so that $\lfloor nt \rfloor \geq \lfloor ns \rfloor + 2$, we use the decomposition

$$\begin{aligned} \mathbb{E}[|Z_t^n - Z_s^n|^2] &\leq c(\mathbb{E}[|Z_t^n - Z_{\lfloor nt \rfloor/n}^n|^2] + \mathbb{E}[|Z_{\lfloor nt \rfloor/n}^n - Z_{(\lfloor ns \rfloor + 1)/n}^n|^2] \\ &\quad + \mathbb{E}[|Z_{(\lfloor ns \rfloor + 1)/n}^n - Z_s^n|^2]). \end{aligned}$$

The tightness follows. This ends the proof of Proposition 3.

Then we need to prove the finite dimensional convergence. Using the moving average representation of the process, we derive

$$\begin{aligned} Z_t^n &= \sum_{i=0}^{\lfloor nt \rfloor} (1 - a_n) \left(\sum_{k=0}^{\lfloor nt \rfloor - i} \psi_k^n + (nt - \lfloor nt \rfloor) \psi_{\lfloor nt \rfloor + 1 - i}^n \right) \frac{\varepsilon_i}{\sqrt{n}} \\ &\quad + (nt - \lfloor nt \rfloor)(1 - a_n) \frac{\varepsilon_{\lfloor nt \rfloor + 1}}{\sqrt{n}}. \end{aligned}$$

This can be rewritten under the following integral form:

$$\begin{aligned} Z_t^n &= \int_0^{\lfloor nt \rfloor/n} (1 - a_n) \left(\sum_{k=0}^{\lfloor nt \rfloor - \lfloor ns \rfloor} \psi_k^n + (nt - \lfloor nt \rfloor) \psi_{\lfloor nt \rfloor + 1 - \lfloor ns \rfloor}^n \right) dW_s^n \\ &\quad + (nt - \lfloor nt \rfloor)(1 - a_n) \frac{\varepsilon_{\lfloor nt \rfloor + 1}}{\sqrt{n}}, \end{aligned}$$

where

$$W_s^n = \frac{1}{\sqrt{n}} \sum_{k=0}^{\lfloor ns \rfloor} \varepsilon_k.$$

From Donsker theorem in Skorohod space, we have that W_s^n converges in law towards a Brownian motion for the Skorohod topology. Furthermore, note that

$$(1 - a_n) \left(\sum_{k=0}^{\lfloor nt \rfloor - \lfloor ns \rfloor} \psi_k^n + (nt - \lfloor nt \rfloor) \psi_{\lfloor nt \rfloor + 1 - \lfloor ns \rfloor}^n \right)$$

belongs to

$$\left[(1 - a_n) \sum_{k=0}^{\lfloor nt \rfloor - \lfloor ns \rfloor} \psi_k^n, (1 - a_n) \sum_{k=0}^{\lfloor nt \rfloor + 1 - \lfloor ns \rfloor} \psi_k^n \right].$$

Thus, from Corollary 1, as a function of s ,

$$(1 - a_n) \left(\sum_{k=0}^{\lfloor nt \rfloor - \lfloor ns \rfloor} \psi_k^n + (nt - \lfloor nt \rfloor) \psi_{\lfloor nt \rfloor + 1 - \lfloor ns \rfloor}^n \right)$$

tends uniformly to $1 - e^{-\frac{\lambda}{m}(t-s)}$ on $[0, t]$. Also, we obviously get that

$$(nt - \lfloor nt \rfloor)(1 - a_n)^{\frac{\varepsilon_{\lfloor nt \rfloor + 1}}{\sqrt{n}}}$$

tends to zero.

Then, using Theorem 2.2 of [23] together with the fact that Skorohod convergence implies pointwise convergence at continuity points of the limit, for given $t \in [0, 1]$, we get the convergence in law of Z_t^n towards

$$\int_0^t (1 - e^{-\frac{\lambda}{m}(t-s)}) dW_s.$$

With the help of Cramer-Wold device, it is easy to extend this result and to show that for any $(t_1, \dots, t_k) \in [0, 1]^k$, we have the convergence in law of $(Z_{t_1}^n, \dots, Z_{t_k}^n)$ towards

$$\left(\int_0^{t_1} (1 - e^{-\frac{\lambda}{m}(t_1-s)}) dW_s, \dots, \int_0^{t_k} (1 - e^{-\frac{\lambda}{m}(t_k-s)}) dW_s \right).$$

Together with the tightness of (Z^n) , this enables us to obtain the weak converges of (Z^n) towards Z .

4.2 The Heavy Tail Case

Let us now place ourselves under Assumption 3 which states that the coefficients sequence has a power law type behavior. Then, using the geometric sums interpretation, we see that the natural “observation scale” of the process is of order $(1 - a_n)^{-1/\alpha}$. This corresponds to Assumption 4. We have the following result.

Theorem 2 Under Assumptions 3 and 4, (Z^n) converges in law towards the process Z defined by

$$Z_t = \int_0^t F^{\alpha, \lambda}(t-s) dW_s,$$

where W is a Brownian motion.

In Sect. 5, we will see that for $\alpha > 1/2$, the limiting process can be viewed as an integrated rough fractional process, whereas for $\alpha < 1/2$, it is close to a fractional Brownian motion with Hurst parameter larger than $1/2$.

To obtain Theorem 2, the same strategy as for the proof of Theorem 1 is used. In particular, we consider the following proposition.

Proposition 4 Under Assumptions 3 and 4, the sequence (Z^n) is tight.

Proof of Proposition 4

Before proving Proposition 4, we need some technical results.

Lemma 2 There exists $c > 0$ such that for any $|z| \leq 1/2$,

$$|1 - \widehat{\phi}(z)| \geq c|z|^\alpha.$$

Proof Since Assumption 3 is satisfied, there exists $c' \neq 0$ and $\delta > 0$ such that

$$1 - \widehat{\phi}(z) \underset{z \rightarrow 0}{\sim} c' z^\alpha$$

and

$$\forall |z| \leq \delta, |1 - \widehat{\phi}(z)| \geq |c'| |z|^\alpha / 2.$$

On $[\delta, 1/2]$, $z \mapsto |1 - \widehat{\phi}(z)|/|z|^\alpha$ is continuous and therefore has a minimum attained in some z_0 : $|1 - \widehat{\phi}(z_0)|/|z_0|^\alpha = M$.

M is strictly positive since $\operatorname{Re}(\widehat{\phi}(z_0)) = \mathbb{E}[\cos(2\pi z_0 X^1)]$, which is strictly smaller than one because, using Assumption 3, $z_0 X^1$ does not almost surely belong to $\mathbb{N}$.

Taking $c = \min(|c'|/2, M)$ ends the proof. $\square$

Lemma 3 For $\alpha > 1/2$, there exists $c > 0$ such that for any n, k, k' ,

$$\mathbb{E}[y_k^n y_{k'}^n] \leq c(1 - a_n)^{1/\alpha-2}.$$

Proof As before, we have

$$\mathbb{E}[(y_k^n)^2] \leq 2 \int_0^{1/2} \frac{1}{|1 - a_n \widehat{\phi}(z)|^2} dz.$$

Therefore, for n large enough, $a_n > 1/2$ and

$$|1 - a_n \widehat{\phi}(z)|^2 = |1 - a_n + a_n(1 - \widehat{\phi}(z))|^2 \geq |1 - a_n|^2 + \frac{|1 - \widehat{\phi}(z)|^2}{4}.$$

Using Lemma 2, this implies that

$$\mathbb{E}[(y_k^n)^2] \leq \int_0^{1/2} \frac{c}{|1 - a_n|^2 + |z|^{2\alpha}} dz.$$

Therefore,

$$\mathbb{E}[(y_k^n)^2] \leq \int_0^{(1-a_n)^{1/\alpha}} \frac{c}{|1 - a_n|^2} dz + \int_{(1-a_n)^{1/\alpha}}^{1/2} \frac{c}{|z|^{2\alpha}} dz,$$

which ends the proof. $\square$

Lemma 4 For $\alpha < 1/2$, there exists $c > 0$ such that for any n, k, k' ,

$$\mathbb{E}[y_k^n y_{k'}^n] \leq \begin{cases} c(1 - a_n)^{2\alpha-1} & \text{if } k = k' \\ c|k - k'|^{2\alpha-1} & \text{if } k \neq k'. \end{cases}$$

Proof The fact that

$$\mathbb{E}[(y_k^n)^2] \leq c(1 - a_n)^{2\alpha-1}$$

is a direct consequence of the proof of Lemma 3 together with the inequality $1 - 2\alpha \geq 2 - 1/\alpha$ for $\alpha < 1/2$.

To prove the other inequality, remark that using the moving average representation of the process, for $k \geq k'$, we get

$$\mathbb{E}[y_k^n y_{k'}^n] = \sum_{j=0}^{k'} \psi_{k-j}^n \psi_{k'-j}^n \leq \sum_{j=0}^{+\infty} \psi_j^n \psi_{j+k-k'}^n.$$

Therefore, using Parseval's theorem together with Lemma 2 and the fact that the ψ_j are non-negative, we obtain

$$\mathbb{E}[y_k^n y_{k'}^n] \leq c \operatorname{Re} \left(\int_0^{1/2} \frac{e^{2\pi i |k-k'|z}}{|1 - a_n \widehat{\phi}(z)|^2} dz \right) \leq c \operatorname{Re} \left(\int_0^{1/2} \frac{e^{2\pi i |k-k'|z}}{|z|^{2\alpha}} dz \right).$$

Thus, using Abel's theorem, we finally obtain

$$\mathbb{E}[y_k^n y_{k'}^n] \leq \frac{c}{|k - k'|^{1-2\alpha}}.$$

□

Lemma 5 *For $\alpha = 1/2$, for any $\varepsilon \in (0, 1)$, there exists $c > 0$ such that for any n, k, k' ,*

$$\mathbb{E}[y_k^n y_{k'}^n] \leq \begin{cases} c(1 - a_n)^{-\varepsilon} & \text{if } k = k' \\ c + c((1 - a_n)^2 |k - k'|)^{-\varepsilon} & \text{if } k \neq k'. \end{cases}$$

Proof Using the same proof as for Lemma 3, we get

$$\mathbb{E}[(y_k^n)^2] \leq c|\log(1 - a_n)|.$$

The first inequality follows. We obtain the second inequality using that for $k > k'$,

$$\begin{aligned} \mathbb{E}[y_k^n y_{k'}^n] &\leq c \operatorname{Re} \left(\int_0^{1/2} \frac{e^{2\pi i |k - k'| z}}{(1 - a_n)^2 + |z|} dz \right) \\ &\leq c \int_0^{(1 - a_n)^2} \frac{1}{(1 - a_n)^2} dz + c \operatorname{Re} \left(\int_{(1 - a_n)^2}^{1/2} \frac{e^{2\pi i |k - k'| z}}{|z|} dz \right) \\ &\leq c + c |\log((1 - a_n)^2 |k - k'|)|. \end{aligned}$$

□

Let us now prove Proposition 4.

In the case where $\alpha > 1/2$, the proof that

$$\mathbb{E}[|Z_t^n - Z_s^n|^2] \leq c|t - s|^2$$

is almost the same as in the light tail case replacing the use of Lemma 1 by that of Lemma 3.

The case $\alpha < 1/2$ is slightly more complicated. We now show that

$$\mathbb{E}[|Z_t^n - Z_s^n|^2] \leq c|t - s|^{1+\eta},$$

for some $\eta > 0$. As before, we begin with the case where $\lfloor nt \rfloor = \lfloor ns \rfloor$. Using Lemma 4 together with Assumption 4 and the fact that $1 - \alpha(2\alpha + 1) \geq 0$, we get

$$\begin{aligned}
\mathbb{E}[|Z_t^n - Z_s^n|^2] &= \frac{(1 - a_n)^2}{n} n^2 |t - s|^2 \mathbb{E}[(y_{\lfloor nt \rfloor + 1}^n)^2] \\
&\leq c \frac{(1 - a_n)^2}{n} n^2 |t - s|^{1+\alpha(2\alpha+1)} \frac{1}{n^{1-\alpha(2\alpha+1)}} (1 - a_n)^{2\alpha-1} \\
&\leq c |t - s|^{1+\alpha(2\alpha+1)}.
\end{aligned}$$

Using the same arguments as for the light tail case, we get a similar bound for $\lfloor nt \rfloor = \lfloor ns \rfloor + 1$.

We now treat the case where $t = k^t/n$ and $s = k^s/n$, where k^t and k^s are integers so that $k^t > k^s$. Using again Lemma 4, we have

$$\mathbb{E}[|Z_t^n - Z_s^n|^2] = \frac{(1 - a_n)^2}{n} \sum_{k_1=k^s+1}^{k^t} \sum_{k_2=k^s+1}^{k^t} \mathbb{E}[y_{k_1}^n y_{k_2}^n].$$

This is smaller than (with obvious notation)

$$\frac{(1 - a_n)^2}{n} \left(\frac{c(k^t - k^s)}{(1 - a_n)^{1-2\alpha}} + c \sum_{\Delta k=1}^{k^t - k^s - 1} \frac{1}{(\Delta k)^{1-2\alpha}} \#\{(k_1, k_2) \in [k^s + 1, k^t]; |k_1 - k_2| = \Delta k\} \right).$$

Therefore, using Assumption 4, we get

$$\begin{aligned}
\mathbb{E}[|Z_t^n - Z_s^n|^2] &\leq \frac{(1 - a_n)^2}{n} \left(\frac{c(k^t - k^s)}{(1 - a_n)^{1-2\alpha}} + c \sum_{\Delta k=1}^{k^t - k^s - 1} \frac{1}{(\Delta k)^{1-2\alpha}} (k^t - k^s) \right) \\
&\leq \frac{(1 - a_n)^2}{n} \left(\frac{c(k^t - k^s)}{(1 - a_n)^{1-2\alpha}} + c(k^t - k^s)^{1+2\alpha} \right) \\
&\leq c((t - s)^{1+\alpha+2\alpha^2} + (t - s)^{1+2\alpha}).
\end{aligned}$$

The result for any s and t in $[0, 1]$ is obtained as in the light tail case.

Finally, the proof for $\alpha = 1/2$ is obtained the same way, using Lemma 5 instead of Lemma 4. This ends the proof of Proposition 4.

The end the proof of Theorem 2 follows as previously, using the decomposition

$$\begin{aligned}
Z_t^n &= \int_0^{\lfloor nt \rfloor / n} (1 - a_n) \left(\sum_{k=0}^{\lfloor nt \rfloor - \lfloor ns \rfloor} \psi_k^n + (nt - \lfloor nt \rfloor) \psi_{\lfloor nt \rfloor + 1 - \lfloor ns \rfloor}^n \right) dW_s^n \\
&\quad + (nt - \lfloor nt \rfloor) (1 - a_n) \frac{\varepsilon_{\lfloor nt \rfloor + 1}}{\sqrt{n}},
\end{aligned}$$

the uniform convergence of

$$(1 - a_n) \sum_{k=0}^{\lfloor nt \rfloor} \psi_k^n$$

towards $F^{\alpha,\lambda}(t)$ (see Corollary 2), and the convergence of W_s^n towards a Brownian motion.

5 Application to Volatility Modeling

Let us now interpret Theorem 2 in terms of volatility modeling on financial markets. Consider that the log-volatility process is driven at discrete times by a nearly unstable autoregressive process with heavy tailed coefficients sequence. We choose to model the log-volatility rather than the volatility itself because it is well established that it is better approximated by a linear autoregressive process than the volatility, see [2, 3, 13, 16]. Such model can reproduce the clustering property of the volatility at multiple time scales.

5.1 The Case $\alpha > 1/2$

When $\alpha > 1/2$, applying the stochastic Fubini's theorem, see [30], we get that the limiting process for the cumulated sums, Z , can be rewritten

$$Z_t = \int_0^t \int_0^s f^{\alpha,\lambda}(s-u) dW_u ds.$$

Therefore, in that case, it is differentiable and its derivative

$$Y_t = \int_0^t f^{\alpha,\lambda}(t-u) dW_u$$

locally behaves as a fractional diffusion with Hurst parameter $H = \alpha - 1/2$. Indeed,

$$f^{\alpha,\lambda}(x) \sim c/x^{1-\alpha}$$

when x is close to zero and recall that a fractional Brownian motion W^H can be written as

$$W_t^H = \int_{-\infty}^t \left[\frac{1}{(t-s)^{1/2-H}} - \frac{1}{(-s)^{1/2-H}} \right] dW_s.$$

In particular, proceeding as in [20], we get that for any $\varepsilon > 0$, Y has Hölder regularity $\alpha - 1/2 - \varepsilon$.

Thus, in this regime, the log-volatility asymptotically behaves as a fractional Brownian motion with Hurst parameter $\alpha - 1/2$. According to [16], this is consistent with empirical measures of the smoothness of the volatility process provided that $\alpha \simeq 0.6$.

5.2 The Case $\alpha < 1/2$

When $\alpha < 1/2$, the behavior of Z is quite different. Proceeding as in [20], we get that for any $\varepsilon > 0$, Z has Hölder regularity $1/2 + \alpha - \varepsilon$. This is not very surprising since the situation $\alpha < 1/2$ is close to that of ARFIMA processes,⁴ which are known to behave as a fractional Brownian motion with Hurst parameter $1/2 + \alpha$ at large time scales, see [15]. In this regime, the log-volatility exhibits apparent long memory, as observed for example in [2].

5.3 The Case $\alpha = 1/2$

In the critical regime $\alpha = 1/2$, we somehow asymptotically retrieve some features of the multifractal model of [3]. Indeed, in [3], the “log-volatility” $\omega_{l,T}(t)$ is written under the form

$$\omega_{l,T}(t) = \int_{-\infty}^t k_{l,T}(t-s) dW_s,$$

where W is a Brownian motion and $k_{l,T}$ is function behaving in the range $l \ll t \ll T$ as

$$k_{l,T}(t) \sim \frac{k_0}{\sqrt{t}},$$

for some model parameters l and T . Therefore, the integrated log-volatility defined as

$$\Omega_{l,T}(t) = \int_0^t \omega_{l,T}(s) ds$$

satisfies

$$\Omega_{l,T}(t) = \int_{-\infty}^t (K_{l,T}(t-s) - K(-s)) dW_s,$$

with

$$K_{l,T}(t) = \mathbb{I}_{t \geq 0} \int_0^t k_{l,T}(s) ds.$$

This behaves as $K_0\sqrt{t}$ in the range $l \ll t \ll T$.

Thus, this “multifractal” regime ($\alpha = 1/2$) appears as the interface between the classical long memory (log-)volatility models ($\alpha < 1/2$) and the more recent rough volatility models ($\alpha > 1/2$).

⁴An ARFIMA process, see [5], can be written as an infinite order autoregressive process whose sum of the coefficients is equal to one and whose coefficients sequence ϕ asymptotically behaves as $\phi_i \sim c/i^{1+\alpha}$ with $0 < \alpha < 1/2$.

References

1. Adas, A.: Traffic models in broadband networks. *IEEE Commun. Mag.* **35**(7), 82–89 (1997)
2. Andersen, T.G., Bollerslev, T., Diebold, F.X., Labys, P.: Modeling and forecasting realized volatility. *Econometrica* **71**(2), 579–625 (2003)
3. Bacry, E., Kozhemyak, A., Muzy, J.: Log-normal continuous cascade model of asset returns: aggregation properties and estimation. *Quant. Finance* **13**(5), 795–818 (2013)
4. Bacry, E., Muzy, J.-F.: Second order statistics characterization of Hawkes processes and non-parametric estimation. arXiv preprint [arXiv:1401.0903](https://arxiv.org/abs/1401.0903) (2014)
5. Beran, J.: *Statistics for Long-Memory Processes*. CRC Press (1994)
6. Billingsley, P.: *Convergence of Probability Measures*. Wiley Online Library (2009)
7. Buchmann, B., Chan, N.H.: Asymptotic theory of least squares estimators for nearly unstable processes under strong dependence. *Ann. Stat.* **35**(5), 2001–2017 (2007)
8. Buchmann, B., Chan, N.H.: Unified asymptotic theory for nearly unstable AR(p) processes. *Stoch. Process. Appl.* **123**(3), 952–985 (2013)
9. Chan, N.H.: The parameter inference for nearly nonstationary time series. *J. Am. Stat. Assoc.* **83**(403), 857–862 (1988)
10. Chan, N.H.: Time series with roots on or near the unit circle. In: *Handbook of Financial Time Series*, pp 695–707. Springer (2009)
11. Chan, N.H., Wei, C.-Z.: Asymptotic inference for nearly nonstationary AR(1) processes. *Ann. Stat.* **15**(3), 1050–1063 (1987)
12. Chan, N.H., Zhang, R.-M.: Inference for nearly nonstationary processes under strong dependence with infinite variance. *Statistica Sinica* **19**(3), 925 (2009)
13. Comte, F., Renault, E.: Long memory in continuous-time stochastic volatility models. *Math. Finance* **8**(4), 291–323 (1998)
14. Cumberland, W., Sykes, Z.: Weak convergence of an autoregressive process used in modeling population growth. *J. Appl. Probab.* **19**(2), 450–455 (1982)
15. Doukhan, P., Oppenheim, G., Taqqu, M.S.: *Theory and Applications of Long-Range Dependence*. Springer (2003)
16. Gatheral, J., Jaisson, T., Rosenbaum, M.: Volatility is Rough. Available at SSRN 2509457 (2014)
17. Hasbrouck, J.: Measuring the information content of stock trades. *J. Finance* **46**(1), 179–207 (1991)
18. Hawkes, A.G.: Point spectra of some mutually exciting point processes. *J. Roy. Stat. Soc. Ser. B (Methodol.)* **33**(3), 438–443 (1971)
19. Jaisson, T., Rosenbaum, M.: Limit theorems for nearly unstable Hawkes processes. *Ann. Appl. Probab.* (2013)
20. Jaisson, T., Rosenbaum, M.: Rough fractional diffusions as scaling limits of nearly unstable heavy-tailed Hawkes processes. Working paper (2014)
21. Jeganathan, P.: On the asymptotic behavior of least-squares estimators in AR time series with roots near the unit circle. *Econometric Theor.* **7**(03), 269–306 (1991)
22. Kalashnikov, V.: *Geometric Sums: Bounds for Rare Events with Applications*. Kluwer Academic Publishers (1997)
23. Kurtz, T.G., Protter, P.: Weak limit theorems for stochastic integrals and stochastic differential equations. *Ann. Probab.* **19**(3), 1035–1070 (1991)
24. Markevičiūtė, J.: Asymptotic results on nearly nonstationary processes. PhD thesis, Vilniaus universitetas (2013)
25. Markevičiūtė, J., Suquet, C., Račkauskas, A.: Functional central limit theorems for sums of nearly nonstationary processes. *Lith. Math. J.* **52**(3), 282–296 (2012)
26. Mathai, A., Haubold, H.J.: Mittag-Leffler functions and fractional calculus. In: *Special Functions for Applied Scientists*, pp 79–134. Springer (2008)
27. Phillips, P.C.: Towards a unified asymptotic theory for autoregression. *Biometrika* **74**(3), 535–547 (1987)

28. Royama, T.: Analytical Population Dynamics. Springer (1992)
29. van der Meer, T., Pap, G., van Zuijlen, M.C.: Asymptotic inference for nearly unstable AR(p) processes. *Econometric Theor.* **15**(02), 184–217 (1999)
30. Veraar, M.: The stochastic Fubini theorem revisited. *Stochast. Int. J. Probab. Stoch. Process.* **84**(4), 543–551 (2012)

Generalised Partial Autocorrelations and the Mutual Information Between Past and Future

Alessandra Luati and Tommaso Proietti

Abstract The paper introduces the generalised partial autocorrelation (GPAC) coefficients of a stationary stochastic process. The latter are related to the generalised autocovariances, the inverse Fourier transform coefficients of a power transformation of the spectral density function. By interpreting the generalised partial autocorrelations as the partial autocorrelation coefficients of an auxiliary process, we derive their properties and relate them to essential features of the original process. Based on a parameterisation suggested by [1] and on Whittle likelihood, we develop an estimation strategy for the GPAC coefficients. We further prove that the GPAC coefficients can be used to estimate the mutual information between the past and the future of a time series.

Keywords Generalised autocovariances · Spectral autoregressive models · Information regularity coefficient

1 Introduction

The spectral density of a covariance stationary process subsumes all the information that is needed for best linear prediction of the future process realisations. In this paper, we consider a class of spectral estimates that is based on a generalised power autoregressive scheme, depending on a finite dimensional parameter set and on a power transformation parameter, which can encompass different spectral estimation methods, among which autoregressive and moving average estimation.

A. Luati (✉)

Department of Statistics, University of Bologna, Bologna, Italy
e-mail: alessandra.luati@unibo.it

T. Proietti

Department of Economics and Finance, University of Rome Tor Vergata,
Roma, Italy
e-mail: tommaso.proietti@uniroma2.it

T. Proietti

CREATES, University of Aarhus, Aarhus, Denmark

In an earlier paper [15], spectral estimation was performed by a generalised set of Yule-Walker equations, based on the generalised autocovariance function, the inverse Fourier transform of a power transformation of the spectrum, which was estimated non-parametrically by a power transformation of the periodogram.

In this chapter, we consider a parameterisation of spectral models based on a set of unrestricted coefficients taking values in the interval $(-1, 1)$, that are interpretable as generalised partial autocorrelations (GPAC). The GPAC are then related to the mutual information between past and future, a measure of uncertainty about the future based on knowledge of the past, considered in [10], and recently revisited by [11, 12]. The novelty here is in the estimation of the mutual information based on the GPAC, via the Whittle likelihood. The computation of the mutual information between past and future entails the availability of the full partial autocorrelation sequence, unless the process is autoregressive, in which case the partial autocorrelation is truncated. Our approach amounts to determining a scale, implied by the power transformation parameter, along which the GPAC sequence is finite.

The chapter is organised as follows. Section 2 reviews the generalised autocovariance function and introduces the GPAC as the partial autocorrelation coefficients of an auxiliary power process. The relation between the GPAC and the mutual information between the past and the future is established in Sect. 4, along with an optimality property of the class of generalised spectral autoregressive models, formerly introduced in Sect. 3 and parameterised by the GPAC, based on [1]. Whittle likelihood estimation is dealt with in Sect. 5. An illustration featuring the time series of the U.S. monthly inflation rate, levels and first differences, is reported in Sect. 6. Finally, Sect. 7 draws some conclusions.

2 Generalised Autocovariances

Let $\{x_t\}_{t \in T}$ be a stationary zero-mean stochastic process indexed by a discrete time set T , with spectral density function $f(\omega)$ such that $\int_{-\pi}^{\pi} \log f(\omega) d\omega > -\infty$, and $\int_{-\pi}^{\pi} f(\omega)^p d\omega < \infty$, $p \in \mathbb{R}$.

The generalised autocovariance (GACV) function is defined [15] as the inverse Fourier transform of the p -th power of the spectral density function,

$$\gamma_{pk} = \frac{1}{2\pi} \int_{-\pi}^{\pi} [2\pi f(\omega)]^p \cos(\omega k) d\omega, \quad (1)$$

for $k = 0, 1, \dots$ and $\gamma_{p,-k} = \gamma_{pk}$. The discrete Fourier transform of γ_{pk} gives

$$[2\pi f(\omega)]^p = \gamma_{p0} + 2 \sum_{k=1}^{\infty} \gamma_{pk} \cos(\omega k).$$

The coefficients γ_{pk} depend on two arguments, the integer lag k and the real power p . As a matter of fact, for $p = 1$, $\gamma_{1k} = \gamma_k$, the autocovariance of the process at lag k ; for $p = 0$, $\gamma_{0k} = 0$, for $k \neq 0$ and $\gamma_{00} = 1$, up to a constant, the autocovariance function of a white noise process; for $p = -1$, $\gamma_{-1k} = \gamma_{ik}$, the inverse autocovariance function ([6]; see also [2]).

The function γ_{pk} lends itself to a further interpretation as the autocovariance function of an auxiliary process derived from x_t . This interpretation turns out to be useful in the derivation of the analytic form of γ_{pk} , as a function of the parameters that govern the process dynamics, by evaluating an expectation in the time domain, rather than solving (1) directly. In addition, the generalised partial autocorrelation function of x_t will be directly defined based on the auxiliary process.

Assuming that $\{x_t\}_{t \in T}$ is purely non-deterministic, its Wold representation is written as

$$x_t = \psi(B)\xi_t, \quad (2)$$

where $\xi_t \sim \text{WN}(0, \sigma^2)$ and $\psi(B) = 1 + \psi_1 B + \psi_2 B^2 + \dots$, with coefficients satisfying $\sum_{j=0}^{\infty} \psi_j^2 < \infty$, and such that all the roots of the characteristic equation $\psi(B) = 0$ are in modulus greater than one; here, $\text{WN}(0, \sigma^2)$ denotes a white noise process, a sequence of zero mean and uncorrelated random variables with constant variance σ^2 and B is the backshift operator, $B^k x_t = x_{t-k}$. The autocovariance function of the linear process (2) is $\gamma_k = \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+k}$ for $k = 0, 1, \dots$, and $\gamma_{-k} = \gamma_k$.

Let us consider the power-transformed process:

$$u_{pt} = \begin{cases} \psi(B)^p \xi_t &= \psi(B)^p \psi(B)^{-1} x_t, & \text{for } p \geq 0 \\ \psi(B^{-1})^p \xi_t &= \psi(B^{-1})^p \psi(B)^{-1} x_t, & \text{for } p < 0. \end{cases} \quad (3)$$

For a real $p > 0$, the power of $\psi(B)$ in (3) is still a power series,

$$\psi(B)^p = \sum_{j=0}^{\infty} \varphi_j B^j,$$

with coefficients given by the recursive relation

$$\varphi_j = \frac{1}{j} \sum_{h=1}^j [h(p+1) - j] \psi_h \varphi_{j-h}, \quad j > 0, \quad \varphi_0 = 1 \quad (4)$$

Gould [9]. When $p < 0$, the coefficients of $\psi(B^{-1})^p = \sum_{j=0}^{\infty} \varphi_j B^{-j}$ are also given by (4).

The spectral density of u_{pt} is $f_u(\omega) = (2\pi)^{-1} |\psi(e^{i\omega})|^{2p} \sigma^2$, and satisfies

$$2\pi f_u(\omega) (\sigma^2)^{p-1} = [2\pi f(\omega)]^p. \quad (5)$$

Equation (5) establishes the relation between the spectrum of the original process and that of the power process u_{pt} .

It follows from (1) and (5) that the generalised autocovariance function of x_t can be interpreted as the autocovariance function of the process u_{pt} , denoted as γ_u ,

$$\gamma_{pk} = (\sigma^2)^{p-1} \gamma_u.$$

It is then straightforward to compute the GACV of x_t as the autocovariance of a linear process,

$$\gamma_{pk} = \sigma^{2p} \sum_{j=0}^{\infty} \varphi_j \varphi_{j+k}.$$

The generalised variance γ_{p0} is related to the variance profile, defined in [13] as the Hölder, or power, mean of the spectrum of x_t :

$$v_p = \left\{ \frac{1}{2\pi} \int_{-\pi}^{\pi} [2\pi f(\omega)]^p d\omega \right\}^{\frac{1}{p}}.$$

Specifically, for $p \neq 0$, $v_p = \gamma_{p0}^{\frac{1}{p}}$. As a particular case, $v_{-1} = \gamma_{-1,0}^{-1}$ is the interpolation error variance $\text{Var}(x_t | \mathcal{F}_{\setminus t})$, where $\mathcal{F}_{\setminus t}$ is the past and future information set excluding the current x_t ; this is also interpreted as the harmonic mean of the spectrum. The limit of v_p for $p \rightarrow 0$ yields the prediction error variance, $\lim_{p \rightarrow 0} v_p = \sigma^2$, which is the geometric average of the spectral density, $\sigma^2 = \exp \left\{ \frac{1}{2\pi} \int_{-\pi}^{\pi} \log 2\pi f(\omega) d\omega \right\}$, i.e. the Szegő-Kolmogorov formula.

By dividing the generalised autocovariance by the generalised variance, one gets the generalised autocorrelation function (GACF), taking values in $[-1, 1]$,

$$\rho_{pk} = \frac{\gamma_{pk}}{\gamma_{p0}}, \quad k = 0, \pm 1, \dots$$

If the GACV of x_t is proportional to the autocovariance function of the auxiliary process u_{pt} , the GACF is equal to the ACF of the auxiliary process. The same holds for the generalised partial autocorrelation coefficients of x_t that are defined here as the sequence of the partial autocorrelation coefficients of u_{pt} and are denoted as π_{pk} .

3 Generalised Autoregressive Spectral Models

The generalised partial autocorrelations are central for estimating the spectral density of a stochastic process according to the following class of models,

$$2\pi f(\omega) = \left[\frac{\sigma_p^2}{\phi_p(e^{-i\omega})\phi_p(e^{i\omega})} \right]^{\frac{1}{p}} \quad (6)$$

where $\phi_p(e^{-i\omega}) = 1 - \phi_{p1}e^{-i\omega} - \phi_{p2}e^{-i\omega 2} - \dots - \phi_{pK}e^{-i\omega K}$. As p varies, (6) defines a set of spectral models encompassing the AR(K) case ($p = 1$), the MA(K) case ($p = -1$), as well as the fractional case (consider, for instance, the combination $K = 1$, $p = 1/d$ and $\phi_{p1} = 1$).

The coefficients $\{\phi_{pk}, k = 1, \dots, K\}$ satisfy the following moment conditions (Yule Walker equations): $\Gamma_{p,K}\phi_{p,K} = \gamma_{p,K}$, where $\Gamma_{p,K}$ denotes the Toeplitz matrix, formed from the GACV, with generic element $\gamma_{p,|h-k|}$, $h, k = 0, \dots, K-1$, $\gamma_{p,K} = (\gamma_{p1}, \dots, \gamma_{pK})'$, and $\phi_{p,K} = (\phi_{p1}, \dots, \phi_{pK})'$; moreover, $\sigma_p^2 = \gamma_{p0} - \phi_{p,K}'\gamma_{p,K}$.

For $p > 0$, the polynomial $\phi_p(B) = 1 - \phi_{p1}B - \dots - \phi_{pK}B^K$ characterises the AR approximation of the process u_{pt} , and provides the spectral factorisation $[2\pi f(\omega)]^p = \sigma_p^2[\phi_p(e^{-i\omega})\phi_p(e^{i\omega})]^{-1}$. By Eqs. (3)–(5), we obtain the AR approximation of the original process, $\pi(B)x_t = \xi_t$, $\pi(B) = [\phi_p(B)]^{1/p}$, or, equivalently, the moving average representation $x_t = \psi(B)\xi_t$, $\psi(B) = [\phi_p(B)]^{-1/p}$. For $p < 0$, the polynomials are in the forward operator B^{-1} .

According to this parameterisation, the GPAC forms a finite sequence. Moreover, as will be stated formally in Sect. 4, processes having spectrum specified as in (6) are optimal with respect to a mutual information criterion.

For a given p , the model (6) features $K + 1$ parameters, which can be estimated by maximising the Whittle likelihood (see details on Sect. 5), under the restriction that the spectral density $[2\pi f(\omega)]^p$ is bounded away from zero and infinity at all frequencies. This is achieved by enforcing the constraints $0 < |\phi_p(e^{-i\omega})|^2 < \infty$, $\forall \omega \in (-\pi, \pi)$ or, equivalently, that the roots of the lag polynomials $\phi_p(B)$ do not lie on the unit circle.

Hence, for the purpose of estimation, we reparameterise the AR coefficients in terms of partial autocorrelations [1]. This is done by a recursive algorithm: letting $\pi_{pk}, k = 1, \dots, K$ denote K coefficients, such that $|\pi_{pk}| < 1$, then for $k = 1, \dots, K$ and $j = 1, \dots, k-1$, compute

$$\phi_{pj}^{(k)} = \phi_{pj}^{(k-1)} - \pi_{pk}\phi_{p,k-j}^{(k-1)}, \quad \phi_{pk}^{(k)} = \pi_{pk}. \quad (7)$$

The final iteration returns coefficients that are in the stationary region.

The coefficients π_{pk} , that are constrained in the range $(-1, 1)$, are in turn obtained as the Fisher inverse transformations of unconstrained real parameters ϑ_{pk} , $k = 1, \dots, K$, e.g. $\pi_{pk} = \frac{\exp(2\vartheta_{pk})-1}{\exp(2\vartheta_{pk})+1}$ for $k = 1, \dots, K$. Also, we set $\vartheta_{p0} = \ln(\sigma_p^2)$.

4 Mutual Information Between Past and Future

The generalised partial autocorrelation coefficients play a role in the estimation of the mutual information between the past and the future of a stationary Gaussian time series. The latter, denoted as I_{p-f} , is a measure of uncertainty about the future $\mathcal{F}_t$, the sigma-algebra generated by $\{x_{t+1}, x_{t+2}, \dots\}$, when the past $\mathcal{P}_t$, the sigma-algebra generated $\{x_t, x_{t-1}, \dots\}$, is known.

More formally, let us consider two stochastic processes, $\{x_t\}_{t \in T}$ and $\{y_s\}_{s \in S}$, defined on a probability space $(\Omega, \mathcal{F}, P)$, and denote as $\mathcal{S}_1$ and $\mathcal{S}_2$ the minimal sigma-algebra generated by $\{x_t\}_{t \in T}$ and $\{y_s\}_{s \in S}$, respectively. The amount of information of the random process $\{x_t\}_{t \in T}$ given by the process $\{y_s\}_{s \in S}$ is (see [10], Chap. IV),

$$I(x, y) = \sup \sum P(A_i \cap B_j) \ln \frac{P(A_i \cap B_j)}{P(A_i)P(B_j)}, \quad (8)$$

where the supremum is taken over all the possible finite partitions of Ω in the non intersecting events $(A_i)_{i=1, \dots, n}$, $(B_j)_{j=1, \dots, m}$, where $A_i \in \mathcal{S}_1$ for all $i = 1, \dots, n$ and $B_j \in \mathcal{S}_2$ for all $j = 1, \dots, m$.

It is easy to verify that the following properties hold: $I(x, y) \geq 0$, with $I(x, y) = 0$ when the sigma-algebra $\mathcal{S}_1$ is independent of $\mathcal{S}_2$; $I(x, y) = I(y, x)$, which motivates the name of mutual information for the quantity in Eq. (8). An important concept related to mutual information is the information regularity of a stochastic process. A stationary random process is said to be information regular if

$$I_\tau = I(\{x_t\}_{t < s}, \{x_t\}_{t \geq s + \tau}) \rightarrow 0, \quad \tau \rightarrow \infty.$$

The value I_τ is sometimes referred to as the information regularity coefficient and I_0 is the mutual information between past and future, that we shall denote here as I_{p-f} .

For Gaussian processes, I_{p-f} may be defined in terms of the cepstral coefficients,

$$c_k = \frac{1}{2\pi} \int_{-\pi}^{\pi} \ln[2\pi f(\omega)] \cos(\omega k) d\omega, \quad k = 1, 2, \dots \quad (9)$$

as [11]

$$I_{p-f} = \frac{1}{2} \sum_{k=1}^{\infty} k c_k^2 \quad (10)$$

and a necessary condition for information regularity is $\sum_{k=1}^{\infty} k c_k^2 < \infty$. Li [11] also shows that the following relation holds between cepstral coefficients and the partial autocorrelation coefficients, $\{\pi_k\}_{k=1, 2, \dots}$, the so called reflectrum identity

$$\sum_{k=1}^{\infty} k c_k^2 = - \sum_{k=1}^{\infty} k \ln(1 - \pi_k^2) \quad (11)$$

and $c_0 = \ln \gamma_0 + \sum_{k=1}^{\infty} \ln(1 - \pi_k^2)$, the latter being a consequence of the Kolmogorov-Szegö formula.

It is shown in [14] that the cepstral coefficients are related to the GACV via

$$c_k = \lim_{p \rightarrow 0} \frac{\gamma_{pk}}{p}, \quad k = 1, 2, \dots$$

The next theorem shows that the mutual information can be written as a function of the generalised partial autocorrelation coefficients.

Theorem 1 *Let π_{pk} denote the generalised partial autocorrelations of the stationary process $\{x_t\}_{t \in T}$. The mutual information between past and future is*

$$I_{p-f} = -\frac{1}{2p^2} \sum_{k=1}^{\infty} k \ln(1 - \pi_{pk}^2)$$

and the equality holds for all p .

Proof Let us denote as I_{p-f}^u the mutual information between past and future of the process u_{pt} , defined by Eq. (3), having spectrum $2\pi f_u(\omega) = [2\pi f(\omega)]^p (\sigma^2)^{1-p}$, as in (5), and prediction error variance σ^2 .

By (10), I_{p-f}^u may be written as a function of the cepstral coefficients of u_{pt} , denoted here as c_{uk} , the inverse Fourier coefficients of the logarithm of the spectral density function of u_{pt} , see Eq. (9),

$$\log 2\pi f_u(\omega) = c_{u0} + 2 \sum_{k=1}^{\infty} c_{uk} \cos(\omega k). \quad (12)$$

It follows from (5) that $\log 2\pi f_u(\omega) = p \log 2\pi f(\omega) + (1-p) \log \sigma^2$ and, from (9),

$$\log 2\pi f_u(\omega) = p \left(c_0 + 2 \sum_{k=1}^{\infty} c_k \cos(\omega k) \right) + (1-p) \log \sigma^2. \quad (13)$$

Combining (12) and (13), and noticing that $c_{u0} = c_0 = \log \sigma^2$, we get the relation between the cepstral coefficients of the original process and the auxiliary process

$$c_{uk} = p c_k$$

which clearly implies that $\sum_{k=1}^{\infty} k c_{uk}^2 = \sum_{k=1}^{\infty} k p^2 c_k^2$ and $p^2 I_{p-f} = I_{p-f}^u$. By the reflectrum identity applied to the process u_{pt} , we are able to conclude that

$$p^2 I_{p-f} = -\frac{1}{2} \sum_{k=1}^{\infty} k \ln(1 - \pi_{pk}^2)$$

which completes the proof of the theorem.

The computation of I_{p-f} entails the availability of the full partial autocorrelation sequence, unless the process is autoregressive, in which case the partial autocorrelation is truncated at K . The approach followed in this paper amounts to determining a scale, determined by the transformation parameter p , along which the GPAC sequence is finite. The following theorem, that generalises Theorem 3.1 of [12],

establishes the optimality of the generalised spectral autoregressive models (6) with respect to the minimum mutual information principle.

Theorem 2 *A process with given generalised autocovariances $\gamma_{pk}, k = 0, 1, \dots, K, p \neq 0$ and minimal information between past and future belongs to the class of spectral models (6).*

Proof The theorem follows straightforwardly from Theorem 3.1 of [12], who state and prove that given $K + 1$ autocovariances forming a positive definite sequence, a Gaussian process achieves the minimum information I_{p-f} when it follows an AR(K) model. The theorem directly applies to an AR(K) auxiliary process with prediction error variance equal to σ_p^2 whose first $K + 1$ autocovariances are $\gamma_{pk}, k = 0, 1, \dots, K$.

In the case when $p = 0$, the process with minimal mutual information belongs to the class of EXP(K) models of [4], obtained by truncating to the lag K the Fourier expansion of the logarithm of the spectral density function (see [12]).

Model selection by the minimum mutual information (MMI) principle is one of the applications of I_{p-f} traced by [12]. As a matter of fact, MMI is related to the maximum entropy principle, by means of the complementarity of I_{p-f} and the entropy rate of a regular stationary and Gaussian process (see Eq. (9) in [11]). Once the class of fitting models is selected through the MMI, a penalised criterion based on I_{p-f} , called LIC, has been proposed by [12] to select the order of the fitting model.

Other applications of the mutual information between past and future arise in connection with the equivalent concept of excess entropy, an intuitive measure of memory stored in a stochastic process: fractional ARMA processes have infinite mutual information. Recently, Hidden Markov models with infinite mutual information between past and future have been investigated and examples of processes with an infinite number of countable states are presented with reference with modelling texts in natural language and complex system analysis [7].

5 Estimation

The main tool for estimating the spectral density function and its functionals is the periodogram.

Let $\{x_t, t = 1, 2, \dots, n\}$ denote a time series, which is a sample realisation from a stationary Gaussian process, and let $\omega_j = \frac{2\pi j}{n}, j = 1, \dots, [n/2]$, be the Fourier frequencies, where $[\cdot]$ denotes the integer part of the argument.

The periodogram (sample spectrum) is defined as

$$I(\omega_j) = \frac{1}{2\pi n} \left| \sum_{t=1}^n (x_t - \bar{x}) e^{-t\omega_j t} \right|^2,$$

where $\bar{x} = n^{-1} \sum_t x_t$.

The following large sample distributional result holds in the short memory case ([5], Chap. 10):

$$\frac{I(\omega_j)}{f(\omega_j)} \sim \text{IID} \frac{1}{2} \chi_r^2, \quad 0 < \omega_j < \pi$$

whereas $\frac{I(\omega_j)}{f(\omega_j)} \sim \chi_1^2$, $\omega_j = 0, \pi$, where χ_r^2 denotes a chi-square random variable with r degrees of freedom; equivalently, $I(\omega_j)$ is exponentially distributed with mean $f(\omega_j)$.

For a given transformation parameter p , the log-likelihood function of the unconstrained parameters ϑ_{pk} , $k = 1, \dots, K$ based on the exponential density for $I(\omega_j)$, for $0 < \omega_j < \pi$ and $j = 1, \dots, N$ where $N = [(n-1)/2]$, is

$$\ell(\vartheta_{p,K}) = - \sum_{j=1}^N \left(\ln f(\omega_j) + \frac{I(\omega_j)}{f(\omega_j)} \right)$$

The latter can be maximised with respect to the unconstrained parameter vector $\vartheta_{p,K}$ by a quasi-Newton optimisation algorithm, using numerical first and second derivatives. The GPAC π_{pk} and the parameters $\phi_{p1}, \dots, \phi_{pK}$ lying in the stationary region are then obtained by (7).

The spectral models for different combinations of K and p can be compared using an information criterion, such as AIC or BIC.

As long as the summability conditions for the GPAC is satisfied, $\sum_{k=1}^{\infty} k \ln(1 - \pi_{pk}^2) < \infty$, i.e. as long as the mutual information as in Theorem 1 is finite, the strong Szegő theorem holds (see [3]) and under the usual regularity conditions on the parameter space and the spectral density function, Theorems 2.1 and 2.2 on Chap. 2 of [8] hold, which imply that $\hat{\vartheta}_{p,K} \rightarrow_p \vartheta_{p,K}$ and $\sqrt{n}(\hat{\vartheta}_{p,K} - \vartheta_{p,K}) \rightarrow_d N(0, V_p)$ where

$$V_p^{-1} = \frac{1}{4\pi} \int_{-\pi}^{\pi} \frac{1}{[2\pi f(\omega)]^{2p}} z(\omega) z(\omega)' d\omega$$

with $z(\omega) = [1, 2 \cos(\omega), 2 \cos(2\omega), \dots, 2 \cos(K\omega)]'$. Given that the function that maps the partial autocorrelation coefficients to the model parameters is one to one and smooth (see [1], Theorem 2), the asymptotic properties of the Whittle estimator continue to hold in the reparameterized model.

6 Illustration

Our illustration deals with the estimation of the mutual information between past and future for the monthly U.S. inflation rate. The latter is computed as the logarithmic change over the previous month of the Consumer Price Index (CPI), multiplied by 1200, and is considered for the period running from January 1960 to December

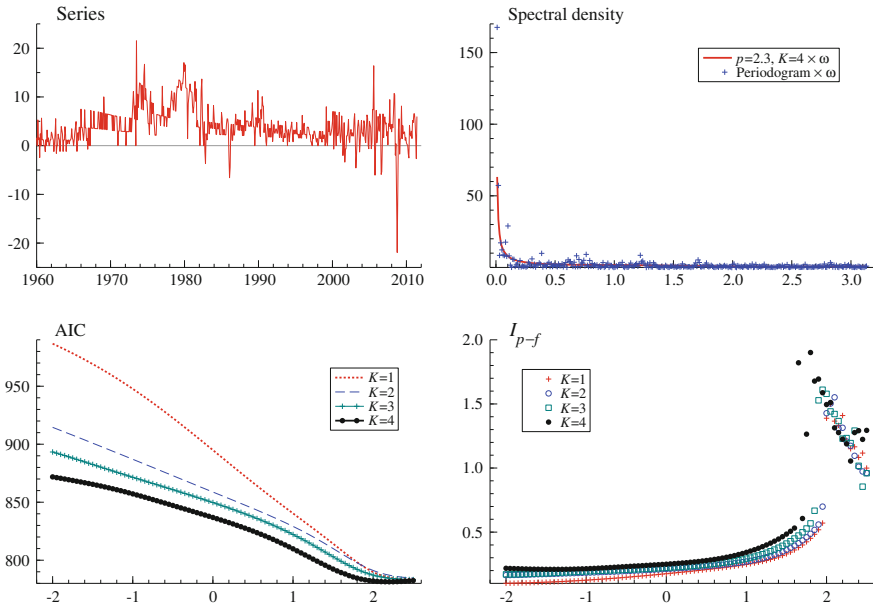


Fig. 1 U.S. monthly inflation rate (annualised rate). Plot of the series (*top left*). Periodogram and fitted GAR spectrum with $p = 2.4$, $K = 1$ (*top right*). Values of the AIC as a function of p for values of $K = 1, 2, 3, 4$ (*bottom left*). Mutual information estimates as a function of p for different values of K (*bottom right*)

2012, for a total of 624 observations. The plot of the series is available in the top right panel of Fig. 1. A widely debated issue deals with the covariance stationarity of the underlying generating process, as the evidence arising from either unit root and stationarity tests is not clear cut. The periodogram, displayed in the right top plot of Fig. 1, takes very large values around the zero frequency.

The left bottom graph displays the behaviour of the Akaike's information criterion, $AIC = -2\ell(\hat{\vartheta}_{p,K}) + 2K$, where $\hat{\vartheta}_{p,K}$ is the Whittle likelihood estimator of the parameters of the generalised AR spectral model (6) fitted to the series for different values of p and K . In particular, for K ranging from 1 to 4, it displays the pattern of the AIC for values of p in the interval $[-2, 2.5]$. The AIC is similar for the values of K considered for p in the range $[2, 2.5]$; when $K = 1$, the estimated generalised autoregressive parameter is equal to 1 for p in this range; similarly, for higher values of K , the generalised polynomial $\phi_p(z)$ has a unit root. The combination of values minimising the AIC is $K = 4$, $p = 2.3$; the estimated GAR polynomial is $\tilde{\phi}_p(e^{-i\omega}) = 1 - 1.0383e^{-i\omega} + 0.2028e^{-2i\omega} + 0.0251e^{-3i\omega} - 0.1896e^{-4i\omega}$; the corresponding GPACs are $\tilde{\pi}_{p1} = 0.9999$, $\tilde{\pi}_{p2} = -0.0612$, $\tilde{\pi}_{p3} = 0.1782$, $\tilde{\pi}_{p4} = 0.1896$. Notice that if $\phi_p(e^{-i\omega}) = (1 - e^{-i\omega})\phi_p^*(e^{-i\omega})$, where $|\phi_p^*(e^{-i\omega})|^2$ is bounded away from zero and finite, and $p > 2$, y_t is long memory process, with a spectral density being $O(|\omega|^{-2d})$ as $|\omega| \rightarrow 0$, with $d = 2/p$. The fitted spectral density is

superimposed to the periodogram in the top left panel. As a result, likelihood inference points decisively towards a nonstationary or a long memory model. Although the AIC is concave in p and has a minimum within the assumed range, the properties of the estimate of p cannot be ascertained. The bottom right graph displays the pattern of the estimated I_{p-f} as p varies in its range and for the same values of K . The plot highlights a neat discontinuity occurring at around $p = 2$, which is a reflection of the fact that the estimated model is long memory.

We then turn our attention to the first differences of the monthly inflation rates, which are plotted in figure in the top left panel of Fig. 2. Model selection performed according to the AIC leads to choosing the GAR specification with $p = -0.8$ and $K = 3$ (see the bottom left plot which depicts the AIC as a function of p for $K = 1, 2, 3, 4, 5$). The selected values of p is not significantly different from -1 , which in turn leads to a moving average model for the spectrum of the time series. The fitted spectral density has a low dynamic range and is superimposed to the periodogram of the series in the top right plot. It can be seen from the same plot that the spectral fit arising from the AR spectrum of the same order, i.e. $p = 1$, $K = 3$, is likely to overfit some features of the periodogram. The estimated GAR polynomial is $\hat{\phi}_p(e^{-i\omega}) = 1 - 0.5419e^{-i\omega} - 0.1989e^{-2i\omega} - 0.0960e^{-3i\omega}$; the corresponding GPACs are $\hat{\pi}_{p1} = 0.7583$, $\hat{\pi}_{p2} = 0.2532$, $\hat{\pi}_{p3} = 0.0960$.

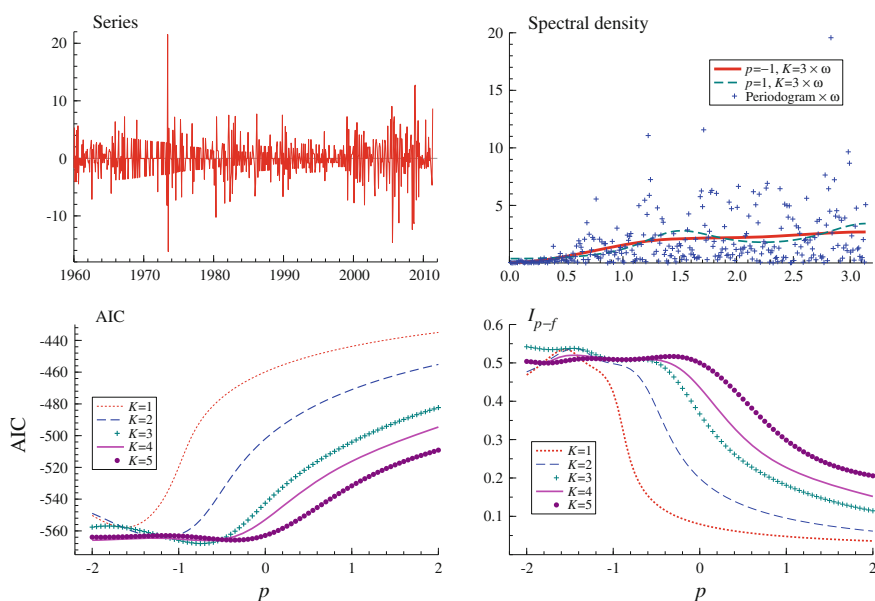


Fig. 2 Changes in U.S. monthly inflation rate (first differences). Plot of the series (*top left*). Periodogram and fitted GAR spectrum with $p = 2.4$, $K = 1$ (*top right*). Values of the AIC as a function of p for values of $K = 1, 2, 3, 4, 5$ (*bottom left*). Mutual information estimates as a function of p for different values of K (*bottom right*)

The behaviour of the estimated mutual information between past and future is displayed in the bottom right panel as p varies in $[-2, 2]$ and for $K = 1, \dots, 5$. The estimates are rather insensitive to p in the negative range (in fact, also the AIC is almost constant), hovering around 0.5.

We conclude this section with a remark on the estimation of I_{p-f} based on the GPAC. When a generalised spectral AR model is fitted to a series, the selection the best pair (p, K) in terms of minimum AIC implies that the estimation of I_{p-f} is based on a *finite* number of generalised partial autocorrelation coefficients, $\tilde{\pi}_{p1}, \dots, \tilde{\pi}_{pK}$. Except in the case when $p = 1$, which identifies a pure autoregressive model in the general class (6), the computation of I_{p-f} based on the ordinary partial autocorrelations or cepstral coefficients involves an infinite number of coefficients and its estimation necessarily requires a truncation of the infinite sequence.

7 Conclusions

In this chapter we have introduced the generalised partial autocorrelation function. This can be useful for parameterising a model for the spectrum of a time series, which encompasses autoregressive and moving average spectral estimation, and for characterising the properties of a random stationary process. We have considered in particular its use for computing (and estimating from a sample time series) the mutual information between the past and the future. The latter is an interesting concept which complements the traditional one step ahead prediction error variance, looking at the joint predictability at the different many horizons that arise for the future.

Acknowledgments We would like to thank the Editors for their kindness during the review process and Robert Stelzer a Referee for the useful comments to the paper.

References

1. Barndorff-Nielsen, O.E., Schou, G.: On the parametrization of autoregressive models by partial autocorrelations. *J. Multivar. Anal.* **3**, 408–419 (1973)
2. Battaglia, F.: Inverse autocovariances and a measure of linear determinism for a stationary process. *J. Time Ser. Anal.* **4**, 79–87 (1983)
3. Bingham, N.H.: Szegő theorems and its probabilistic descendants. *Probab. Surv.* **9**, 287–324 (2012)
4. Bloomfield, P.: An exponential model for the spectrum of a scalar time series. *Biometrika* **60**, 217–226 (1973)
5. Brockwell, P.J., Davis, R.A.: *Time Series: Theory and Methods*. Springer-Verlag, New York (1991)
6. Cleveland, W.S.: The inverse autocorrelations of a time series and their applications. *Technometrics* **14**, 277–293 (1972)
7. Debowski, L.: On hidden markov processes with infinite excess entropy. *J. Theor. Probab.* **27**, 539–551 (2014)

8. Dzhaparidze, K.: *Parameter Estimation and Hypothesis Testing in Spectral Analysis of Stationary Time Series*. Springer-Verlag, New York (1986)
9. Gould, H.W.: Coefficient identities for powers of Taylor and Dirichlet series. *Am. Math. Monthly* **81**, 3–14 (1974)
10. Ibragimov, I.A., Rozanov, Y.A.: *Gaussian Random Processes*. Springer-Verlag, New York (1978)
11. Li, L.M.: Some notes on mutual information between past and future. *J. Time Ser. Anal.* **27**, 309–322 (2005)
12. Li, L.M., Xie, Z.: Model selection and order determination for time series by information between the past and the future. *J. Time Ser. Anal.* **17**, 65–84 (1996)
13. Luati, A., Proietti, T., Reale, M.: The variance profile. *J. Am. Stat. Assoc.* **107**, 607–621 (2012)
14. Proietti, T., Luati, A.: The exponential model for the spectrum of a time series: extensions and applications. MPRA working paper 45280 (2014)
15. Proietti, T., Luati, A.: The generalised autocovariance function. *J. Econometrics* **187**, 245–257 (2015)

Efficient Estimation of Integrated Volatility in Presence of Infinite Variation Jumps with Multiple Activity Indices

Jean Jacod and Viktor Todorov

Abstract In a recent paper [6], we derived a rate efficient (and in some cases variance efficient) estimator for the integrated volatility of the diffusion coefficient of a process in presence of infinite variation jumps. The estimation is based on discrete observations of the process on a fixed time interval with asymptotically shrinking equidistant observation grid. The result in [6] is derived under the assumption that the jump part of the discretely-observed process has a finite variation component plus a stochastic integral with respect to a stable-like Lévy process with index $\beta > 1$. Here we show that the procedure of [6] can be extended to accommodate the case when the jumps are a mixture of finitely many integrals with respect to stable-like Lévy processes with indices $\beta_1 > \dots > \beta_M \geq 1$.

Keywords Central limit theorem · Integrated volatility · Itô semimartingale · Jumps · Jump activity · Quadratic variation · Stable process

AMS 2000 Subject Classifications 60F05 · 60F17 · 60G51 · 60G07

1 Introduction

In this paper we revisit the question of efficient, or at least rate-efficient, estimation of the integrated volatility of the diffusion coefficient of a one-dimensional process X which is observed at discrete times. The process X is an Itô semimartingale, whose continuous martingale part is a stochastic integral $\int_0^t \sigma_s dW_s$ with respect to a standard Brownian motion W , and the integrated volatility over a given time interval $[0, T]$ is $C_T = \int_0^t \sigma_s^2 ds$.

J. Jacod (✉)

Institut de Mathématiques de Jussieu, CNRS – UMR 7586,
Université Pierre et Marie Curie–P6, 4 Place Jussieu, 75252 Paris-cedex 05, France
e-mail: jean.jacod@upmc.fr

V. Todorov

Department of Finance, Northwestern University, Evanston 60208-2001, USA
e-mail: v-todorov@northwestern.edu

The problem of estimating the integrated volatility C_T is arguably the most studied one in the literature dealing with inference based on high-frequency observations. The volatility inference problem has been analyzed in various settings. This includes the case when X has jumps, when the process X is observed with additional microstructure noise, and when the observations take place at irregularly spaced—and possibly random—times.

In this paper, we discard microstructure and irregular spacing, and concentrate on the case when X is perfectly observed at times $i\Delta_n$ for $i = 0, 1, \dots$ within a fixed finite time interval $[0, T]$, and when the time lag Δ_n is small, and eventually goes to 0 (the “high-frequency” setting). We put emphasis on the case when X has jumps with a “high” degree of activity, meaning that those jumps are not locally summable, since the case when they are summable (equivalently, with finite variation) is well established already (with references given below).

In the above setting we answer the following two questions. First, knowing that in the continuous case, and also when there are summable jumps (plus a few appropriate technical assumptions), one can exhibit estimators which converge at the optimal (or, efficient) rate $1/\sqrt{\Delta_n}$, as $\Delta_n \rightarrow 0$, is it possible to obtain estimators with the same efficient rate in the non-summable jumps case? Second, if this holds, can one find estimators that are variance-efficient, in the sense that their (normalized) asymptotic variance is the same efficient variance as the one found in the continuous case?

These two questions have been studied in the paper [6], when the jump part of X can be split into a sum $Z + Z'$ as follows: Z'_t is an absolutely convergent sum $\sum_{s \leq t} \Delta Z'_s$ (here, $\Delta V_s = V_s - V_{s-}$ is the jump size at time s for any càdlàg process V), and $Z_t = \int_0^t \gamma_s dY_s$ for some (unknown) process γ and a stable or “stable-like” process Y with index $\beta \in (1, 2)$ (we refer to [6] for the formal definition of “stable-like” processes—they include many processes used in various applications such as the tempered stable processes). Namely, in this case in [6] we have constructed estimators which converge at the rate $1/\sqrt{\Delta_n}$, and which are variance-efficient when further the process Y is symmetric, whereas if Y is not symmetric the asymptotic variance is twice the efficient variance.

Our goal in this paper is to extend this result in two directions: first we replace the assumption $Z_t = \int_0^t \gamma_s dY_s$ by the fact that the “spot Lévy measures” of Y behave like a stable Lévy measure with index β near the origin, with a multiplicative coefficient which is a stochastic process a_t (see below for the precise statement); this is much weaker than assuming $Z_t = \int_0^t \gamma_s dY_s$. Second, we replace the single process Z by a sum $\sum_{m=1}^M Z^m$, where each Z^m is as Z above with an index β_m , and $\beta_1 > \beta_2 > \dots > \beta_M$, and the β_m are “regularly spaced”, in the sense that they lie on a grid of the form $\{2 - j\rho : j = 1, \dots, [2/\rho]\}$ for some (unknown) number $\rho > 0$ (the assumption is in fact even weaker than this but too technical to be stated in the introduction). Then, we exhibit estimators for C_T which converge with the efficient rate $1/\sqrt{\Delta_n}$, and with the asymptotic variance equal to twice the efficient one. Should we have $Z_t^m = \int_0^t \gamma_s^m dY_s^m$ with Z^m stable-like with index β_m and symmetric, it would also be possible to construct estimators achieving the efficient variance bound, but for simplicity we will not do this here.

The method used here is basically the same as in [6], and relies on estimating locally the volatility (diffusion coefficient) from the empirical characteristic function of the increments of the process over blocks of decreasing length but containing an increasing number of observations, and then summing the local volatility estimates. The last step consists in iterating the de-biasing procedure introduced in that paper.

The rest of the paper is organized as follows. In Sect. 2 we describe the general setting, and present a short review of the methods which have been proposed in the literature for jump-robust volatility estimation. Section 3 is devoted to recalling the estimators of [6] and the associated CLT. In Sect. 4 we solve the problem of estimating integrated volatility with a rate $1/\sqrt{\Delta_n}$ under our generalized assumptions. Proofs are given in Sect. 5.

2 Estimating Integrated Volatility: A Review

2.1 The General Setting

The underlying process X is a one-dimensional Itô semimartingale defined on a filtered space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$, so it can be written as

$$X_t = X_0 + \int_0^t b_s ds + \int_0^t \sigma_s dW_s + \delta * (p - q)_t + \delta' * p_t, \quad (2.1)$$

where W is a standard Brownian motion and p is a Poisson measure on $\mathbb{R}_+ \times E$ with (deterministic) compensator $q(dt, dz) = dt \otimes \lambda(dz)$. Here E is a Polish space and λ is a σ -finite measure on E . For a function ϕ on $\Omega \times \mathbb{R}_+ \times E$ and a random measure ν on $\mathbb{R}_+ \times E$, the notation $\phi * \nu_t$ stands for the double (ordinary or stochastic) integral $\int_0^t \int_E \phi(s, z) \nu(ds, dz)$. The processes b_t and σ_t are optional, the functions δ and δ' on $\Omega \times \mathbb{R}_+ \times E$ are predictable, with $|\delta| \leq 1$ and they all are such that the integrals in (2.1) make sense (this will be implied by our assumptions below).

It is perhaps more natural to write the last two terms in (2.1) as

$$(\delta 1_{\{|\delta| \leq 1\}}) * (p - q)_t + (\delta 1_{\{|\delta| > 1\}}) * p_t,$$

with the same function δ in both terms. However, given the structure that we will introduce on the infinite variation jumps below (and hence on δ), the formulation (2.1) offers more flexibility. For example, it allows us to add to X a term such as $\delta' * p$ with a function δ' having no regularity in time.

We will further assume that the volatility process σ_t is itself an Itô semimartingale, which can thus be written as

$$\begin{aligned} \sigma_t = & \sigma_0 + \int_0^t b_s^\sigma ds + \int_0^t H_s^\sigma dW_s + \int_0^t H_s^{\prime\sigma} dW_s' \\ & + (\delta^\sigma(s, z) 1_{\{|\delta^\sigma(s, z)| \leq 1\}}) * (p - q)_t + (\delta^\sigma(s, z) 1_{\{|\delta^\sigma(s, z)| > 1\}}) * p_t. \end{aligned} \quad (2.2)$$

Choosing the same Poisson measure $\underline{p}$ to drive both X and σ is not a restriction, and W' in (2.2) is another Brownian motion independent of W . Note that we need both W and W' to allow for general dependence between the diffusion components of X and σ .

Our aim in this paper is to estimate the integrated volatility which is formally defined as

$$C_T = \int_0^T c_s ds, \quad \text{where } c_t = \sigma_t^2.$$

The estimation is done on the basis of the observations $X_{i\Delta_n}$ for $i = 0, \dots, [T/\Delta_n]$. We assume that $\Delta_n \rightarrow 0$, and the observed increments are denoted as $\Delta_i^n X = X_{i\Delta_n} - X_{(i-1)\Delta_n}$.

In order to describe the assumptions we need for our asymptotic results, we first introduce a property relative to a generic process U . This will be applied to U being b in (2.1) or some of the coefficients appearing in (2.2). This property is as follows: there is a constant γ such that, for all $s, t \geq 0$, we have

$$\mathbb{E}(|U_{s+t} - U_s|^2 \mid \mathcal{F}_s) \leq \gamma t. \quad (2.3)$$

When U is an Itô semimartingale with locally bounded characteristics, it satisfies this property “locally” (in the sense that for a sequence τ_n of stopping times increasing to ∞ the stopped process $(U_{\tau_n \wedge t})_{t \geq 0}$ satisfies (2.2)). It is also satisfied when the paths of U are Hölder with index $1/2$.

Our first assumption, quite standard in almost all high-frequency situations, is as follows

Assumption (A₀) We have (2.1) and (2.2), and there exist two numbers $r \in [0, 1]$ and $r' \in (1, 2)$, a sequence τ_n of stopping times increasing to infinity, two sequences J_n, J'_n of $[0, 1]$ -valued Borel functions on E such that $\int J_n(z)^{r'} \lambda(dz) < \infty$ and $\int J'_n(z) \lambda(dz) < \infty$, and a sequence Γ_n of numbers, such that

$$t < \tau_n \Rightarrow \begin{cases} |b_t|, |\sigma_t|, |b_t^\sigma|, |H_t^\sigma|, |H_t^{\prime\sigma}| \leq \Gamma_n \\ |\delta(t, z)| \leq J_n(z) \\ |\delta'(t, z)|^r \wedge 1 \leq J'_n(z), \quad |\delta^\sigma(t, z)|^2 \wedge 1 \leq J'_n(z), \end{cases} \quad (2.4)$$

and the stopped processes $H_{t \wedge \tau_n}^\sigma$ and $b_{t \wedge \tau_n}$ satisfy (2.3) with $\gamma = \Gamma_n$.

2.2 Some Known Results

As mentioned in the introduction, the question of estimating C_T for a given $T > 0$, on the basis of the discrete observations within the time interval $[0, T]$, has already been thoroughly studied. We refer to [2] for many bibliographical comments, and only provide a brief sketch below.

Let us start with the continuous case, that is, $\delta = \delta' = 0$ in (2.1). The most natural estimator at stage n is the approximate quadratic variation (or, realized volatility)

$$\widehat{C}_T^n = \sum_{i=1}^{\lfloor T/\Delta_n \rfloor} (\Delta_i^n X)^2.$$

The normalized sequence $\sqrt{\Delta_n} (\widehat{C}_T^n - C_T)$ converges (in the sense of stable convergence in law) to a limit Z_T which is defined on an extension of the space, and which conditionally on $\mathcal{F}$ is a centered Gaussian variable with (conditional) variance $V_T = 2 \int_0^T c_s^2 ds$, as $\Delta_n \rightarrow 0$. For this, we do not need (A₀): on top of $\delta = \delta' = 0$, having V_T and $\int_0^T b_s^2 ds$ finite-valued is enough, and we do have functional (in T) convergence, even.

Moreover, when σ_t is non-random (so, apart from the drift term, the process X has independent increments, and is Lévy when further $\sigma_t = \sigma$ is constant), the LAN property holds for the (deterministic) parameter C_T , and V_T above is the inverse of the Fisher information, so indeed the estimators $\widehat{C}_T^n$ are asymptotically efficient. The same efficiency holds, due to the LAMN property now, when σ_t has the form $\sigma_t = \sigma(t, X_t)$ for a smooth enough function σ . Efficiency also holds, due to an infinite-dimensional Hajek type convolution theorem for more general stochastic volatility models, see [5].

These considerations lead us to postulate that “efficiency” for a sequence $\widehat{C}_T^n$ of estimators (for X either continuous or discontinuous) means that the sequence $\sqrt{\Delta_n} (\widehat{C}_T^n - C_T)$ converges stably in law to a limit which is $\mathcal{F}$ -conditionally centered Gaussian with variance V_T , whereas rate-efficiency means the same with again the rate $1/\sqrt{\Delta_n}$, but with a conditional variance possibly bigger than V_T .

The above described asymptotic properties of the approximate quadratic variation, $\widehat{C}_T^n$, badly fail when X has jumps since, to begin with, $\widehat{C}_T^n$ no longer converges to C_T but to $C_T + \sum_{s \leq T} (\Delta X_s)^2$. However, one can replace $\widehat{C}_T^n$ by a truncated version:

$$\widehat{C}_T^m = \sum_{i=1}^{\lfloor T/\Delta_n \rfloor} (\Delta_i^n X)^2 1_{\{|\Delta_i^n X| \leq u_n\}}.$$

This is the *truncated realized volatility* introduced by Mancini [8, 9]. For a properly chosen sequence of truncation level u_n (going to 0), the $\widehat{C}_T^m$ above enjoy exactly the same properties as $\widehat{C}_T^n$ does in the continuous case, so they are efficient, under the following assumption: we have (2.1) with $\delta = 0$, and b is locally bounded, and σ is càdlàg, and δ satisfies (2.4) (recall $r < 1$). We do not need (2.2), but the key properties are that the jump part of X reduces to the last term in (2.1) and that $r < 1$.

Another way to annihilate the role of the jumps is to use the *multipower variations* introduced by Barndorff-Nielsen and Shephard, see [3, 4]:

$$\widehat{C}_T^m = \sum_{i=1}^{\lfloor T/\Delta_n \rfloor - k + 1} \prod_{m=1}^k |\Delta_{i+m-1}^n X|^{2/k},$$

where $k \geq 2$ is an integer. Then, if $k \geq 3$ and under (A_0) with $\delta = 0$ and $r \leq 2/k$, plus the fact that σ never vanishes, these estimators are rate efficient, although the asymptotic variance is $\alpha(k)V_T$ for some known constant $\alpha(k) > 1$ (so we loose a bit of variance-efficiency). If instead we use the truncated bipower variation:

$$\sum_{i=1}^{\lfloor T/\Delta_n \rfloor - 1} |\Delta_i^n X \Delta_{i+1}^n X| 1_{\{|\Delta_i^n X| \leq u_n\}} 1_{\{|\Delta_{i+1}^n X| \leq u_n\}}$$

for a properly chosen sequence u_n , we obtain the same result, except that we can take any $r \leq 1$ in (A_0) (and still $\delta = 0$, of course), see [11].

Other methods are also available, although less widely used, and they all necessitate $\delta = 0$, or equivalently that the degree of activity of jumps (see below for this notion) is at most 1.

What happens in the presence of infinite variation jumps? We can examine for example the behavior of the truncated realized volatility $\widehat{C}_T^n$ (the same would apply to the other estimators) when we allow $\delta \neq 0$ in (2.1). Under (A_0) , no central limit theorem holds, but we have $\Delta_n^{\varepsilon - (2-r')/2} (\widehat{C}_T^n - C_T) \xrightarrow{\mathbb{P}} 0$, for any $\varepsilon > 0$ (and a sequence u_n depending on ε): so the rate can approach $1/\Delta_n^{(2-r')/2}$, but not quite reach it, and this is slower than $1/\sqrt{\Delta_n}$.

There is a good reason for this restriction, because of a minimax-type result proven in [7]: for *any* sequence of estimators $\widehat{C}_t^n$ and positive reals w_n , if the variables $w_n (\widehat{C}_t^n - C_t)$ are bounded in probability, uniformly in n and also in X ranging through all Itô semimartingales satisfying (A_0) for some fixed sequences Γ_n and J_n , then necessarily $w_n \leq K \left(\frac{\log(1/\Delta_n)}{\Delta_n} \right)^{(2-r')/2}$ for a constant K . This minimax rate is thus slower than $1/\sqrt{\Delta_n}$ (and becomes much slower when r' approaches (2)).

Note that in (2.4) we assume $r' > 1$; of course, it may happen that one can choose J_n satisfying $\int J_n(z)^r \lambda(dz) < \infty$ as well, but in this case we could replace the pair (δ, δ') by the pair $(0, \delta + \delta')$. When we only have $\int J_n(z)^{r'} \lambda(dz) < \infty$ for some $r' > 1$, the minimax result stated above seems to rule out the existence of estimators converging to C_T at the rate $1/\sqrt{\Delta_n}$, and it certainly does in a non-parametric setting.

However, in a parametric or semi-parametric setting, things could be different. For a simpler heuristic discussion, assume that X is a Lévy process. Then $b_t = b$ and $\sigma_t = \sigma$ are constants and (A_0) amounts to saying that the Lévy measure F satisfies $\int (|x|^{r'} \wedge 1) F(dx) < \infty$: since r' is arbitrarily close to 2, this is almost the most general possible Lévy measure. When F is *known*, an analysis of the Fisher information for c , or more appropriately here for the parameter $C_T = cT$, shows that there should exist estimators with rate $1/\sqrt{\Delta_n}$ and asymptotic variance $2Tc^2$ (hence, efficient in the previous sense), see [1]. Such estimators do exist in principle (although they are not explicit), because one could show in this case that the LAN property holds. The same theoretical result holds much more generally, for Lévy processes whose Lévy measures are sums of a known measure F_0 , plus any number of stable Lévy measures with indices $\beta_1 > \dots > \beta_M$, even when those are unknown parameters, and one could prove that this holds also when F_0 is unspecified, but

subject to $\int (|x|^r \wedge 1) F_0(dx) < \infty$ for some $r < 1$. The reason is that we are then in a semi-parametric setting, with nuisance parameters b , the β_m 's and also the scale and skewness parameters of the stable Lévy measures, plus a non-parametric nuisance term F_0 which plays no role at the end because $r < 1$.

These heuristic considerations motivate the additional assumptions which are stated in the next section, and which are far more general than the Lévy setting of the previous paragraph. Under these assumptions we will be able to construct estimators which are at least rate-efficient even though $r' > 1$ in (A_0) .

2.3 Additional Assumptions

The additional hypotheses which are needed are expressed in terms of the spot Lévy measures of the purely discontinuous martingale term in (2.1), that is of

$$\bar{X}_t = \int_0^t \int_{\mathbb{R}} \delta(s, z) (p(ds, dz) - q(ds, dx)). \quad (2.5)$$

These spot Lévy measures are the measures $F_t = F_{t,\omega}$ such that the predictable compensator of the jump measure of $\bar{X}$ takes the form occurring in the $dt \otimes F_{t,\omega}(dx)$ (this factorization exists because X is an Itô semimartingale). Equivalently, one may take for $F_{t,\omega}$ the restriction to $\mathbb{R} \setminus \{0\}$ of the image of the measure λ by the map $z \mapsto \delta(\omega, t, z)$. Its symmetric tail for $x > 0$ is thus

$$\bar{F}_t(x) := F_t((-\infty, -x) \cup (x, \infty)) = \lambda(\{z : |\delta(t, z)| > x\}). \quad (2.6)$$

With this notation, our main “semi-parametric” assumption is as follows:

Assumption (A) We have (A_0) , and also an integer $M \geq 0$, a finite family $2 > \beta_1 > \dots > \beta_M > 0$ of numbers, and M nonnegative predictable càdlàg or càglàd processes $a_t^1, \dots, a_t^M$ such that, with τ_n and Γ_n as in (A_0) , the stopped processes $(a_{t \wedge \tau_n}^m)^{1/\beta_m}$ and $\delta(t \wedge \tau_n, z)/J_n(z)$ for all z satisfy (2.3) with $\gamma = \Gamma_n$, and moreover

$$t < \tau_n \Rightarrow \left| \bar{F}_t(x) - \sum_{m=1}^M \frac{a_t^m}{x^{\beta_m}} \right| \leq \frac{\Gamma_n}{x^r}. \quad (2.7)$$

Note that the number r' occurring in (A_0) should be bigger than β_1 . When $\beta_1 < 1$ (2.7) implies for each $r'' \in (\beta_1 \vee r, 1)$ the existence of λ -integrable functions J_n'' such that at $|\delta(t, z)|^{r''} \leq J_n''(z)$ when $t < \tau_n$, so in this case we could use the pair $(0, \delta + \delta')$ instead of (δ, δ') and replace r by r'' : we would be on known grounds with well-established efficient estimators for C_T , as explained in the previous section. The same comment *a fortiori* applies when $M = 0$. So what follows has some interest

only when $\beta_1 \geq 1$ and $M \geq 1$, but we use the assumption (A) as stated above for convenience in the proof. Note that, when $M = 1$, (A) is comparable, although significantly weaker, than the assumption made in [6].

(A) implies that the spot Blumenthal-Gettoor index of $\bar{X}$ (BG index, for short, also called “degree of jump activity”) at time t is β_m if $a_t^1 = \dots = a_t^{m-1} = 0 < a_t^m$, and analogously the global BG index of X on the time interval $[0, T]$ is β_m if $A_T^1 = \dots = A_T^{m-1} = 0 < A_T^m$, where we have set

$$A_T^m = \int_0^T a_s^m ds, \quad (2.8)$$

and the same holds for the original process X as well, upon substituting β_m with $r \vee \beta_m$. One can view the variable A_T^m as the “integrated intensity” of the jumps of $\bar{X}$ with activity degree β_m .

Remark 1 Since we split the jumps of X into two separate components δ and δ' , in order to assume some kind of regularity in time for the component δ only, it seems natural to use in (A) the spot Lévy measures of $\bar{X}$. However, since the same number r occurs in (2.7) and also in (2.4) to control δ' , it is easy to see that (2.7) holds for the spot Lévy measures of $\bar{X}$, if and only if it holds for the spot Lévy measures of the process X itself (those are the images of λ by the map $z \mapsto \delta(t, z) + \delta'(tz)$).

Our representation of the jumps in X in (2.1) as integrals with respect to Poisson random measures is rather general. It accommodates specifications via Lévy-driven SDEs which are often used in applications. Indeed, any process of the form

$$X_t = X_0 + \int_0^t b_s ds + \int_0^t \sigma_s dW_s + \sum_{m=1}^M \int_0^t \gamma_{s-}^m dY_s^m + \delta' * p_t \quad (2.9)$$

with W and p as in (2.1), Y^m being Lévy processes without Gaussian part and having arbitrary dependencies with p and with one another, and γ_t^m being a càdlàg adapted process, can be represented as in (2.1).

In the setting of (2.9), Assumption (A) concerning the jumps of X is satisfied as soon as the stopped processes $\gamma_{t \wedge \tau_n}^m$ satisfy (2.3) and the (non random and non time-dependent) Lévy measure F^m of Y^m satisfy $|\bar{F}^m(x) - 1/x^{\beta_m}| \leq K/x^r$ for some constant K . Note that here one has an implicit (innocuous) standardization of Z^m , and the connection with (2.7) is that $a_t^m = |\gamma_{t-}^m|^{\beta_m}$. This explains why we introduce a regularity assumption on $(a_t^m)^{1/\beta_m}$ in (A), rather than a condition on a_t itself.

If we replace $\int_0^t \gamma_{s-}^m dY_s^m$ by $\int_0^t \gamma_{s-}^{m,+} dY_s^{m,+} + \int_0^t \gamma_{s-}^{m,-} dY_s^{m,-}$ in (2.9) with Lévy processes $Y^{m,\pm}$ having only positive jumps, the same comments above apply and therefore the setting of [6], which uses the above spectrally positive Lévy processes, is a special case of the present one.

It turns out that we will need some more structure on the values of the successive BG indices β_m , as given by the next assumption:

Assumption (B) The numbers $2 - \beta_m$ all belong to the set $\{j\rho : j = 1, 2, \dots\}$ for some unknown constant $\rho \in (0, 1)$ (so necessarily $M \leq [2/\rho]$).

We will heavily use the following integral for β equal to one of the β_m 's (it is convergent for all $\beta > 0$, but absolutely convergent when $\beta > 1$ only):

$$\beta > 0 \mapsto \chi(\beta) = \int_0^\infty \frac{\sin y}{y^\beta} dy. \quad (2.10)$$

Finally, the reader should be aware that, in the context of general Itô semimartingales, (A) and (B) are nevertheless restrictive for at least two distinct reasons:

1. Equation (2.7) stipulates an expansion of the tail function $\overline{F}_t(x)$ near 0, and in particular that $\overline{F}_t(x) \sim a_t^1/x^{\beta_1}$. This is *not the case* for a typical Lévy measure: when F is a Lévy measure, there is a unique β ($= \beta_1$ here) such that $0 < \limsup_{x \rightarrow 0} x^\beta \overline{F}(x) < \infty$, but $\liminf_{x \rightarrow 0} x^p \overline{F}(x)$ can vanish for $p = \beta$, and even for all $p \in (0, \beta]$. Dealing with Lévy measures F_t having such an erratic behavior near 0 seems hopeless with the method developed in [6] or below. So, even in the Lévy case, (2.7) is quite restrictive. Note, however, that the expansion $\sum_{m=1}^M a_t^m/x^{\beta_m}$ could be replaced by $\sum_{m=1}^M a_t^m L(x)/x^{\beta_m}$, where L is a slowly varying function (unspecified, but the same for all (ω, t) , and also all m). This would necessitate rather obvious changes in the proofs in [6] and here.
2. In (2.7) a_t^m may be time-dependent and random. This is not the case for the β_m 's, which *should be constants*. A more natural assumption would be to assume (2.7) with $\beta_m = \beta_t^m(\omega)$ being time varying and random. Unfortunately, even under strong regularity assumptions on the maps $t \mapsto \beta_t^m$, the method used here does not work at all in such a generalized setting. However, upon taking M large and ρ small in (B), we can obtain results for a reasonable approximation of random and times-varying BG indices. This is perhaps a first step toward a more general theory.

3 Initial Estimators

In this section we recall the characteristic function based method for efficient estimation of integrated volatility that we proposed in [6], and we further derive an infeasible CLT for the volatility estimators in presence of infinite variation jumps with multiple BG indices. We will use the developed limit theory in the next section to propose feasible and rate efficient inference for the integrated volatility in the general setting of assumptions (A) and (B).

The volatility estimator of [6] is built from the real part of the “local” empirical characteristic functions of increments, taken at point $u_n/\sqrt{\Delta_n}$ for some sequence $u_n > 0$ going to 0. Here, “local” means that the empirical characteristic function is computed on asymptotically shrinking windows of time length v_n , where $v_n = 2k_n\Delta_n$ and $k_n \geq 1$ is a suitable sequence of integers going to infinity, to be specified later.

In particular, for each $u > 0$ we set

$$L(u)_j^n = \frac{1}{k_n} \sum_{l=0}^{k_n-1} \cos(u(\Delta_{2jk_n+1+2l}^n X - \Delta_{2jk_n+2+2l}^n X)/\sqrt{\Delta_n}). \quad (3.1)$$

The reason for forming $L(u)_j^n$ on the basis of the first differences of the increments is to “symmetrize” the jump measure around zero as the difference of the increments of a Lévy process is equal in distribution to an increment of a Lévy process with symmetric Lévy measure. Of course, if the jumps in X are “essentially symmetric” (see [6] for formal statements), one does not need to difference the increments in the construction of $L(u)_j^n$. Note that $L(u)_j^n$ is a local version of the realized Laplace transform of volatility studied by [10].

Under (A) we have approximately

$$\mathbb{E}(L(u)_j^n \mid \mathcal{F}_{jv_n}) \approx \exp\left(-u^2 c_{jv_n} - 2 \sum_{m=1}^M \chi(\beta_m) u_n^{\beta_m} \Delta_n^{1-\beta_m/2} a_{jv_n}^m\right), \quad (3.2)$$

with formal results given in the proof. Since we are interested in the integrated volatility, we therefore form

$$\widehat{c}(u)_j^n = -\frac{1}{u^2} \log\left(L(u)_j^n \vee \frac{1}{\log(1/\Delta_n)}\right), \quad (3.3)$$

which satisfy $0 \leq \widehat{c}(u)_j^n \leq \frac{\log \log(1/\Delta_n)}{u^2}$ since $L(u)_j^n$ are not bigger than 1. $\widehat{c}(u)_j^n$ serves as a local estimator of the volatility on the block, more precisely of the average of c_t over the interval $(jv_n, (j+1)v_n]$. The associated estimators for integrated volatility are then:

$$\widehat{C}(u)_t^n = v_n \sum_{j=0}^{\lfloor t/v_n \rfloor - 1} \left(\widehat{c}(u)_j^n - \frac{1}{u^2 k_n} (\sinh(u^2 \widehat{c}(u)_j^n))^2 \right), \quad (3.4)$$

where $\sinh(x) = \frac{1}{2}(e^x - e^{-x})$ is the hyperbolic sine. The presence of the term involving the function $\sinh$ in the volatility estimator is to eliminate the effect of an asymptotic bias which arises from the nonlinear transformation of the local empirical characteristic function on the blocks.

Under appropriate assumptions on the sequence u_n we will see that $\widehat{C}(u_n)_T$ converges to C_T , and there is an associated Central Limit Theorem with the convergence rate $1/\sqrt{\Delta_n}$. However, this CLT exhibits usually a non-negligible bias, and to account for it we consider the following normalized error processes (recall (2.10) for the function $\chi(\beta)$):

$$Z(u)_t^n = \frac{1}{\sqrt{\Delta_n}} \left(\widehat{C}(u)_t^n - C_t - \sum_{m=1}^M A^m(u)_t^n \right), \quad \text{where} \quad (3.5)$$

$$A^m(u)_t^n = 2\chi(\beta_m) u_n^{\beta_m-2} \Delta_n^{1-\beta_m/2} A_t^m.$$

The next theorem presents the CLT for $Z(u)_t^n$, and also for the differences $Z(yu)_t^n - Z(u)_t^n$ when $y > 0$. The reason for giving a CLT for these differences is that they will play a key role in the de-biasing procedure developed in the next section.

Theorem 1 *Assume (A) and let $\mathcal{Y}$ be any finite subset of $(0, \infty)$. Choose the two sequences k_n and u_n in such a way that*

$$k_n \sqrt{\Delta_n} \rightarrow 0, \quad k_n \Delta_n^{1/2-\varepsilon} \rightarrow \infty \quad \forall \varepsilon > 0, \quad u_n \rightarrow 0, \quad \frac{k_n \sqrt{\Delta_n}}{u_n^2} \rightarrow 0. \quad (3.6)$$

Then, we have the (functional) stable convergence in law:

$$\left(Z(u_n)^n, \left(\frac{1}{u_n^2} (Z(yu_n)^n - Z(u_n)^n) \right)_{y \in \mathcal{Y}} \right) \xrightarrow{\mathcal{L}-s} (Z, ((y^2 - 1)\bar{Z})_{y \in \mathcal{Y}}), \quad (3.7)$$

where the limit is defined on an extension $(\tilde{\Omega}, \tilde{\mathcal{F}}, (\tilde{\mathcal{F}}_t)_{t \geq 0}, \tilde{\mathbb{P}})$ of the original space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$ and can be written as

$$Z_t = 2 \int_0^t c_s dW_s^{(1)}, \quad \bar{Z}_t = \frac{2}{\sqrt{3}} \int_0^t c_s^2 dW_s^{(2)}. \quad (3.8)$$

where $W^{(1)}$ and $W^{(2)}$ are two independent Brownian motions, independent of the σ -field $\mathcal{F}$.

Theorem 1 extends a corresponding result in [6] for $\widehat{C}(u)_t^n$ to the setting of jumps with multiple BG indices. We note that for the above theorem we assume only assumption (A) and we do not need assumption (B) for the relationship between the values of the multiple BG indices.

The two limiting processes Z and $\bar{Z}$ can be equivalently characterized by the fact that, defined on the extended space, they are globally $\mathcal{F}$ -conditional centered Gaussian martingales, $\mathcal{F}$ -conditionally independent, with $\mathcal{F}$ -conditional variance

$$\tilde{\mathbb{E}}((Z_t)^2 \mid \mathcal{F}) = 4 \int_0^t c_s^2 ds, \quad \tilde{\mathbb{E}}((\bar{Z}_t)^2 \mid \mathcal{F}) = \frac{4}{3} \int_0^t c_s^4 ds. \quad (3.9)$$

We note that since $u_n \rightarrow 0$, (3.7) implies that the difference $Z(yu_n)^n - Z(u_n)^n$ for some $y \neq 1$ is of higher asymptotic order. We will make use of this fact when developing the debiasing procedure in the next section. A consequence of the above result is the following degenerate asymptotic behavior:

$$(Z(y_l u_n)^n)_{1 \leq l \leq L} \xrightarrow{\mathcal{L}-s} (Z, \dots, Z). \quad (3.10)$$

Finally, concerning the choices of the sequences k_n and u_n , one can set $k_n \asymp 1/\sqrt{\Delta_n} (\log(1/\Delta_n))^\alpha$ and $u_n \asymp 1/(\log(1/\Delta_n))^{\alpha'}$, for any reals α, α' such that $0 < \alpha' < \frac{\alpha}{2}$.

4 Efficient Estimators

When $M = 0$ or $\beta_1 < 1$ the estimators $\widehat{C}(u_n)_t^n$ converge to C_t at rate $1/\sqrt{\Delta_n}$, and there is no asymptotic bias. When $M \geq 1$ and $\beta_1 \geq 1$ we still have the convergence, but the rate now is arbitrarily close to, although slower than, $1/\Delta_n^{(2-\beta_1)/2}$ (when expressed as a power of Δ_n): so the estimators almost achieve the minimax bound found in [7], and the rate $1/\sqrt{\Delta_n}$ is not achieved. In view of Theorem 1, the reason for this slow rate of convergence is the presence of biases coming from the infinite jump variation part of the process. One can therefore estimate the biases and perform the following bias correction which restores rate efficient estimation of integrated volatility. We fix the time horizon $T > 0$, pick some $\zeta > 1$, and then set

$$\widehat{C}(u, \zeta)_T^n = \widehat{C}(u)_T^n - \frac{(\widehat{C}(\zeta u)_T^n - \widehat{C}(u)_T^n)^2}{\widehat{C}(\zeta^2 u)_T^n - 2\widehat{C}(\zeta u)_T^n + \widehat{C}(u)_T^n}, \quad (4.1)$$

with the convention that the ratio above is 0 when its denominator vanishes. In [6], we proved the following result for the bias-corrected estimator $\widehat{C}(u, \zeta)_T^n$. However, as already mentioned, the assumption in that paper is stronger than (A), even when $M = 1$, and it is also assumed that $\beta_1 > 1$, so we will have to revisit the proof of this theorem:

Theorem 2 *Assume (A) with $M = 1$ and also that $C_T > 0$ a.s. Assume that (3.6) holds for k_n, u_n . Then the variables $\frac{1}{\sqrt{\Delta_n}}(\widehat{C}(u_n, \zeta)_T^n - C_T)$ converge stably in law to the variable Z_T .*

Notice the additional assumption $C_T > 0$ a.s. in Theorem 2. This assumption is necessary in order for the denominators in (4.1) to be “non-degenerate” in an appropriate sense. These denominators could indeed go to 0 with an uncontrolled speed on the set $\{C_T = 0\}$ on which the Brownian part of X is not active between times 0 and T , so indeed if $\mathbb{P}(C_T > 0) < 1$ the above result holds in restriction to the set $\{C_T > 0\}$ only.

Of course, when $\beta_1 < 1$ there is no need to de-bias the initial estimators $\widehat{C}(u_n)_T^n$, but a priori we do not know whether $\beta_1 < 1$ or not. As a matter of fact, we do not know either whether $M = 1$ or not, and when $M \geq 2$ the above de-biasing is in general not sufficient to remove all the bias due to the infinite variation jumps which is of higher order than $1/\sqrt{\Delta_n}$. Under the additional assumption (B), however, this can be done by “iterating” the previous de-biasing method. The generalized de-biasing procedure goes as follows:

Step 1—initialization: Choose a real $\zeta > 1$ and an integer $k \geq 1$, and put $\widehat{C}(u, \zeta, 0)_T^n = \widehat{C}(u)_T^n$. Choose also some $\xi > 0$ (typically small).

Step 2—iteration: Assuming $\widehat{C}_n(u, \zeta, j-1)$ known for some integer j between 1 and k , define $\widehat{C}_n(u, \zeta, j)$ as

$$\widehat{C}(u, \zeta, j)_T^n = \widehat{C}_n(u, \theta, j-1)_T^n + u^2 \sqrt{\Delta_n} \xi + \frac{(\widehat{C}_n(\zeta u, \theta, j-1)_T^n - \widehat{C}(u, \zeta, j-1)_T^n)^2}{\widehat{C}(\zeta^2 u, \zeta, j-1)_T^n - 2\widehat{C}(\theta u, \zeta, j-1)_T^n + \widehat{C}(u, \theta, j-1)_T^n}, \quad (4.2)$$

with again 0 for the above ratio when the denominator vanishes.

The asymptotic result is given in the following theorem.

Theorem 3 Assume (A) and (B) with $\rho \geq \rho_0$ for some $\rho_0 \in (0, 1)$, and also that $C_T > 0$ a.s. Let N be the biggest integer such that $N\rho_0 \leq 1$. Take any $\zeta > 1$ and choose the sequences k_n and u_n satisfying (3.6). Then the variables $\frac{1}{\sqrt{\Delta_n}} (\widehat{C}(u_n, \zeta, N)_T^n - C_T)$ converge stably in law to the variable Z_T .

In practice, of course, we don't know ρ , but we can decide the number of iterations we are ready to undertake. This amounts to choosing a priori the value of N , and then assume that (B) holds with some $\rho \geq 1/N$. If the “real” ρ is bigger than $1/N'$ for some integer $N' < N$, then we have iterated the procedure too many times than is actually necessary. This does not harm from an asymptotic viewpoint.

However, in finite samples de-biasing can make the estimator very unstable. Therefore, in practice it is desirable to have a data-driven adaptive choice of N which does not perform more de-biasing than is actually needed. Developing such a method can be based on the difference $\widehat{C}_n(u, \theta, j)_T^n - \widehat{C}(\zeta u, \zeta, j)_T^n$ and its asymptotic distribution when $j\rho_0 \leq 1$ which given our previous results is asymptotically mixed-normal. In an adaptive method for de-biasing, one will proceed debiasing until the difference $\widehat{C}_n(u, \theta, j)_T^n - \widehat{C}(\zeta u, \zeta, j)_T^n$ becomes smaller than some high quantile of its limit distribution. If j is too small, then the bias term will dominate the limiting behavior of $\widehat{C}_n(u, \theta, j)_T^n - \widehat{C}(\zeta u, \zeta, j)_T^n$ and hence will eventually exceed the above-mentioned high quantile of the limit distribution. We leave the formal analysis of such adaptive method for future work.

Remark 2 The estimators $\widehat{C}(u_n, \zeta, N)_T^n$ are rate-efficient, but not variance-efficient, since their asymptotic variance is twice the (supposedly) optimal one $2 \int_0^T c_s^2 ds$, see Sect. 2.2. This comes of course from the fact that in (3.1) we take the differences of two successive increments. So, we really use about $\lceil T/2\Delta_n \rceil$ variables (the aforementioned differences) instead of all the $\lceil T/\Delta_n \rceil$ increments.

As explained in [6], it is possible to use

$$L(u)_j^n = \frac{1}{k_n} \sum_{l=0}^{k_n-1} \cos(u \Delta_{j k_n + 1 + l}^n X / \sqrt{\Delta_n}) \quad (4.3)$$

instead of (3.1), and accordingly modify the de-biasing terms and the normalization in (3.3), in the case the jumps are “essentially symmetric” about 0: this means that we have the same expansion, with the same processes $a_t^m/2$, for the left and right tails $\overline{F}_t^-(x) = F_t((-\infty, -x))$ and $\overline{F}_t^+(x) = F_t((x, \infty))$. This would result in getting

estimators achieving the optimal asymptotic variance.

On the other hand, if we use (4.3) instead of (3.1) when this symmetry assumption fails, we still have Theorem 1, but with a centering term in (3.5) having a much more complicated structure. When there is a single index ($M = 1$) it is possible to do an iteration procedure resulting in a theorem analogous to the previous one, but now with the optimal conditional variance. Otherwise, since the normalizing factors $u_n^{\beta_m-2} \Delta_n^{1-\beta_m/2}$ in (3.5) and those occurring in the new additional centering terms are not commensurable, a simple iteration procedure as the one described above is not going to work.

5 Proofs

A standard localization procedure shows that it is enough to prove all results of this paper when (A) is substituted with

Assumption (SA) We have (A) with $\tau_1 \equiv \infty$, we write $J = J_1$ and $J' = J'_1$, and moreover we have $|\delta'(\cdot, z)|^r \leq \Gamma J(z)$ and $|\delta^\sigma(\cdot, z)|^2 \leq \Gamma J(z)$ for some constant Γ (implying in particular that δ' and δ^σ are bounded).

Below, this strengthened assumption is in force. Up to modifying b^σ without altering its boundedness, one may incorporate the last integral in (2.2) into the previous compensated sum of jumps term, and also use Itô's formula, to get for suitable processes and function b^c, H^c, H'^c, δ^c :

$$\begin{aligned} \sigma_t &= \sigma_0 + \int_0^t b_s^\sigma ds + \int_0^t H_s^\sigma dW_s + \int_0^t H_s'^\sigma dW'_s + \delta^\sigma(s, z) * (p - q)_t \\ c_t &= c_0 + \int_0^t b_s^c ds + \int_0^t H_s^c dW_s + \int_0^t H_s'^c dW'_s + \delta^c(s, z) * (p - q)_t. \end{aligned} \quad (5.1)$$

Then (SA) implies, for some constant $\Gamma \geq 1$,

$$\begin{aligned} &|b_t|, |\sigma_t|, |c_t|, |b_t^\sigma|, |b_t^c|, |H_t^\sigma|, |H_t'^\sigma|, |H_t^c|, |H_t'^c|, a_t \leq \Gamma \\ &|\delta(t, z)|^{r'} \leq J(z), \quad |\delta'(t, z)|^r, |\delta^\sigma(t, z)|^2, |\delta^c(t, z)|^2 \leq \Gamma J'(z) \\ &V_t = \tilde{b}_t, H_t^\sigma, H_t^c, \frac{\delta(t, z)}{J(z)} \Rightarrow \mathbb{E}(|V_{s+t} - V_s|^2 | \mathcal{F}_s) \leq \Gamma t \\ &V_t = \sigma_t, c_t \Rightarrow |\mathbb{E}(V_{s+t} - V_s | \mathcal{F}_s)| \leq \Gamma t, \quad \mathbb{E}(|V_{s+t} - V_s|^2 | \mathcal{F}_s) \leq \Gamma t \\ &\bar{F}_t(x) \leq \frac{\bar{K}}{x^{\beta_1}}, \quad |\bar{F}_t(x) - \sum_{m=1}^M \frac{a_t^m}{x^{\beta_m}}| \leq \frac{\Gamma}{x^r} \\ &\mathbb{E}(|a_{s+t}^m - a_s^m|^{2/(\beta_m \wedge 1)} | \mathcal{F}_s) \leq \Gamma t \end{aligned} \quad (5.2)$$

(for the last line we use that, if $x, y \in [-\alpha, \alpha]$, we have $||x|^\beta - |y|^\beta| \leq |x - y|^\beta$ when $0 < \beta \leq 1$, whereas $||x|^\beta - |y|^\beta| \leq \beta \alpha^{\beta-1} |x - y|$ when $\beta > 1$).

Below, y, y' implicitly are in the fixed finite subset $\mathcal{Y}$ of $(0, \infty)$, which is supposed to contain 1. We denote a generic constant, changing from line to line, as K , and it possibly depends on r, M, β_m, Γ , and sometimes on some extra parameter q such as a power or on the set $\mathcal{Y}$, but never on n and the various indices $i, j, \dots$ or variables $u, y, \dots$ which may occur. Analogously, ϕ_n stands for a generic sequence decreasing to 0.

We will also use the following trick: if $\beta_1 \leq 1$, we add a fictitious index $\beta_0 \in (1, 2)$, with the associated process a_t^0 vanishing identically. The conditions about $\bar{F}_t$ in (5.2) are satisfied as well if we replace β_1 by β_0 for the first one, and the sum $\sum_{m=1}^M$ by $\sum_{m=0}^M$ for the second one, whereas the processes $A^0(u)$ in (3.5) also vanish identically, so that the sum $\sum_{m=M_0}^M A^m(u)_t$ is the same with $M_0 = 1$ and $M_0 = 0$. Note also that β_0 can be chosen arbitrarily in $(1, 2)$. Hence, we can and will assume, without restriction, that

$$\frac{3}{2} < \beta_1 < 2. \quad (5.3)$$

5.1 Estimates

The two sequences k_n, u_n , satisfying (3.6), are fixed. As in [6] the following processes play a key role:

$$\begin{aligned} U(u)_t &= e^{-u^2 c_t}, \quad \bar{U}(m; u)_t^n = e^{-2\Delta_n^{1-\beta_m/2} \chi(\beta_m) u^{\beta_m} a_t^m} \\ \mathcal{U}(u)_t^n &= U(u)_t \prod_{m=1}^M \bar{U}(m; u)_t^n. \end{aligned} \quad (5.4)$$

Since $0 \leq c_t \leq \Gamma$ and $0 \leq a_t^m \leq \Gamma$ we see that, upon increasing Γ if necessary, we have for all $u \in (0, 1]$:

$$\frac{1}{\Gamma} \leq U(u)_t \leq 1, \quad \frac{1}{\Gamma} \leq \bar{U}(m; u)_t^n \leq 1, \quad \frac{1}{\Gamma} \leq \mathcal{U}(u)_t^n \leq 1. \quad (5.5)$$

According to (6.12) and (6.13) of [6], plus a trivial change due to the fact that our assumptions on a_t^m yield that $|\mathbb{E}(a_{t+s}^m - a_s | \mathcal{F}_s)|$ is smaller than $K t^{(\beta_m \wedge 1)/2}$ instead of $K t^{\beta_m/2}$, we have for all $q \geq 2$ and $u \in (0, 1]$:

$$\begin{aligned} |\mathbb{E}(U(u)_{t+s} - U(u)_t | \mathcal{F}_t)| &\leq K u^2 s \\ \mathbb{E}(|U(u)_{t+s} - U(u)_t|^q | \mathcal{F}_t) &\leq K u^{2q} s \\ |\mathbb{E}(\bar{U}^m(u)_{t+s}^n - \bar{U}^m(u)_t^n | \mathcal{F}_t)| &\leq K \Delta_n^{1-\beta_m/2} u^{\beta_m} s^{(\beta_m \wedge 1)/2} \\ \mathbb{E}(|\bar{U}^m(u)_{t+s}^n - \bar{U}^m(u)_t^n|^q | \mathcal{F}_t) &\leq K \Delta_n^{q(1-\beta_m/2)} u^{q\beta_m} s^{(q\beta_m/2) \wedge 1} \end{aligned} \quad (5.6)$$

and thus

$$\begin{aligned}
 |\mathbb{E}(\mathcal{U}(u)_{t+s}^n - \mathcal{U}(u)_t^n \mid \mathcal{F}_t)| &\leq K(u^2 s + \sum_{m=1}^M \Delta_n^{1-\beta_m/2} u^{\beta_m} s^{\beta_m/2}) \\
 \mathbb{E}(|\mathcal{U}(u)_{t+s}^n - \mathcal{U}(u)_t^n|^q \mid \mathcal{F}_t) &\leq K(u^{2q} s + \sum_{m=1}^M \Delta_n^{q(1-\beta_m/2)} u^{q\beta_m} s^{1\wedge(q\beta_m/2)}) \\
 \mathbb{E}(|\mathcal{U}(u)_{t+s}^n - \mathcal{U}(u)_t^n - (U(u)_{t+s} - U(u)_t) \prod_{m=1}^M u(m; u)_t^n|^q \mid \mathcal{F}_t) \\
 &\leq K \sum_{m=1}^M s^{1\wedge(q\beta_m/2)} \Delta_n^{q(1-\beta_m/2)} u^{q\beta_m}.
 \end{aligned} \tag{5.7}$$

Next, with the notation $\psi_t^{n,i} = 1_{(i\Delta_n, (i+1)\Delta_n]}(t) - 1_{((i+1)\Delta_n, (i+2)\Delta_n]}(t)$, we set

$$\begin{aligned}
 \rho_i^n &= \frac{u_n}{\sqrt{\Delta_n}} \sigma_{i\Delta_n} (\Delta_{i+1}^n W - \Delta_{i+2}^n W) = \frac{u_n}{\sqrt{\Delta_n}} \int_{R_+} \sigma_{i\Delta_n} \psi_t^{n,i} dW_t \\
 \rho_i^{\prime\prime n} &= \frac{u_n}{\sqrt{\Delta_n}} \int_{R_+ \times E} \delta(i\Delta_n, z) \psi_t^{n,i} (\underline{p} - \underline{q})(dt, dz) \\
 \rho_i^n &= \rho_i^{\prime\prime n} + \bar{\rho}_i^{\prime\prime n}, \quad \bar{\rho}_i^n = \frac{1}{u_n} (\Delta_{i+1}^n X - \Delta_{i+2}^n X),
 \end{aligned}$$

and also

$$\begin{aligned}
 \xi(y)_j^{w,n} &= \begin{cases} \frac{1}{k_n} \sum_{l=0}^{k_n-1} (\cos(y\rho_{1+2jk_n+2l}^n) - \mathcal{U}(yu_n)_{2(jk_n+l)\Delta_n}^n) & \text{if } w = 1 \\ \frac{1}{k_n} \sum_{l=0}^{k_n-1} (\cos(y\bar{\rho}_{1+2jk_n+2l}^n) - \cos(y\rho_{1+2jk_n+2l}^n)) & \text{if } w = 2 \\ \frac{1}{k_n} \sum_{l=0}^{k_n-1} (\mathcal{U}(yu_n)_{2(jk_n+l)\Delta_n}^n - \mathcal{U}(yu_n)_{jv_n}^n) & \text{if } w = 3. \end{cases} \\
 \xi(y)_j^n &= \frac{1}{\mathcal{U}(yu_n)_{jv_n}^n} \sum_{w=1}^3 \xi(y)_j^{w,n} \\
 \Omega(y)_{n,t} &= \bigcap_{0 \leq j < [t/v_n]} \{|\xi(y)_j^n| \leq \tfrac{1}{2}\}, \quad \Omega_{n,t} = \bigcap_{y \in \mathcal{Y}} \Omega(y)_{n,t}.
 \end{aligned} \tag{5.8}$$

Lemma 1 For all $q \geq 2$ and $y \in \mathcal{Y}$ we have

$$\begin{aligned}
 |\mathbb{E}(\cos(y\bar{\rho}_i^n) - \cos(y\rho_i^n) \mid \mathcal{F}_{i\Delta_n})| &\leq u_n^4 \sqrt{\Delta_n} \phi_n \\
 \mathbb{E}(|\cos(y\bar{\rho}_i^n) - \cos(y\rho_i^n)|^q \mid \mathcal{F}_{i\Delta_n}) &\leq u_n^4 \sqrt{\Delta_n} \phi_n.
 \end{aligned} \tag{5.9}$$

Proof This is basically Lemma 11 of [6], with a few changes which we explain below. We have $y\bar{\rho}_i^n = \sum_{k=1}^5 \theta(k)_i^n$, where

$$\begin{aligned}
 \theta(1)_i^n &= y\rho_i^n, \quad \theta(2)_i^n = \frac{yu_n}{\sqrt{\Delta_n}} \int_{R_+} (\sigma_s - \sigma_s^{n,i}) \psi_s^{n,i} dW_s, \\
 \theta(3)_i^n &= \frac{yu_n}{\sqrt{\Delta_n}} \int_{R_+} (b_s - b_{i\Delta_n}) \psi_s^{n,i} ds \\
 \theta(4)_i^n &= \frac{yu_n}{\sqrt{\Delta_n}} \int_{R_+ \times E} (\delta(s, z) - \delta(i\Delta_n, z)) \psi_s^{n,i} (\underline{p} - \underline{q})(ds, dz) \\
 \theta(5)_i^n &= \frac{yu_n}{\sqrt{\Delta_n}} \int_{R_+ \times E} \delta'(s, z) \psi_s^{n,i} \underline{p}(ds, dz).
 \end{aligned}$$

The only difference with [6] is in the definition of the term $\theta(4)_i^n$, so the proof goes through, provided we have the following estimate:

$$\mathbb{E}(|\theta(4)_i^n| \wedge 1 \mid \mathcal{F}_{i\Delta_n}) \leq u_n^4 \sqrt{\Delta_n} \phi_n. \quad (5.10)$$

Let $v \in (0, 1]$, and recall $|\delta(s, z)| \leq J(z) \leq 1$. We have $\theta(4)_i^n = M(v)^n + N(v)^n - B(v)^n$, where

$$\begin{aligned} M(v)^n &= \frac{yu_n}{\sqrt{\Delta_n}} \int_{R_+ \times \{z: J(z) \leq v\}} (\delta(s, z) - \delta(i\Delta_n, z)) \psi_s^{n,i} (p - q)(ds, dz) \\ B(v)^n &= \frac{yu_n}{\sqrt{\Delta_n}} \int_{R_+ \times \{z: J(z) > v\}} (\delta(s, z) - \delta(i\Delta_n, z)) \psi_s^{n,i} q(ds, dz) \\ N(v)^n &= \frac{yu_n}{\sqrt{\Delta_n}} \int_{R_+ \times \{z: J(z) > v\}} (\delta(s, z) - \delta(i\Delta_n, z)) \psi_s^{n,i} p(ds, dz), \end{aligned}$$

First, upon using (5.2) and $\int_E J(z)^{2-\varepsilon} \lambda(dz) < \infty$ for some $\varepsilon > 0$, we get

$$\begin{aligned} \mathbb{E}((M(v)^n)^2 \mid \mathcal{F}_{i\Delta_n}) &= \frac{y^2 u_n^2}{\Delta_n} \mathbb{E} \left(\int_{i\Delta_n}^{(i+2)\Delta_n} ds \int_{\{z: J(z) \leq v\}} \frac{|\delta(s, z) - \delta(i\Delta_n, z)|^2}{J(z)^2} J(z)^2 \lambda(dz) \mid \mathcal{F}_{i\Delta_n} \right) \\ &\leq K u_n^2 \Delta_n v^\varepsilon \int_E J(z)^{2-\varepsilon} \lambda(dz) \leq K u_n^2 \Delta_n v^\varepsilon. \end{aligned}$$

Second, we have

$$\begin{aligned} \mathbb{E}(|B(v)^n| \mid \mathcal{F}_{i\Delta_n}) &\leq \frac{yu_n}{\sqrt{\Delta_n}} \mathbb{E} \left(\int_{i\Delta_n}^{(i+2)\Delta_n} ds \int_{\{z: J(z) > v\}} \frac{|\delta(s, z) - \delta(i\Delta_n, z)|}{J(z)} J(z) \lambda(dz) \mid \mathcal{F}_{i\Delta_n} \right) \\ &\leq K \frac{u_n \Delta_n}{v^{1-\varepsilon}} \int_E J(z)^{2-\varepsilon} \lambda(dz) \leq K \frac{u_n \Delta_n}{v^{1-\varepsilon}}. \end{aligned}$$

Third, the process $p([0, t] \times \{z : J(z) > v\})$ is a Poisson process with parameter $\lambda(\{z : J(z) > v\}) \leq K/v^{2-\varepsilon}$, hence

$$\mathbb{P}(N(v)^n \neq 0 \mid \mathcal{F}_{i\Delta_n}) \leq K \frac{\Delta_n}{v^{2-\varepsilon}}.$$

Therefore, the left hand side of (5.10) is not bigger than

$$\mathbb{E}(|M(v)^n| \mid \mathcal{F}_{i\Delta_n}) + \mathbb{E}(|B(v)^n| \mid \mathcal{F}_{i\Delta_n}) + \mathbb{P}(N(v)^n \neq 0 \mid \mathcal{F}_{i\Delta_n}),$$

which in turn is smaller than $K(v^{\varepsilon/2} u_n \sqrt{\Delta_n} + v^{\varepsilon-2} \Delta_n)$ by the previous estimates and the Cauchy-Schwarz inequality. Upon taking $v = v_n = \Delta_n^{1/(4-\varepsilon)}$ and using the consequence of (3.6) which is $\Delta_n^\eta \leq K u_n$ for any $\eta > 0$, we deduce (5.10). $\square$

Lemma 2 For all $q \geq 2$ we have

$$|\mathbb{E}(\cos(y\rho_i^n) \mid \mathcal{F}_{i\Delta_n}) - \mathcal{U}(yu_n)_{i\Delta_n}^n| \leq u_n^4 \sqrt{\Delta_n} \phi_n \quad (5.11)$$

$$\left| \mathbb{E}(\cos(y\rho_i^n) \cos(y'\rho_i^n) \mid \mathcal{F}_{i\Delta_n}) - \frac{\mathcal{U}((y+y')u_n)_{i\Delta_n}^n + \mathcal{U}(|y-y'|u_n)_{i\Delta_n}^n}{2} \right| \leq u_n^4 \sqrt{\Delta_n} \phi_n \quad (5.12)$$

$$\mathbb{E}(|\cos(y\rho_i^n) - \mathcal{U}(yu_n)_{i\Delta_n}^n|^q \mid \mathcal{F}_{i\Delta_n}) \leq Ku_n^4. \quad (5.13)$$

Proof In view of the definition of ρ_i^n and of the property that F_t is the image of λ by the map $z \mapsto \delta(t, z)$, we see that, with the notation $G(w)_t = \int (1 - \cos(wx)) F_t(dx)$, we have

$$\mathbb{E}(\cos(y\rho_i^n) \mid \mathcal{F}_{i\Delta_n}) = U(yu_n)_{i\Delta_n} e^{-2\Delta_n G(yu_n/\sqrt{\Delta_n})_{i\Delta_n}}. \quad (5.14)$$

The analysis of the asymptotic behavior of $G(w)_t$ as $w \rightarrow \infty$ is similar to what is done in the proof of Lemma 12 of [6]. Letting $\zeta \in (0, 1]$ and using (5.2) we first observe that

$$0 \leq \int_{\{x: |x| > \zeta\}} (1 - \cos(wx)) F_t(dx) \leq 2\bar{F}_t(\zeta) \leq \frac{K}{\zeta^{\beta_1}}.$$

Next, Fubini's theorem yields

$$\int_{\{x: |x| \leq \zeta\}} (1 - \cos(wx)) F_t(dx) = \int_0^{\zeta w} \bar{F}_t\left(\frac{z}{w}\right) \sin(z) dz - \int_0^{\zeta w} \bar{F}_t(\zeta) \sin(z) dz,$$

and the absolute value of the last term above is again smaller than K/ζ^{β_1} because $|\int_0^x \sin z dz| \leq 2$ for all x . To evaluate the first term we use (5.2) again to get

$$\left| \int_0^{\zeta} \bar{F}_t\left(\frac{z}{w}\right) \sin(z) dz - \sum_{m=1}^M w^{\beta_m} \chi(\beta_m) a_t^m \right| \leq \sum_{m=1}^M \left| w^{\beta_m} \int_{\zeta w}^{\infty} \frac{\sin z}{z^{\beta_m}} dz \right| + K \int_0^{\zeta w} \frac{w^r}{z^r} dz.$$

We have $\int_x^{\infty} \frac{\sin z}{z^{\beta_m}} dz = \frac{\cos x}{x^{\beta_m}} - \beta_m \int_x^{\infty} \frac{\cos z}{z^{1+\beta_m}} dz$ by integration by parts, yielding $|\int_x^{\infty} \frac{\sin z}{z^{\beta_m}} dz| \leq 2/x^{\beta_m}$, whereas $\int_0^{\zeta w} \frac{w^r}{z^r} dz = \frac{1}{1-r} w \zeta^{1-r}$ (recall $r < 1$). Putting all these together yields

$$\left| G(w)_t - \sum_{m=1}^M w^{\beta_m} \chi(\beta_m) a_t^m \right| \leq K \left(\zeta^{-\beta_1} + \sum_{m=1}^M \zeta^{-\beta_m} + w \zeta^{1-r} \right) \leq K (\zeta^{-\beta_1} + w \zeta^{1-r})$$

because $\beta_m \leq \beta_1$. We apply this with $w = yu_n/\sqrt{\Delta_n}$ and $\zeta = \Delta_n^{1/2(1-r+\beta_1)}$ to get

$$\left| \Delta_n G(yu_n/\sqrt{\Delta_n})_t - \sum_{m=1}^M \Delta_n^{1-\beta_m/2} (yu_n)^{\beta_m} \chi(\beta_m) a_t^m \right| \leq K \Delta_n^{\frac{2-2r+\beta_1}{2-2r+2\beta_1}} \leq Ku_n^4 \sqrt{\Delta_n}$$

because again $\Delta_n^\eta \leq Ku_n$ for any $\eta > 0$, whereas $\frac{2-2r+\beta_1}{2-2r+2\beta_1} > \frac{1}{2}$. Since $|e^x - e^y| \leq |x - y|$ if $x, y \leq 0$, and recalling (5.4) and (5.14), we deduce (5.11).

At this stage, (5.12) and (5.13) follow from (5.11) exactly as in Lemma 12 of [6]. $\square$

Below, we use the simplifying notation

$$V(y, y')_t^n = \mathcal{U}((y + y')u_n)_t^n + \mathcal{U}(|y - y'|u_n)_t^n - 2\mathcal{U}(yu_n)_t^n \mathcal{U}(y'u_n)_t^n. \quad (5.15)$$

Lemma 13 of [6] is unchanged, except for (6.38) which becomes here for any $q \geq 2$:

$$\begin{aligned} |\mathbb{E}(\xi(y)_j^{3,n} | \mathcal{F}_{jv_n})| &\leq K u_n^4 \sqrt{\Delta_n} \phi_n \\ \mathbb{E}(|\xi(y)_j^{3,n}|^q | \mathcal{F}_{jv_n}) &\leq K (u_n^{2q} v_n + \sum_{m=1}^M \Delta_n^{q(1-\beta_m/2)} u_n^{q\beta_m} v_n^{1 \wedge (q\beta_m/2)}). \end{aligned} \quad (5.16)$$

However, when we assume (5.3), that is $\beta_1 > \frac{3}{2}$, it is clear that

$$\Delta_n^{q(1-\beta_m/2)} u_n^{q\beta_m} v_n^{1 \wedge (q\beta_m/2)} \leq K \Delta_n^{q(1-\beta_1/2)} u_n^{q\beta_1} v_n$$

for all $m \geq 1$. So, we indeed have (6.38) of [6], without change at all, and the following key lemma follows:

Lemma 3 *For all $q \geq 2$ we have for $j < [t/v_n]$:*

$$\begin{aligned} |\mathbb{E}(\xi(y)_j^n | \mathcal{F}_{jv_n})| &\leq K u_n^4 \sqrt{\Delta_n} \phi_n \\ \left| \mathbb{E}(\xi(y)_j^n \xi(y')_j^n | \mathcal{F}_{jv_n}) - \frac{1}{2k_n} \frac{V(y, y')_{jv_n}^n}{\mathcal{U}(yu_n)_{jv_n}^n \mathcal{U}(y'u_n)_{jv_n}^n} \right| &\leq K u_n^4 \sqrt{\Delta_n} \phi_n \\ \mathbb{E}(|\xi(y)_j^n|^q | \mathcal{F}_{jv_n}) &\leq K \left(\frac{u_n^4}{k_n^{q/2}} + u_n^{2q} v_n + \Delta_n^{q(1-\beta_1/2)} u_n^{q\beta_1} v_n \right). \end{aligned} \quad (5.17)$$

5.2 Proof of Theorem 1

The proof is once more the same as in [6], with a single change: namely, in (6.19) and (6.20), and in the subsequent definitions such as for the processes $V^{\kappa, n, \theta}$ (notation of that paper), we replace $h_{\kappa, \theta u_n}(a_t, a'_t)$ by the following (recall that here we are in the situation where $\kappa = 2$):

$$2 \sum_{m=1}^M u_n^{\beta_m-2} y^{\beta_m-2} \chi(\beta_m) a_t^m.$$

All the rest is unchanged, and thus Theorem 1 is proved. Moreover, Theorem 2 is deduced from Theorem 1 as in [6] again.

5.3 Proofs for the Rate-Efficient Estimators

Theorem 2 could be deduced from Theorem 1 as in [6] again. However, we need a sharper result for the iteration procedure of Theorem 3 to work. This method could be applied to other similar situations, but we explain it in our setting. A consequence will be Theorem 3.

We fix the time horizon $T > 0$. We are given a number $\rho \in (0, 1]$, and M is the biggest integer such that $M\rho \leq 1$. We suppose that for each $u > 0$ we have a sequence of estimators $\tilde{C}_n(u)$, with a specific asymptotic behavior described by the following, where $u_n > 0$ is a sequence with $u_n \rightarrow 0$ and $u_n/\Delta_n^\varepsilon \rightarrow \infty$ for all $\varepsilon > 0$, as in (3.6), and where $\mathcal{Y}$ is some finite subset of $(0, \infty)$:

Property (P) *There is a family $\Phi = (\Phi_m : 1 \leq m \leq M)$ of variables on $(\Omega, \mathcal{F}, \mathbb{P})$ and two variables $\tilde{Z}$ and U on an extension of this space, such that if we set*

$$\tilde{Z}_n(u) = \frac{1}{\sqrt{\Delta_n}} \left(\tilde{C}_n(u) - C_T - \sum_{m=1}^M u^{-m\rho} \Delta_n^{m\rho/2} \Phi_m \right) \quad (5.18)$$

for all $u > 0$, we have the stable convergence in law

$$\left(\tilde{Z}_n(u_n), \left(\frac{1}{u_n^2} (\tilde{Z}_n(yu_n) - \tilde{Z}_n(u_n)) \right)_{y \in \mathcal{Y}} \right) \xrightarrow{\mathcal{L}-s} (\tilde{Z}, ((y^2 - 1)U)_{y \in \mathcal{Y}}). \quad (5.19)$$

To emphasize the ingredients in (P) we sometimes write it as $P(\Phi, \tilde{Z}, U, \mathcal{Y})$. We associate with Φ the sets

$$\begin{aligned} \Omega(\Phi)_m &= \begin{cases} \{\Phi_1 \neq 0\} & \text{if } m = 1 \\ \{\Phi_1 = \dots = \Phi_{m-1} = 0 \neq \Phi_m\} & \text{if } 2 \leq m \leq M \\ \{\Phi_1 = \dots = \Phi_M = 0\} & \text{if } m = M + 1 \end{cases} \\ \Omega'(\Phi)_m &= \Omega(\Phi)_m \cap \{\Phi_{m+1} = \dots = \Phi_M = 0\}. \end{aligned}$$

We pick some $\xi \in \mathbb{R}$ and set, for any $\zeta > 1$,

$$\begin{aligned} G_n(u, \zeta) &= \tilde{C}_n(\zeta u) - \tilde{C}_n(u) \\ G'_n(u, \zeta) &= \tilde{C}_n(\zeta^2 u) - 2\tilde{C}_n(\zeta u) + \tilde{C}_n(u) \\ \tilde{C}'_n(u, \zeta, \xi) &= \tilde{C}_n(u) - \frac{G_n(u, \zeta)^2}{G'_n(u, \zeta)} 1_{\{G'_n(u, \zeta) \neq 0\}} + u_n^2 \sqrt{\Delta_n} \xi. \end{aligned} \quad (5.20)$$

We then have the following result:

Proposition 1 *Let ζ and the finite set $\mathcal{Y}$ be given, and set $\mathcal{Y}' = \{y\zeta^j : y \in \mathcal{Y}, j = 0, 1, 2\}$. If the estimators $\tilde{C}_n(u)$ satisfy $P(\Phi, \tilde{Z}, U, \mathcal{Y}')$ with $U \neq 0$ a.s., then the estimators $\tilde{C}'_n(u, \zeta, \xi)$ satisfy $P(\Phi', \tilde{Z}, U', \mathcal{Y})$, where U' and $\Phi' = (\Phi'_m)_{1 \leq m \leq M}$ satisfy the following, for all $m = 1, \dots, M$ and $k = 1, \dots, M + 1$:*

$$\begin{aligned} \Omega(\Phi)_m &\subset \Omega(\Phi')_{m+1}, & \Omega'(\Phi)_m &\subset \Omega(\Phi')_{M+1} \\ U' = \xi + h_k U &\text{ on } \Omega(\Phi)_k, \text{ where } h_k = \begin{cases} \left(\frac{\zeta^{2+k\rho}-1}{\zeta^{k\rho}-1}\right)^2 & \text{if } k \leq M \\ 0 & \text{if } k = M+1. \end{cases} \end{aligned} \quad (5.21)$$

Here, the variable U' explicitly depends on the numbers ζ and ξ , but the variables Φ'_m will be seen to also depend on ζ .

On the set $\Omega(\Phi)_{M+1}$ we have $\frac{1}{\sqrt{\Delta_n}}(\tilde{C}_n(u_n) - C_T) \xrightarrow{\mathcal{L}-s} \tilde{Z}$, and the same holds for $\tilde{C}'_n(u_n, \zeta, \xi)$, but of course in this case there is no reason to use these new estimators. However, although one might loose some kind of stability for finite samples, asymptotically it does not hurt us to use the new estimators. This is quite an important property, because in practice the variables Φ_m are not observable, so neither is the set $\Omega(\Phi)_{M+1}$.

Proof (1) We assume $P(\Phi, \tilde{Z}, U)$ with $U \neq 0$ a.s. We simplify our notation by setting

$$\begin{aligned} Z_n &= \tilde{Z}_n(u_n), & \hat{Z}_n^y &= \frac{1}{u_n^2}(\tilde{Z}_n(yu_n) - \tilde{Z}_n(u_n)) \\ Y_n^y &= \frac{1}{u_n^2}(\tilde{Z}_n(\zeta yu_n) - \tilde{Z}_n(yu_n)) = \hat{Z}_n^{\zeta y} - \hat{Z}_n^y \\ Y_n'^y &= \frac{1}{u_n^2}(\tilde{Z}_n(\zeta^2 yu_n) - 2\tilde{Z}_n(\zeta yu_n) + \tilde{Z}_n(yu_n)) = \hat{Z}_n^{\zeta^2 y} - 2\hat{Z}_n^{\zeta y} + \hat{Z}_n^y. \end{aligned}$$

Then (P) implies

$$\begin{aligned} (Z_n, (\hat{Z}_n^y)_{y \in \mathcal{Y}}, (Y_n^y)_{y \in \mathcal{Y}}, (Y_n'^y)_{y \in \mathcal{Y}}) &\xrightarrow{\mathcal{L}-s} \\ (\tilde{Z}, ((y^2 - 1)U)_{y \in \mathcal{Y}}, (y^2(\zeta^2 - 1)U)_{y \in \mathcal{Y}}, (y^2(\zeta^2 - 1)^2 U)_{y \in \mathcal{Y}}). \end{aligned} \quad (5.22)$$

We also write

$$\psi_m = \zeta^{-m\rho} - 1, \quad w_{n,y} = \frac{\sqrt{\Delta_n}}{yu_n},$$

and we can rephrase the claim as follows: find Φ'_m and U' satisfying (5.21) and such that, if

$$\tilde{Z}'_n(yu_n) := \frac{1}{\sqrt{\Delta_n}} \left(\tilde{C}'_n(yu_n, \zeta) - C_T - \sum_{m=1}^M w_{n,y}^{m\rho} \Phi'_m \right)$$

and $\hat{Z}_n'^y = \frac{1}{u_n^2}(\tilde{Z}'_n(yu_n) - \tilde{Z}'_n(u_n))$, we have the following stable convergence in law:

$$(\tilde{Z}'_n(u_n), (\hat{Z}_n'^y)_{y \in \mathcal{Y}}) \xrightarrow{\mathcal{L}-s} (\tilde{Z}, ((y^2 - 1)U')_{y \in \mathcal{Y}}). \quad (5.23)$$

(2) In view of (5.18) and (5.20), we have

$$\begin{aligned}\tilde{C}_n(yu_n) &= C_T + \sum_{m=1}^M (w_{n,y})^{m\rho} \Phi_m + \sqrt{\Delta_n} Z_n + u_n^2 \sqrt{\Delta_n} \widehat{Z}_n^y \\ G_n(yu_n, \zeta) &= \sum_{m=1}^M \psi_m(w_{n,y})^{m\rho} \Phi_m + u_n^2 \sqrt{\Delta_n} Y_n^y \\ G'_n(yu_n, \zeta) &= \sum_{m=1}^M \psi_m^2(w_{n,y})^{m\rho} \Phi_m + u_n^2 \sqrt{\Delta_n} Y_n'^y.\end{aligned}$$

We will in fact argue on each set $\Omega_m = \Omega(\Phi)_m$ for $m = 1, \dots, M+1$ separately, since they form a partition of Ω . On the set Ω_{M+1} we simply have

$$\tilde{C}'_n(yu_n, \zeta) = C_T + \sqrt{\Delta_n} Z_n + u_n^2 \sqrt{\Delta_n} \left(\widehat{Z}_n^y - \frac{(Y_n^y)^2}{Y_n'^y} 1_{\{Y_n'^y \neq 0\}} + y^2 \xi \right).$$

(5.22) and $U \neq 0$ a.s. imply $\mathbb{P}(Y_n'^y = 0) \rightarrow 0$, so a simple calculation using (5.22) again shows that (5.23) holds in restriction to Ω_{M+1} , with

$$\Phi'_1 = \dots = \Phi'_M = 0, \quad U' = \xi \quad \text{on } \Omega_{M+1}. \quad (5.24)$$

(3) Now, we suppose that we are on the set Ω_m for some $m = 1, \dots, M$. We easily check that (with an empty sum set to 0, and since $\Phi_m \neq 0$):

$$\begin{aligned}\tilde{C}'_n(yu_n, \zeta) &= C_T \\ &+ \sum_{j=m}^M (w_{n,y})^{j\rho} \Phi_j + \sqrt{\Delta_n} Z_n + u_n^2 \sqrt{\Delta_n} (\widehat{Z}_n^y + y^2 \xi) - \theta_l^{-m\rho} (w_{n,y})^{m\rho} \Phi_m \frac{N_n(y)^2}{D_n(y)},\end{aligned}$$

where
$$\begin{cases} N_n(y) = 1 + \sum_{j=m+1}^M (w_{n,y})^{(j-m)\rho} \frac{\psi_j \Phi_j}{\psi_m \Phi_m} + u_n^2 \sqrt{\Delta_n} (w_{n,y})^{-m\rho} \frac{Y_n^l}{\psi_m \Phi_m} \\ D_n(y) = 1 + \sum_{j=m+1}^M (w_{n,y})^{(j-m)\rho} \frac{\psi_j^2 \Phi_j}{\psi_m^2 \Phi_m} + u_n^2 \sqrt{\Delta_n} (w_{n,y})^{-m\rho} \frac{Y_n'^y}{\psi_m^2 \Phi_m}. \end{cases}$$

We make an expansion of the ratio $N_n(y)^2/D_n(y)$, in such a way that we keep all “significant” terms which are of order as big as $u_n^2 \sqrt{\Delta_n}$, once multiplied by $w_{n,y}^{m\rho}$. For this, we observe that $w_{n,y} \rightarrow 0$ and $w_{n,y}^{j\rho} = o(u_n^2 \sqrt{\Delta_n})$ if and only if $j > M$. Then, after some (tedious) computation, we end up with

$$\tilde{C}'_n(yu_n, \zeta) = C_T + \sum_{k=1}^5 H_n^y(k) + \sqrt{\Delta_n} Z_n + u_n^2 \sqrt{\Delta_n} \left(\widehat{Z}_n^y - \frac{2Y_n^y}{\psi_m} + \frac{Y_n'^y}{\psi_m^2} + y^2 \xi \right),$$

where, with $J_r^m(k)$ denoting the set of all r -uples $\{j_i\}$ of integers with $j_i \geq m+1$ and $\sum_{i=1}^r j_i = k$,

$$\begin{aligned}
 H_n^y(1) &= \sum_{j=m+1}^M (w_{n,y})^{j\rho} \left(1 - \frac{2\psi_j}{\psi_m}\right) \Phi_j \\
 H_n^y(2) &= - \sum_{j=m+2}^M (w_{n,y})^{j\rho} \sum_{(s,k) \in J_2^m(j)} \frac{\psi_s \psi_k}{\psi_m^2} \frac{\Phi_s \Phi_k}{\Phi_m} \\
 H_n^y(3) &= \sum_{j=m+1}^M (w_{n,y})^{j\rho} \sum_{r=1}^{j-m} (-1)^{r+1} \sum_{\{j_i\} \in J_r^m(j+rm-m)} \frac{\prod_{i=1}^r (\psi_{j_i}^2 \Phi_{j_i})}{\psi_m^r \Phi_m^{r-1}} \\
 H_n^y(4) &= 2 \sum_{j=m+2}^M (w_{n,y})^{j\rho} \sum_{r=1}^{j-m-1} (-1)^{r+1} \sum_{k=m+1}^{j-r} \sum_{\{j_i\} \in J_r^m(j+rm-k)} \frac{\psi_k \Phi_k \prod_i (\psi_{j_i}^2 \Phi_{j_i})}{\psi_m^{r+1} \Phi_m^r} \\
 H_n^y(5) &= \sum_{j=m+3}^M (w_{n,y})^{j\rho} \sum_{r=1}^{j-m-2} (-1)^{r+1} \sum_{l=2m+2}^{j+m-r} \sum_{(s,k) \in J_2^m(l)} \sum_{\{j_i\} \in J_r^m(j+rm+m-z)} \frac{\psi_s \Phi_s \psi_k \Phi_k \prod (\psi_{j_i}^2 \Phi_{j_i})}{\psi_m^{r+2} \Phi_m^{r+1}}.
 \end{aligned}$$

It is then easy, although somehow tedious, to deduce (5.23) in restriction to Ω_m from (5.22), upon taking

$$\begin{aligned}
 j = 0, \dots, m &\Rightarrow \Phi'_j = 0 \\
 j = m+1, \dots, M &\Rightarrow \Phi'_j = \left(1 - \frac{2\psi_j}{\psi_m}\right) \Phi_j - \sum_{(s,k) \in J_2^m(j)} \frac{\psi_s \psi_k}{\psi_m^2} \frac{\Phi_s \Phi_k}{\Phi_m} \\
 &+ \sum_{r=1}^{j-m} (-1)^{r+1} \sum_{\{j_i\} \in J_r^m(j+rm-m)} \frac{\prod_{i=1}^r (\psi_{j_i}^2 \Phi_{j_i})}{\psi_m^r \Phi_m^{r-1}} \\
 &+ 2 \sum_{r=1}^{j-m-1} (-1)^{r+1} \sum_{k=m+1}^{j-r} \sum_{\{j_i\} \in J_r^m(j+rm-k)} \frac{\psi_k \Phi_k \prod_i (\psi_{j_i}^2 \Phi_{j_i})}{\psi_m^{r+1} \Phi_m^r} \\
 &+ \sum_{r=1}^{j-m-2} (-1)^{r+1} \sum_{l=2m+2}^{j+m-r} \sum_{(s,k) \in J_2^m(l)} \sum_{\{j_i\} \in J_r^m(j+rm+m-z)} \frac{\psi_s \Phi_s \psi_k \Phi_k \prod (\psi_{j_i}^2 \Phi_{j_i})}{\psi_m^{r+2} \Phi_m^{r+1}} \\
 U' &= \xi + \left(1 - \frac{2(\zeta^2-1)}{\psi_m} + \frac{(\zeta^2-1)^2}{\psi_m^2}\right) U = \xi + h_m U.
 \end{aligned} \tag{5.25}$$

At this stage, if we define Ψ'_m on Ω and U' on the extended space by (5.24) in restriction to Ω_{M+1} and by (5.25) in restriction to Ω_m for any $m = 1, \dots, M$, and upon using standard properties of the stable convergence in law, we deduce (5.23), whereas (5.21) is obvious. $\square$

Proof (of Theorem 2) We define $\tilde{C}_n(u)$ to be $\hat{C}(u)_T^n$, so $\hat{C}(u, \zeta)_T^n$ is exactly $\tilde{C}'_n(u, \zeta, 0)$, as given by (5.21) with $\xi = 0$. We also set $\rho = 2 - \beta_1$, so Theorem 1 implies that $P(\Phi, \tilde{Z}, U)$ holds with $\Phi_1 = A_T^1$ and $\Phi_2 = \dots = \Phi_M = 0$ and with $\tilde{Z} = Z_T$ and $U = \bar{Z}_T$. Note that $U \neq 0$ a.s. because $C_T > 0$ a.s. by assumption. Theorem 1 and the previous proposition yield that the estimators $\tilde{C}'_n(u, \zeta, 0)$ satisfy $P(0, \tilde{Z}, U', \{1\})$, and (5.19) gives us the result. $\square$

Proof (of Theorem 3) We assume (B) here, with $\rho > \rho_0$ and $\rho_0 \in [0, 1)$, so we have $\beta_m = 2 - \nu_m \rho$ for integers $1 \leq \nu_1 < \dots < \nu_M \leq 1/\rho$. We also take $\xi > 0$.

We start with some simple changes in the setting. First, in the de-biasing sum in (3.5) one can drop out all terms such that $\beta_m < 1$, because those terms multiplied by $1/\sqrt{\Delta_n}$ and even by $1/u_n^2 \sqrt{\Delta_n}$, go to 0, hence can be removed without altering Theorem 1. In other words, we may assume $\beta_m \geq 1$ for all m . Observe that, now, all ν_m defined above are at most equal to biggest integer M' such that $M' \rho_0 \leq 1$, because $\rho \geq \rho_0$.

Second, we add “fictitious” indices, so that the indices fill in the whole set $\{2 - m\rho : m = 1, \dots, M'\}$, and we set the associated process a^m to be identically 0 for all those fictitious indices. This does not affect the de-biasing term in (3.5), so Theorem 1 stays valid with these new indices. Therefore, without loss of generality we may and will assume that we have $\beta_m = 2 - m\rho$ for all $m \leq M$ (observe that now we may again have $\beta_m < 1$ for some values of m , but we will no longer suppress those).

(2) We will prove by induction on j the following claim, where $\mathcal{Y}_m$ is the set $\{\zeta^j : j = 0, 1, \dots, 2m + 2\}$:

Claim: The estimators $\widehat{C}(u, \zeta, j)_T^n$ satisfy $P(\Phi^j, \widetilde{Z}, U^j, \mathcal{Y}_{M-j})$; we have $\Phi_1^j = \dots = \Phi_{j \wedge M}^j = 0$; furthermore U^j is $\mathcal{F}$ -conditionally Gaussian non-degenerate if $j = 0$ or in restriction to the complement of the set $\Omega(\Phi^{j-1})_{M+1}$ if $j \geq 1$, and $U^j = \xi$ otherwise.

When $j = 0$, this claim is simply Theorem 1 with $\mathcal{Y} = \mathcal{Y}_{2M}$, upon taking

$$\Phi_m^0 = A_T^m, \quad \widetilde{Z} = Z_T, \quad U^0 = \overline{Z}_T, \quad (5.26)$$

and U^0 is $\mathcal{F}$ -conditionally Gaussian non-degenerate because $C_T > 0$ a.s.

Suppose now the claim holds for some $j \geq 0$. We will apply Proposition 1 to $\widetilde{C}_n(u, \zeta, j) = \widehat{C}(u, \zeta, j)_T^n$. Indeed, the associated estimators in (5.20) are exactly $\widehat{C}(u, \zeta, j+1)_T^n$. Note also that $U^j \neq 0$ a.s. on Ω because our claim holds for j . Then this proposition tells us that the claim for $j+1$ holds (use (5.21) to obtain that $\Phi_1^{j+1} = \dots = \Phi_{(j+1) \wedge M}^{j+1} = 0$, and also that $U^{j+1} = \xi$ on the set $\Omega(\Phi^{j-1})_{M+1}$ and that U^{j+1} is $\mathcal{F}$ -conditionally Gaussian non-degenerate on the complement $\Omega(\Phi^{j-1})_{M+1}^c$, because the numbers h_k are positive for all $k \leq M$).

(3) What precedes shows that the claim holds for all j , up to $M+1$. Now, M is the number N of iterations made in the theorem, that is, $\widetilde{C}_n(u, \zeta, M+1) = \widehat{C}(u, \zeta, N)_T^n$. On the other hand, if $j = M+1$ all components of Φ^j vanish identically. Hence the estimators $\widehat{C}(u, \zeta, N)_T^n$ satisfy $P(0, \widetilde{Z}, U^N, \{1\})$, and we recall that $\widetilde{Z}$ is given in (5.26). This gives us the result. $\square$

References

1. Ait-Sahalia, Y., Jacod, J.: Fisher's information for discretely sampled Lévy processes. *Econometrica* **76**, 727–761 (2008)
2. Ait-Sahalia, Y., Jacod, J.: *High-Frequency Financial Econometrics*. Princeton University Press (2014)
3. Barndorff-Nielsen, O.E., Shephard, N.: Power and bipower variation with stochastic volatility and jumps. *J. Finan. Econometrics* **2**, 1–48 (2004)
4. Barndorff-Nielsen, O.E., Shephard, N., Winkel, M.: Limit theorems for multipower variation in the presence of jumps. *Stochast. Processes Appl.* **116**, 796–806 (2006)
5. Clément, E., Delattre, S., Gloter, A.: An infinite dimensional convolution theorem with applications to the efficient estimation of the integrated volatility. *Stochast. Processes Appl.* **123**, 2500–2521 (2013)
6. Jacod, J., Todorov, V.: Efficient estimation of integrated volatility in presence of infinite variation jumps. *Ann. Stat.* **42**, 1029–1069 (2014)
7. Jacod, J., Reiss, M.: A remark on the rates of convergence for the integrated volatility estimation in the presence of jumps. *Ann. Stat.* **42**, 1131–1144 (2014)
8. Mancini, C.: Disentangling the jumps of the diffusion in a geometric jumping Brownian motion. *Giornale dell'Istituto Italiano degli Attuari*. **LXIV**, 19–47 (2001)
9. Mancini, C.: The speed of convergence of the threshold estimator of integrated variance. *Stochast. Processes Appl.* **121**, 845–855 (2011)
10. Todorov, V., Tauchen, G.: The realized laplace transform of volatility. *Econometrica* **80**, 1105–1127 (2012)
11. Vetter, M.: Limit theorems for bipower variation of semimartingales. *Stochast. Processes Appl.* **120**, 22–38 (2010)

Model Selection for Volatility Prediction

Masayuki Uchida and Nakahiro Yoshida

*Dedicated to Professor Ole Barndorff-Nielsen on the occasion
of his 80th birthday*

Abstract We consider a stochastic regression model defined by stochastic differential equations. Based on an expected Kullback-Leibler information for the approximated distributions, we propose an information criterion for selection of volatility models. We show that the information criterion is asymptotically unbiased for the expected Kullback-Leibler information. We also give examples and simulation results of model selection.

Keywords Non-ergodic diffusions · Stable convergence · Stochastic differential equation · Volatility · Model selection · Information criterion

This work was in part supported by Japan Society for the Promotion of Science Grants-in-Aid for Scientific Research Nos. 24340015 and 24300107 (Scientific Research), Nos. 24654024, 24650148 and 26540011 (Challenging Exploratory Research); CREST Japan Science and Technology Agency; NS Solutions Corporation; and by a Cooperative Research Program of the Institute of Statistical Mathematics. The authors thank the referee for valuable comments.

M. Uchida

Graduate School of Engineering Science, Osaka University, Osaka, Japan

N. Yoshida (✉)

Graduate School of Mathematical Sciences, University of Tokyo, Tokyo, Japan

e-mail: nakahiro@ms.u-tokyo.ac.jp

M. Uchida · N. Yoshida

CREST, Japan Science and Technology Agency, Tokyo, Japan

© Springer International Publishing Switzerland 2016

M. Podolskij et al. (eds.), *The Fascination of Probability,*

Statistics and their Applications, DOI 10.1007/978-3-319-25826-3_16

1 Introduction

We consider an m -dimensional stochastic integral equation of stochastic regression model

$$Y_t = Y_0 + \int_0^t b_s ds + \int_0^t \sigma(X_s, \theta) dw_s, \quad t \in [0, T], \quad (1)$$

where w is an r -dimensional standard Wiener process on a stochastic basis $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in [0, T]}, P)$, b and X are progressively measurable processes with values in $\mathbb{R}^m$ and $\mathbb{R}^d$, respectively, σ is an $\mathbb{R}^m \otimes \mathbb{R}^r$ -valued function defined on $\mathbb{R}^d \times \Theta$, and Θ is a bounded domain in $\mathbb{R}^p$. The data set consists of discrete observations $\mathbf{X}_n = (X_{t_j}, Y_{t_j})_{0 \leq j \leq n}$ with $t_j = jh$ for $h = h_n = T/n$. The process b is completely unobservable and unknown. The data $\mathbf{X}_n$ are high frequency data and the asymptotics will be considered for $n \rightarrow \infty$. The unknown true value of θ is denoted by θ^* .

We consider functional estimation of the spot volatility $\sigma(x, \theta^*)$. By the quasi likelihood method, it is possible to construct an estimator $\hat{\theta}_n$ for θ that possesses consistency and asymptotic mixed normality when $n \rightarrow \infty$. Both the quasi maximum likelihood estimator (QMLE) and the quasi Bayes estimator (QBE) perform nicely and can be a candidate of such an estimator. They also have polynomial-order of decay of tail distributions and hence admit convergence of moments, see Uchida and Yoshida [40]. Thus we obtain an estimator $\sigma(x, \hat{\theta}_n)$ of $\sigma(x, \theta^*)$. Our goal is to predict the spot volatility $\sigma(X_{T'}, \theta^*)$, given $X_{T'}$ for $T' > T$. The state $X_{T'}$ is assumed to be given when prediction is considered. This matches the two facts that the drift is not consistently estimable on the finite time interval $[0, T]$, and that practitioners in finance often have a feeling that the historical data is less informative than the present price.¹

We will consider a model selection problem for the purpose of good prediction of the spot volatility $\sigma(x, \theta^*)$. Suppose that we have two stochastic models (1) for $\sigma(x, \theta) = \sigma^{(i)}(x, \theta)$ and $b_s = b_s^{(i)}$, $i = 1, 2$. Call these models $\mathbf{M}^{(i)}$ for $i = 1, 2$. More rigorously, each model $\mathbf{M}^{(i)}$ is equipped with an estimator $\hat{\theta}_n^{(i)}$ and we regard the pair $(\mathbf{M}^{(i)}, \hat{\theta}_n^{(i)})$ as the i th model $\mathbf{M}^{(i)}$, $i = 1, 2$. Based on the model $\mathbf{M}^{(i)}$, the prediction of the spot volatility $\sigma(x, \theta^*)$ is given by $\sigma^{(i)}(x, \hat{\theta}_n^{(i)})$. We then have two predictors $\sigma^{(1)}(x, \hat{\theta}_n^{(1)})$ and $\sigma^{(2)}(x, \hat{\theta}_n^{(2)})$ for $\sigma(x, \theta^*)$. A question is to select a better predictor that gives better approximation to $\sigma(x, \theta^*)$ than the other. The selection rule depends on x . This dependency is natural because the performance of each predictor depends on the state x at which one is considering the prediction problem. Thus it can occur that, for example, the model $\mathbf{M}^{(1)}$ is selected at $x = x_1$ but the model $\mathbf{M}^{(2)}$ is selected at a different state $x = x_2$ though these estimators use the same data. Namely, two predictions for $\sigma(x_1, \theta^*)$ and $\sigma(x_2, \theta^*)$ are different questions. In this paper, we will give an information criterion sVIC_x for volatility model selection.

¹However, as a matter of fact, it is incorrect as for estimation of volatility.

As measurement and storage technologies developed, statistical inference for high frequency data has been attracting attention since around 1990: Prakasa Rao [29, 30], Dacunha-Castelle and Florens-Zmirou [12], Florens-Zmirou [14], Yoshida [42, 46], Genon-Catalot and Jacod [15], Bibby and Soerensen [11], Kessler [21], Shimizu and Yoshida [33], Uchida [35], Ogihara and Yoshida [26, 27], Uchida and Yoshida [39, 40] and Masuda [25] among others.

For real data, a new trend emerged in statistical analysis as Andersen and Bollerslev [3], Andersen et al. [4], Barndorff-Nielsen and Shephard [6, 7]. A great number of studies on various variations were induced in this trend.

It is well known that model selection is of fundamental importance in data analysis and statisticians experienced a remarkable success of AIC (Akaike [1, 2]) and various information criteria thereafter (Schwarz [32], Takeuchi [34], Konishi and Kitagawa [22] and many others). From a mathematical aspect, AIC is a bias-correction of an estimator of the Kullback-Leibler divergence between the predictive distribution and the true distribution. It is possible to view this problem in the frame of higher-order asymptotic theory; cf. Konishi and Kitagawa [22] for the generalized information criterion, Uchida and Yoshida [36, 38] for information criteria for mixing processes, Uchida and Yoshida [37] for small diffusions.

After discovery of non standard behavior, i.e. asymptotic mixed normality, of the estimators (White [41], Anderson [5], Rao [31], Keiding [19, 20]), the necessity of extension of the classical asymptotic decision theory was recognized and the non-ergodic statistics was formulated around 1980 (Basawa and Koul [8], Basawa and Prakasa Rao [9], Jeganathan [18], and Basawa and Scott [10]). The nesting condition with Rényi mixing was applied to deduce mixed normal limits for classical objects of non-ergodic statistics; see e.g. Feigin [13].

Inference for volatility parameters under the finite time horizon is a theme of non-ergodic statistics. There, asymptotic mixed normality is featured, as mentioned above. For this type of limit theorems, the martingale problem approach as in Genon-Catalot and Jacod [15], Jacod [16], and also convergence of stochastic integrals in Jakubowski et al. [17] and Kurtz and Protter [23] are available to obtain stable convergence. Later on, it will be necessary to carry out computations in higher-order than ordinary limit theorems. When the process has a mixing property or ergodicity, the mixing expansion [24, 45] and the classical martingale expansion [43, 44] served in derivation of information criteria for stochastic processes. However, we cannot apply classical approaches because our present question is in non-ergodic statistics where limit theorems are non normal due to lack of ergodicity. Asymptotic expansion for martingales with a mixed normal limit is in Yoshida [48, 49]. In this article, we will apply a method based on the Malliavin calculus used there. In connection with this subject, we refer the reader to [47] for a distributional expansion of a quadratic form of an Itô process, including realized volatility, and to Podolskij and Yoshida [28] for an expansion of the p -variation.

2 Quasi Likelihood Analysis for Volatility

Suppose that Θ is a bounded domain in $\mathbb{R}^p$ with a locally Lipschitz boundary, which means that Θ has the strong local Lipschitz condition. Let $C_{\uparrow}^{k,l}(\mathbb{R}^d \times \Theta; \mathbb{R}^m)$ denote the space of all functions f satisfying the following conditions: (i) $f(x, \theta)$ is an $\mathbb{R}^m$ -valued function on $\mathbb{R}^d \times \Theta$, (ii) $f(x, \theta)$ is continuously differentiable with respect to x up to order k for all θ and (iii) for $|\mathbf{n}| = 0, 1, \dots, k$, $\partial_x^{\mathbf{n}} f(x, \theta)$ is continuously differentiable with respect to θ up to order l for all x . Moreover, for $|\mathbf{v}| = 0, 1, \dots, l$ and $|\mathbf{n}| = 0, 1, \dots, k$, $\partial_{\theta}^{\mathbf{v}} \partial_x^{\mathbf{n}} f(x, \theta)$ is of at most polynomial growth in x uniformly in θ .² Here $\mathbf{n} = (n_1, \dots, n_d)$ and $\mathbf{v} = (v_1, \dots, v_p)$ are multi-indices, $\mathbf{p} = \dim(\Theta)$, $|\mathbf{n}| = n_1 + \dots + n_d$, $|\mathbf{v}| = v_1 + \dots + v_p$, $\partial_x^{\mathbf{n}} = \partial_{x_1}^{n_1} \dots \partial_{x_d}^{n_d}$, $\partial_{x_i} = \partial/\partial x_i$, and $\partial_{\theta}^{\mathbf{v}} = \partial_{\theta_1}^{v_1} \dots \partial_{\theta_p}^{v_p}$, $\partial_{\theta_i} = \partial/\partial \theta_i$. We denote by $\rightarrow^p$ and $\rightarrow^{d_s(\mathcal{F})}$ the convergence in probability and the $\mathcal{F}$ -stable convergence in distribution, respectively. For matrices A and B of the same size, we write $A^{\otimes 2} = AA^*$ and $A[B] = \text{Tr}(AB^*)$, where $\star$ means the transpose. More generally, $T^{(1)}[T^{(2)}] = \sum_{i_1, \dots, i_k} T_{i_1, \dots, i_k}^{(1)} T_{i_1, \dots, i_k}^{(2)}$ for two tensors $T^{(1)} = (T_{i_1, \dots, i_k}^{(1)})$ and $T^{(2)} = (T_{i_1, \dots, i_k}^{(2)})$. A tensor is identified with a multi-linear form, that is, $T^{(1)}[T^{(2)}, T^{(3)}] = T^{(1)}[T^{(2)} \otimes T^{(3)}]$ for tensors $T^{(i)}$ ($i = 1, 2, 3$) of adequate sizes. Set $S(x, \theta) = \sigma(x, \theta)^{\otimes 2}$ and $\Delta_j Y = Y_{t_j} - Y_{t_{j-1}}$. Denote by I_m the $m \times m$ -identity matrix. We assume that the function σ admits a continuous extension over $\mathbb{R}^d \times \bar{\Theta}$, and denote it by σ . Let

$$\chi_0 = \inf_{\theta \neq \theta^*} \frac{-\mathbb{Y}(\theta)}{|\theta - \theta^*|^2}, \quad (2)$$

where

$$\mathbb{Y}(\theta) = -\frac{1}{2T} \int_0^T \left\{ \log \left(\frac{\det S(X_t, \theta)}{\det S(X_t, \theta^*)} \right) + \text{Tr} \left(S^{-1}(X_t, \theta) S(X_t, \theta^*) - I_m \right) \right\} dt.$$

The index χ_0 expresses the degree of separation of the model for θ from the true model for θ^* .

We make the following assumptions.

[A1] $\sigma \in C_{\uparrow}^{2,4}(\mathbb{R}^d \times \Theta; \mathbb{R}^m \otimes \mathbb{R}^r)$ and $\inf_{x, \theta} \det S(x, \theta) > 0$.

[A2] $\sup_{0 \leq t \leq T} \|b_t\|_p < \infty$ for all $p > 1$.

[A3] The process X admits a representation

$$X_t = X_0 + \int_0^t \hat{b}_s ds + \int_0^t a_s dw_s + \int_0^t \tilde{a}_s d\tilde{w}_s,$$

where $\hat{b}$, a and $\tilde{a}$ are progressively measurable processes taking values in $\mathbb{R}^d$, $\mathbb{R}^d \otimes \mathbb{R}^r$ and $\mathbb{R}^d \otimes \mathbb{R}^{r_1}$, respectively, satisfying

²That is, there exists a constant C such that $|\partial_{\theta}^{\mathbf{v}} \partial_x^{\mathbf{n}} f(x, \theta)| \leq C(1 + |x|)^C$ for all $(x, \theta) \in \mathbb{R}^d \times \Theta$.

$$\|X_0\|_p + \sup_{t \in [0, T]} (\|\hat{b}_t\|_p + \|a_t\|_p + \|\tilde{a}_t\|_p) < \infty$$

for every $p > 1$, and $\tilde{w}$ is an r_1 -dimensional Wiener process independent of w .
[A4] For every $L > 0$, there exists $c_L > 0$ such that

$$P[X_0 \leq r^{-1}] \leq \frac{c_L}{r^L}$$

for all $r > 0$.

Condition [A1] seems theoretically restrictive because a simple Black-Scholes model does not satisfies it though considering the logarithmic transform is another natural possibility, of course. However, practically we can modify the model outside of a realistic region of values of X so that the coefficient satisfies the nondegeneracy condition. Estimation will not be affected by this modification when the path of X is staying in that region. If analysis in degenerate cases is of interest, the problem becomes a different one in character.

Since the transition density does not have an explicit form, the inference is carried out by a quasi likelihood function. Let

$$\mathbb{H}_n(\theta) = -\frac{1}{2} \sum_{k=1}^n \{ \log \det S(X_{t_{j-1}}, \theta) + h^{-1} S^{-1}(X_{t_{j-1}}, \theta) [(\Delta_j Y)^{\otimes 2}] \}.$$

Then the maximum likelihood type estimator $\hat{\theta}_n$ is any estimator that takes values in $\bar{\Theta}$ and satisfies

$$\mathbb{H}_n(\hat{\theta}_n) = \sup_{\theta \in \bar{\Theta}} \mathbb{H}_n(\theta). \quad (3)$$

Let $\Gamma(\theta^*) = (\Gamma^{ij}(\theta^*))_{i,j=1,\dots,p}$ with

$$\Gamma^{ij}(\theta^*) = \frac{1}{2T} \int_0^T \text{Tr}((\partial_{\theta_i} S) S^{-1} (\partial_{\theta_j} S) S^{-1}(X_t, \theta^*)) dt,$$

where all factors in the trace are evaluated at (X_t, θ^*) . Let ζ be a p -dimensional standard normal random variable independent of $\Gamma(\theta^*)$.

Theorem 1 (Uchida and Yoshida [40]) *Suppose that Conditions [A1], [A2], [A3] and [A4] are fulfilled. Then $\sqrt{n}(\hat{\theta}_n - \theta^*) \rightarrow_{d_s(\mathcal{F})} \Gamma(\theta^*)^{-1/2} \zeta$ and*

$$E \left[\mathbf{f}(\sqrt{n}(\hat{\theta}_n - \theta^*)) \right] \rightarrow \mathbb{E} \left[\mathbf{f}(\Gamma(\theta^*)^{-1/2} \zeta) \right]$$

as $n \rightarrow \infty$ for all continuous functions $\mathbf{f}$ of at most polynomial growth.

Remark 1 This is Theorem 4 of Uchida and Yoshida [40], where Condition [A4] is referred to as Condition [H2]. The same paper verified Condition [H2] by combining

non-degeneracy of X and the non-degeneracy of the function

$$Q(x, \theta, \theta^*) = \text{Tr} \left(S(x, \theta)^{-1} S(x, \theta^*) - I_m \right) - \log \det \left(S(x, \theta)^{-1} S(x, \theta^*) \right).$$

An analytic criterion and an geometric criterion are discussed there. The geometric criterion says the non degeneracy of χ_0 holds for non degenerate diffusion process X if the null set of $Q(\cdot, \theta, \theta^*)$ is covered by a union of locally Lipschitz graphs possibly depending on θ . The analytic criterion is more complicated. Asymptotic results of Bayes type estimator are given in Theorems 2 and 5 in the same paper. It turns out that $\hat{\theta}_n$ is unique on an event A_n with $P[A_n] \rightarrow 1$ thanks to the convexity of $\mathbb{H}_n$.

3 Volatility Information Criterion

For a matrix S_0 and a non-singular matrix S_1 , let

$$U(S_1, S_0) = \frac{1}{2} \log \det S_1 + \frac{1}{2} S_1^{-1} [S_0].$$

Set $\Gamma_n(\hat{\theta}_n) = (\Gamma_n^{ij}(\hat{\theta}_n))_{i,j=1,\dots,p}$ with

$$\Gamma_n^{ij}(\hat{\theta}_n) = \frac{1}{2n} \sum_{k=1}^n \text{Tr} \left((\partial_{\theta_i} S) S^{-1} (\partial_{\theta_j} S) S^{-1} (X_{t_{j-1}}, \hat{\theta}_n) \right).$$

$S(x, \theta) = \sigma(x, \theta)^{\otimes 2}$ is a more basic quantity than $\sigma(x, \theta)$ from the aspect of modeling of distribution of Y . The significance of the “spot volatility” $\sigma(x, \theta^*)$ or $S(x, \theta^*)$ is that it determines, as the first approximation, the behavior of Y in distributional sense in very short time. The distribution of $Y_{t+\delta t} - Y_t$ is approximated by the normal density $N_m(0, \delta S(x, \theta^*))$ when δ is small. So it is natural to adopt the Kullback-Leibler divergence between $N_m(0, \delta S(x, \theta^*))$ and $N_m(0, \delta S(x, \hat{\theta}_n))$ to measure the discrepancy between $\sigma(x, \theta^*)$ and $\sigma(x, \hat{\theta}_n)$. Let $\phi_m(z; 0, A)$ denote the m -dimensional normal density with mean 0 and covariance matrix A .

In this way, we select a model that attains the minimum K-L divergence

$$\begin{aligned} & K(N_m(0, \delta S(x, \theta^*)), N_m(0, \delta S(x, \hat{\theta}_n))) \\ &= \int \log \frac{\phi_m(z; 0, \delta S(x, \theta^*))}{\phi_m(z; 0, \delta S(x, \hat{\theta}_n))} \phi_m(z; 0, \delta S(x, \theta^*)) dz \\ &= U(S(x, \hat{\theta}_n), S(x, \theta^*)) - U(S(x, \theta^*), S(x, \theta^*)) \end{aligned}$$

among all statistical models in question. Only $U(S(x, \hat{\theta}_n), S(x, \theta^*))$ differs among correctly specified models.

Since $U(S(x, \hat{\theta}_n), S(x, \theta^*))$ involves the unknown $S(x, \theta^*)$, we estimate this value by

$$U(S(x, \hat{\theta}_n), S(x, \hat{\theta}_n)) = \frac{1}{2} \log \det S(x, \hat{\theta}_n) + \frac{1}{2} \mathbf{m}.$$

However, this estimator has a bias. We will evaluate the bias to derive an information criterion.

Let

$$\mathbb{H}_n^\dagger(\theta) = -\frac{1}{2} \sum_{k=1}^{[n/2]} \left\{ \log \det S(X_{t_{2(k-1)}}, \theta) + (2h)^{-1} S^{-1}(X_{t_{2(k-1)}}), \theta \right\} [(Y_{t_{2k}} - Y_{t_{2(k-1)}})^{\otimes 2}].$$

Then the maximum likelihood type estimator $\hat{\theta}_n^\dagger$ with half of the data is any estimator that satisfies

$$\mathbb{H}_n^\dagger(\hat{\theta}_n^\dagger) = \sup_{\theta \in \tilde{\Theta}} \mathbb{H}_n^\dagger(\theta). \quad (4)$$

For $x \in \mathbb{R}^d$, define sVIC_x (spot volatility information criterion at x) by

$$\begin{aligned} \text{sVIC}_x &= \log \det S(x, \hat{\theta}_n) + \partial_\theta(S^{-1})(x, \hat{\theta}_n)[S(x, \hat{\theta}_n) \otimes (\hat{\theta}_n^\dagger - \hat{\theta}_n)] \\ &\quad + \frac{1}{2n} \partial_\theta^2(S^{-1})(x, \hat{\theta}_n)[S(x, \hat{\theta}_n) \otimes \Gamma_n^{-1}(\hat{\theta}_n)]. \end{aligned}$$

A model that minimizes sVIC_x is selected for prediction of the volatility $\sigma(x, \theta^*)$ at state x . We will see bias-correction by sVIC_x in estimation of the Kullback-Leibler divergence validates the information criterion, as it was done by the standard Akaike information criterion in ergodic statistics.

Let $r_0 = r_1 + r$. Hereafter we will consider an r_0 -dimensional Wiener space $(\mathbb{W}, \mathbb{B}(\mathbb{W}), P^{\mathbb{W}})$, where $\mathbb{W} = C([0, T]; \mathbb{R}^{r_0})$ equipped with the sup-norm, the Borel σ -field $\mathbb{B}(\mathbb{W})$ and an r_0 -dimensional Wiener measure $P^{\mathbb{W}}$. Denote by $\hat{w} = (\tilde{w}, w)$ the $\mathbb{R}^{r_1} \times \mathbb{R}^r$ -valued canonical process on $\mathbb{W}$. We suppose that $(\Omega, \mathcal{F}, P) = (\Omega', \mathcal{F}', P') \times (\mathbb{W}, \mathbb{B}(\mathbb{W}), P^{\mathbb{W}})$ for a probability space $(\Omega', \mathcal{F}', P')$ and consider the filtration $\mathbb{F} = (\mathcal{F}_t)$ with $\mathcal{F}_t = \cap_{s>t} (\mathcal{F}' \times \mathcal{F}_s^{\mathbb{W}})$ for $\mathcal{F}_t^{\mathbb{W}} = \sigma[\hat{w}_s; s \leq t]$. Functions on Ω' or $\mathbb{W}$ are naturally regarded as functions on Ω . The standard Malliavin calculus on $\mathbb{W}$ is extended on Ω . We will denote by D and D^* the H -derivative and the divergence operator, respectively. The (s, p) -space of E -valued functionals on Ω is denoted by $\mathbb{D}_{s,p}(E)$ for a differentiability index $s \in \mathbb{R}$, an integrability index $p > 1$, and a separable Hilbert space E . The (s, p) -norm is denoted by $\|\cdot\|_{s,p}$.

We will work with the following conditions in place of [A2] and [A3].

[A2[#]] Process $b = (b_t)_{t \in [0, T]}$ admits a representation

$$b_t = b_0 + \int_0^t b_s^{[0]} ds + \int_0^t b_s^{[1]} d\hat{w}_s$$

for some (vector/matrix-valued) càdlàg adapted processes $b^{[0]}$ and $b^{[1]}$ such that

$$\|b_0\|_p + \sup_{t \in [0, T]} \sum_{i=0,1} \|b_t^{[i]}\|_p < \infty$$

for every $p > 1$.

[A3[#]] (i) The process X admits a representation

$$X_t = X_0 + \int_0^t \hat{b}_s ds + \int_0^t \hat{a}_s d\hat{w}_s,$$

where $\hat{b} = (\hat{b}_t)_{t \in [0, T]}$ and $\hat{a} = (\hat{a}_t)_{t \in [0, T]}$ are càdlàg adapted processes satisfying

$$\|X_0\|_p + \sup_{t \in [0, T]} (\|\hat{b}_t\|_p + \|\hat{a}_t\|_p) < \infty$$

for every $p > 1$.

(ii) $X_t \in \cap_{p>1} \mathbb{D}_{2,p}(\mathbb{R}^d)$ and $\sup_{(s,t) \in [0, T]^2} \|D_s X_t\|_p + \sup_{(r,s,t) \in [0, T]^3} \|D_{r,s} X_t\|_p < \infty$ for every $p > 1$, and there exist stochastic processes $\Psi_t^{(i)}$ ($i = 1, 2$) such that

$$\lim_{\delta \downarrow 0} \sup_{t \in [0, T]} \left(\sup_{s: s < t < (s+\delta) \wedge T} \|D_s X_t - \Psi_t^{(1)}\|_p + \sup_{r, s: r < s < (r+\delta) \wedge T} \|D_{r,s} X_t - \Psi_t^{(2)}\|_p \right) = 0$$

for every $p > 1$.

Condition [A3[#]] can be verified when X is a diffusion process satisfying a stochastic differential equation with sufficiently smooth coefficients.

The following result validates sVIC_x.

Theorem 2 Suppose that Conditions [A1], [A2[#]], [A3[#]] and [A4] are fulfilled. Then for each $x \in \mathbb{R}^d$,

$$E \left[\text{sVIC}_x + \mathfrak{m} - 2U(S(x, \hat{\theta}_n), S(x, \theta^*)) \right] = o \left(\frac{1}{n} \right).$$

4 Proof of Theorem 2

In Lemma 1 below, we consider a sequence of points $(t_j)_{j=0,\dots,m}$ for $n \in \mathbb{N}$ such that $0 \leq t_0 < t_1 < \dots < t_m \leq T$. Let $I_j = (t_{j-1}, t_j]$ and $|I_j| = t_j - t_{j-1}$.

Lemma 1 *Suppose that the following conditions are fulfilled.*

(i) $\Phi \in \mathbb{D}_{k,2}(\mathbb{R})$ has densities such that

$$\sup_{u_1, \dots, u_\ell \in [0, T]} \|D_{u_1, \dots, u_\ell} \Phi\|_2 < \infty \quad (1 \leq \ell \leq k).$$

(ii) $a = (a(s))_{s \in [0, T]}$ is a progressively measurable process satisfying $\sup_{t \in [0, T]} \|a(t)\|_2 < \infty$.

Let

$$\mathbb{M}[v_1 \otimes \dots \otimes v_k] = \sum_{j=1}^m a(t_{j-1}) \int_{t_{j-1}}^{t_j} v_1 d\hat{w}_{s_1} \int_{t_{j-1}}^{s_1} v_2 d\hat{w}_{s_2} \dots \int_{t_{j-1}}^{s_{k-1}} v_k d\hat{w}_{s_k}$$

for $v_1, \dots, v_k \in \mathbb{R}^{r_0}$. Then

$$E[\mathbb{M}\Phi] = \sum_{j=1}^m \int_{t_{j-1}}^{t_j} \int_{t_{j-1}}^{s_1} \dots \int_{t_{j-1}}^{s_k} E\left[a(t_{j-1}) D_{s_k, \dots, s_2, s_1} \Phi\right] ds_k \dots ds_2 ds_1.$$

Moreover, if $t \mapsto a(t)$ is L_2 -continuous and if

$$\lim_{\delta \downarrow 0} \sup_{t \in [0, T]} \sup_{s_k, \dots, s_1 \in \mathcal{T}(t, \delta)} \|D_{s_k, \dots, s_1} \Phi - \Psi_t\|_2 = 0$$

for some stochastic process $\Psi = (\Psi_t)_{t \in [0, T]}$, where $\mathcal{T}(t, \delta) = \{(s_k, \dots, s_1) \in [0, T]^k; s_k < \dots < s_1 < t < (s_k + \delta) \wedge T\}$, then

$$E[\mathbb{M}\Phi] = \sum_{j=1}^m \frac{|I_j|^{k-1}}{k!} \int_{t_{j-1}}^{t_j} E[a(t) \Psi_t] dt + o\left(\sum_{j=1}^m |I_j|^k\right)$$

in $\mathbb{R}^{\otimes r_0}$ as $\max_{j=1, \dots, m} |I_j| \rightarrow 0$.

Proof It is possible to prove this lemma by applying duality between the H -derivative D and the divergence operator D^* . $\square$

For a sequence (R_n) of random vectors and a sequence (r_n) of positive numbers, we write $R_n = O_L(r_n)$ if $\sup_n \|r_n^{-1} R_n\|_{L^p} < \infty$ for every $p > 1$, and write $R_n = o_L(r_n)$ if $\lim_n \|r_n^{-1} R_n\|_{L^p} = 0$ for every $p > 1$. By Theorem 1, $\hat{\theta}_n - \theta^* = O_L(n^{-1/2})$.

Proof of Theorem 2 For an open ball $B(\theta^*) \subset \Theta$, $\partial_\theta \mathbb{H}_n(\hat{\theta}_n) = 0$ when $\hat{\theta}_n \in B(\theta^*)$. Then it is possible to obtain a stochastic expansion

$$\sqrt{n}(\hat{\theta}_n - \theta^*) = \Gamma^{-1} M_n^{(0)} + \frac{1}{\sqrt{n}} N_n. \quad (5)$$

This formula is obtained first by using Taylor's formula on an event A_n such that $P(A_n^c) = O(n^{-K})$ as $n \rightarrow \infty$ for every $K > 0$; this estimate of A_n is provided by the boundedness of L_p -norms of $\sqrt{n}(\hat{\theta}_n - \theta^*)$ thanks to Theorem 1. In this procedure, we remark that the residual term included in N_n is negligible in L_p -sense due to $P(A_n^c) = O(n^{-K})$ and the boundedness of Θ . In other words, we have a global expansion taking the form of (5). More precisely,

$$M_n^{(0)} = -\frac{1}{2\sqrt{n}} \sum_j (\sigma'_{t_{j-1}} (\partial_\theta S_{t_{j-1}}^{-1}) \sigma_{t_{j-1}}) [h^{-1}(\Delta_j w)^{\otimes 2} - I_r], \quad (6)$$

where $\sigma_t = \sigma(X_t, \theta^*)$ and $\partial_\theta S_t^{-1} = (\partial_\theta S^{-1})(X_t, \theta^*)$. See [50] for details of the stochastic expansion. Though the random variable N_n has an involved expression, it is possible to verify

$$E[N_n] \rightarrow c \quad (n \rightarrow \infty) \quad (7)$$

for some constant vector c . Moreover, by using Lemma 1 under [A3[#]], we obtain

$$\sqrt{n} E[\Gamma^{-1} M_n^{(0)}] \rightarrow c_1 \quad (n \rightarrow \infty) \quad (8)$$

for some constant vector c_1 .³ Then from (5), (7) and (8),

$$E[\hat{\theta}_n - \theta^*] = \frac{1}{n} C + o\left(\frac{1}{n}\right) \quad (9)$$

as $n \rightarrow \infty$ for some constant vector C .

The half estimator admits a stochastic expansion similar to (5) with (6). Though the expression of sums $\sum_j \cdots$ appearing in the stochastic expansion is slightly different for odd n , it is observed that the last summand does not effect asymptotically. Therefore we can conclude that

$$E[\hat{\theta}_n^\div - \theta^*] = \frac{2}{n} C + o\left(\frac{1}{n}\right) \quad (10)$$

as $n \rightarrow \infty$.

³The integrability condition can be relaxed, in fact.

Write $\hat{S}_n = S(x, \hat{\theta}_n)$ and $S^* = S(x, \theta^*)$ for simplicity. We will evaluate the bias of $U(\hat{S}_n, \hat{S}_n)$, that is,

$$E[U(\hat{S}_n, \hat{S}_n) - U(\hat{S}_n, S^*)] = \frac{1}{2}E[\mathbf{m} - \hat{S}_n^{-1}[S^*]].$$

It follows from Theorem 1 that $\sqrt{n}(\hat{\theta}_n - \theta^*) \rightarrow_{d_s(\mathcal{F})} \Gamma(\theta^*)^{-1/2}\zeta$. Furthermore, $\Gamma_n(\hat{\theta}_n) \rightarrow^p \Gamma(\theta^*)$. We have

$$\begin{aligned} \hat{S}_n^{-1} &= S(x, \hat{\theta}_n)^{-1} \\ &= S(x, \theta^*)^{-1} + \partial_\theta(S^{-1})(x, \theta^*)[\hat{\theta}_n - \theta^*] \\ &\quad + \frac{1}{2}\partial_\theta^2(S^{-1})(x, \theta^*)[(\hat{\theta}_n - \theta^*)^{\otimes 2}] + O_L(n^{-3/2}). \end{aligned}$$

From (9) and (10), we have

$$E[\hat{\theta}_n^\div - \hat{\theta}_n] = \frac{C}{n} + o\left(\frac{1}{n}\right).$$

Now

$$\begin{aligned} &2E[U(\hat{S}_n, \hat{S}_n) - U(\hat{S}_n, S^*)] \\ &= E[\mathbf{m} - \hat{S}_n^{-1}[S^*]] \\ &= -\partial_\theta(S^{-1})(x, \theta^*)[n^{-1}C, S^*] - E\left[\frac{1}{2n}\partial_\theta^2(S^{-1})(x, \theta^*)[S^* \otimes \Gamma^{-1}]\right] + o(n^{-1}) \end{aligned}$$

Then obviously,

$$\begin{aligned} &E\left[\left\{\log \det \hat{S}_n + \mathbf{m} + \partial_\theta(S^{-1})(x, \theta^*)\left[S^* \otimes (\hat{\theta}_n^\div - \hat{\theta}_n)\right] \right. \right. \\ &\quad \left. \left. + \frac{1}{2n}\partial_\theta^2(S^{-1})(x, \theta^*)[S^* \otimes \Gamma^{-1}]\right\} - 2U(\hat{S}_n, S^*)\right] = o(n^{-1}). \end{aligned}$$

As a consequence, we obtain

$$\begin{aligned} &E\left[\left\{\log \det \hat{S}_n + \mathbf{m} + \partial_\theta(S^{-1})(x, \hat{\theta}_n)[\hat{S}_n \otimes (\hat{\theta}_n^\div - \hat{\theta}_n)] \right. \right. \\ &\quad - \partial_\theta(S^{-1})(x, \theta^*)[\partial_\theta S(x, \theta^*)[\hat{\theta}_n - \theta^*] \otimes (\hat{\theta}_n^\div - \hat{\theta}_n)] \\ &\quad - \partial_\theta^2(S^{-1})(x, \theta^*)[S^* \otimes (\hat{\theta}_n^\div - \hat{\theta}_n) \otimes (\hat{\theta}_n - \theta^*)] \\ &\quad \left. \left. + \frac{1}{2n}\partial_\theta^2(S^{-1})(x, \hat{\theta}_n)[S^* \otimes \Gamma_n^{-1}]\right\} - 2U(\hat{S}_n, S^*)\right] = o(n^{-1}). \end{aligned}$$

By stochastic expansions,

$$\begin{aligned}
 & \sqrt{n}(\hat{\theta}_n - \theta^*) \\
 &= -\frac{1}{2\sqrt{n}}\Gamma^{-1} \sum_{j=1}^n \sigma' \partial_{\theta}(S^{-1})\sigma(X_{t_{j-1}}, \theta^*)[h^{-1}(\Delta_j w)^{\otimes 2} - I_r] + O_L(n^{-1/2}) \\
 &= -\frac{1}{2\sqrt{n}}\Gamma^{-1} \sum_{k=1}^{[n/2]} \sigma' \partial_{\theta}(S^{-1})\sigma(X_{t_{2(k-1)}}, \theta^*)[h^{-1}(\Delta_{2k-1} w)^{\otimes 2} - I_r] + \{h^{-1}(\Delta_{2k} w)^{\otimes 2} - I_r\} \\
 &\quad + O_L(n^{-1/2})
 \end{aligned}$$

For the half estimator $\hat{\theta}_n^{\div}$,

$$\begin{aligned}
 & \sqrt{[n/2]}(\hat{\theta}_n^{\div} - \theta^*) \\
 &= -\frac{1}{2\sqrt{[n/2]}}\Gamma^{-1} \sum_{k=1}^{[n/2]} \sigma' \partial_{\theta}(S^{-1})\sigma(X_{t_{2(k-1)}}, \theta^*)[(2h)^{-1}(w_{t_{2k}} - w_{t_{2(k-1)}})^{\otimes 2} - I_r] \\
 &\quad + O_L(n^{-1/2}).
 \end{aligned}$$

Noticing that the step-size for the Riemann approximation is $2h$, we have

$$\begin{aligned}
 & E[\sqrt{n}(\hat{\theta}_n - \theta^*) \otimes \sqrt{n/2}(\hat{\theta}_n^{\div} - \theta^*)] \\
 &= \frac{1}{2\sqrt{2}T}E\left[\Gamma^{-1} \int_0^T (\partial_{\theta}(S^{-1})S\partial_{\theta}(S^{-1})S)(X_t, \theta^*)dt(\Gamma^{-1})'\right] + o(1) \\
 &= \frac{1}{\sqrt{2}}E[\Gamma^{-1}] + o(1),
 \end{aligned}$$

in other words,

$$E[(\hat{\theta}_n - \theta^*) \otimes (\hat{\theta}_n^{\div} - \theta^*)] = \frac{1}{n}E[\Gamma^{-1}] + o\left(\frac{1}{n}\right),$$

As a result,

$$\begin{aligned}
 E[(\hat{\theta}_n - \theta^*) \otimes (\hat{\theta}_n^{\div} - \hat{\theta}_n)] &= E[(\hat{\theta}_n - \theta^*) \otimes (\hat{\theta}_n^{\div} - \theta^*)] - E[(\hat{\theta}_n - \theta^*) \otimes (\hat{\theta}_n - \theta^*)] \\
 &= o\left(\frac{1}{n}\right).
 \end{aligned}$$

As a matter of fact, this is the optimality of $\hat{\theta}_n$. Consequently

$$\begin{aligned}
 & E\left[\left\{\log \det \hat{S}_n + \mathfrak{m} + \partial_{\theta}(S^{-1})(x, \hat{\theta}_n)[\hat{S}_n \otimes (\hat{\theta}_n^{\div} - \hat{\theta}_n)]\right.\right. \\
 & \quad \left.\left. + \frac{1}{2n}\partial_{\theta}^2(S^{-1})(x, \hat{\theta}_n)[\hat{S}_n \otimes \Gamma_n^{-1}]\right\} - 2U(\hat{S}_n, S^*)\right] = o\left(\frac{1}{n}\right).
 \end{aligned}$$

This completes the proof. $\square$

Remark 2 (i) We select a model that minimizes sVIC_x at state x . Here we note that $\mathfrak{m}$ is removed from sVIC_x since it is the dimension of Y common for all statistical models. (ii) Minimizing the expectation of sVIC_x in x is another possibility, that is, select a model that minimizes

$$\text{sVIC} = E_{\theta^*}[\text{sVIC}(X_{T'})|\mathcal{D}_n]|_{\theta^*=\hat{\theta}_n}$$

for the data $\mathcal{D}_n$. However, we need to introduce additional conditions on the setting by considering a small ϵ model, a short time T , or others, as well as the Markovian property in practice. (iii) An ad hoc idea is to minimize the integral

$$\text{sVIC}_\mu = \int \text{sVIC}_x \mu(dx)$$

with some measure μ chosen by the user.

5 Simulation Studies

As an example of model selection with sVIC_x , the following setting is studied. The data $\mathbf{X}_n = (X_{t_j})_{0 \leq j \leq n}$ are obtained from the one-dimensional diffusion process defined by

$$(\text{true model}) \quad dX_t = -\frac{1}{2}(X_t - 10)dt + \sqrt{1 + 5(X_t)^2} dw_t, \quad t \in [0, 1], \quad X_0 = 0.$$

We consider the model selection problem for the following four statistical models.

$$(\text{model 1}) \quad dX_t = b_t^{(1)}dt + \theta_0^{(1)} dw_t, \quad t \in [0, 1], \quad X_0 = 0,$$

$$(\text{model 2}) \quad dX_t = b_t^{(2)}dt + \sqrt{\theta_0^{(2)} + \theta_1^{(2)}(X_t)^2} dw_t, \quad t \in [0, 1], \quad X_0 = 0,$$

$$(\text{model 3}) \quad dX_t = b_t^{(3)}dt + \sqrt{\theta_0^{(3)} + \theta_1^{(3)}X_t + \theta_2^{(3)}(X_t)^2} dw_t, \quad t \in [0, 1], \quad X_0 = 0,$$

$$(\text{model 4}) \quad dX_t = b_t^{(4)}dt + \sqrt{\theta_0^{(4)} + \theta_1^{(4)}X_t + \theta_2^{(4)}(X_t)^2 + \theta_3^{(4)}(\cos X_t)^2} dw_t, \quad t \in [0, 1], \quad X_0 = 0.$$

The existence of a solution to these models is not necessarily ensured for all values of parameters, however, these are only models not generating the data. When for $m = 1, 2, 3, 4$, $b_t^{(m)}$ is the same drift term as the true model, one of the possibilities for the parameter space is as follows: $\theta_0^{(1)} > 0$; $\theta_0^{(2)} > 0$, $\theta_1^{(2)} > 0$; $\theta_0^{(3)} > 0$, $\theta_2^{(3)} > 0$, $(\theta_1^{(3)})^2 < 4\theta_0^{(3)}\theta_2^{(3)}$; $\theta_0^{(4)} > 0$, $\theta_2^{(4)} > 0$, $\theta_3^{(4)} \geq 0$, $(\theta_1^{(4)})^2 < 4\theta_0^{(4)}\theta_2^{(4)}$ or $\theta_0^{(4)} > 0$, $\theta_2^{(4)} > 0$, $\theta_3^{(4)} < 0$, $(\theta_1^{(4)})^2 < 4\theta_2^{(4)}(\theta_0^{(4)} + \theta_3^{(4)})$. We may apply a criterion in [40].

We set

$$\text{sVIC}_\mu = \int_{\mathbb{R}} \text{sVIC}_x \phi(x; X_0, 1) dx.$$

Akaike type information criterion (AIC) is defined as $\text{AIC} = -2\{\mathbb{U}_n(\hat{\theta}_n) - \dim(\Theta)\}$, where $\mathbb{U}_n(\hat{\theta}_n) = \sup_{\theta \in \Theta} \mathbb{U}_n(\theta)$ and $\mathbb{U}_n(\theta) = -\frac{1}{2} \sum_{k=1}^n \{\log S(X_{t_{j-1}}, \theta) + \frac{(X_{t_j} - X_{t_{j-1}})^2}{hS(X_{t_{j-1}}, \theta)}\}$. We examine the number of models selected by sVIC_x for each state $x = 1, 2, 3, 4, 5, 10, 100, 1000, 10000$, sVIC_μ and AIC among the competing models for 1000 independent simulated sample paths through simulations for $h = 1/390$ (Table 1).

Next, we consider another example which the data are obtained by the one-dimensional diffusion model defined by

$$(\text{true model}) \quad dX_t = (1 + 2X_t)dw_t, \quad t \in [0, 1], \quad X_0 = 3.$$

Since the process $1 + 2X_t$ is an exponential martingale, we pretend that the non-generacy of the diffusion coefficient is practically satisfied for this model.

The candidates are the following four statistical models.

$$(\text{model 1}) \quad dX_t = b_t^{(1)}dt + \theta_0^{(1)}dw_t, \quad t \in [0, 1], \quad X_0 = 3,$$

$$(\text{model 2}) \quad dX_t = b_t^{(2)}dt + (\theta_0^{(2)} + \theta_1^{(2)}X_t)dw_t, \quad t \in [0, 1], \quad X_0 = 3,$$

$$(\text{model 3}) \quad dX_t = b_t^{(3)}dt + (\theta_0^{(3)} + \theta_1^{(3)}X_t + \theta_2^{(3)}(X_t)^2)dw_t, \quad t \in [0, 1], \quad X_0 = 3,$$

$$(\text{model 4}) \quad dX_t = b_t^{(4)}dt + (\theta_0^{(4)} + \theta_1^{(4)}X_t + \theta_2^{(4)}(X_t)^2 + \theta_3^{(4)}(\cos(X_t))^2)dw_t, \quad t \in [0, 1], \quad X_0 = 3.$$

In case that $b_t^{(m)} = 0$ for $m = 1, 2, 3, 4$, one of the possible parameter spaces is as follows: $\theta_0^{(1)} > 0$; $\theta_0^{(2)} > 0, \theta_1^{(2)} > 0$; $\theta_0^{(3)} > 0, \theta_1^{(3)} > 0$ if $\theta_2^{(3)} = 0$; $(\theta_1^{(3)})^2 < 4\theta_0^{(3)}\theta_2^{(3)}$ if $\theta_2^{(3)} \neq 0$; the same as the model 3 if $\theta_3^{(4)} = 0$; $\theta_1^{(4)} > 0$ if $\theta_2^{(4)} = 0$; $\theta_2^{(4)}\theta_3^{(4)} > 0, (\theta_1^{(4)})^2 < 4\theta_2^{(4)}(\theta_0^{(4)} + \theta_3^{(4)})$; $\theta_2^{(4)}\theta_3^{(4)} < 0, (\theta_1^{(4)})^2 < 4\theta_2^{(4)}\theta_0^{(4)}$.

In the same way as the previous example, we did simulation studies on the number of models selected by sVIC_x for each state $x = 1, 2, 3, 4, 5, 10, 100, 1000, 10000$, sVIC_μ and AIC among the competing models for 1000 independent simulated sample paths through simulations for $h = 1/390$ (Table 2).

Though inconsistency of the proposed criteria can be seen, it is not crucial since the consistency is different from the precision of prediction, as is often remarked. If two models give an equivalent prediction, then both models can be adopted.

Table 1 The number of selected models among four models 1–4

	$x = 1$	$x = 2$	$x = 3$	$x = 4$	$x = 5$	$x = 10$	$x = 100$	$x = 1000$	$x = 10000$	$sVIC_{\mu}$	AIC
Model 1	0	0	0	0	0	0	0	0	0	0	1000
Model 2	941	895	953	975	993	1000	1000	1000	1000	998	0
Model 3	59	77	47	25	4	0	0	0	0	2	0
Model 4	0	28	0	0	3	0	0	0	0	0	0

Table 2 The number of selected models among four models 1–4

	$x = 1$	$x = 2$	$x = 3$	$x = 4$	$x = 5$	$x = 10$	$x = 100$	$x = 1000$	$x = 10000$	sVIC _{μ}	AIC
Model 1	0	0	0	0	0	0	0	0	0	1	1000
Model 2	937	962	963	964	976	992	1000	1000	1000	994	0
Model 3	62	37	37	34	24	7	0	0	0	5	0
Model 4	1	1	0	2	0	1	0	0	0	0	0

References

1. Akaike, H.: Information theory and an extension of the likelihood ratio principle. In: Petrov, B.N., Csaki, F. (eds.) *Proceedings of the Second International Symposium of Information Theory*, pp. 267–281 (1973)
2. Akaike, H.: A new look at the statistical model identification. *IEEE Trans. Autom. Control* **19**(6), 716–723 (1974)
3. Andersen, T.G., Bollerslev, T.: Answering the skeptics: yes, standard volatility models do provide accurate forecasts. *Int. Econ. Rev.* **39**, 885–905 (1998)
4. Andersen, T.G., Bollerslev, T., Diebold, F.X., Labys, P.: The distribution of realized exchange rate volatility. *J. Amer. Statist. Assoc.* **96**(453), 42–55 (2001)
5. Anderson, T.W.: On asymptotic distributions of estimates of parameters of stochastic difference equations. *Ann. Math. Statist.* **30**, 676–687 (1959)
6. Barndorff-Nielsen, O.E., Shephard, N.: Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **64**(2), 253–280 (2002)
7. Barndorff-Nielsen, O.E., Shephard, N.: Econometric analysis of realized covariation: high frequency based covariance, regression, and correlation in financial economics. *Econometrica* **72**(3), 885–925 (2004)
8. Basawa, I.V., Koul, H.L.: Asymptotic tests of composite hypotheses for nonergodic type stochastic processes. *Stoch. Process. Appl.* **9**(3), 291–305 (1979). doi:[10.1016/0304-4149\(79\)90051-6](https://doi.org/10.1016/0304-4149(79)90051-6)
9. Basawa, I.V., Prakasa Rao, B.L.S.: Statistical inference for stochastic processes. In: *Probability and Mathematical Statistics*. Academic Press Inc. [Harcourt Brace Jovanovich Publishers], London (1980)
10. Basawa, I.V., Scott, D.J.: Asymptotic optimal inference for nonergodic models. In: *Lecture Notes in Statistics*, vol. 17. Springer, New York (1983)
11. Bibby, B.M., Sørensen, M.: Martingale estimation functions for discretely observed diffusion processes. *Bernoulli* **1**(1–2), 17–39 (1995)
12. Dacunha-Castelle, D., Florens-Zmirou, D.: Estimation of the coefficients of a diffusion from discrete observations. *Stochastics* **19**(4), 263–284 (1986)
13. Feigin, P.D.: Stable convergence of semimartingales. *Stoch. Process. Appl.* **19**(1), 125–134 (1985). doi:[10.1016/0304-4149\(85\)90044-4](https://doi.org/10.1016/0304-4149(85)90044-4)
14. Florens-Zmirou, D.: Approximate discrete-time schemes for statistics of diffusion processes. *Statistics* **20**(4), 547–557 (1989)
15. Genon-Catalot, V., Jacod, J.: On the estimation of the diffusion coefficient for multi-dimensional diffusion processes. *Ann. Inst. H. Poincaré Probab. Statist.* **29**(1), 119–151 (1993)
16. Jacod, J.: On continuous conditional Gaussian martingales and stable convergence in law. In: *Séminaire de Probabilités, XXXI, Lecture Notes in Mathematics*, vol. 1655, pp. 232–246. Springer, Berlin (1997)
17. Jakubowski, A., Mémin, J., Pages, G.: Convergence en loi des suites d'intégrales stochastiques sur l'espace 1 de skorokhod. *Probab. Theory Relat. Fields* **81**(1), 111–137 (1989)
18. Jeganathan, P.: On the asymptotic theory of estimation when the limit of the log-likelihood ratios is mixed normal. *Sankhyā Ser. A* **44**(2), 173–212 (1982)
19. Keiding, N.: Correction to: “Estimation in the birth process” (*Biometrika* 61 (1974), 71–80). *Biometrika* **61**, 647 (1974)
20. Keiding, N.: Maximum likelihood estimation in the birth-and-death process. *Ann. Statist.* **3**, 363–372 (1975)
21. Kessler, M.: Estimation of an ergodic diffusion from discrete observations. *Scand. J. Statist.* **24**(2), 211–229 (1997)
22. Konishi, S., Kitagawa, G.: Generalised information criteria in model selection. *Biometrika* **83**(4), 875–890 (1996)
23. Kurtz, T.G., Protter, P.: Weak limit theorems for stochastic integrals and stochastic differential equations. *Ann. Probab.* **19**(3), 1035–1070 (1991)

24. Kusuoka, S., Yoshida, N.: Malliavin calculus, geometric mixing, and expansion of diffusion functionals. *Probab. Theory Relat. Fields* **116**(4), 457–484 (2000)
25. Masuda, H.: Convergence of gaussian quasi-likelihood random fields for ergodic lévy driven sde observed at high frequency. *Ann. Stat.* **41**(3), 1593–1641 (2013)
26. Ogihara, T., Yoshida, N.: Quasi-likelihood analysis for the stochastic differential equation with jumps. *Stat. Infer. Stoch. Process.* **14**(3), 189–229 (2011)
27. Ogihara, T., Yoshida, N.: Quasi-likelihood analysis for stochastic regression models with non-synchronous observations (2012). arXiv preprint [arXiv:1212.4911](https://arxiv.org/abs/1212.4911)
28. Podolskij, M., Yoshida, N.: Edgeworth expansion for functionals of continuous diffusion processes (2013). arXiv preprint [arXiv:1309.2071](https://arxiv.org/abs/1309.2071)
29. Prakasa Rao, B.L.S.: Asymptotic theory for nonlinear least squares estimator for diffusion processes. *Math. Operationsforsch. Statist. Ser. Statist.* **14**(2), 195–209 (1983)
30. Prakasa Rao, B.L.S.: Statistical inference from sampled data for stochastic processes. In: *Statistical Inference from Stochastic Processes* (Ithaca, NY, 1987), Contemporary Mathematics, vol. 80, pp. 249–284. American Mathematical Society, Providence (1988)
31. Rao, M.M.: Consistency and limit distributions of estimators of parameters in explosive stochastic difference equations. *Ann. Math. Statist.* **32**, 195–218 (1961)
32. Schwarz, G., et al.: Estimating the dimension of a model. *Ann. Stat.* **6**(2), 461–464 (1978)
33. Shimizu, Y., Yoshida, N.: Estimation of parameters for diffusion processes with jumps from discrete observations. *Stat. Infer. Stoch. Process.* **9**(3), 227–277 (2006). doi: [10.1007/s11203-005-8114-x](https://doi.org/10.1007/s11203-005-8114-x)
34. Takeuchi, K.: Distribution of information statistics and criteria for adequacy of models. *Math. Sci.* **153**, 12–18 (1976). (In Japanese)
35. Uchida, M.: Contrast-based information criterion for ergodic diffusion processes from discrete observations. *Ann. Inst. Stat. Math.* **62**(1), 161–187 (2010)
36. Uchida, M., Yoshida, N.: Information criteria in model selection for mixing processes. *Stat. Infer. Stoch. Process.* **4**(1), 73–98 (2001)
37. Uchida, M., Yoshida, N.: Information criteria for small diffusions via the theory of Malliavin-Watanabe. *Stat. Infer. Stoch. Process.* **7**(1), 35–67 (2004)
38. Uchida, M., Yoshida, N.: Asymptotic expansion and information criteria. *SUT J. Math.* **42**(1), 31–58 (2006)
39. Uchida, M., Yoshida, N.: Adaptive estimation of an ergodic diffusion process based on sampled data. *Stoch. Process. Appl.* **122**(8), 2885–2924 (2012)
40. Uchida, M., Yoshida, N.: Quasi likelihood analysis of volatility and nondegeneracy of statistical random field. *Stoch. Process. Appl.* **123**(7), 2851–2876 (2013)
41. White, J.S.: The limiting distribution of the serial correlation coefficient in the explosive case. *Ann. Math. Statist.* **29**, 1188–1197 (1958)
42. Yoshida, N.: Estimation for diffusion processes from discrete observation. *J. Multivar. Anal.* **41**(2), 220–242 (1992)
43. Yoshida, N.: Malliavin calculus and asymptotic expansion for martingales. *Probab. Theory Relat. Fields* **109**(3), 301–342 (1997)
44. Yoshida, N.: Malliavin calculus and martingale expansion. *Bull. Sci. Math.* **125**(6–7), 431–456 (2001). *Rencontre Franco-Japonaise de Probabilités* (Paris, 2000)
45. Yoshida, N.: Partial mixing and conditional edgeworth expansion for diffusions with jumps. *Probab. Theory Relat. Fields* **129**, 559–624 (2004)
46. Yoshida, N.: Polynomial type large deviation inequalities and quasi-likelihood analysis for stochastic differential equations. *Ann. Inst. Stat. Math.* **63**(3), 431–479 (2011)
47. Yoshida, N.: Asymptotic expansion for the quadratic form of the diffusion process (2012). arXiv preprint [arXiv:1212.5845](https://arxiv.org/abs/1212.5845)
48. Yoshida, N.: Martingale expansion in mixed normal limit (2012). arXiv preprint [arXiv:1210.3680v3](https://arxiv.org/abs/1210.3680v3)
49. Yoshida, N.: Martingale expansion in mixed normal limit. *Stoch. Process. Appl.* **123**(3), 887–933 (2013). doi: [10.1016/j.spa.2012.10.007](https://doi.org/10.1016/j.spa.2012.10.007)
50. Yoshida, N.: Stochastic expansion of the quasi maximum likelihood estimator for volatility (2015). to be available at arXiv

A Markov Chain Estimator of Multivariate Volatility from High Frequency Data

Peter Reinhard Hansen, Guillaume Horel, Asger Lunde
and Ilya Archakov

Abstract We introduce a multivariate estimator of financial volatility that is based on the theory of Markov chains. The Markov chain framework takes advantage of the discreteness of high-frequency returns. We study the finite sample properties of the estimation in a simulation study and apply it to high-frequency commodity prices.

Keywords Markov chain · Multivariate volatility · Quadratic variation · Integrated variance · Realized variance · High frequency data

JEL Classification C10 · C22 · C80

The first and third authors acknowledge support from CREATES—Center for Research in Econometric Analysis of Time Series (DNRF78), funded by the Danish National Research Foundation.

P.R. Hansen (✉)

European University Institute and CREATES,
Villa San Paolo, Via Della Piazzuola 43, 50133 Firenze, Italy
e-mail: reinhardhansen@gmail.com

G. Horel

Serenitas Capital,
175 Varick St, New York 10014, USA
e-mail: guillaume.horel@gmail.com

A. Lunde

University of Aarhus and CREATES,
Fuglesangs All 4, 8210 Aarhus V, Denmark
e-mail: alunde@econ.au.dk

I. Archakov

European University Institute,
Villa San Paolo, Via Della Piazzuola 43, 50133 Firenze, Italy
e-mail: ilya.archakov@eui.eu

1 Introduction

This paper introduces the Markov chain estimator of multivariate volatility. Our analysis builds on the results by [19] who proposed the univariate Markov chain estimator. The multivariate extension poses new challenges related to asynchronicity and the potential need to enforce the estimator to be positive semidefinite.

The availability of high-frequency financial data has made it possible to estimate volatility over relatively short periods of time, such as an hour or a day. The main obstacle in obtaining precise estimators is the fact that high-frequency returns do not conform with conventional no-arbitrage models. The reason is that there is a great deal of autocorrelation in tick-by-tick returns. The apparent contradiction can be explained by market microstructure noise, which gives rise to the notion that the observed price is a noisy measure of the efficient price. In this paper, we introduce a multivariate volatility estimator that is built on the theory of Markov chains. The estimator utilizes the discreteness of high-frequency data, and the framework implicitly permits a high degree of serial dependence in the noise as well as dependence between the efficient price and the noise.

The use of high-frequency data for volatility estimation has been very active over the past two decades, since [3] used the realized variance to evaluate GARCH models. The realized variance is simply the sum of squared intraday returns, and its properties were detailed in [11], for the case where the semimartingale is observed without noise, which was extended to the multivariate context in [12]. The noise in high-frequency returns motivated a number of robust estimators, including the two-scale estimator by [33], the realized kernels by [8], and the pre-average estimator by [25]. Empirical features of the market microstructure noise were detailed in [22], which documented that the noise is both serially dependent and endogenous, in the sense that there is dependence between the underlying semimartingale and the noise. These empirical features motivated the development of the multivariate realized kernel in [9], which is an estimator that permits the noise to have both of these features.

An attractive feature of the Markov framework is that serially dependent and endogenous noise is a natural part of the framework. Moreover, the Markov chain estimator is simple to compute and the same is the case for the estimator of its asymptotic variance. It only takes basic matrix operations to compute the estimator and its confidence intervals.

To illustrate our estimator consider the case with two assets. The bivariate sequence of high-frequency returns is denoted by $\{\Delta X_t\}_{t=1}^n$, and we define the $S \times 2$ matrix $\mathbf{x}$, where S is the number of states for ΔX_t , and each row of $\mathbf{x}$ corresponds to a possible realization of ΔX_t . For instance, the s -th row of $\mathbf{x}$ may equal $x_{s,\cdot} = (2, -1)$ that is the state where the first asset increased by 2 units, while the second asset went down by one unit. The $S \times S$ transition matrix, P , for a Markov chain of order $k = 1$ is given by

$$P_{r,s} = \Pr(\Delta X_{t+1} = x_{s,\cdot} | \Delta X_t = x_{r,\cdot}), \quad r, s = 1, \dots, S,$$

and its stationary distribution, $\pi = (\pi_1, \dots, \pi_S)'$, is characterized by $\pi'P = \pi'$. We define $\Lambda_\pi = \text{diag}(\pi_1, \dots, \pi_S)$ and the *fundamental matrix* $Z = (I - P + \Pi)^{-1}$ where $\Pi = \iota\pi'$ with $\iota = (1, \dots, 1)' \in \mathbb{R}^S$. From the maximum likelihood estimator of P we deduced estimates of π and Z , denoted $\hat{\pi}$ and $\hat{Z}$, see Sect. 3 for details. The multivariate Markov chain estimator is given by

$$\text{MC} = nD^{-1} \left\{ \mathbf{x}'(\Lambda_{\hat{\pi}}\hat{Z} + \hat{Z}'\Lambda_{\hat{\pi}} - \hat{\pi}\hat{\pi}' - \Lambda_{\hat{\pi}})\mathbf{x} \right\} D^{-1},$$

where $D = \text{diag}(\delta_1, \delta_2)$ and $\delta_j^2 = n^{-1} \sum_{t=1}^n X_{j,t}^2$ is the sample average of the squared price of the j -th asset, $j = 1, 2$. The expression inside the curly brackets is the estimator of the long-run variance of a finite Markov chain, see [20]. The scaling involving D , is a transformation needed for the estimator to be an estimator of volatility of logarithmic prices. The scaling with the sample size, n , relates to the local-to-zero asymptotic scheme that arises under in-fill asymptotics.

Hansen and Horel [19] showed that filtering can resolve the problems caused by market microstructure noise under weak assumptions that essentially amounts to the noise process to be ergodic with finite first moment. This result is theoretical in nature, because the ideal filter requires knowledge about the data generating process. In order to turn the theoretical filtering result into an actual estimator, one needs to adopt a statistical model, and our approach is to model the increments of the process with a Markov chain model, which is a natural starting point given the discrete nature of high-frequency data.

The discreteness of financial data is a product of the so-called *tick size*, which defines the coarseness of the grid that prices are confined to. For example, the tick-size is currently 1 cent for most of the stocks that are listed on the New York Stock Exchange. The implication is that all transaction and quoted prices are in whole cents. The Markov estimator can also be applied to time series that do not live on a grid, by forcing the process onto a grid. While this will introduce rounding error, it will not affect the long-run variance of the process. Delattre and Jacod [16] studied the effect of rounding on realized variances for a standard Brownian motion, and [28] extended this analysis to log-normal diffusions.

The present paper adds to a growing literature on volatility estimation using high-frequency data, dating back to [34, 35]. Well known estimators include the realized variance, see [5, 8]; the two-scale and multi-scale estimators, see [32, 33]; the realized kernels, see [8, 9]. The finite sample properties of these estimators are analyzed in [6, 7], and the close relation between multi-scale estimators and realized kernels is established in [10]. Other estimators include those based on moving average filtered returns, see [4, 21, 30]; the range-based estimator, see [14]; the pre-averaging estimator, see [25]; the quantile-based estimator [13]; and the duration-based estimator, see [2].

The stochastic properties of market microstructure noise are very important in this context. Estimators that are robust to iid noise can be adversely affected by dependent noise. Hansen and Lunde [22] analyzed the empirical features of market microstructure noise and showed that serial dependence and endogenous noise are pronounced

in high-frequency stock prices. Endogenous noise refers to the dependence between the noise and the efficient price. A major advantage of the Markov chain estimator is that dependent and endogenous noise is permitted in the framework. In fact, dependent and endogenous noise arises naturally in this context, see [18]. Thus estimation and inference are done under a realistic set of assumptions in regard to the noise.

The present paper is an extension of [19] to the multivariate context. This extension posed new challenges that are specific to the multivariate context. For instance, different assets are typically not traded at synchronous times. This non-synchronicity leads to the so-called Epps effect, which manifests itself by a bias towards zero for the realized covariance as the sampling frequency increases. See [31] for a study of the determinants of the Epps effect. Another issue that may arise in the multivariate context is a need for the estimator to be positive semidefinite, which is not guaranteed by all multivariate estimators. The asynchronicity poses few obstacles for the Markov chain estimator, albeit a large order of the Markov chain, or another remedy, may be needed if an illiquid asset is paired with a liquid asset.

The outline of this paper is as follows. The Markov chain framework is presented in Sect. 2, and the estimator in Sect. 3. In Sect. 4 we present two composite estimators that estimate every element of the matrix separately. In Sect. 5 we propose a novel projection methods that may be needed to ensure that the composite estimators are positive semidefinite. The properties of the estimators are evaluated in Sect. 6 with a simulation study, and an empirical application to commodity prices is presented in Sect. 7.

2 The Markov Chain Framework

Let $\{X_t\}$ denote a d -dimensional process, whose returns ΔX_t can take S distinct values in $\mathbb{R}^d$. For notational convenience we take ΔX_t to be a row-vector. The possible states for the k -tuple, $\Delta \mathcal{X}_t = \{\Delta X_{t-k+1}, \dots, \Delta X_t\}$, are indexed by $s = 1, \dots, S^k$, where the s -th state corresponds to the case where $\Delta \mathcal{X}_t = \mathbf{x}_s$, which is an $1 \times kd$ vector. See the example below.

We make the following assumption about the increments of the process.

Assumption 1 The increments $\{\Delta X_t\}_{t=1}^n$ are ergodic and distributed as a homogeneous Markov chain of order $k < \infty$, with $S < \infty$ states.

The homogeneity assumption is unlikely to be valid in the context of high-frequency data. Fortunately the assumption is not critical for our results, because by increasing the order, k , of the homogeneous Markov chain that is imposed on the high-frequency returns, the resulting estimator becomes robust to inhomogeneity, see [19]. This feature of the Markov chain estimator is demonstrated in our simulation study in Sect. 6.

The transition matrix, P , is given by

$$P_{r,s} = \Pr(\Delta \mathcal{X}_{t+1} = \mathbf{x}_s | \Delta \mathcal{X}_t = \mathbf{x}_r), \quad \text{for } r, s = 1, \dots, S^k,$$

and the corresponding *stationary distribution*, $\pi \in \mathbb{R}^{S^k}$, which is unique under Assumption 1, is defined by $\pi'P = \pi'$. The *fundamental matrix* by [26] is defined by

$$Z = (I - P + \Pi)^{-1},$$

where $\Pi = \iota\pi'$ with $\iota = (1, \dots, 1)' \in \mathbb{R}^{S^k}$ so that each row of Π is simply π' .

The $S^k \times d$ matrix, f , is defined to be the last d columns of $\mathbf{x}$. So f_s is the value that (the latest observation of) ΔX_t has in state s . (Recall that a state represents a realization of k consecutive returns). Finally, we define the diagonal matrix $\Lambda_\pi = \text{diag}(\pi_1, \dots, \pi_{S^k})$.

The following example illustrates the multivariate Markov chain estimation in the case where $d = 2$ and $S = 2$, and $k = 1, 2$.

Example 1 Suppose that we have two assets and that all price changes are up or down by a single unit. If the order of the Markov chain is $k = 1$, then the transition matrix, P , is a 4×4 matrix, and we can define the state matrix as

$$\mathbf{x} = f = \begin{pmatrix} 1 & 1 \\ 1 & -1 \\ -1 & 1 \\ -1 & -1 \end{pmatrix}.$$

If, instead, the order is $k = 2$, then we have $S^2 = 16$ states, and consequently P will be an 16×16 matrix and f an 16×2 matrix. For instance, we may order the states as below, so that a row of $\mathbf{x}$ corresponds to a state value for $(\Delta X_{t-1}, \Delta X_t)$ and the corresponding row of f will have just the state value for ΔX_t :

$$\mathbf{x} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & -1 \\ 1 & 1 & -1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & -1 & -1 & -1 \\ -1 & 1 & 1 & 1 \\ -1 & 1 & 1 & -1 \\ -1 & 1 & -1 & 1 \\ -1 & 1 & -1 & -1 \\ -1 & -1 & 1 & 1 \\ -1 & -1 & 1 & -1 \\ -1 & -1 & -1 & 1 \\ -1 & -1 & -1 & -1 \end{pmatrix} \quad f = \begin{pmatrix} 1 & 1 \\ 1 & -1 \\ -1 & 1 \\ -1 & -1 \\ 1 & 1 \\ 1 & -1 \\ -1 & 1 \\ -1 & -1 \\ 1 & 1 \\ 1 & -1 \\ -1 & 1 \\ -1 & -1 \\ 1 & 1 \\ 1 & -1 \\ -1 & 1 \\ -1 & -1 \end{pmatrix}.$$

Although the transition matrix is a 16×16 matrix, it has at most four non-zero elements in each row. The reason is that many transitions are impossible. For, instance if $\Delta \mathcal{X}_t = \{(1, 1), (1, 1)\}$ then the next state will have to be $\{(1, 1), (*, *)\}$, and a transition to, $\{(-1, -1), (1, 1)\}$ say, is impossible, and thus have probability zero. So the transition matrix, P , will be increasingly sparse as k increases.

The underlying idea of the Markov chain estimator is a martingale decomposition of

$$X_t = Y_t + \mu_t + U_t,$$

where $\{Y_t, \mathcal{F}_t\}$ is a martingale with increments $\Delta Y'_t = e'_{s_t} Z f - e'_{s_{t-1}} P Z f$, $\mu_t = t\mu$ with $\mu = E(\Delta X_t)$, and U_t is a stationary ergodic bounded process.

The asymptotic scheme that will be used in the present context is the following:

$$f = n^{-1/2} \xi, \quad \text{with } \xi \in \mathbb{R}^{S^k \times d} \text{ fixed.} \quad (1)$$

This local-to-zero asymptotic scheme is similar to those used in [16, 28], and is natural under in-fill asymptotics. In the present context, it guarantees almost sure convergence of the estimator.

Under this scheme, it follows from [20] (and the ergodic theorem) that

Proposition 1 *Suppose that Assumption 1 holds, then under the asymptotic scheme (1), we have*

$$\sum_{t=1}^n \Delta Y_t \Delta Y'_t \xrightarrow{a.s.} \xi' Z' (\Lambda_\pi - P' \Lambda_\pi P) Z \xi = \xi' (\Lambda_\pi Z + Z' \Lambda_\pi - \pi \pi' - \Lambda_\pi) \xi,$$

as $n \rightarrow \infty$.

Proof By Assumption 1 it follows that $\vartheta'_t = e'_{s_t} Z \xi - e'_{s_{t-1}} P Z \xi$ is an ergodic Markov chain (of order $k+1$) with

$$E \vartheta'_t = \xi' Z' (\Lambda_\pi - P' \Lambda_\pi P) Z \xi = \xi' (\Lambda_\pi Z + Z' \Lambda_\pi - \pi \pi' - \Lambda_\pi) \xi,$$

where the first identity follows from [18, Theorem 2] and the second from [20, Lemma 1]. By the ergodic theorem it follows that $\frac{1}{n} \sum_{t=1}^n \vartheta'_t = \sum_{t=1}^n \Delta Y_t \Delta Y'_t$ converges to $\xi' (\Lambda_\pi Z + Z' \Lambda_\pi - \pi \pi' - \Lambda_\pi) \xi$ almost surely (and in mean).

An implication of Proposition 1 is that

$$\Sigma^\# = \xi' (\Lambda_\pi Z + Z' \Lambda_\pi - \pi \pi' - \Lambda_\pi) \xi,$$

is the quadratic variation of the martingale component. The same quantity is also the long-run variance of $\Delta X'_t$ in the sense that

$$\text{var}(X'_n - X'_0) = \text{var}\left(\sum_{t=1}^n \Delta X'_t\right) \rightarrow \Sigma^\#, \quad \text{as } n \rightarrow \infty.$$

There are different ways to construct a multivariate volatility estimator using Markov chain methods, and we shall present three distinct estimators and highlight each of their advantages.

3 The Markov Estimator

Let $\hat{P}$ be the maximum likelihood estimator of P and let $\hat{\pi}$ be its corresponding eigenvector, $\hat{\pi}'\hat{P} = \hat{\pi}'$. Furthermore, let $\hat{\Gamma} = \iota\hat{\pi}'$ and $\hat{Z} = (I - \hat{P} + \hat{\Gamma})^{-1}$. The expression for the long-run variance of the Markov chain motivates the estimator

$$\text{MC}^\# = nf'(\Lambda_{\hat{\pi}}\hat{Z} + \hat{Z}'\Lambda_{\hat{\pi}} - \hat{\pi}\hat{\pi}' - \Lambda_{\hat{\pi}})f,$$

for which we have the following asymptotic distribution.

Proposition 2 Suppose that Assumption 1 holds, then under the asymptotic scheme (I), we have

$$n^{1/2}(\text{MC}^\# - \Sigma^\#) \xrightarrow{d} N(0, \Omega),$$

where the asymptotic covariance between the (i, j) th and (l, m) th elements is

$$\Omega_{ij,kl} = \sum_{r,s,v} [V(r)]_{s,v} [\Xi(r,s)]_{i,j} [\Xi(r,v)]_{l,m}, \quad (2)$$

with $V(s) = \frac{1}{\pi_r}(\Lambda_{e'_r P} - P'e_r e'_r P)$ and

$$\begin{aligned} \Xi(r,s) &= \pi_r \xi' Z' (\Lambda_{z_s} - P' \Lambda_{z_s} P) Z \xi \\ &\quad + \pi_r \xi' (\pi \pi' - 2\pi z'_s - 2z_s \pi') \xi + \xi' [\Lambda_\pi Z e_r z'_s + z_s e'_r Z' \Lambda_\pi] \xi, \end{aligned}$$

and where $z'_s = e'_s Z$ is the s -th row of Z .

Proof Follows from [20, Theorem 2] by adapting their expressions (substitute ξ for f and $\xi'\pi$ for μ).

Remark 1 We note that in the univariate case, $\text{MC}^\#$ simplifies to $nf' \Lambda_{\hat{\pi}} (2\hat{Z} - \hat{\Gamma} - I)f$, which was the estimator proposed by [19].

3.1 Volatility of Logarithmic Prices

The object of interest is, typically, the volatility of log-prices, rather than $\Sigma^\#$, which is the volatility of the price process in levels. An exact estimator can be obtained with the Markov framework, by first extracting the Martingale component of X_t , however, for the univariate case [19] show that the following estimator,

$$MC = \frac{MC^\#}{\frac{1}{n} \sum_{t=1}^n X_{T_t}^2},$$

is virtually identical to the realized variance of the filtered logarithmic prices that are deduced from the estimated Markov chain. The resulting approximate estimator of the quadratic variation has several advantages, such as computational simplicity. In the present multivariate context, we adopt the following estimator:

$$MC = D^{-1} MC^\# D^{-1}, \quad (3)$$

with $D = \text{diag}(\delta_1, \dots, \delta_d)$ and $\delta_j^2 = n^{-1} \sum_{t=1}^n X_{j,t}^2$, $j = 1, \dots, d$. Our simulation in Sect. 6 shows that this approximate estimator is more accurate than other realized measures.

Alternatively one could use the estimator $MC_{i,j}^\# / \frac{1}{n} \sum_{t=1}^n X_{i,t} X_{j,t}$, for $i, j = 1, \dots, d$, but we prefer (3) because positive definiteness of $MC^\#$ is passed onto MC , and in practice $\delta_i \delta_j \simeq \frac{1}{n} \sum_{t=1}^n X_{i,t} X_{j,t}$ because the prices do not vary much in relation to their average level over the estimation window, which is typically a trading day.

4 Composite Markov Estimators

The number of possible states increases exponentially with the dimension of the process, d . Consequently, the dimension of P can become unmanageable even with moderate values of S , k , and d . For instance, with $d = 10$ assets, and price changes ranging from -4 to 4 cents, $S = 9$, and a Markov chain of order $k = 2$, the transition matrix would be $(S^d)^k \times (S^d)^k = 9^{20} \times 9^{20}$, which is impractical.

As an alternative, one can construct a composite estimator, that combines lower dimensional Markov estimator, which is in the spirit of [23, 29]. In this section we consider two such estimators. The first is constructed from univariate estimators, using a simple transformation for the estimation of covariances. The second estimator is constructed from bivariate Markov estimators, which has the advantage that standard errors of each element will be readily available. We will make use of these standard errors in the next section.

4.1 The 1-Composite Markov Estimator

In this section we introduce a composite estimator that is based on univariate Markov estimators. The identity

$$\text{cov}(X, Y) = \frac{\text{var}(X + Y) - \text{var}(X - Y)}{4}$$

motivates the estimator

$$\text{MC}_{i,j}^{\#1} = \frac{1}{4}(\text{MC}_{X_i+X_j}^{\#} - \text{MC}_{X_i-X_j}^{\#}),$$

where $\text{MC}_{X_i+X_j}^{\#}$ and $\text{MC}_{X_i-X_j}^{\#}$ are the univariate Markov chain estimator, applied to the time series $X_{i,t} + X_{j,t}$ and $X_{i,t} - X_{j,t}$, respectively. Note that the diagonal terms, $i = j$, simplifies to $\frac{1}{4}\text{MC}_{2X_i}^{\#} = \text{MC}_{X_i}^{\#}$. This approach to polarization-based estimation of the covariance is well known. In the context of high-frequency data it was first used in [24, Sect. 3.6.1] who also explored related identities. More recently it has been used in [1].

The 1-Composite estimator is mapped into estimators of the volatility of log-returns using the same diagonal matrix, D , as in (3), thus $\text{MC}^1 = D^{-1}\text{MC}^{\#1}D^{-1}$.

4.1.1 Pre-Scaling

If one seeks to estimate the covariance of two assets, whose increments are on different grid sizes, it can be advantageous to use differentiated scaling of the assets, specifically

$$\text{cov}(X, Y) = \frac{\text{var}(aX + bY) - \text{var}(aX - bY)}{4ab},$$

where a and b are constants. This can, in some cases, greatly reduce the number of states, which is computationally advantageous.

4.2 The 2-Composite Markov Estimator

In this subsection we introduce a composite estimator that uses bivariate $\text{MC}^{\#}$ estimates. For all pairs of assets we compute the correlation along with an estimate of its asymptotic variance, which will be used in the next section.

We simply estimate the bivariate Markov process $(X_{i,t}, X_{j,t})'$, and obtain the estimator of Sect. 3, $\text{MC}^{\#}$, which is a 2×2 matrix. The covariance terms we seek is the lower-left (or upper-right) element

$$\text{MC}_{i,j}^{\#_2} = \begin{cases} \text{MC}^{\#}(X_i) & \text{if } i = j, \\ \text{MC}_{1,2}^{\#}(X_i, X_j) & \text{if } i \neq j, \end{cases}$$

where $\text{MC}_{1,2}^{\#}(X_i, X_j)$ is the upper right element of the 2×2 matrix $\text{MC}^{\#}$, for the bivariate process, (X_i, X_j) . In contrast to the covariance estimated with the 1-composite estimator, the standard error of $\text{MC}_{1,2}^{\#}(X_i, X_j)$ is readily available from (2).

Analogous to the other estimators, the 2-composite estimator is mapped into estimators of the volatility of log-returns with $\text{MC}^2 = D^{-1}\text{MC}^{\#_2}D^{-1}$.

4.3 Advantages and Drawbacks of Composite Estimators

The advantages of the composite estimators are threefold.

- Computational: The state space for a univariate series is smaller than that of a multivariate.
- Dimension: Enables the construction of covariance matrices of any dimension, whereas the multivariate approach is limited to relatively low dimensions.
- No need to synchronize the observation times for each of the asset, e.g. by refresh time, see [9].

The drawbacks of the composite estimators include:

- Positive semidefinite estimate is not guaranteed
- Estimate of asymptotic variance is not readily available.

The dimension of the transition matrix (and fundamental matrix) increases rapidly with the dimension of the process d , and at some point it becomes computationally impossible to manipulate the relevant expressions that are needed for the computation of the Markov estimator. In our empirical analysis with $k = 5$, the dimension of P was about 500–1000 for $d = 1$, about 3000–5000 for $d = 2$, and about 8000–10,000 for $d = 3$. The problem with non-psd appears to be relatively rare in practice when d is small. We have only seen one case where a 5×5 estimate was non-psd estimate. The occurrence is more common in higher dimensions. Of the 251 14×14 estimators we obtained for 2013, 14 of them were non-psd.

5 Enforcing Positivity

While $\text{MC}^{\#}$ is a quadratic form that yields a positive semidefinite estimator, there is no reason to expect that the composite estimators, $\text{MC}^{\#_1}$ and $\text{MC}^{\#_2}$, will be positive semidefinite (PSD) in finite samples. This problem is often encountered in estimation of high-dimensional variance-covariance matrices.

One can project a non-PSD estimate, by solving the following semi-definite program for the variable Σ

$$\min_{\Sigma} \|\Sigma - A\|_{\text{Fro}} \quad \text{subject to} \quad \Sigma \succeq 0. \quad (4)$$

The solution can be found efficiently by computing the spectral decomposition of the matrix A , and drop all negative eigenvalues, i.e. map the symmetric matrix, $A = Q \text{diag}(\lambda_1, \dots, \lambda_d) Q'$ into $Q \text{diag}(\lambda_1^+, \dots, \lambda_d^+) Q'$, where $\lambda_1, \dots, \lambda_d$ are the eigenvalues of A and $x^+ = \max(x, 0)$. Such an estimator will, due to the zero eigenvalues, be on the boundary of the space of psd matrices, which motivated [27] to impose an additional constraint, $\text{diag}(\Sigma) = \text{diag}(A)$.

In this paper we propose a novel projection that takes advantage of standard errors of the individual elements of the matrix A when these are available. Thus let ω_{ij} be (an estimate of) the standard errors of A_{ij} . Then we solve the following program

$$\min_{\Sigma} \sum_{i,j=1}^d \left(\frac{\Sigma_{ij} - A_{ij}}{\omega_{ij}} \right)^2 \quad \text{subject to} \quad \Sigma \succeq 0. \quad (5)$$

The solution can be obtained using semidefinite programming solvers that are readily available, including the `cvx` software for Matlab by [17]. The optimization problem can be supplemented with the constraint $\text{diag}(\Sigma) = \text{diag}(A)$, which would produce a constrained estimate with strictly positive eigenvalues, except in pathological cases, e.g. if A is psd with zero eigenvalues to begin with.

The projection in (5) is appealing because it attempts to influence accurately measured elements of A less than those that are relatively inaccurate. An even more appealing projection along these lines would also account for correlations across elements. In the present context, such cross correlations are only available for the estimator $\text{MC}^\#$. However, since this estimator, $\text{MC}^\#$, is psd per construction, there is no need for a projection of this estimator.

6 Simulation

In this section we compare the 1-composite Markov estimator against some benchmark. Diagonal elements are compared with the realized variance (RV) and the realized kernel (RK). Off-diagonal elements are compared with the realized covariance (RC).

6.1 Efficient Price

Our simulations are based on two designs for the latent price process, Y_t . In the first design, Y_t is simply sampled from a Brownian motion with constant volatility. In the second design, Y_t is drawn from a stochastic volatility model, which is known as the Dothan model in the literature on interest rates, similar to that used in [8]. Specifically we simulate

$$\log Y_{i,t} = \log Y_{i,t} + \sigma_{i,t} V_{i,t}, \quad i = 1, 2,$$

where $V_{i,t} = \gamma Z_{i,t} + \sqrt{1 - \gamma^2} W_{i,t}$ with $(Z_{1,t}, Z_{2,t}, W_{1,t}, W_{2,t})$ being iid Gaussian, all having unit variance and zero correlation, with the exception that $\text{cov}(W_{1,t}, W_{2,t}) = \rho$.

In the design with stochastic volatility, the volatility, $\sigma_{i,t}$, correlates with $Z_{i,t}$, so that γ controls the leverage effect of the volatility on the stock prices. Specifically,

$$\sigma_{i,t} = \sqrt{\Delta} \left\{ \exp(\beta_0 + \beta_1 \tau_{i,t}) \right\},$$

where $\tau_{i,t} = \exp(\alpha \Delta) \tau_{i,t-1} + \sqrt{\frac{\exp(2\alpha \Delta) - 1}{2\alpha}} Z_{i,t}$, with $\tau_{i,1}$ drawn from its unconditional distribution, and $\Delta = \frac{1}{N}$ with $N = 23,400$. Additional details about the specification is given in the Appendix.

The values of the parameters in both designs are summarized in Table 1.

6.2 Noise

We will use two specifications for the noise. The first is pure rounding noise, so that

$$X_t = \delta[Y_t/\delta],$$

where $[a]$ is the rounding of a to the integers so that the parameter δ controls the coarseness of the rounding error.

The second specification has an additive noise component in addition to the rounding error, specifically

$$X_t = \delta[(Y_t + U_t)/\delta],$$

Table 1 Parameters values for simulating the efficient price process, Y_t

	β_0	β_1	α	ρ	γ
Constant volatility	0	0	–	–0.3	0
Stochastic volatility	–0.3125	0.125	–0.025	–0.3	0.5

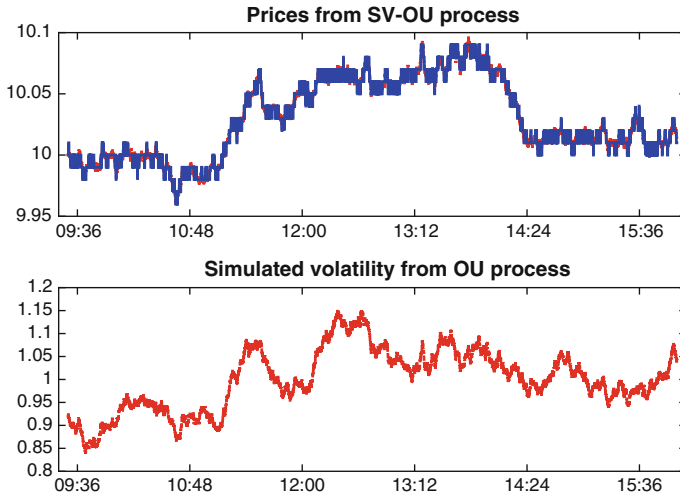


Fig. 1 A typical sample path of the simulated stochastic volatility process. The *upper panel* displays the price process, Y_t , and the observed process, X_t , that is subject to noise and rounding error. The *lower panel* displays the corresponding volatility process, $\sigma^2(t)$

where U_t are iid and uniformly distributed. The idea is that it would more closely resemble the bid ask bounce (due to the additional jitter introduced by U_t , we will either round up or down).

In our simulation study we use $\delta = 0.01$ to emulate rounding errors to a grid, and the noise is $U_t \sim \text{iid}U[-\frac{1}{3}, \frac{1}{3}]$ which adds additional (mean-reverting) jitter to the returns.

In Fig. 1 we show an example of a realization of the process with stochastic volatility using the design in Table 1. The upper panel has Y_t and X_t , where the latter is clearly identified by it being confined to the grid values. The lower panel displays the corresponding volatility process, specifically we plot $\sigma^2(t) = \sigma_{i,t}^2/\Delta$.

6.3 Estimators and Tuning Parameters

We consider the realized variance computed with different sampling frequencies. To imitate a 24 h period with second-by-second price observations, we generate 23,400 noisy high-frequency returns in each simulation.

The realized variance (RV) and the realized covariance (RC) is computed using different sampling frequencies. The choice of sampling frequency entails a bias-variance trade-off, because the bias arising from the noise is most pronounced at high sampling frequencies, while the variance of the estimator increases as the sampling frequency is lowered. Thus for the RV and the RC we sample every H -th price observation where $H \in \{1, 3, 5, 10, 15, 30, 60, 120, 240\}$.

The multivariate realized kernel (MRK) follows the implementation in [9], which is based on the Parzen kernel, and an automatic selection of the bandwidth parameter. This estimator is also applied to high-frequency returns based on the various sample frequencies. The MRK should, in principle, be most accurate when based on returns sampled at the highest frequency, $H = 1$.

The tuning parameter for the Markov chain estimator is the order of the Markov chain, k , and we apply this estimator for $k \in \{1, \dots, 5\}$.

6.4 Simulation Results

We report bias and the root mean squared error (RMSE) for each of the estimators using the various choices for their respective tuning parameters. The results are based on 10,000 simulations. The results are presented in Tables 2 and 3 for the case with constant volatility and stochastic volatility, respectively.

Consider first the case with constant volatility in Table 2. With pure rounding error we note that the Markov chain estimator tend to outperform both the kernel estimator and the realized variance in terms of the mean squared error. Similarly for the covariance, the MC 1-composite estimator dominates the RC and performs on par with the MRK. The Markov estimator is somewhat insensitive to the choice of k , so even with a non-optimal choice for k , the Markov estimator is fairly accurate. The MRK is similarly insensitive to the choice for H . In contrast, the RV and the RC are very sensitive to the choice of H , and suffer from large biases when H is small.

Turning to the case with both additive noise and rounding error. This design generates increments with rather different features. While $k = 1$ was the optimal choice with rounding error, the best configuration is now $k = 3$ or $k = 4$. The RV performs even worse in this design, the RC just as bad as in the previous design, whereas MRK performs as well as in the previous design, and is on par with the Markov estimator. This comparison is again made with hindsight as assume that relatively good choices of tuning parameters, for k and H , respectively, are used. For the covariance, we observe that the RMSE of the Markov estimator is predominantly driven by a bias.

Next we turn to the result in Table 3 which is for the case with stochastic volatility. The Markov chain estimator is based on fitting a homogeneous Markov chain to the observed increment. For this reason it might be expected that the Markov estimator is not well suited for the design with time varying volatility, see Fig. 1. However, even in the case with stochastic volatility that induced an inhomogeneous model for the increments, we see that MC performs well. The RMSEs are, as expected, a bit larger. Interestingly, it is the design with pure rounding errors that results in the largest RMSEs. Both the Markov estimator and the MRK appear to benefit from the additional layer of noise that is added prior to the rounding error.

Table 2 Simulation results for the case with constant volatility

<i>Panel A: Constant volatility and pure rounding error</i>								
	Variance				Covariance			
	MC				MC			
k	Bias	Rmse			Bias	Rmse		
1	0.002	0.109			-0.008	0.075		
2	0.004	0.144			-0.053	0.093		
3	0.007	0.174			-0.030	0.098		
4	0.005	0.202			-0.020	0.107		
5	0.003	0.221			-0.010	0.120		
H	RV		MRK		RC		MRK	
	Bias	Rmse	Bias	Rmse	Bias	Rmse	Bias	Rmse
1	3.752	3.763	0.140	0.182	-0.450	0.451	-0.008	0.074
3	3.445	3.455	0.058	0.136	-0.373	0.377	-0.007	0.088
5	3.136	3.145	0.037	0.136	-0.313	0.320	-0.007	0.097
10	2.486	2.494	0.017	0.146	-0.215	0.229	-0.008	0.110
15	2.014	2.021	0.010	0.155	-0.156	0.178	-0.009	0.119
30	1.222	1.229	0.002	0.175	-0.082	0.120	-0.011	0.135
60	0.646	0.657	-0.005	0.200	-0.042	0.100	-0.014	0.154
120	0.324	0.350	-0.015	0.234	-0.020	0.104	-0.018	0.182
240	0.161	0.231	-0.024	0.283	-0.010	0.128	-0.019	0.224
<i>Panel B: Stochastic volatility with noise and rounding error</i>								
	Variance				Covariance			
	MC				MC			
k	Bias	Rmse			Bias	Rmse		
1	0.281	0.302			-0.066	0.115		
2	0.140	0.173			-0.137	0.151		
3	0.056	0.131			-0.078	0.104		
4	0.025	0.130			-0.076	0.103		
5	0.011	0.141			-0.049	0.092		
H	RV		MRK		RC		MRK	
	Bias	Rmse	Bias	Rmse	Bias	Rmse	Bias	Rmse
1	9.697	9.708	0.112	0.159	-0.451	0.456	-0.006	0.075
3	8.086	8.095	0.044	0.131	-0.374	0.389	-0.005	0.090
5	6.829	6.838	0.027	0.135	-0.314	0.338	-0.006	0.099
10	4.720	4.727	0.013	0.149	-0.213	0.252	-0.007	0.113
15	3.488	3.495	0.009	0.160	-0.158	0.204	-0.008	0.122
30	1.859	1.866	0.005	0.181	-0.081	0.140	-0.011	0.138
60	0.938	0.949	0.002	0.208	-0.039	0.114	-0.015	0.160
120	0.467	0.491	-0.007	0.243	-0.020	0.115	-0.018	0.188
240	0.233	0.293	-0.016	0.291	-0.009	0.137	-0.021	0.230

^aPanel A has simulation results for the case where the underlying volatility is constant and the observed process is only subject to rounding error. Panel B presents the corresponding results for the case with both noise and rounding error

Table 3 Simulation results for the case with stochastic volatility

<i>Panel A: Stochastic volatility and pure rounding error</i>								
	Variance				Covariance			
	MC				MC			
<i>k</i>	Bias	Rmse			Bias	Rmse		
1	0.005	0.125			-0.020	0.091		
2	0.003	0.165			-0.024	0.084		
3	0.003	0.214			-0.019	0.089		
4	0.002	0.260			-0.010	0.098		
5	-0.001	0.327			-0.007	0.108		
	RV		MRK		RC		MRK	
<i>H</i>	Bias	Rmse	Bias	Rmse	Bias	Rmse	Bias	Rmse
1	3.070	3.325	0.132	0.196	-0.325	0.443	-0.007	0.072
3	2.812	3.032	0.058	0.168	-0.270	0.370	-0.006	0.086
5	2.555	2.741	0.037	0.180	-0.227	0.313	-0.007	0.094
10	2.017	2.141	0.018	0.198	-0.155	0.221	-0.008	0.105
15	1.636	1.721	0.011	0.222	-0.114	0.173	-0.009	0.113
30	1.016	1.056	0.000	0.249	-0.059	0.112	-0.011	0.129
60	0.571	0.622	-0.010	0.295	-0.031	0.098	-0.014	0.151
120	0.304	0.371	-0.022	0.374	-0.016	0.103	-0.016	0.181
240	0.157	0.314	-0.031	0.498	-0.009	0.127	-0.016	0.223
<i>Panel B: Stochastic volatility with noise and rounding error</i>								
	Variance				Covariance			
	MC				MC			
<i>k</i>	Bias	Rmse			Bias	Rmse		
1	0.219	0.272			-0.055	0.143		
2	0.117	0.185			-0.093	0.131		
3	0.055	0.175			-0.053	0.097		
4	0.034	0.193			-0.050	0.094		
5	0.021	0.212			-0.034	0.090		
	RV		MRK		RC		MRK	
<i>H</i>	Bias	Rmse	Bias	Rmse	Bias	Rmse	Bias	Rmse
1	9.619	9.639	0.103	0.202	-0.333	0.461	-0.007	0.073
3	8.010	8.028	0.042	0.170	-0.275	0.392	-0.005	0.087
5	6.757	6.773	0.026	0.177	-0.230	0.340	-0.006	0.095
10	4.656	4.668	0.013	0.197	-0.160	0.256	-0.007	0.108
15	3.432	3.443	0.008	0.214	-0.116	0.205	-0.007	0.117
30	1.825	1.836	0.002	0.253	-0.062	0.142	-0.009	0.134
60	0.924	0.942	-0.004	0.311	-0.032	0.116	-0.011	0.157
120	0.462	0.508	-0.013	0.389	-0.017	0.113	-0.014	0.185
240	0.232	0.357	-0.023	0.488	-0.009	0.132	-0.016	0.226

^aPanel A has simulation results for the case where the underlying volatility is stochastic and the observed process is only subject to rounding error. Panel B presents the corresponding results for the case with both noise and rounding error

7 Empirical Analysis

7.1 Data Description

We apply the Markov chain estimator to high-frequency commodity prices, that have previously been used in [15]. We confine our empirical analysis to 2013 data and consider in our study high frequency data for 14 assets. The 14 assets include the exchange traded fund, SPY, that tracks the S&P 500 index, and 13 commodity futures. We refer to [15] for detailed information about the data, including the procedures used for cleaning the high frequency data for outliers and other anomalies. Summary statistics for the 14 assets are presented in (Table 4).

Of the 15 commodities analyzed in [15], we drop two of these series for computational reasons. Specifically, we dropped “Heating Oil” (HO) because it has an unusually large number of distinct second-to-second price increments and “Feeder Cattle” (FC) because it is substantially less liquid compared with the other commodities. Thus, in addition to the SPY, we use the following 13 commodities in our empirical analysis: Crude Light (CL), Natural Gas (NG), Gold (GC), Silver (SV), Copper (HG), Live Cattle (LC), Lean Hogs (LH), Coffee (KC), Sugar (SB), Cotton (CT), Corn (CN), Soybeans (SY) and Wheat (WC).

We exclusively apply the Markov estimators to high-frequency data from the time interval 10:00–14:00 eastern standard time, because all assets are actively traded in this period. The high-frequency prices for eight selected assets for March 18th, 2013 are displayed in Fig. 2.

7.2 Empirical Results

First we present detailed results for March 18th, 2013 (to celebrate the occasion for writing this paper). Daily estimates (for the 10:00–14:00 interval) for all trading days in 2013 will be presented in figures.

7.2.1 Daily Estimates for March 18, 2013

In Table 5 we present five estimators of the volatility matrix for five assets. There are relatively large discrepancies between the two realized variances, which may be due to sampling error or market microstructure noise. The Markov estimators are largely in agreement about the correlations, but the full estimator yields a smaller estimate of the diagonal elements in some cases. This may be caused by the estimator being somewhat unreliable, as it is based on $n = 8,700$ observations and the underlying transition matrix is an $8,600 \times 8,600$ matrix in this case. Further research is needed to characterize the limitations of the full estimator in practice.

Table 4 Summary statistics for the 251 trading days in 2013

Data summary statistics														
	SPY	CL	NG	GC	SV	HG	LC	LH	KC	SB	CT	CN	SY	WC
Transactions per day (full trading day)	19178	21993	9761	21671	9607	9738	2835	2775	2922	4013	3325	7058	8882	4684
Transactions per day: n (10:00AM–2:00PM)	10117	8634	4495	6320	3024	2706	2315	2213	1613	2125	1940	4362	5011	2895
Primitive states:(10:00AM–2:00PM)	11	19	11	17	11	8	8	8	11	7	23	11	13	9
Tick size	0.01	0.01	0.001	0.1	0.5	0.05	0.025	0.025	0.05	0.01	0.01	0.25	0.25	0.25
Volatility share: κ (10AM–2 PM)/24 h	0.31	0.36	0.40	0.24	0.20	0.20	0.47	0.25	0.44	0.37	0.46	0.24	0.25	0.52
Annual volatility (2013)	11 %	19 %	31 %	20 %	32 %	19 %	12 %	24 %	27 %	18 %	21 %	33 %	25 %	22 %

The average number of price observations per day, and within the 10:00 AM to 2:00 PM window, n , are reported in the first two rows, followed by the average number of primitive states, S_i , and the tick size for each of the assets. The second last row reports κ —the fraction of daily volatility that occurs during the 10:00 AM to 2:00 PM window on average. The average volatility for 2013 is reported in the last row

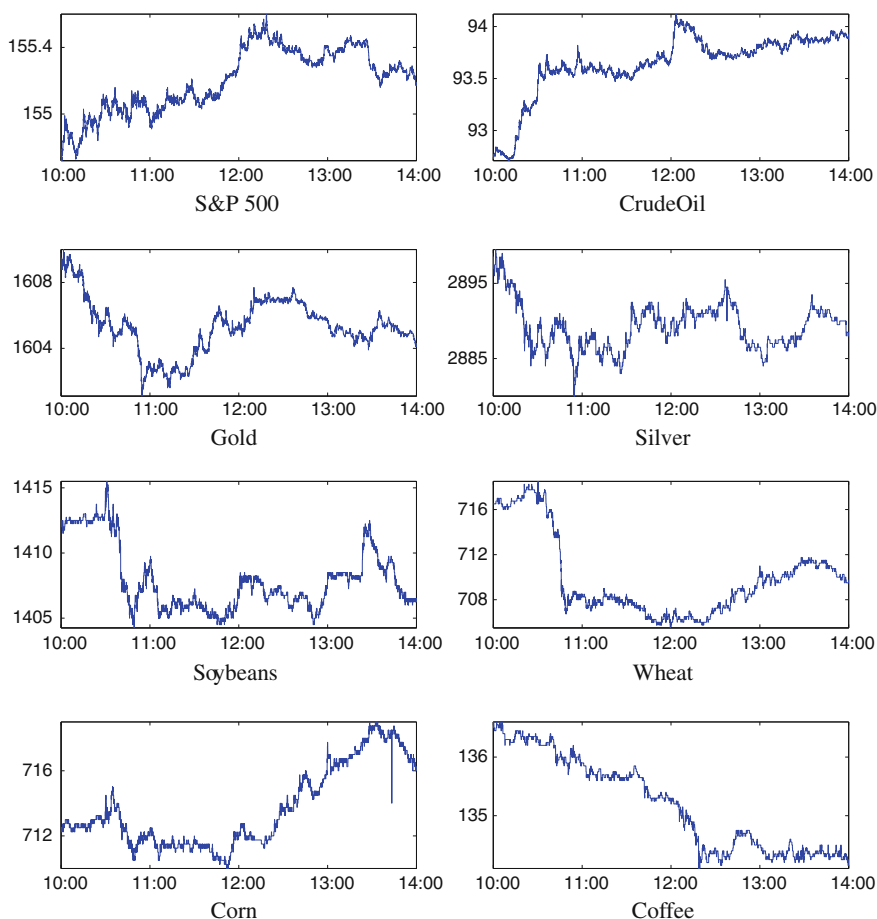


Fig. 2 High frequency prices for eight selected commodities on March 18th, 2013 during the period from 10:00 AM to 2:00 PM

In Tables 6 and 7 we present estimates of the full 14×14 matrix. The realized variances are in Table 6 and the two composite Markov estimators are in Table 7. Joint estimation of the full 14×14 covariance matrix is not expected to be precise because the number of observed states (and the dimensionality of the transition matrix) is equal to the number of observations in that case. As in the previous example, the two realized variance estimators produce quite different values whereas the composite Markov estimators produce rather similar results. In general, signs and magnitudes of the elements of Markov covariance matrices are largely in agreement with those of the realized variances.

Table 5 Two realized variances and the three variants of the Markov chain estimator are presented

<i>RV_{5min}: Realized variance with 5-min sampling</i>					
	SPY	CL	GC	SV	KC
S&P 500	81.16	79.39	−8.26	−20.03	−15.24
Light crude	<i>0.48</i>	344.08	−36.52	−1.39	−2.29
Gold	−0.08	−0.16	146.11	164.45	−47.74
Silver	−0.13	−0.00	<i>0.77</i>	316.22	−36.77
Coffee “C”	−0.08	−0.01	−0.20	−0.10	407.69
<i>RV_{10min}: Realized variance with 10-min sampling</i>					
	SPY	CL	GC	SV	KC
S&P 500	68.92	90.57	1.98	−20.04	−40.79
Light crude	<i>0.59</i>	342.05	−69.66	−57.07	−11.96
Gold	<i>0.03</i>	−0.43	78.43	50.73	−35.05
Silver	−0.20	−0.26	<i>0.48</i>	139.93	−17.48
Coffee “C”	−0.23	−0.03	−0.19	−0.07	446.70
<i>MC: Markov chain estimator (full)</i>					
	SPY	CL	GC	SV	KC
S&P 500	80.65	88.17	−24.05	−77.56	−1.89
Light crude	<i>0.49</i>	407.24	−75.01	−124.63	49.36
Gold	−0.26	−0.35	109.99	155.36	−32.78
Silver	−0.45	−0.32	<i>0.77</i>	370.69	−12.07
Coffee “C”	−0.01	<i>0.15</i>	−0.19	−0.04	257.15
<i>MC¹: Markov chain estimator 1-composite</i>					
	SPY	CL	GC	SV	KC
S&P 500	116.84	75.77	−10.69	4.58	21.01
Light crude	<i>0.35</i>	391.86	14.34	61.58	41.60
Gold	−0.09	<i>0.07</i>	116.87	151.25	−16.83
Silver	<i>0.02</i>	<i>0.16</i>	<i>0.72</i>	380.17	−0.71
Coffee “C”	<i>0.09</i>	<i>0.10</i>	−0.08	−0.00	421.94
<i>MC²: Markov chain 2-composite</i>					
	SPY	CL	GC	SV	KC
S&P 500	116.84	80.35	−12.92	−7.44	13.76
Light Crude	<i>0.38</i>	391.86	−1.45	39.30	29.37
Gold	−0.11	−0.01	116.87	149.72	−27.00
Silver	−0.04	<i>0.10</i>	<i>0.71</i>	380.17	−59.90
Coffee “C”	<i>0.06</i>	<i>0.07</i>	−0.12	−0.15	421.94

The estimators are for the 10:00 AM–2:00 PM period on March 18, 2013 for five selected assets. Variances and covariances are annualized and further scaled by 10^4 . Correlations are in the lower triangle in italic font

Table 6 The table presents two realized variances for the 14 assets

RV_{5min}	SPY	CL	NG	GC	SV	HG	LC	LH	KC	SB	CT	CN	SY	WC
S&P 500	81.16	79.39	18.76	-8.26	-20.03	44.26	-23.18	-1.03	-15.24	23.70	46.58	12.24	16.52	-21.65
Light crude	0.48	344.08	80.92	-36.52	-1.39	73.65	-7.29	-69.62	-2.29	-20.68	11.62	5.80	32.45	23.80
Natural gas	0.07	0.15	805.41	-4.63	60.09	159.46	-94.66	-21.79	53.54	-50.70	-54.01	-85.28	22.49	37.54
Gold	-0.08	-0.16	-0.01	146.11	164.45	6.66	-8.35	-4.19	-47.74	-39.13	11.51	-43.91	-21.55	-32.41
Silver	-0.13	-0.00	0.12	0.77	316.22	59.49	-21.80	9.80	-36.77	-51.15	16.38	-61.21	-46.76	-22.01
Copper	0.30	0.24	0.34	0.03	0.20	269.51	-70.43	13.29	89.35	13.11	24.30	-53.07	68.05	8.94
Live cattle	-0.19	-0.03	-0.25	-0.05	-0.09	-0.32	180.30	18.30	-23.07	-16.29	-26.96	70.81	-6.47	80.24
Lean hogs	-0.00	-0.15	-0.03	-0.01	0.02	0.03	0.05	618.27	-54.91	-7.75	9.10	-44.27	-88.09	-1.78
Coffee "C"	-0.08	-0.01	0.09	-0.20	-0.10	0.27	-0.09	-0.11	407.69	34.62	12.53	123.43	62.26	180.13
Sugar #1	0.16	-0.07	-0.11	-0.20	-0.17	0.05	-0.07	-0.02	0.10	274.87	65.41	-29.63	28.12	46.86
Cotton #2	0.24	0.03	-0.09	0.04	0.04	0.07	-0.09	0.02	0.03	0.18	478.52	87.19	102.05	53.99
Corn	0.05	0.01	-0.12	-0.14	-0.14	-0.13	0.21	-0.07	0.24	-0.07	0.16	640.54	288.11	412.80
Soybeans	0.09	0.09	0.04	-0.09	-0.13	0.20	-0.02	-0.17	0.15	0.08	0.23	0.56	415.85	232.42
Wheat	-0.10	0.05	0.06	-0.11	-0.05	0.02	0.25	-0.00	0.38	0.12	0.10	0.69	0.48	553.00

(continued)

Table 6 (continued)

RV _{10min}	SPY	CL	NG	GC	SV	HG	LC	LH	KC	SB	CT	CN	SY	WC
S&P 500	68.92	90.57	-7.81	1.98	-20.04	26.06	3.25	24.47	-40.79	-5.38	35.40	-3.91	-24.18	-32.89
Light crude	0.59	342.05	98.06	-69.66	-57.07	90.71	9.94	21.31	-11.96	-29.12	88.15	-8.19	20.25	18.49
Natural gas	-0.04	0.22	578.93	45.71	101.82	160.88	-78.18	-104.34	104.28	-42.88	47.80	-36.26	66.09	51.45
Gold	0.03	-0.43	0.21	78.43	50.73	9.61	-17.97	-8.42	-35.05	-2.83	-28.70	-38.85	-18.87	-32.80
Silver	-0.20	-0.26	0.36	0.48	139.93	54.77	-23.71	-15.59	-17.48	-13.00	-33.72	-39.72	-6.86	-23.42
Copper	0.18	0.28	0.38	0.06	0.26	317.27	-102.54	114.21	123.27	35.28	-37.42	-78.76	48.24	49.66
Live cattle	0.03	0.04	-0.25	-0.16	-0.16	-0.45	164.68	30.43	-25.38	2.54	88.65	151.23	65.74	148.58
Lean hogs	0.14	0.06	-0.21	-0.05	-0.06	0.31	0.12	418.34	-94.24	41.35	31.23	-113.20	-66.97	4.08
Coffee "C"	-0.23	-0.03	0.21	-0.19	-0.07	0.33	-0.09	-0.22	446.70	60.91	-62.30	108.18	100.85	279.64
Sugar #1	-0.04	0.10	0.11	0.02	0.07	0.13	0.01	0.13	0.19	242.43	145.50	26.27	-11.14	108.60
Cotton #2	0.19	0.21	0.09	-0.14	-0.13	-0.09	0.31	0.07	-0.13	0.42	503.18	54.13	73.48	147.24
Corn	-0.02	-0.02	-0.06	-0.17	-0.13	-0.17	0.45	-0.21	0.19	-0.06	0.09	699.55	403.33	542.96
Soybeans	-0.14	0.05	0.13	-0.10	-0.03	0.13	0.24	-0.15	0.22	-0.03	0.15	0.71	463.78	452.48
Wheat	-0.14	0.03	0.07	-0.13	-0.07	0.10	0.39	0.01	0.45	0.24	0.22	0.70	0.72	860.64

Variances and covariances are annualized and further scaled by 10⁴. Correlations are in the lower triangle in italic font

Table 7 The table presents two Markov estimates of the variance-covariance for the 14 assets

MC ¹	SPY	CL	NG	GC	SV	HG	LC	LH	KC	SB	CT	CN	SY	WC
S&P 500	116.84	75.77	-34.48	-10.69	4.58	73.32	2.39	-5.64	21.01	-6.10	2.00	-4.96	2.38	1.81
Light crude	0.35	391.86	-23.61	14.34	61.58	101.52	11.42	-31.64	41.60	5.07	11.34	14.14	5.57	3.12
Natural gas	-0.11	-0.04	921.74	-1.88	-14.76	30.77	-26.48	8.46	29.60	-10.02	-61.60	52.74	-6.90	-1.04
Gold	-0.09	0.07	-0.01	116.87	151.25	9.18	-4.74	7.84	-16.83	10.41	-36.42	28.66	5.55	19.28
Silver	0.02	0.16	-0.02	0.72	380.17	31.04	30.64	30.20	-0.71	-3.73	85.57	-1.61	5.51	-17.91
Copper	0.35	0.27	0.05	0.04	0.08	365.15	-8.35	-21.68	-4.71	-36.98	13.93	-5.18	-23.51	-7.07
Live cattle	0.02	0.05	-0.07	-0.04	0.13	-0.04	142.36	40.37	4.44	8.11	-1.72	13.95	-18.30	27.09
Lean hogs	-0.02	-0.06	0.01	0.03	0.06	-0.05	0.14	606.71	4.55	-2.74	-22.64	-0.43	6.19	-19.86
Coffee "C"	0.09	0.10	0.05	-0.08	-0.00	-0.01	0.02	0.01	421.94	10.27	91.99	-1.41	39.64	-39.64
Sugar #1	-0.03	0.01	-0.02	0.05	-0.01	-0.10	0.04	-0.01	0.03	345.21	-37.34	3.76	-6.25	32.86
Cotton #2	0.01	0.02	-0.09	-0.14	0.19	0.03	-0.01	-0.04	0.19	-0.09	544.65	134.15	68.07	73.68
Corn	-0.02	0.03	0.08	0.12	-0.00	-0.01	0.05	-0.00	-0.00	0.01	0.26	475.69	204.56	276.23
Soybeans	0.01	0.01	-0.01	0.03	0.01	-0.06	-0.08	0.01	0.10	-0.02	0.15	0.49	365.44	132.19
Wheat	0.01	0.01	-0.00	0.09	-0.05	-0.02	0.11	-0.04	-0.10	0.09	0.16	0.63	0.34	407.53

(continued)

Table 7 (continued)

MC ²	SPY	CL	NG	GC	SV	HG	LC	LH	KC	SB	CT	CN	SY	WC
S&P 500	116.84	80.35	-12.55	-12.92	-7.44	76.88	8.40	-5.69	13.76	1.67	42.68	-9.93	10.89	0.47
Light crude	0.38	391.86	7.67	-1.45	39.30	125.36	6.38	-11.56	29.37	-7.56	89.63	-25.61	22.06	-3.37
Natural gas	-0.04	0.01	921.74	1.64	-44.28	84.25	-32.41	-10.85	74.22	-53.72	-56.59	34.81	-12.94	-22.93
Gold	-0.11	-0.01	0.00	116.87	149.72	15.23	-2.08	25.73	-27.00	-4.72	-37.49	0.15	5.05	-3.15
Silver	-0.04	0.10	-0.07	0.71	380.17	23.70	4.60	16.03	-59.90	23.58	-60.49	-36.39	-1.79	-38.08
Copper	0.37	0.33	0.15	0.07	0.06	365.15	-29.91	17.98	24.53	0.16	80.92	-7.86	-34.35	-16.89
Live cattle	0.07	0.03	-0.09	-0.02	0.02	-0.13	142.36	18.71	-20.08	4.14	-15.27	9.25	-19.75	31.61
Lean hogs	-0.02	-0.02	-0.01	0.10	0.03	0.04	0.06	606.71	69.63	7.01	426.46	-27.19	19.11	35.20
Coffee "C"	0.06	0.07	0.12	-0.12	-0.15	0.06	-0.08	0.14	421.94	23.83	5.64	-15.59	54.31	-39.85
Sugar #1	0.01	-0.02	-0.10	-0.02	0.07	0.00	0.02	0.02	0.06	345.21	95.41	11.69	-28.41	29.00
Cotton #2	0.17	0.19	-0.08	-0.15	-0.13	0.18	-0.05	0.74	0.01	0.22	544.65	83.64	19.19	68.24
Corn	-0.04	-0.06	0.05	0.00	-0.09	-0.02	0.04	-0.05	-0.03	0.03	0.16	475.69	236.91	320.47
Soybeans	0.05	0.06	-0.02	0.02	-0.00	-0.09	-0.09	0.04	0.14	-0.08	0.04	0.57	365.44	162.80
Wheat	0.00	-0.01	-0.04	-0.01	-0.10	-0.04	0.13	0.07	-0.10	0.08	0.14	0.73	0.42	407.53

Variances and covariances are annualized and further scaled by 10⁴. Correlations are in the lower triangle in italic font

7.2.2 Estimates for Pooled March Data

Finally we have pooled the high-frequency data for all of March and estimated the 14×14 matrix that reflects the volatility in March, 2013 that occurred during the 10:00AM–2:00PM trading periods. With 20 trading days in March, 2013, this adds up to 80h of high-frequency data. Precision is expected to improve with the larger sample size, although the dimensions of the underlying transition matrices are expected to increase as a larger number of states and transitions will be observed in a larger sample, and the latter can potentially cause computational difficulties. For the 2-composite estimator with $k = 4$ we observed between 15000 and 30000 distinct states in the pooled data set. Another challenges for the Markov estimator in the pooled sample is that a larger degree of inhomogeneity may be expected. Hansen and Horel [19] showed that an inhomogeneous Markov process can be approximated by a homogeneous Markov process, by increasing the order of the Markov chain. So a larger k may be needed in the pooled data, which also poses computational challenges.

In Tables 8 and 9 we report estimates for the 14×14 covariance matrix computed with the realized variances and the two composite Markov estimators. In contrast to the data for March 18th, 2013, the realized variances are largely in agreement for the pooled data. Albeit some differences are observed between the 5- and 10-min realized variances. The composite Markov estimators are in disagreement in some cases, which we attributed to the different order of the Markov chain that were used. The 1-composite estimator was computed with $k = 5$ whereas 2-composite was estimated with $k = 3$, for computational reasons. Naturally, one could use a higher order to compute the diagonal elements, but we used the same order for all entries of the 2-composite estimator to illustrate the differences that arise in this case. The 1-composite Markov estimator produces estimates that are generally in agreement with those of the realized variance, both in terms of magnitude and signs of covariances.

7.2.3 Daily Estimates for 2013

We have estimated variances and covariances for the 10:00–14:00 interval for all trading days during 2013. Some selected series are presented in Figs. 3 and 4.

In Fig. 3 we plot annualized volatilities for SPY, Crude Oil, Gold, and Wheat based on the Markov estimator with $k = 5$, and these are benchmarked with those of the realized variance with 10-min sampling. The estimates are quite similar, both for the very liquid assets, SPY, Crude Oil, and Gold, and the relatively illiquid assets, Wheat, whose high-frequency data have pronounced bid-ask bounces.

In Fig. 4 we plot daily estimates of the correlations for both the 1- and 2-composite Markov estimators. These are benchmarked with the realized correlations based on 10-min sampling. We observed that Gold/Silver are highly correlated and its correlation is highly persistent over the year 2013. More moderate correlations are observed for SPY/Crude Oil and Soybeans/Wheat, and these series exhibit a higher

Table 8 The table presents two realized variances for the 14 assets for the month of March, 2013

RV_{5min}	SPY	CL	NG	GC	SV	HG	LC	LH	KC	SB	CT	CN	SY	WC
S&P 500	98.20	72.55	7.53	-29.49	-40.80	59.85	11.82	8.20	14.35	-0.77	11.53	12.48	6.01	14.05
Light crude	0.42	304.64	41.55	4.72	27.94	106.73	17.78	34.32	46.45	5.08	15.23	45.38	25.09	37.27
Natural gas	0.03	0.08	828.53	-0.90	5.96	43.62	5.33	5.39	3.84	-21.71	4.43	18.29	28.41	-0.53
Gold	-0.27	0.02	-0.00	123.93	201.43	21.89	-5.48	-18.43	-10.53	-3.25	-3.96	8.03	14.74	6.89
Silver	-0.18	0.07	0.01	0.80	512.90	81.49	2.16	-45.85	15.06	13.05	-0.98	4.98	30.76	8.66
Copper	0.38	0.38	0.09	0.12	0.22	257.37	7.16	-10.27	49.36	11.44	-2.78	40.49	38.99	52.05
Live cattle	0.08	0.07	0.01	-0.03	0.01	0.03	199.14	153.29	4.09	-6.71	8.24	11.19	10.78	15.68
Lean hogs	0.03	0.06	0.01	-0.05	-0.07	-0.02	0.36	935.02	9.53	-60.11	20.26	7.31	35.86	-1.77
Coffee "C"	0.06	0.10	0.01	-0.04	0.03	0.12	0.01	0.01	657.53	74.46	16.09	63.27	22.37	61.59
Sugar #1	-0.00	0.01	-0.03	-0.01	0.02	0.03	-0.02	-0.08	0.12	551.36	2.53	79.07	42.36	72.83
Cotton #2	0.05	0.04	0.01	-0.02	-0.00	-0.01	0.03	0.03	0.03	0.00	488.85	-69.86	-17.14	-85.78
Corn	0.04	0.08	0.02	0.02	0.01	0.08	0.03	0.01	0.08	0.11	-0.10	996.64	335.43	711.51
Soybeans	0.03	0.06	0.04	0.06	0.06	0.10	0.03	0.05	0.04	0.08	-0.03	0.45	552.12	270.99
Wheat	0.05	0.07	-0.00	0.02	0.01	0.11	0.04	-0.00	0.08	0.11	-0.14	0.79	0.40	818.54

(continued)

Table 8 (continued)

RV _{10min}	SPY	CL	NG	GC	SV	HG	LC	LH	KC	SB	CT	CN	SY	WC
S&P 500	103.38	83.04	1.75	-35.00	-43.92	67.72	18.68	17.29	13.72	1.82	14.03	11.27	13.61	20.55
Light crude	0.45	329.37	59.02	3.89	37.04	127.49	17.70	46.91	68.14	12.95	7.74	50.17	29.61	36.87
Natural gas	0.01	0.12	771.89	14.50	45.49	41.19	-2.69	8.37	0.61	-54.02	22.02	-18.33	4.83	-0.14
Gold	-0.32	0.02	0.05	117.30	193.10	14.79	-16.56	-28.31	-4.43	7.41	-6.04	2.63	11.57	2.25
Silver	-0.19	0.09	0.07	0.79	510.20	84.45	-18.71	-84.96	34.49	30.12	-0.59	1.41	21.38	14.52
Copper	0.42	0.44	0.09	0.09	0.23	255.99	16.10	-0.08	58.08	18.61	-18.35	39.15	40.07	63.66
Live cattle	0.13	0.07	-0.01	-0.11	-0.06	0.07	206.36	182.41	13.66	-7.21	26.27	21.69	24.43	18.48
Lean hogs	0.06	0.09	0.01	-0.09	-0.13	-0.00	0.43	870.97	-8.22	-37.37	1.44	38.99	50.73	23.20
Coffee "C"	0.05	0.14	0.00	-0.02	0.06	0.14	0.04	-0.01	689.21	75.36	9.40	96.86	77.15	83.21
Sugar #1	0.01	0.03	-0.09	0.03	0.06	0.05	-0.02	-0.06	0.13	502.96	17.81	92.54	76.18	91.50
Cotton #2	0.07	0.02	0.04	-0.03	-0.00	-0.05	0.09	0.00	0.02	0.04	442.55	-12.93	19.53	0.36
Corn	0.04	0.09	-0.02	0.01	0.00	0.08	0.05	0.04	0.12	0.13	-0.02	999.24	356.24	647.75
Soybeans	0.05	0.06	0.01	0.04	0.04	0.10	0.07	0.07	0.11	0.13	0.04	0.44	653.10	265.91
Wheat	0.08	0.08	-0.00	0.01	0.02	0.15	0.05	0.03	0.12	0.16	0.00	0.78	0.40	683.20

Variances and covariances are annualized and further scaled by 10⁴. Correlations are in the lower triangle in italic font

Table 9 The table presents two Markov estimates of the variance-covariance for the 14 assets

MC ¹	SPY	CL	NG	GC	SV	HG	LC	LH	KC	SB	CT	CN	SY	WC
S&P 500	106.38	69.59	-1.52	-12.53	-9.17	49.64	5.09	-3.36	10.86	-1.08	5.40	11.11	12.02	8.20
Light crude	0.37	326.26	14.89	9.34	28.47	82.09	6.53	2.76	16.30	-5.62	5.13	2.92	16.26	14.55
Natural gas	-0.00	0.03	932.63	7.18	18.48	10.08	12.13	-5.79	22.71	-0.61	-9.77	-22.36	14.87	1.44
Gold	-0.10	0.04	0.02	149.93	219.11	26.36	-0.83	1.71	1.00	-2.37	-1.74	-0.26	2.41	-0.02
Silver	-0.04	0.07	0.03	0.76	557.24	73.28	0.12	-5.14	6.68	-13.18	1.91	4.63	5.79	-1.43
Copper	0.30	0.28	0.02	0.13	0.19	259.93	3.97	-6.91	18.82	-1.84	9.30	16.38	9.38	10.22
Live cattle	0.03	0.02	0.03	-0.00	0.00	0.02	213.33	128.46	11.34	-1.90	16.85	-2.27	1.36	5.90
Lean hogs	-0.01	0.01	-0.01	0.00	-0.01	-0.01	0.31	830.83	9.27	-5.11	9.64	3.63	11.97	6.54
Coffee "C"	0.04	0.04	0.03	0.00	0.01	0.05	0.03	0.01	636.57	36.08	7.91	2.97	7.70	8.80
Sugar #1	-0.00	-0.01	-0.00	-0.01	-0.02	-0.00	-0.01	-0.01	0.06	528.18	-1.88	33.55	15.23	28.24
Cotton #2	0.02	0.01	-0.01	-0.01	0.00	0.03	0.05	0.02	0.01	-0.00	490.22	12.67	11.50	13.48
Corn	0.04	0.01	-0.03	-0.00	0.01	0.04	-0.01	0.00	0.00	0.05	0.02	835.96	299.02	523.38
Soybeans	0.05	0.04	0.02	0.01	0.01	0.02	0.00	0.02	0.01	0.03	0.02	0.43	578.95	220.34
Wheat	0.03	0.03	0.00	-0.00	-0.00	0.02	0.02	0.01	0.01	0.05	0.02	0.70	0.35	666.82

(continued)

Table 9 (continued)

MC ²	SPY	CL	NG	GC	SV	HG	LC	LH	KC	SB	CT	CN	SY	WC
S&P 500	116.84	80.35	-12.55	-12.92	-7.44	76.88	8.40	-5.69	13.76	1.67	42.68	-9.93	10.89	0.47
Light crude	0.38	391.86	7.67	-1.45	39.30	125.36	6.38	-11.56	29.37	-7.56	89.63	-25.61	22.06	-3.37
Natural gas	-0.04	0.01	921.74	1.64	-44.28	84.25	-32.41	-10.85	74.22	-53.72	-56.59	34.81	-12.94	-22.93
Gold	-0.11	-0.01	0.00	116.87	149.72	15.23	-2.08	25.73	-27.00	-4.72	-37.49	0.15	5.05	-3.15
Silver	-0.04	0.10	-0.07	0.71	380.17	23.70	4.60	16.03	-59.90	23.58	-60.49	-36.39	-1.79	-38.08
Copper	0.37	0.33	0.15	0.07	0.06	365.15	-29.91	17.98	24.53	0.16	80.92	-7.86	-34.35	-16.89
Live cattle	0.07	0.03	-0.09	-0.02	0.02	-0.13	142.36	18.71	-20.08	4.14	-15.27	9.25	-19.75	31.61
Lean hogs	-0.02	-0.02	-0.01	0.10	0.03	0.04	0.06	606.71	69.63	7.01	426.46	-27.19	19.11	35.20
Coffee "C"	0.06	0.07	0.12	-0.12	-0.15	0.06	-0.08	0.14	421.94	23.83	5.64	-15.59	54.31	-39.85
Sugar #1	0.01	-0.02	-0.10	-0.02	0.07	0.00	0.02	0.02	0.06	345.21	95.41	11.69	-28.41	29.00
Cotton #2	0.17	0.19	-0.08	-0.15	-0.13	0.18	-0.05	0.74	0.01	0.22	544.65	83.64	19.19	68.24
Corn	-0.04	-0.06	0.05	0.00	-0.09	-0.02	0.04	-0.05	-0.03	0.03	0.16	475.69	236.91	320.47
Soybeans	0.05	0.06	-0.02	0.02	-0.00	-0.09	-0.09	0.04	0.14	-0.08	0.04	0.57	365.44	162.80
Wheat	0.00	-0.01	-0.04	-0.01	-0.10	-0.04	0.13	0.07	-0.10	0.08	0.14	0.73	0.42	407.53

Variances and covariances are annualized and further scaled by 10^4 . The 1-composite estimator is computed with $k = 5$ and the 2-composite estimator is computed with $k = 3$. Correlations are in the lower triangle in italic font

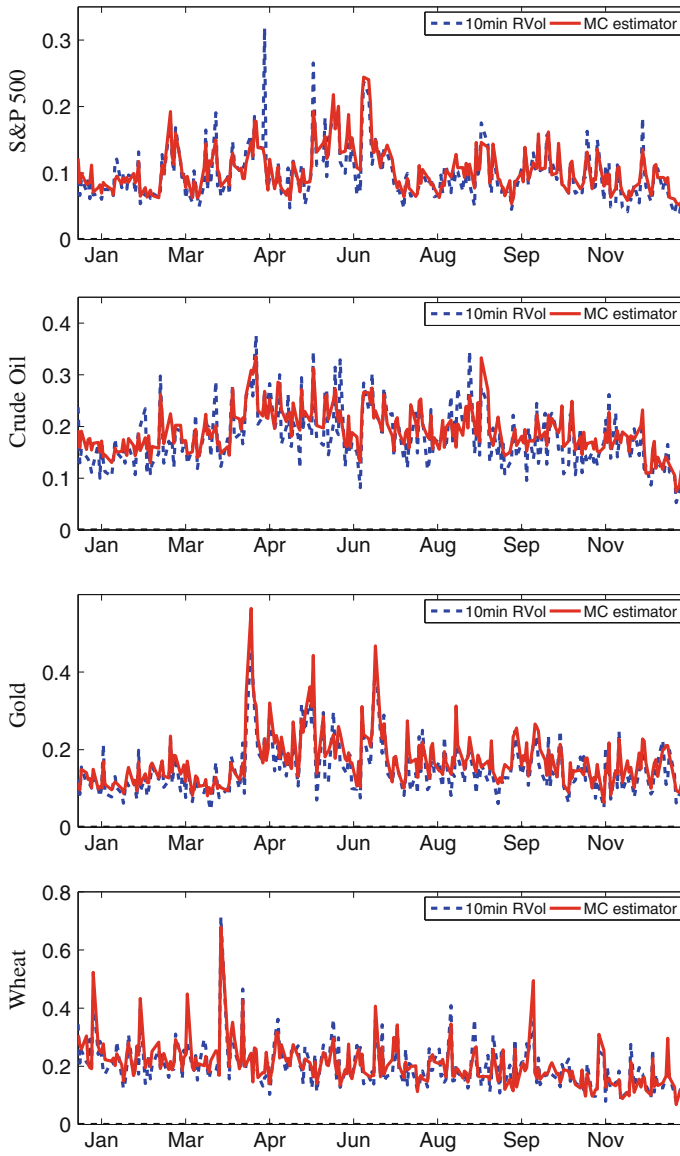


Fig. 3 Realized volatility based on 10 min returns against volatility computed with MC estimator (order $k = 5$) in 2013. Estimated values are annualized

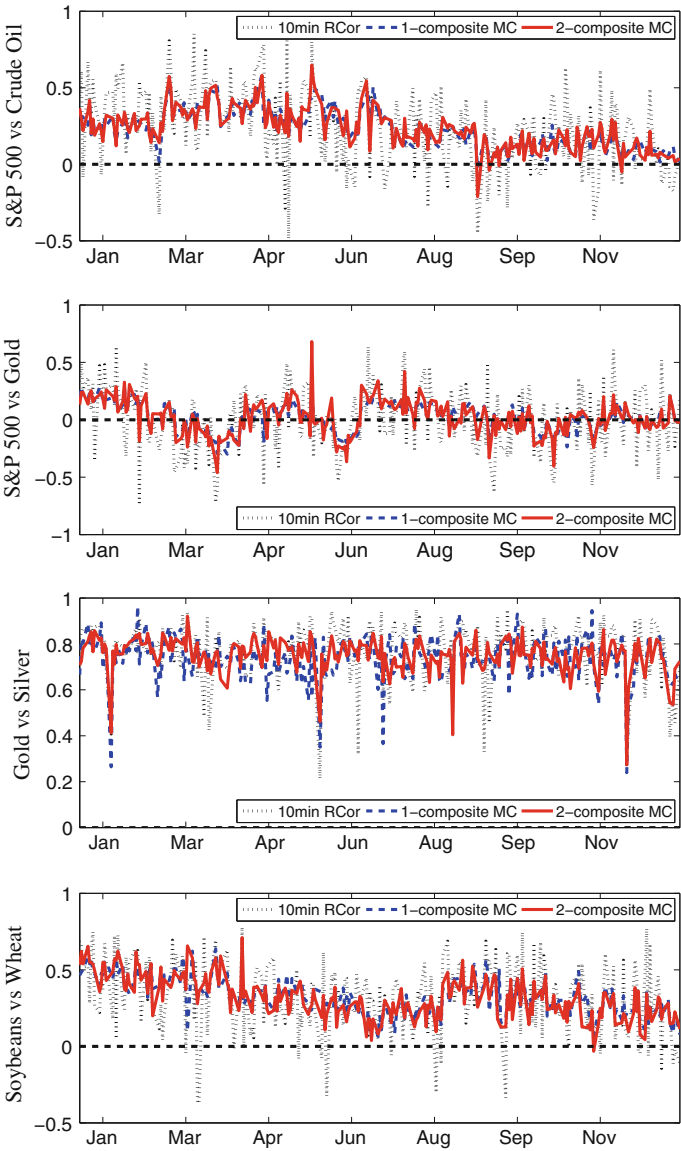


Fig. 4 Realized correlation based on 10min returns against correlation computed with MC estimator (order $k = 5$) in 2013

degree of time-variation. In the case of SPY/Gold we observe a less stable correlation that changes sign several times during the year. The general patterns are successfully captured by the composite Markov estimators, and while the realized correlation is in agreement about the general trends, it exhibits far more day-to-day variation which suggests that it is less accurate. The smooth and persistent behavior of the Markov estimators may be attributed to these estimators being more accurate.

8 Conclusion

In this paper we have proposed a multivariate volatility estimator that is based on the theory of finite Markov chains. The Markov chain estimator takes advantage of the fact that high-frequency prices are confined to a grid. This is the first robust multivariate estimator for which standard errors are readily available. Previous estimators include the multivariate realized kernel estimator, whose standard error also requires an estimate of the long-run variance of the noise, which is difficult to estimate because the noise is, in practice, small, serially dependent, and endogenous. The multivariate kernel estimator (MRK) converges at rate $n^{1/5}$. In contrast, the Markov estimator converges at rate $n^{1/2}$ owing to the specification assumed for the high-frequency data. These rates are, however, not directly comparable for practical situation, as the order of the Markov chain may be required to increase with n , in order to accommodate inhomogeneity resulting from time-varying volatility. Our simulation design suggests that the Markov estimator and the MRK performs similarly in practice, so the major advantage of the Markov estimator is the readily available standard error.

The estimator performs well in a simulation design, and is relatively insensitive to the choice of the order of the Markov chain, k , which is the tuning parameter that must be chosen in practice.

A potential limitation of the estimator is the high-dimensional objects that the estimator is computed from. For the full estimator the dimension can be as large as $(S^d)^k$, where S is the number of primitive states for the individual series, d is the dimension of the process and k is the order of the Markov chain. The dimension will typically be much smaller in practice because many states are not observed in a given sample, and the transition matrix will be very sparse, because most transitions between states are unobserved. So there is a need to further analyze the finite sample properties of the full Markov estimator, and to characterize its limitations.

The two composite Markov estimators alleviate the challenges with high dimensional objects, but may require a projection to guarantee a positive semidefinite estimate. For this purpose we have proposed a novel projection that makes use of the standard errors of the elements of the matrix being projected. Since these are readily available for the 2-composite estimator it is appealing to incorporate these, so that a projection leaves accurately estimated elements relatively unchanged.

The empirical analysis of commodity prices illustrated the three Markov estimators, and benchmarked them against conventional realized variances. The estimates were largely in agreement, but the Markov estimators fare particularly well with

regards to estimating correlations. While the time series of daily correlation estimates based on the realized variance were somewhat erratic, those of the Markov estimators were more stable.

Appendix: Details on the Simulation Design with Stochastic Volatility

For comparison with the simulation design with constant volatility we seek to have the integrated variance be 1, in expectation. This is achieved as follows. Note that

$$E((d \log Y_{i,t})^2) = E[\exp\{2(\beta_0 + \beta_1 \tau_{i,t})\} \Delta],$$

and since we (approximately) have $\tau_{i,t} \sim N(0, a^2)$ with

$$a^2 = \frac{\frac{1 - \exp(2\alpha \Delta)}{-2\alpha}}{1 - \exp(\alpha \Delta)^2} = \frac{1}{-2\alpha},$$

it follows that

$$2(\beta_0 + \beta_1 \tau_{i,t}) \sim N(2\beta_0, 4\beta_1^2 \frac{1}{-2\alpha}).$$

Hence

$$E[\exp\{2(\beta_0 + \beta_1 \tau_{i,t})\}] = \exp\left(2\beta_0 + \beta_1^2 \frac{1}{-\alpha}\right),$$

which will be equal to 1 if we set $\beta_0 = \beta_1^2/(2\alpha)$.

References

1. Ait-Sahalia, Y., Fan, J., Xiu, D.: High-frequency covariance estimates with noisy and asynchronous financial data. *J. Am. Stat. Assoc.* **105**(492), 1504–1517 (2010)
2. Andersen, T., Dobrev, D., Schaumburg, E.: Duration-based volatility estimation. working paper (2008)
3. Andersen, T.G., Bollerslev, T.: Answering the skeptics: yes, standard volatility models do provide accurate forecasts. *Int. Econ. Rev.* **39**(4), 885–905 (1998)
4. Andersen, T.G., Bollerslev, T., Diebold, F.X., Ebens, H.: The distribution of realized stock return volatility. *J. Finan. Econ.* **61**(1), 43–76 (2001)
5. Andersen, T.G., Bollerslev, T., Diebold, F.X., Labys, P.: The distribution of exchange rate volatility. *J. Am. Stat. Assoc.* **96**(453), 42–55 (2001). Correction published in 2003, vol. 98, p. 501
6. Bandi, F.M., Russell, J.R.: Market microstructure noise, integrated variance estimators, and the accuracy of asymptotic approximations. working paper (2006)
7. Bandi, F.M., Russell, J.R.: Microstructure noise, realized variance, and optimal sampling. *Rev. Econ. Stud.* **75**, 339–369 (2008)

8. Barndorff-Nielsen, O.E., Hansen, P.R., Lunde, A., Shephard, N.: Designing realised kernels to measure the ex-post variation of equity prices in the presence of noise. *Econometrica* **76**, 1481–536 (2008)
9. Barndorff-Nielsen, O.E., Hansen, P.R., Lunde, A., Shephard, N.: Multivariate realised kernels: consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading. *J. Econometrics* **162**, 149–169 (2011)
10. Barndorff-Nielsen, O.E., Hansen, P.R., Lunde, A., Shephard, N.: Subsampling realised kernels. *J. Econometrics* **160**, 204–219 (2011)
11. Barndorff-Nielsen, O.E., Shephard, N.: Econometric analysis of realised volatility and its use in estimating stochastic volatility models. *J. R. Stat. Soc. B* **64**, 253–280 (2002)
12. Barndorff-Nielsen, O.E., Shephard, N.: Econometric analysis of realised covariation: high frequency based covariance, regression and correlation in financial economics. *Econometrica* **72**, 885–925 (2004)
13. Christensen, K., Oomen, R.C., Podolskij, M.: Realised quantile-based estimation of the integrated variance. Working paper (2008)
14. Christensen, K., Podolskij, M.: Realized range-based estimation of integrated variance. *J. Econometrics* **141**, 323–349 (2007)
15. Christoffersen, P., Lunde, A., Olesen, K.V.: Factor structure in commodity futures return and volatility. Working paper (2014)
16. Delattre, S., Jacod, J.: A central limit theorem for normalized functions of the increments of a diffusion process, in the presence of round-off errors. *Bernoulli* **3**, 1–28 (1997)
17. Grant, M.C., Boyd, S.P., Ye, Y.: CVX: Matlab Software for Disciplined Convex Programming (2014)
18. Hansen, P.R.: A martingale decomposition of discrete Markov chains. Working paper (2014)
19. Hansen, P.R., Horel, G.: Quadratic variation by Markov chains. Working paper (2009)
20. Hansen, P.R., Horel, G.: Limit theory for the long-run variance of finite Markov chains. Working paper (2014)
21. Hansen, P.R., Large, J., Lunde, A.: Moving average-based estimators of integrated variance. *Econometric Rev.* **27**, 79–111 (2008)
22. Hansen, P.R., Lunde, A.: Realized variance and market microstructure noise. *J. Bus. Econ. Stat.* **24**, 127–218 (2006). The 2005 Invited Address with Comments and Rejoinder
23. Hautsch, N., Kyj, L.M., Oomen, R.: A blocking and regularization approach to high dimensional realized covariance estimation. *J. Appl. Econometrics* **27**, 625–645 (2012)
24. Horel, G.: Estimating integrated volatility with markov chains. Ph.D. thesis, Stanford University (2007)
25. Jacod, J., Li, Y., Mykland, P.A., Podolskij, M., Vetter, M.: Microstructure noise in the continuous case: the pre-averaging approach. *Stochast. Process. Their Appl.* **119**, 2249–2276 (2009)
26. Kemeny, J., Snell, J.: Finite markov chains. Springer, New York (1976)
27. Ledoit, O., Santa-Clara, P., Wolf, M.: Flexible multivariate GARCH modeling with an application to international stock markets. *Rev. Econ. Stat.* **85**, 735–747 (2003)
28. Li, Y., Mykland, P.A.: Determining the volatility of a price process in the presence of rounding errors. Working paper (2006)
29. Lunde, A., Shephard, N., Sheppard, K.: Econometric analysis of vast covariance matrices using composite realized kernels. Working paper (2014)
30. Maheu, J.M., McCurdy, T.H.: Nonlinear features of realized FX volatility. *Rev. Econ. Stat.* **84**, 668–681 (2002)
31. Renò, R.: A closer look at the epps effect. *Int. J. Theor. Appl. Finance* **6**, 87–102 (2003)
32. Zhang, L.: Efficient estimation of stochastic volatility using noisy observations: a multi-scale approach. *Bernoulli* **12**, 1019–1043 (2006)
33. Zhang, L., Mykland, P.A., Ait-Sahalia, Y.: A tale of two time scales: determining integrated volatility with noisy high frequency data. *J. Am. Stat. Assoc.* **100**, 1394–1411 (2005)
34. Zhou, B.: High-frequency data and volatility in foreign exchange rates. *J. Bus. Econ. Stat.* **14**(1), 45–52 (1996)
35. Zhou, B.: Parametric and nonparametric volatility measurement. In: Dunis, C.L., Zhou, B. (eds.) *Nonlinear Modelling of High Frequency Financial Time Series*, Chap. 6, pp. 109–123. Wiley, New York (1998)

Dependence Uncertainty for Aggregate Risk: Examples and Simple Bounds

Paul Embrechts and Edgars Jakobsons

Abstract Over the recent years, numerous results have been derived in order to assess the properties of regulatory risk measures (in particular VaR and ES) under dependence uncertainty. In this paper we complement this mainly methodological research by providing several numerical examples for both homogeneous as well as inhomogeneous portfolios. In particular, we investigate under which circumstances the so-called worst-case VaR can be well approximated by the worst-case (i.e. comonotonic) ES. We also study best-case values and simple lower bounds.

Keywords Value-at-risk · Expected shortfall · Risk aggregation · Dependence uncertainty · Inhomogeneous portfolio · Rearrangement algorithm

1 Introduction

Recent regulatory discussions in the realm of banking and insurance have brought the following quantitative aspects very much to the forefront:

- (i) The choice of risk measure for the calculation of regulatory capital; examples include Value-at-Risk (VaR) and Expected Shortfall (ES).
- (ii) The properties of statistical estimators of such risk measures, especially when based on extreme tail observations.
- (iii) The issue of Model Uncertainty (MU). The latter can be interpreted in many ways: statistical, numerical, functional, ...

In this paper we will mainly concentrate on a combination of (i) and (iii), and discuss the (functional) model uncertainty of risk measures VaR and ES for linear portfolios in both the homogeneous as well as inhomogeneous case. The interpretation of MU concerns portfolios with known marginal distributions but unknown

P. Embrechts (✉) · E. Jakobsons
Department of Mathematics, ETH Zurich, 8092 Zurich, Switzerland
e-mail: embrechts@math.ethz.ch

E. Jakobsons
e-mail: edgars.jakobsons@math.ethz.ch

interdependence. The latter is often referred to as Dependence Uncertainty (DU). It plays a crucial role in various examples throughout the banking and insurance literature. Going forward, several relevant examples will be given. One general purpose reference is [17].

The considered setting is a static one-period model, however, the results are also applicable to the dynamic case, using marginal distributions conditioned on the available information, e.g. conditional mean and volatility. While conditioning on the volatility may lead to light-tailed marginal distributions, it is not always the case, as shown in [23], where methods from extreme value theory are applied to GARCH-filtered (and still heavy-tailed) one-dimensional time series in order to obtain point estimates of VaR and ES. Moreover, [7] argue that the volatility of financial time series is forecastable only up to 10 days ahead (relevant for market risk) but not for longer time scales (such as 1 year, relevant for insurance). The effect of tail-heaviness of the marginal distributions on the portfolio risk will therefore be analyzed. Furthermore, the conditional dependence between the margins may change over time, as illustrated by [10], where a parametric copula approach is used to model and detect changes in the dependence structure. This demonstrates some of the difficulties in estimating the dependence. The mainly numerical results presented in our paper aim at understanding better the estimation of the best and worst possible risk measure values under DU. The paper only offers a first step on the way to this goal, and numerous alternative approaches as well as results are available or can be obtained. We very much hope that this paper incites other researchers to look at these and related problems of Quantitative Risk Management.

2 Homogeneous Portfolios

We consider random variables $X_i \sim F_i$, $i = 1, \dots, d$ and $S_d = X_1 + \dots + X_d$. For the purpose of this paper we assume the marginal distribution functions (dfs) F_i to be known. If $F_i = F$, $i = 1, \dots, d$, we refer to the *homogeneous* case, otherwise we refer to the *inhomogeneous* case. For most of the paper we will concentrate on the homogeneous case, though for many of the results information on the inhomogeneous case can be obtained; see Sect. 3 for a brief discussion. The joint distribution of $(X_1, \dots, X_d)$ is not specified, and thus the aggregate value-at-risk $\text{VaR}_\alpha(S_d)$ and expected shortfall $\text{ES}_\alpha(S_d)$ are not uniquely determined. In the following we consider the range of possible values these aggregate risk measures can take, under fixed marginal distributions, but unspecified joint model. This framework in risk management is referred to as *dependence uncertainty*; see [18]. In this section we focus on the homogeneous case and assume throughout that the support of F is bounded below. Define the generalized inverse

$$F^{-1}(p) = \inf\{x : F(x) \geq p\}, \quad p \in (0, 1],$$

and $F^{-1}(0) = \inf\{x : F(x) > 0\}$ (the left endpoint of the support); for properties of generalized inverses, see [11]. Since ES and VaR are translation equivariant, for simplicity also assume $F^{-1}(0) = 0$ (via translation). To give the main results in the literature on DU bounds on risk measures in the following sections, a definition from [33] will be useful.

Definition 1 ([33]) For $d \geq 1$, a distribution function F is called *d-completely mixable* (*d-CM*) if there exist rvs $X_1, \dots, X_d \sim F$ and a constant k such that $X_1 + \dots + X_d = dk$ a.s.

Examples of distributions that are completely mixable for $d \geq 2$ include uniform, Gaussian, Cauchy and other unimodal symmetric continuous distributions; for higher values of d also distributions with monotone or concave densities on a bounded support; see [30] for further examples.

Some notation from [3] will also be needed. With respect to a df G (to be specified), we introduce functions $H(c)$ and $D(c)$, $c \in [0, 1]$ and a constant c_d :

$$\begin{aligned} H(c) &= (d-1)G^{-1}((d-1)c/d) + G^{-1}(1-c/d), \\ D(c) &= \frac{1}{1-c} \int_c^1 H(t)dt = \frac{d}{1-c} \int_{(d-1)c/d}^{1-c/d} G^{-1}(t)dt, \\ c_d &= \inf\{c \in [0, 1] : H(c) \leq D(c)\}, \end{aligned} \quad (1)$$

where for $c = 1$ we set $D(1) := D(1-) = H(1) = dG^{-1}(1-1/d)$. Also, introduce two conditions

- (A) H is non-increasing on $[0, c_d]$,
- (B) The conditional distribution of G on $[(d-1)c_d/d, 1-c_d/d]$ is *d-completely mixable*.

These conditions will imply validity and sharpness of some of the bounds stated in the following sections. A special case in which these conditions are satisfied is given in the following lemma.

Lemma 1 ([3]) *If the df G admits a non-increasing density on its support, then conditions (A) and (B) hold.*

Furthermore, [3] motivate numerically that these conditions are satisfied also in some other cases, using examples with Lognormal and Gamma dfs.

2.1 Upper Bound on VaR

For a random variable $X \sim F_X$ representing a loss, VaR at confidence level $\alpha \in (0, 1)$ is defined as the α -quantile,

$$\text{VaR}_\alpha(X) = F_X^{-1}(\alpha).$$

In turn, the upper bound on $\text{VaR}_\alpha(S_d)$ over all joint models is defined as

$$\overline{\text{VaR}}_\alpha(S_d) := \sup\{\text{VaR}_\alpha(X_1 + \cdots + X_d) : X_i \sim F_i, i = 1, \dots, d\}.$$

In the homogeneous case $F_i = F, i = 1, \dots, d$, the upper bound $\overline{\text{VaR}}_\alpha(S_d)$ can be obtained by solving the integral inequality for c_d in (1) with respect to the conditional distribution in the tail (defined below).

Proposition 1 ([12, 27, 34]) *Define H , D and c_d in (1) with respect to F_α , the conditional distribution of F on $[F^{-1}(\alpha), \infty)$. If F_α admits a non-increasing density on its support, then*

$$\overline{\text{VaR}}_\alpha(S_d) = D(c_d).$$

Remark 1 A non-increasing density above the α -quantile is a natural assumption for high values of α , which holds for essentially all distributions used in practice. The proof of Proposition 1 uses Lemma 1 to verify conditions (A) and (B), which are sufficient for the bound in Proposition 1 to be sharp.

Before giving the next proposition, we recall the definition of expected shortfall,

$$\text{ES}_\alpha(X) = \frac{1}{1-\alpha} \int_\alpha^1 \text{VaR}_q(X) dq.$$

Let $\text{VaR}_\alpha^+(S_d)$ and $\text{ES}_\alpha^+(S_d)$ denote $\text{VaR}_\alpha(S_d)$ and $\text{ES}_\alpha(S_d)$ respectively, when the X_i are comonotonic. It is well-known that the *comonotonic* dependence structure gives the worst-case expected shortfall; see [9] and p. 251 in [24].

$$\overline{\text{ES}}_\alpha(S_d) := \sup\{\text{ES}_\alpha(X_1 + \cdots + X_d) : X_i \sim F_i, i = 1, \dots, d\} = \sum_{i=1}^d \text{ES}_\alpha(X_i) =: \text{ES}_\alpha^+(S_d).$$

Moreover, as d increases, $\overline{\text{VaR}}_\alpha(S_d)$ is asymptotically equivalent to $\text{ES}_\alpha^+(S_d)$. A first result in this direction is:

Proposition 2 ([28]) *Suppose $(X_d)_{d \geq 1}$ is a sequence of rvs identically distributed as F , where F is integrable and has a decreasing density on $[F^{-1}(\alpha), \infty)$. Then*

$$\lim_{d \rightarrow \infty} \frac{\overline{\text{VaR}}_\alpha(X_1 + \cdots + X_d)}{\text{ES}_\alpha^+(X_1 + \cdots + X_d)} = 1.$$

Remark 2 Proposition 2 holds under much more general conditions, also in the heterogeneous case; see [18]. Results of this type are relevant for regulatory practice for both the banking world (the so-called Basel framework) as well as insurance (Solvency 2); see [18].

In Table 1 the smallest number of margins d is given for which $\overline{\text{VaR}}_\alpha(S_d)$ is within 10 % of $\text{ES}_\alpha^+(S_d)$, i.e. after which the asymptotically equivalent sequence gives a

Table 1 Thresholds for the number of margins d for which $\overline{\text{VaR}}_\alpha(S_d)$ is within 10 % of $\text{ES}_\alpha^+(S_d)$

Pareto(a), $a =$	10.0	5.0	3.0	2.5	2.0	1.7	1.6	1.5
$\overline{\text{VaR}}_{0.95}$	3	3	4	4	6	12	17	29
LogN($0, \sigma^2$), $\sigma =$	0.75	0.90	1.13	1.26	1.49	1.71	1.81	1.94
$\overline{\text{VaR}}_{0.95}$	3	3	3	4	5	6	7	9
Pareto(a), $a =$	10.0	5.0	3.0	2.5	2.0	1.7	1.6	1.5
$\overline{\text{VaR}}_{0.99}$	2	3	3	4	6	11	16	27
LogN($0, \sigma^2$), $\sigma =$	0.70	0.89	1.20	1.37	1.66	1.95	2.08	2.23
$\overline{\text{VaR}}_{0.99}$	2	3	3	4	5	6	7	9
Pareto(a), $a =$	10.0	5.0	3.0	2.5	2.0	1.7	1.6	1.5
$\overline{\text{VaR}}_{0.999}$	2	3	3	4	6	11	15	27
LogN($0, \sigma^2$), $\sigma =$	0.67	0.92	1.33	1.56	1.93	2.28	2.44	2.63
$\overline{\text{VaR}}_{0.999}$	2	2	3	4	5	7	8	11

The levels $\alpha \in \{0.95, 0.99, 0.999\}$ are listed for comparison. The parameter σ for LogN($0, \sigma^2$) is chosen to match the ratio $\text{VaR}_\alpha(X_1)/\text{ES}_\alpha(X_1)$ with that of Pareto(a), i.e. $P[X_1 > x] = (x+1)^{-a}$, $x \geq 0$

reasonable approximation. Notice that in all but the most heavy-tailed cases, $d \approx 10$ was sufficient. Hence, if the number of margins is greater than that, we may use the easily calculated $\text{ES}_\alpha^+(S_d)$ as a reasonable estimate of the conservative $\overline{\text{VaR}}_\alpha(S_d)$.

The convergence rate in Proposition 2 is also known.

Proposition 3 ([18]) *If $\mathbb{E}[|X_1 - \mathbb{E}[X_1]|^k]$ is finite for some $k > 1$ and $\text{ES}_\alpha(X_1) > 0$, then, as $d \rightarrow \infty$,*

$$\frac{\overline{\text{VaR}}_\alpha(S_d)}{\text{ES}_\alpha^+(S_d)} = 1 - O(d^{-1+1/k}).$$

So in particular, if all moments are finite, then the convergence rate is $O(d^{-1})$, and if the distribution has a regularly varying tail (see Sect. 7.3 in [24]) with index $-\rho$, for example $X_1 \sim \text{Pareto}(\rho)$, then the convergence is slower, since $k < \rho < \infty$. In Fig. 1 the differences $1 - \overline{\text{VaR}}_\alpha(S_d)/\text{ES}_\alpha^+(S_d)$ as d increases are plotted on a logarithmic scale for different Pareto and Lognormal distributions. We observe that, although initially the rate of convergence is slower, for large d it seems faster than the theoretical one. Moreover, for small d the rates are not very different between Pareto and Lognormal. Tail-heaviness is often measured using the notion of regular variation, and also Proposition 3 determines the convergence rate according to the highest existing moment. However, regular variation concerns quantiles asymptotically approaching 1, and Proposition 3 holds for d tending to infinity. In order to analyze $\overline{\text{VaR}}_\alpha(S_d)$ at a fixed level, say, $\alpha = 0.99$, and for small d , we need a measure of tail-heaviness that captures the behavior for moderate α and d . We refrain from using skewness, kurtosis and higher moments, as they may not exist for power tails. Instead, we will define a different measure of tail-heaviness in the following section.

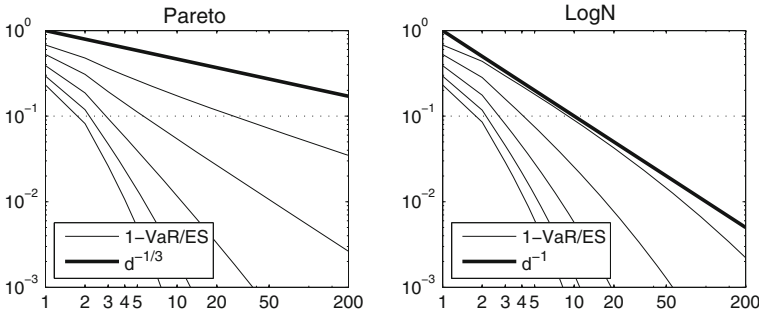


Fig. 1 Relative differences $1 - \overline{\text{VaR}}_\alpha(S_d)/\text{ES}_\alpha^+(S_d)$ for $\alpha = 0.99$ on the vertical axis versus d on the horizontal, on a log-log scale. Below the dotted line the relative difference is smaller than 10 %. The left panel contains Pareto(a) distributions, $a = 1.5, 2, 3, 5, 10$ from top down. The bold line shows the theoretical convergence rate $O(d^{-1/3})$ for Pareto(1.5) according to Proposition 3. In the right panel the Lognormal $\text{LogN}(0, \sigma^2)$ case is plotted, $\sigma = 2.23, 1.66, 1.20, 0.89, 0.70$ from top down, chosen to match the ratio $\text{VaR}_\alpha(X_1)/\text{ES}_\alpha(X_1)$ with that of Pareto. The bold line shows the theoretical convergence rate $O(d^{-1})$ for distributions with all moments finite

2.1.1 The Normalized Mean-Median Ratio

For $X \sim F$ define the *normalized mean-median ratio* as

$$M_\alpha(F) := \frac{\text{ES}_\alpha(X) - \text{VaR}_\alpha(X)}{\text{MS}_\alpha(X) - \text{VaR}_\alpha(X)},$$

where MS is the *median shortfall*, $\text{MS}_{1-p}(X) := \text{VaR}_{1-p/2}(X)$; see [21]. $M_\alpha(F)$ is a measure of tail-heaviness defined for regularly varying, sub-exponential, as well as exponential tails, and only requires the first moment to exist. $M_\alpha(F)$ is invariant under scaling and translation, since the risk measures used in the definition are positively homogeneous and translation equivariant. In Fig. 2 M_α is plotted for Gamma, Lognormal and Pareto distributions, as a function of the shape parameter.

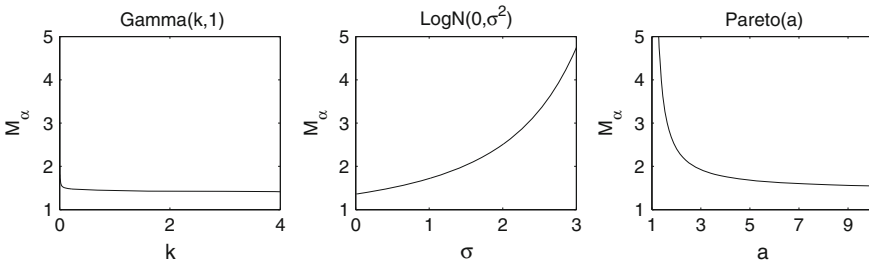


Fig. 2 $M_{0.99}$ for $\Gamma(k, 1)$, $\text{LogN}(0, \sigma^2)$ and Pareto(a) distribution, depending on the shape parameters k, σ, a , respectively

For all three families M_α diverges to infinity as the tail becomes heavier, and converges to a finite value for light tails.

For the Gamma distribution $\Gamma(k, 1)$ the values are computed numerically; to reach values in the typical range for Lognormal and Pareto, a very small parameter k is required (e.g. $k = 0.0027$ for $M_\alpha(\Gamma) = 3$). For $k = 1$ we recover the exponential distribution $\mathcal{E}(1)$, for which

$$M_\alpha(\mathcal{E}) = \frac{-\log(1-\alpha) + 1 + \log(1-\alpha)}{-\log((1-\alpha)/2) + \log(1-\alpha)} = \frac{1}{\log(2)} \approx 1.4427, \quad (2)$$

independent of α (and scale). And as $k \rightarrow \infty$, $\Gamma(k, 1)$ converges (under scaling and translation) to the standard Normal distribution Φ , so

$$\lim_{k \rightarrow \infty} M_\alpha(\Gamma(k, 1)) = \frac{\varphi(\Phi^{-1}(\alpha))/(1-\alpha) - \Phi^{-1}(\alpha)}{\Phi^{-1}(1-(1-\alpha)/2) - \Phi^{-1}(\alpha)} \quad (\approx 1.3583 \text{ for } \alpha = 0.99), \quad (3)$$

using the analytic formula for ES of the Normal distribution; see p. 45 in [24].

For the Lognormal distribution $\text{LogN}(\mu, \sigma^2)$,

$$M_\alpha(\text{LogN}(\mu, \sigma^2)) = \frac{e^{\sigma^2/2}(1 - \Phi(\Phi^{-1}(\alpha) - \sigma))/(1-\alpha) - e^{\sigma\Phi^{-1}(\alpha)}}{e^{\sigma\Phi^{-1}(1-(1-\alpha)/2)} - e^{\sigma\Phi^{-1}(\alpha)}},$$

independent of the scale e^μ . Since for $Z \sim \Phi$ we have $e^{\sigma Z} \sim \text{LogN}(0, \sigma^2)$, as $\sigma \rightarrow 0$, using the Taylor series of e^x about $x = 0$ we recover the same limit 1.3583 as in (3).

For the Pareto distribution $\mathcal{P}(a)$, $a > 1$,

$$M_\alpha(\mathcal{P}(a)) = \frac{\frac{a}{a-1}(1-\alpha)^{-1/a} - (1-\alpha)^{-1/a}}{((1-\alpha)/2)^{-1/a} - (1-\alpha)^{-1/a}} = \frac{1}{(a-1)(2^{-1/a} - 1)},$$

independent of the level α . As $a \rightarrow \infty$, we obtain $1/\log(2)$ as the limit, the same value as in (2). This is due the fact that if $X \sim \mathcal{E}(1)$, then $e^{X/a} - 1 \sim \mathcal{P}(a)$, and we can again use the Taylor series of e^x about $x = 0$.

We shall use M_α as a measure of tail-heaviness in order to compare the tails of different families of distributions, and from that deduce how close $\overline{\text{VaR}}_\alpha(S_d)$ is to $\text{ES}_\alpha^+(S_d)$ for small values of d . Notice that $M_\alpha(F)$ exactly determines the worst-case VaR when $d = 2$ and the density is decreasing beyond $F^{-1}(\alpha) = \text{VaR}_\alpha(X_1)$. This is because for $d = 2$ the worst-case dependence structure for VaR_α is *counter-monotonicity* above the α -quantile of X_1 and X_2 , viz. $X_2 = F^{-1}(\alpha + (1 - F(X_1)))$ when $F(X_1) \geq \alpha$; see [15, 22]. Furthermore, when the density is decreasing, the minimal sum $X_1 + X_2$ in the tail is attained when $X_1 = X_2 = \text{MS}_\alpha(X_1)$. Hence $\overline{\text{VaR}}(S_2) = 2 \text{MS}_\alpha(X_1)$. See also (4) and Fig. 3 in the next section.

2.1.2 The Normalized VaR Bound

Since VaR is a positively homogeneous and translation equivariant risk measure, it is easy to recalculate VaR under these operations. The upper bound on VaR depends only on the conditional distribution F on the interval $[F^{-1}(\alpha), \infty)$. Moreover, it is easy to see that

$$\overline{\text{VaR}}_\alpha(S_d) \in [\text{VaR}_\alpha^+(S_d), \text{ES}_\alpha^+(S_d)].$$

So, in order to focus on the issue of specifying *where* in this interval $\overline{\text{VaR}}_\alpha$ lies, we define the *normalized VaR bound*

$$\delta_\alpha(S_d) := \frac{\overline{\text{VaR}}_\alpha(S_d) - \text{VaR}_\alpha^+(S_d)}{\text{ES}_\alpha^+(S_d) - \text{VaR}_\alpha^+(S_d)} \in [0, 1].$$

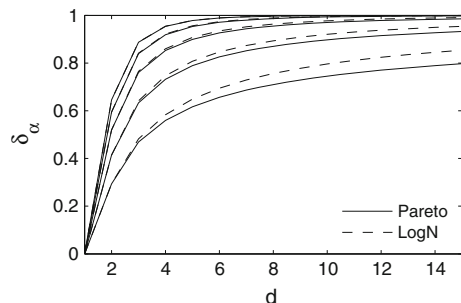
Note that $\delta_\alpha(S_d)$ does not depend on the scale and location of F . In Fig. 3 the values of $\delta_\alpha(S_d)$ are plotted in dependence on d . The parameters for Pareto and Lognormal distributions are chosen so that the M_α values match, thus, as explained in the previous section,

$$\delta_\alpha(S_2) = \frac{2 \text{MS}_\alpha(X_1) - 2 \text{VaR}_\alpha(X_1)}{2 \text{ES}_\alpha(X_1) - 2 \text{VaR}_\alpha(X_1)} = 1/M_\alpha(F), \quad (4)$$

and the upper VaR bounds agree for $d = 2$. For $d > 2$ the bounds for the matched Lognormal and Pareto dfs converge at a similar rate, especially for lighter tails, when convergence to 1 is rather fast. The difference $1 - \delta_\alpha$ seems to decrease as a power of d , so in Fig. 4 we plot this difference versus the dimension d on a logarithmic scale. We observe that, apart from the “kink” at $d = 2$, the dependence is approximately linear on the log-log scale. Based on this numerical evidence, and fixing the value at $d = 2$ according to (4), we consider an approximate model

$$1 - \delta_\alpha \approx (1 - 1/M_\alpha) \left(\frac{d}{2}\right)^{\lambda(M_\alpha)}. \quad (5)$$

Fig. 3 $\delta_\alpha(S_d)$ plotted in dependence on d . Pareto(a) with $a = 10, 5, 3, 2, 1.5$ from *top down*, with tail-heaviness $M_\alpha = 1.55, 1.68, 1.92, 2.41, 3.40$, respectively, and LogN(μ, σ^2) with the parameters chosen to match the $M_\alpha, \alpha = 0.99$



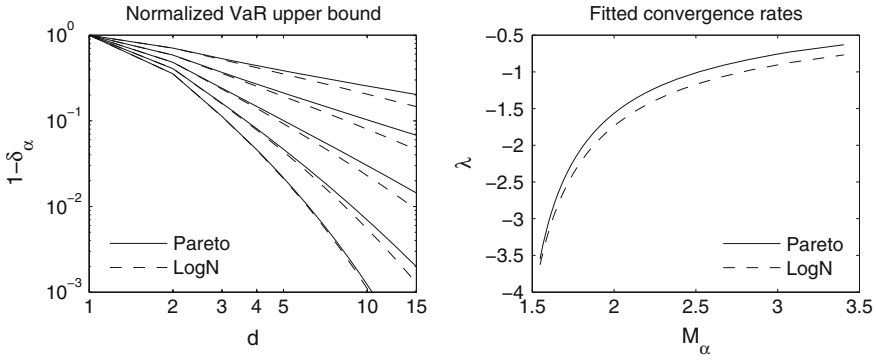


Fig. 4 Left panel normalized VaR bound δ_α , $\alpha = 0.99$, depending on the dimension d . Solid lines correspond to Pareto(a) distributions with $a = 1.5, 2, 3, 5, 10$ and $M_\alpha = 3.40, 2.41, 1.92, 1.68, 1.55$ from top down. Dashed lines correspond to Lognormal distributions with parameters selected to match the M_α values. Right panel the fitted values of $\lambda(M_\alpha)$ obtained by linear regression for different tail-weight levels M_α

Taking logs from both sides we obtain a linear dependence, and using linear regression for $d \in \{2 \dots 15\}$ we fit the slope parameter λ (the intercept at $d = 2$ is fixed) for each considered M_α . The obtained values of λ are plotted on the right panel of Fig. 4. They show that the convergence rate is indeed determined by the tail-heaviness M_α , and the values of $\lambda(M_\alpha)$ for the Pareto and Lognormal families are similar. Note also that e.g. for Pareto(1.5) the convergence rate according to Proposition 3 is $O(d^{-1/3})$ as $d \rightarrow \infty$, whereas from Fig. 4 we read that for small d the difference $1 - \delta_\alpha$ decreases approximately as $d^{-2/3}$ for the corresponding $M_\alpha = 3.4$. Finally, the fitted curves for δ_α using λ values from the linear regression are plotted in Fig. 5. The close match shows that the model (5) works well for small dimensions.

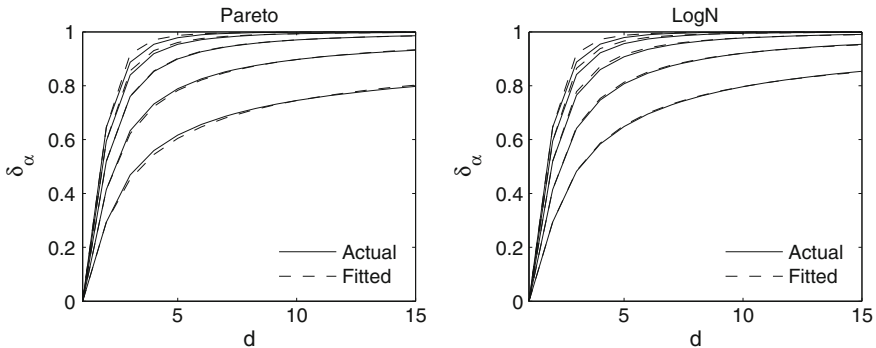


Fig. 5 Normalized VaR bound δ_α , $\alpha = 0.99$ (solid lines) compared to the fitted values (dashed) according to the model (5), depending on the dimension d . Left panel Pareto(a) distributions with $a = 1.5, 2, 3, 5, 10$ and $M_\alpha = 3.40, 2.41, 1.92, 1.68, 1.55$ from bottom up. Right panel Lognormal distributions with parameters selected to match the M_α values

2.2 Lower Bound on VaR

The lower bound on $\text{VaR}_\alpha(S_d)$ over all joint models is defined as

$$\underline{\text{VaR}}_\alpha(S_d) := \inf\{\text{VaR}_\alpha(X_1 + \cdots + X_d) : X_i \sim F_i, i = 1, \dots, d\}.$$

Before providing the basic result for the lower bound on VaR in the homogeneous case, we define the *left-tail expected shortfall*

$$\text{LES}_\alpha(X) = \frac{1}{\alpha} \int_0^\alpha \text{VaR}_p(X) \, dp,$$

and the comonotonic $\text{LES}_\alpha^+(S_d) = \sum_{i=1}^d \text{LES}_\alpha(X_i) = d \, \text{LES}_\alpha(X_1)$ (since $X_i \sim F$, $\forall i$).

Proposition 4 ([3]) *If the support of F is bounded below, then*

$$\underline{\text{VaR}}_\alpha(X_1 + \cdots + X_d) \geq \max\{(d-1)F^{-1}(0) + F^{-1}(\alpha), d \, \text{LES}_\alpha(X_1)\}. \quad (6)$$

Furthermore, define H and D as in (1) with respect to F^α , the conditional distribution of F on $[F^{-1}(0), F^{-1}(\alpha)]$. If conditions (A) and (B) with respect to F^α are satisfied, then (6) holds with equality, i.e.

$$\underline{\text{VaR}}_\alpha(S_d) = \max\{H(0), D(0)\}.$$

Since we assume that the essential support of F is the positive half-axis, (6) reduces to $\underline{\text{VaR}}_\alpha(X_1 + \cdots + X_d) \geq \max\{F^{-1}(\alpha), d \, \text{LES}_\alpha(X_1)\}$. In contrast to Proposition 1, in Proposition 4 we include the conditions (A) and (B) explicitly. Recall that previously these were implied by the non-increasing density assumption. While it is reasonable to assume this for the upper tail, in the lower tail this assumption holds for some distributions (e.g. Pareto) but fails for other (e.g. Lognormal). In such cases, one may resort to testing (A) numerically and testing for complete mixability in (B) using the method in [29]. Note, however, that if $F^{-1}(\alpha) > d \, \text{LES}_\alpha(X_1)$ then the bound can be sharp also without the d -CM condition. Typically, however, letting $d^* = F^{-1}(\alpha) / \text{LES}_\alpha(X_1)$, we may use as a good approximation

$$\begin{aligned} \underline{\text{VaR}}_\alpha(S_d) &\approx F^{-1}(\alpha) & \text{for } d < d^*, \\ \underline{\text{VaR}}_\alpha(S_d) &\approx d \, \text{LES}_\alpha(X_1) & \text{for } d \geq d^*. \end{aligned} \quad (7)$$

An alternative numerical approach for computing an approximation of the sharp bounds on VaR under DU is given by the Rearrangement Algorithm (RA); see [16] or [26, 27] for earlier formulations.¹ In Table 2 the values for Pareto, Lognormal

¹ A website set up by Giovanni Puccetti with the title “The Rearrangement Algorithm project” provides full details and recent developments on the RA; see <https://sites.google.com/site/rearrangementalgorithm/>.

Table 2 Thresholds for the number of margins d for which $\text{VaR}_\alpha(S_d)/\text{LES}_\alpha^+(S_d)$ has converged to 1

$\alpha = 0.99$	d^*	2.36	3.42	5.59	6.55	8.19	11.00	14.91
Pareto(a),	$a =$	—	—	10	5	3	2	1.5
	d	—	—	6	7	9	11	15
LogN($0, \sigma^2$),	$\sigma =$	0.40	0.59	0.88	0.98	1.13	1.35	1.60
	d	3	4	6	7	9	11	15
Gamma($k, 1$),	$k =$	5.00	2.00	0.77	0.58	0.40	0.25	0.16
	d	3	4	6	7	9	11	15

Parameters for the distributions are chosen to match the ratio $d^* = \text{VaR}_\alpha(X_1)/\text{LES}_\alpha(X_1)$

and Gamma distributed variables are listed for which the ratio $\text{VaR}_\alpha(S_d)/\text{LES}_\alpha^+(S_d)$ has reached the limit 1. Due to a non-increasing density, in the Pareto case (6) is always sharp and the limit is reached at $d = \lceil d^* \rceil$. As a side note, for the Pareto(a) distribution at level $p = \alpha$, as $a \rightarrow \infty$,

$$d^* = \frac{(1 - p)^{-1/a} - 1}{(1 - (1 - p)^{1-1/a})/((1 - 1/a)/p) - 1} \longrightarrow \frac{-\log(1 - p)}{1 + ((1 - p)/p) \log(1 - p)},$$

so for $p = 0.99$ we have $d^* > 4.8298$, i.e. no parameter a gives a lower value of d^* , hence they are not listed in Table 2. For the Lognormal and Gamma dfs, the values were obtained by the RA, using as discretization parameter $N = 10^5$ and the stopping condition $\varepsilon = 10^{-4}$. In Fig. 6 the normalized lower VaR bounds are plotted. The approximate bounds from (7) are the same for the three families. These are sharp for the Pareto, but differ from the sharp bounds (here, computed using the RA) for the Lognormal and Gamma families in the lighter-tailed cases, since the densities are not decreasing and the conditions in Proposition 4 are not satisfied. In particular, for small

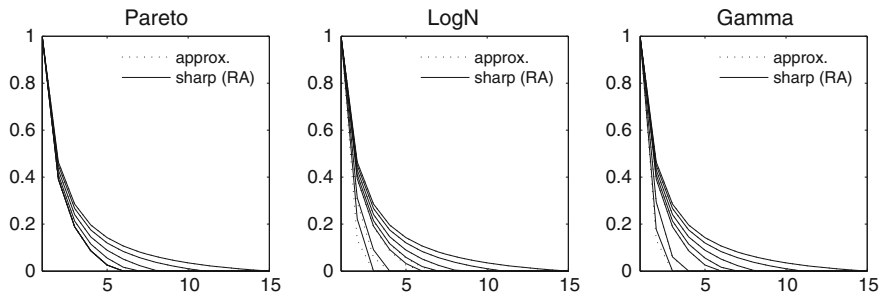


Fig. 6 $(\text{VaR}_\alpha(S_d) - \text{LES}_\alpha^+(S_d))/(\text{VaR}_\alpha^+(S_d) - \text{LES}_\alpha^+(S_d))$ plotted in dependence on d for Pareto(a), LogN($0, \sigma^2$) and $\Gamma(k, 1)$ distributions with the parameters chosen to match the d^* (see Table 2). For Pareto, the approximate bounds (7) are sharp, whereas for LogN and Gamma there are small deviations for the lighter-tailed cases in comparison to the sharp bounds given by the RA

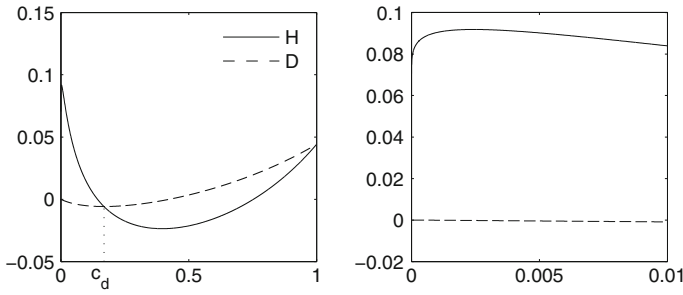


Fig. 7 Functions H and D with respect to F^α , the conditional distribution of $F = \text{LogN}(0, 0.59^2)$ on $[0, F^{-1}(\alpha)]$, $d = 3$, $\alpha = 0.99$. The vertical scale is normalized to show the interval $[\text{LES}_\alpha^+(S_d), \text{VaR}_\alpha^+(S_d)]$ as $[0, 1]$. *Left panel* full range $c \in [0, 1]$. *Right panel* close-up of the range $[0, 0.01]$ shows that the function H is increasing initially. $H(0) = 0.058$, while the sharp lower bound (from the RA) is 0.091

d it is the condition (A) that matters, and is not satisfied e.g. for the $\text{LogN}(0, 0.59^2)$ df with $d = 3$; see Fig. 7. With parameter $k \leq 1$, however, $\Gamma(k, 1)$ has a decreasing density, so by Lemma 1 satisfies the conditions and the approximate bound (6) is sharp. Overall, (6) gives an easily calculated lower bound on aggregate VaR, which is close to the sharp bound.

2.3 Lower Bound on ES

The lower bound on $\text{ES}_\alpha(S_d)$ under dependence uncertainty is defined as

$$\underline{\text{ES}}_\alpha(S_d) := \inf\{\text{ES}_\alpha(X_1 + \dots + X_d) : X_i \sim F_i, i = 1, \dots, d\}.$$

In the literature, few results on $\underline{\text{ES}}_\alpha$ are available. The two main references are [25] (using the RA) and [3]. The latter provides the following result in the homogeneous case.

Proposition 5 ([3]) *Define H , D and c_d as in (1) with respect to F . If condition (A) is satisfied, then*

$$\underline{\text{ES}}_\alpha(S_d) \geq \begin{cases} \frac{1}{1-\alpha} [\mathbb{E}[S_d] - \alpha D(c_d)] & \text{if } \alpha \leq 1 - c_d, \\ \frac{1}{1-\alpha} \int_0^{1-\alpha} H(t) dt & \text{if } \alpha > 1 - c_d. \end{cases} \quad (8)$$

Furthermore, if condition (B) is satisfied, then (8) holds with equality.

Remark 3 Due to Lemma 1, Proposition 5 is most useful in the case of a decreasing density, e.g. for Pareto or Exponential df. In other cases the conditions (A) and (B) would need to be verified numerically. Alternatively, one may apply the RA.

The latter, however, uses a discretization of the marginal distributions, which may underestimate the integral over the infinite tail of the aggregate risk.

A simple lower bound on the aggregate ES can be calculated as follows.

Proposition 6 ([6]) *If $F^{-1}(0) \geq 0$, then*

$$\underline{\text{ES}}_{1-p}(S_d) \geq \text{ES}_{1-p/d}(X_1). \quad (9)$$

Furthermore, if $F(0) \geq (d-1)/d$, then (9) holds with equality.

For this bound to be sharp, a sufficient mass at 0 (at least $p(d-1)/d$) is required for the marginal distributions, in which case Proposition 5 gives the same bound. In Fig. 8 all three bounds are plotted for different Pareto and Lognormal dfs. First, notice that the convergence to the limit ($\mathbb{E}[S_d]$ in this case) is much slower than for the VaR bounds. For the Pareto df, bounds from Proposition 5 are sharp, while the error in the RA bound due to discretization is clearly visible in the heavy-tailed cases (discretization parameter $N = 10^5$ and stopping condition $\varepsilon = 10^{-4}$ were used). The non-sharp bound from Proposition 6, however, is very close, because the Pareto distribution has a large mass near 0. For the Lognormal dfs, the density is small near 0, hence there is a visible approximation error for the lighter tails. The bound from the RA again underestimates the tail-expectation in the heavy-tailed cases due to discretization. Overall, Proposition 5 seems to give the best bound (due to infinite tails, condition (A) is likely to hold).

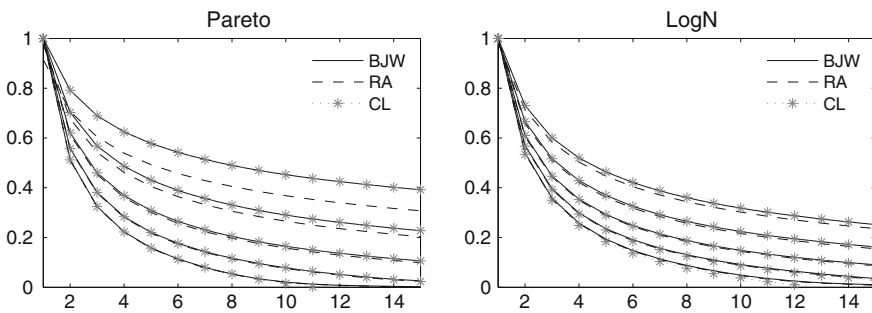


Fig. 8 $(\text{ES}_\alpha(S_d) - \mathbb{E}[S_d]) / (\text{ES}_\alpha^+(S_d) - \mathbb{E}[S_d])$ plotted in dependence on d . Solid lines are the bound from Proposition 8 ([3]), dashed lines correspond to the RA ([25]) and dotted lines to Proposition 6 ([6]). Left panel Pareto(a) distributions with $a = 1.5, 2, 3, 5, 10$ from top down. Right panel LogN(μ, σ^2) distributions with the parameters chosen to match $\mathbb{E}[X_1]$ and $\text{ES}_\alpha(X_1)$ with those of Pareto

3 Inhomogeneous Portfolios

In this section suppose $X_i \sim F_i, i = 1, \dots, d$, where the F_i 's are not necessarily the same, and hence $S_d = X_1 + \dots + X_d$ is the aggregate loss of a possibly inhomogeneous portfolio. In order to investigate the key determinants of the DU bounds for VaR and ES of an inhomogeneous portfolio, we sample at random different families and parameters for the marginal distributions. In particular, we choose F_i independently for $i = 1, \dots, d$ from each of the following families with probability $1/3$:

- Gamma, $\Gamma(k, \mu/k)$ with $k \sim 1 + \text{Poisson}(1)$, $\mu \sim \mathcal{E}(1/5)$,
- Lognormal, $\text{LogN}(\mu, \sigma^2)$ with $\mu \sim \mathcal{N}(0.2, 0.4^2)$, $\sigma \sim \Gamma(8, 0.1)$,
- Generalized Pareto, $\text{GPD}(k, v, 0)$ with $k \sim 1/(1.5 + \mathcal{E}(1/5))$, $v \sim \Gamma(4, 0.3)$.

For the latter Pareto-type df from Extreme Value Theory, see [14]. In this manner we sample the marginal distributions of 100 portfolios for each $d = 2, \dots, 15$; in total 1400 portfolios. Then, using only basic properties of these randomly selected marginal distributions and their respective risk quantities, such as $\text{VaR}_\alpha(X_i)$, $\text{MS}_\alpha(X_i)$ and $\text{ES}_\alpha(X_i)$, we will approximate the dependence uncertainty bounds for the aggregated portfolio risk measures.

3.1 Upper Bound on VaR

In this section we focus on approximating the upper bound for $\text{VaR}_\alpha(S_d)$. As the two key drivers, similarly to the homogeneous case, we look at the dimension and the mean-median ratio, properly adapted. Since the marginal distributions can be different, the scale and shape of each margin affects the aggregate VaR. In Fig. 9, the left panel, we observe that the nominal portfolio dimension d is a poor predictor of

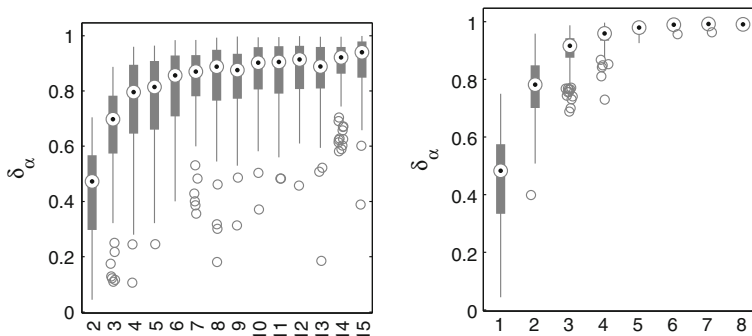


Fig. 9 Boxplots of $\delta_{0.99}(S_d)$ for portfolios with randomly sampled marginal distributions in dependence on the nominal d on the *left panel*, and in dependence on the effective $\tilde{d}$ (rounded) on the *right panel*

$\delta_\alpha(S_d)$. The reason is that the different scales of the marginals have the numerical effect of a dimension reduction for the underlying portfolio; for example, if one marginal has a much larger scale than the others, then the VaR of this marginal gives a good estimate of the aggregate VaR. In this case the dependence uncertainty is small and δ_α is close to 0 (similar to $d = 1$). If the marginals are of a similar scale, then the dependence uncertainty is greater (similar to the homogeneous case d) and δ_α is closer to 1.

Let $\text{VaR}^+(S_d)$, $\text{MS}^+(S_d)$ and $\text{ES}^+(S_d)$ denote $\text{VaR}(S_d)$, $\text{MS}(S_d)$ and $\text{ES}(S_d)$ respectively, when the X_i are comonotonic. Define the *effective dimension* as

$$\tilde{d}(S_d) := \frac{\text{ES}_\alpha^+(S_d) - \text{VaR}_\alpha^+(S_d)}{\max_{i=1,\dots,d} \{\text{ES}_\alpha(X_i) - \text{VaR}_\alpha(X_i)\}}.$$

On the right panel of Fig. 9 we observe that the effective dimension $\tilde{d}$ determines the normalized VaR-bound δ_α better. In order to take into account also the tail-heaviness of the marginal distributions, define the *average mean-median ratio* as

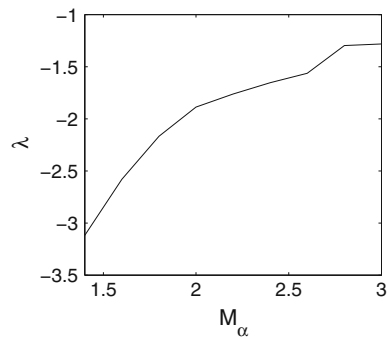
$$\tilde{M}_\alpha(S_d) := \frac{\text{ES}_\alpha^+(S_d) - \text{VaR}_\alpha^+(S_d)}{\text{MS}_\alpha^+(S_d) - \text{VaR}_\alpha^+(S_d)}.$$

We speculate that, similar to the homogeneous case, the rate of convergence depends on the tail-heaviness $\tilde{M}_\alpha(S_d)$, and consider an approximate model

$$\delta_\alpha \approx 1 - \tilde{d}^{\lambda(\tilde{M}_\alpha)}.$$

Rearranging and taking log, we obtain a linear dependence $\log(1 - \delta_\alpha) \approx \lambda \log(\tilde{d})$ for which we estimate λ using linear regression. The estimated parameters are plotted in Fig. 10. As expected, for light tails the convergence is fast, $O(\tilde{d}^{-3})$, and slower for heavy tails. The fitted lines for δ_α for three levels of the average tail-weight $\tilde{M}_\alpha$, as well as the true VaR bound for portfolios with similar tail-weight, are plotted in Fig. 11. We observe that the two key determinants $\tilde{d}$ and $\tilde{M}_\alpha$ provide a reasonable

Fig. 10 The fitted values of $\lambda(\tilde{M}_\alpha)$ obtained by linear regression for different tail-weight levels $\tilde{M}_\alpha$



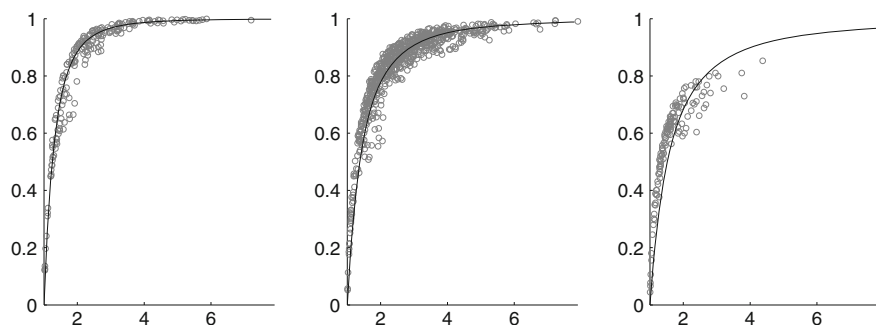


Fig. 11 Fitted curves $\delta_{.99} = 1 - \tilde{d}^\lambda$ for tail-heaviness levels $\tilde{M}_\alpha = 1.4, 1.8, 2.4$, superimposed on scatterplots of the true $\delta_{.99}$ of those sampled portfolios with tail-heaviness within 10 % of these levels.

approximation of $\delta_\alpha(S_d)$ and hence of $\overline{\text{VaR}}_\alpha(S_d)$. Of course, these conclusions are to be understood in the specific context of the example. An out-of-sample test would be required to determine whether this is a good fit in general.

3.2 Lower Bound on VaR

Similar to the homogeneous case (6), an easily calculated but non-sharp bound is

$$\underline{\text{VaR}}_\alpha(S_d) \geq \max\{\text{VaR}_\alpha(X_i), i = 1, \dots, d\} \vee \text{LES}_\alpha^+(S_d). \quad (10)$$

In Fig. 12 the error of this approximation is shown, relative to the length of the possible interval $\underline{\text{VaR}}_\alpha(S_d) \in [\text{LES}_\alpha^+(S_d), \text{VaR}_\alpha^+(S_d)]$. We see that of the 1400 portfolios, only

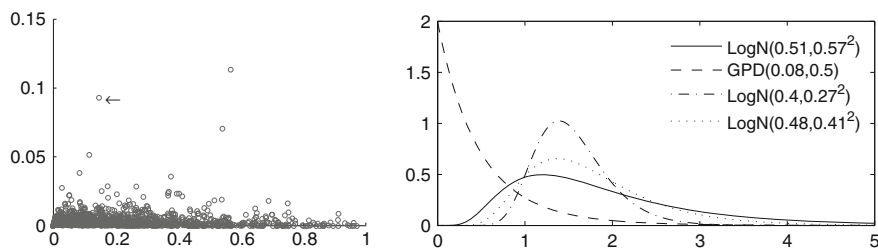


Fig. 12 *Left panel* True $\underline{\text{VaR}}_\alpha(S_d)$ on the horizontal axis (computed using the RA) versus the error of the approximate bound (10) on the vertical axis. For each portfolio the values in $[\text{LES}_\alpha^+(S_d), \text{VaR}_\alpha^+(S_d)]$ are normalized to lie within $[0, 1]$. The arrow indicates an example portfolio, for which the marginal densities are plotted on the *Right panel*. Notice the light left tail of $\text{LogN}(0.4, 0.27^2)$. Here $\underline{\text{VaR}}_\alpha(S_4) = 7.18$, while $\max\{\text{VaR}_\alpha(X_i), i = 1, \dots, 4\} = 6.22$. The possible range is $[\text{LES}_\alpha^+(S_4), \text{VaR}_\alpha^+(S_4)] = [5.67, 16.01]$, hence the error is 9 % of this range

4 have a relative error above 5 %. The marginal densities for one of these portfolios is plotted in the right panel of Fig. 12. Notice the light left tails of the Lognormal margins (especially for $\sigma < 0.5$); these do not have enough mass to compensate for the heaviest right tail, and hence the approximate bound (10) is not sharp. This is in agreement with the results in the homogeneous case; see Fig. 6. Keeping this observation in mind, we can in most cases use (10) as a good approximation for the most favorable aggregate VaR.

3.3 Lower Bound on ES

In the inhomogeneous case a lower bound on the ES is given in [20], based on a similar construction to that of (1). In this case, however, due to the lack of symmetry, one cannot combine the tails of the margins at the same rate; a dynamical weighting needs to be found by solving a system of functional equations. Also, to check the validity and sharpness of the obtained bounds, conditions similar to (A) and (B) would need to be verified. Instead, based on the good performance of the approximate bounds in Sect. 2.3, we consider the following generalization, based on Theorem 4.1 in [6].

Theorem 1 *For rvs $X_i \geq 0$, let $\zeta = \min\{x : \sum_{i=1}^d F_i(x) \geq d - (1 - \alpha)\}$. Then*

$$\underline{\text{ES}}_\alpha(S_d) \geq \frac{1}{1 - \alpha} \sum_{i=1}^d \mathbb{E}[X_i \mathbf{I}_{\{X_i > \zeta\}}]. \quad (11)$$

Furthermore, if $\sum_{i=1}^d F_i(0) \geq d - 1$ and F_i are continuous at ζ , then (11) holds with equality.

Proof Let $S'_d = \sum_{i=1}^d X_i \mathbf{I}_{\{X_i > \zeta\}}$. Clearly $S_d \geq S'_d$, so by monotonicity of ES,

$$\text{ES}_\alpha(S_d) \geq \text{ES}_\alpha(S'_d) \geq \text{ES}_\alpha(S_d^M) = \frac{1}{1 - \alpha} \sum_{i=1}^d \mathbb{E}[X_i \mathbf{I}_{\{X_i > \zeta\}}],$$

where $S_d^M = \sum_{i=1}^d Y_i$ for $Y_i \stackrel{d}{=} X_i \mathbf{I}_{\{X_i > \zeta\}}$ such that $Y_1, \dots, Y_d$ are mutually exclusive and hence $S_d^M \leq_{st} S'_d$ (stop-loss order); see [8]. If $\sum_{i=1}^d F_i(0) \geq d - 1$, then there exist mutually exclusive $X_i \sim F_i$, $i = 1, \dots, d$ and (11) is an equality. $\square$

Also in the inhomogeneous case the RA ([25]) can be applied, with the caveat that the obtained bound may underestimate the expectation in the infinite tail. We computed both bounds at the level $\alpha = 0.99$ for the sampled portfolios, the results are plotted in Fig. 13. First, notice that the bound for higher-dimensional portfolios is closer to $\mathbb{E}[S_d]$. However, for a fixed d , the RA underestimates the bound (due to discretization) when $\underline{\text{ES}}_\alpha(S_d)$ is closer to $\text{ES}_\alpha^+(S_d)$, since this is often due to one or more heavy-tailed margins (e.g. the GPD with parameter $k > 0.5$, which has infinite variance). On the

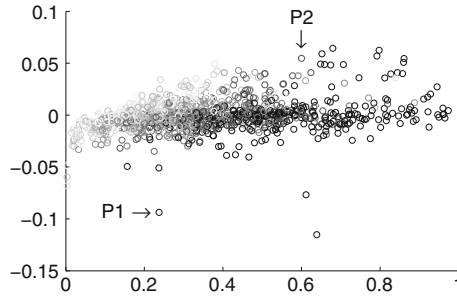
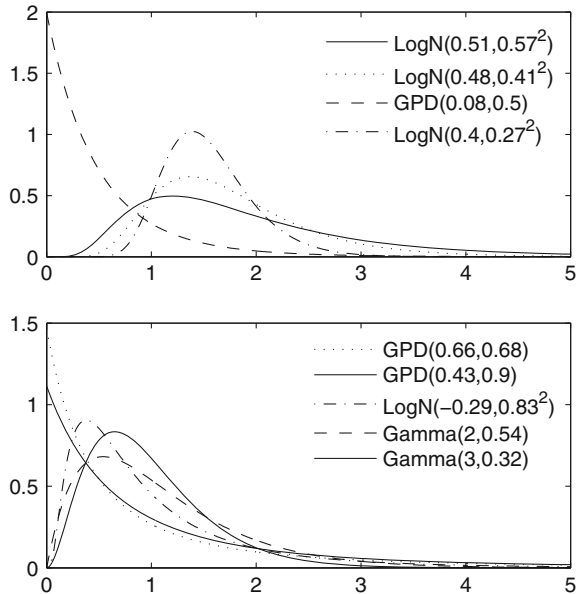


Fig. 13 On the *horizontal axis* $\text{ES}_\alpha(S_d)$ of the sampled portfolios as approximated by the RA (with discretization parameter $N = 10^5$ and stopping condition $\varepsilon = 10^{-4}$). *Vertical axis* the difference between the non-sharp bound (11) and the RA bound. For portfolio labeled P1 the error is negative, so (11) is away from sharpness, while for P2 the error is positive, so the RA bound underestimates the tail integral. The marginal densities for these portfolios are plotted in Fig. 14. The color corresponds to the dimension from $d = 2$ (dark) to $d = 15$ (light). For each portfolio the values in $[\mathbb{E}[S_d], \text{ES}_\alpha^+(S_d)]$ are normalized to lie within $[0, 1]$

other hand, the RA gives a sharper bound than (11) in the presence of marginals with light left tails (e.g. $\text{LogN}(\mu, \sigma^2)$ with $\sigma < 0.5$). Marginal densities for example portfolios in each of these two cases are given in Fig. 14. Out of the 1400 portfolios,

Fig. 14 *Top panel* marginal densities of portfolio P1 from Fig. 13. Notice the light left tails of the Lognormal margins. RA yields $\text{ES}_\alpha(S_4) \geq 8.98$ while (11) $\text{ES}_\alpha(S_4) \geq 7.73$. The possible range is $[\mathbb{E}[S_4], \text{ES}_\alpha^+(S_4)] = [5.81, 19.20]$, hence the difference is -9% of this range. *Bottom panel* marginal densities of portfolio P2. Notice the heavy right tails of the GPD margins. RA yields $\text{ES}_\alpha(S_5) \geq 64.63$ while (11) $\text{ES}_\alpha(S_5) \geq 69.90$. The possible range is $[6.73, 103.19]$, hence the relative difference is 5%



in only 7 cases the RA was worse than (11) by more than 5 % points, and in only 6 cases the RA was better by more than 5 % points (relative to the possible interval $\text{ES}_\alpha(S_d) \in [\mathbb{E}[S_d], \text{ES}_\alpha^+(S_d)]$). Hence, depending on the portfolio, either of the two bounds may give a value closer to the sharp lower bound on ES, but in most cases the difference is small (and both of the bounds are close to sharpness).

4 Conclusions and Further Research

Model uncertainty in general and dependence uncertainty in particular have taken center stage of the regulatory discourse of financial and insurance regulation. In the present paper we have mainly looked at some recent results from the latter, i.e. properties of extremal regulatory risk measures, where we assume full knowledge of the marginal risk distributions, but we do not have, or want, to make assumptions on their interdependence. Whereas numerous analytic results have recently appeared on this issue, there is an increasing need for a better understanding of the underlying numerical problems. This paper provides a first (small) step in this direction. Comparing simple approximate DU bounds for VaR and ES with more advanced ones, we observe that the former often give reasonable estimates of the sharp bounds. Moreover, we identify the cases when the approximate bounds are far from the sharp ones. We also note that the convergence rate of $\bar{\text{VaR}}$ as the portfolio dimension increases is different for small dimensions compared to the theoretical asymptotic rate. The above examples should serve as illustrations for the recent analytic results in the literature, build intuition and motivate further research in the area of dependence uncertainty.

An important direction of research is investigating what partial information on the dependence structure helps to obtain narrower bounds, which under complete dependence uncertainty tend to be very wide. An early approach considers the case when some multivariate marginals are known, and it is found that it can lead to strongly improved bounds; see [13, 16, 31]. Another approach is taken in [5], where the full-dimensional joint density of the assets in portfolio is assumed to be known, but only on a “trusted region”, which is a subset of the support. They find that the bounds quickly deteriorate as the trusted region deviates from the entire support. Bernard et al. [4] argue that variance of the sum of marginals (portfolio variance) is often available, and that this dependence information often leads to substantially narrower VaR bounds. Recently, dynamic factor multivariate GARCH models have been proposed for forecasting financial time series; see [2, 32]. Consequently, a new and relevant approach would be using factor models in order to reduce the DU to a smaller dimension, namely, the number of factors. Recall the observations from Sect. 3, which show that for small-dimensional portfolios the bounds are closer to the comonotonic case (narrower). Gordy [19] even considers the portfolio VaR in a one-factor model, eliminating DU entirely. Of course, any additional assumptions on the

structure should be justified, so it is equally important to understand what additional information on the dependence is typically available in practice. An illuminating case study in the applied setting of economic capital computation for a bank is [1], where the DU bounds are calculated, compared with the value corresponding to a t -copula, and other practical issues discussed.

Acknowledgments The authors would like to thank the Swiss Finance Institute for financial support. They also thank the referee and the editor for valuable comments, and Ruodu Wang and Giovanni Puccetti for numerous discussions on issues related to the broader research area of dependence uncertainty within Quantitative Risk Management.

Appendix

See (Figs. 15, 16, 17, 18 and 19).

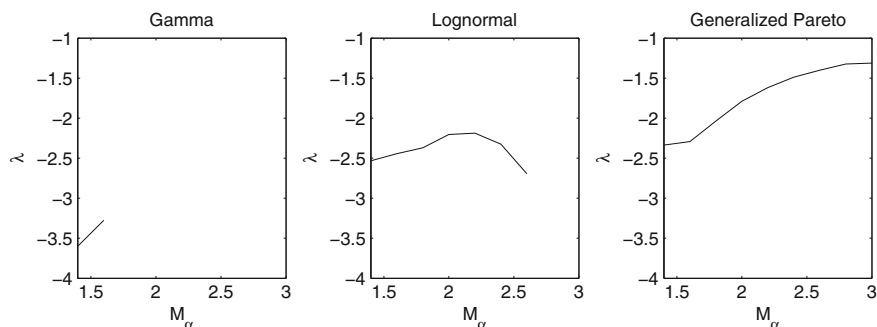


Fig. 15 The fitted values of $\lambda(\tilde{M}_\alpha)$, $\alpha = 0.99$ obtained by regression for different tail-weight levels $\tilde{M}_\alpha$, for three collections of portfolios: either only Gamma, Lognormal, or Generalized Pareto marginal distributions

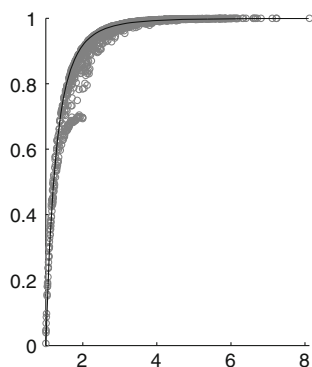


Fig. 16 Fitted curve $\delta_\alpha = 1 - \tilde{d}^\lambda$, $\alpha = 0.99$ for tail-heaviness level $\tilde{M}_\alpha = 1.4$, superimposed on scatterplots of the true $\delta_\alpha(S_d)$ of the sampled portfolios. Portfolios with only Gamma marginal distributions

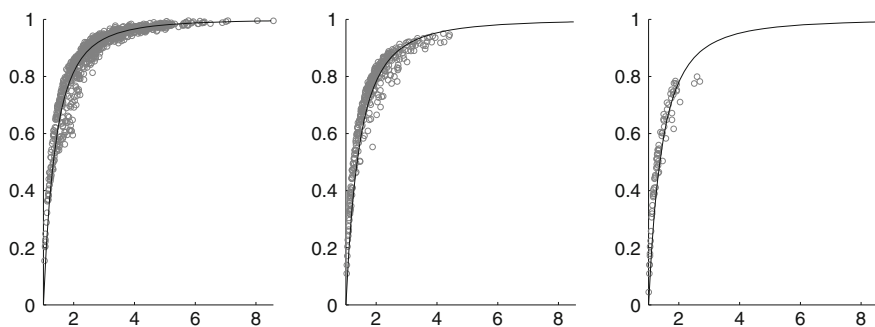


Fig. 17 Fitted curves $\delta_\alpha = 1 - \tilde{d}^\lambda$, $\alpha = 0.99$ for tail-heaviness levels $\tilde{M}_\alpha = 1.6, 2.0, 2.2$, superimposed on scatterplots of the true $\delta_\alpha(S_d)$ of those sampled portfolios with tail-heaviness within 10 % of these levels. Portfolios with only Lognormal marginal distributions

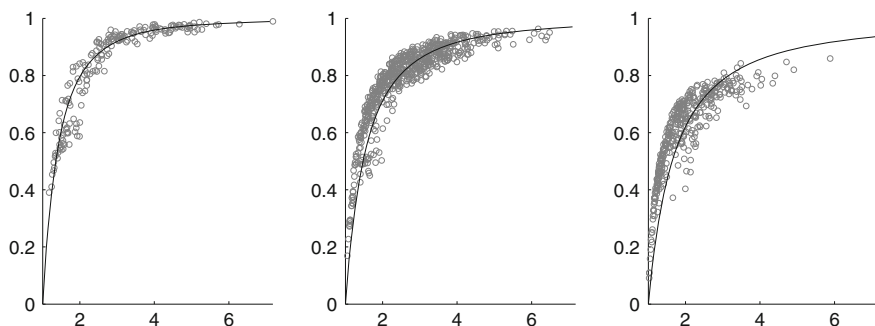


Fig. 18 Fitted curves $\delta_\alpha = 1 - \tilde{d}^\lambda$, $\alpha = 0.99$ for tail-heaviness levels $\tilde{M}_\alpha = 1.6, 2.0, 2.6$, superimposed on scatterplots of the true $\delta_\alpha(S_d)$ of those sampled portfolios with tail-heaviness within 10 % of these levels. Portfolios with only Generalized Pareto marginal distributions

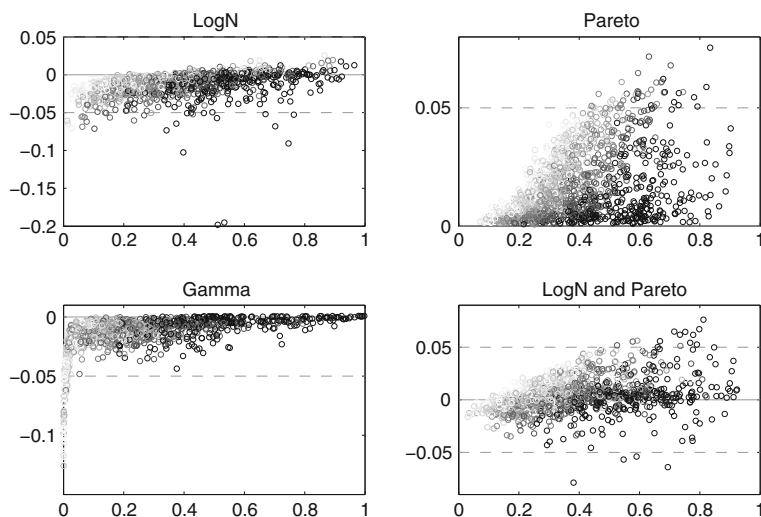


Fig. 19 On the horizontal axis $ES_\alpha(S_d)$ of the sampled portfolios as approximated by the RA (with discretization parameter $N = 10^5$ and stopping condition $\varepsilon = 10^{-4}$). Vertical axis the difference between the non-sharp bound (11) and the RA bound. For each portfolio the values in $[\mathbb{E}[S_d], ES_\alpha^+(S_d)]$ are normalized to lie within $[0, 1]$. The color corresponds to the dimension from $d = 2$ (dark) to $d = 15$ (light). The marginal dfs are sampled from the families given above each plot. For LogN and Gamma distributions, light lower tails are possible, so (11) is away from sharpness and gives a worse bound than the RA. For Pareto, (11) always gives a better bound due to heavy upper tails. For LogN and Pareto mixed, either of the two bounds may be sharper

References

1. Aas, K., Puccetti, G.: Bounds on total economic capital: the DNB case study. *Extremes* **17**(4), 693–715 (2014)
2. Alessi, L., Barigozzi, M., Capasso, M.: Estimation and forecasting in large datasets with conditionally heteroskedastic dynamic common factors. ECB Working Paper No. 1115 (2009)
3. Bernard, C., Jiang, X., Wang, R.: Risk aggregation with dependence uncertainty. *Insur. Math. Econ.* **54**, 93–108 (2014)
4. Bernard, C., Rüschendorf, L., Vanduffel, S.: Value-at-Risk bounds with variance constraints. *J. Risk Insur.* (2015) (to appear)
5. Bernard, C., Vanduffel, S.: A new approach to assessing model risk in high dimensions. *J. Bank. Finance*. **58**, 166–178 (2015)
6. Cheung, K.C., Lo, A.: General lower bounds on convex functionals of aggregate sums. *Insur. Math. Econ.* **53**(3), 884–896 (2013)
7. Christoffersen, P.F., Diebold, F.X.: How relevant is volatility forecasting for financial risk management? *Rev. Econ. Stat.* **82**(1), 12–22 (2000)
8. Dhaene, J., Denuit, M.: The safest dependence structure among risks. *Insur. Math. Econ.* **25**(1), 11–21 (1999)
9. Dhaene, J., Denuit, M., Goovaerts, M.J., Kaas, R., Vyncke, D.: The concept of comonotonicity in actuarial science and finance: theory. *Insur. Math. Econ.* **31**(1), 3–33 (2002)
10. Dias, A., Embrechts, P.: Testing for structural changes in exchange rates' dependence beyond linear correlation. *Euro. J. Finance* **15**(7–8), 619–637 (2009)

11. Embrechts, P., Hofert, M.: A note on generalized inverses. *Math. Methods Oper. Res.* **77**(3), 423–432 (2013)
12. Embrechts, P., Puccetti, G.: Bounds for functions of dependent risks. *Finance Stochast.* **10**(3), 341–352 (2006)
13. Embrechts, P., Puccetti, G.: Bounds for the sum of dependent risks having overlapping marginals. *J. Multivar. Anal.* **101**(1), 177–190 (2010)
14. Embrechts, P., Klüppelberg, C., Mikosch, T.: *Modelling Extremal Events for Insurance and Finance*, vol. 33. Springer (1997)
15. Embrechts, P., Höing, A., Puccetti, G.: Worst VaR scenarios. *Insur. Math. Econ.* **37**(1), 115–134 (2005)
16. Embrechts, P., Puccetti, G., Rüschendorf, L.: Model uncertainty and VaR aggregation. *J. Bank. Finance* **37**(8), 2750–2764 (2013)
17. Embrechts, P., Puccetti, G., Rüschendorf, L., Wang, R., Beleraj, A.: An academic response to Basel 3.5. *Risks* **2**(1), 25–48 (2014)
18. Embrechts, P., Wang, B., Wang, R.: Aggregation-robustness and model uncertainty of regulatory risk measures. *Finance Stochast.* **19**(4), 763–790 (2015)
19. Gordy, M.B.: A risk-factor model foundation for ratings-based bank capital rules. *J. Finan. Intermediation* **12**(3), 199–232 (2003)
20. Jakobsons, E., Han, X., Wang, R.: General convex order on risk aggregation. *Scand. Actuarial J.* (2015) (to appear)
21. Kou, S., Peng, X., Heyde, C.C.: External risk measures and Basel accords. *Math. Oper. Res.* **38**(3), 393–417 (2013)
22. Makarov, G.D.: Estimates for the distribution function of a sum of two random variables when the marginal distributions are fixed. *Theor. Probab. Appl.* **26**(4), 803–806 (1982)
23. McNeil, A.J., Frey, R.: Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. *J. Empir. Finance* **7**(3), 271–300 (2000)
24. McNeil, A.J., Frey, R., Embrechts, P.: *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton University Press (2005)
25. Puccetti, G.: Sharp bounds on the expected shortfall for a sum of dependent random variables. *Stat. Probab. Lett.* **83**(4), 1227–1232 (2013)
26. Puccetti, G., Rüschendorf, L.: Computation of sharp bounds on the distribution of a function of dependent risks. *J. Comput. Appl. Math.* **236**(7), 1833–1840 (2012)
27. Puccetti, G., Rüschendorf, L.: Sharp bounds for sums of dependent risks. *J. Appl. Probab.* **50**(1), 42–53 (2013)
28. Puccetti, G., Rüschendorf, L.: Asymptotic equivalence of conservative value-at-risk-and expected shortfall-based capital charges. *J. Risk* **16**(3), 3–22 (2014)
29. Puccetti, G., Wang, R.: Detecting complete and joint mixability. *J. Comput. Appl. Math.* **280**, 174–187 (2015)
30. Puccetti, G., Wang, B., Wang, R., et al.: Advances in complete mixability. *J. Appl. Probab.* **49**(2), 430–440 (2012)
31. Rüschendorf, L.: Bounds for distributions with multivariate marginals. In: Mosler, K. and Scarsini, M. (eds.) *Stochastic Orders and Decisions*, vol. 19, pp. 285–310. JSTOR. IMS Lecture Notes (1991)
32. Santos, A.A.P., Moura, G.V.: Dynamic factor multivariate GARCH model. *Comput. Stat. Data Anal.* **76**, 606–617 (2014)
33. Wang, B., Wang, R.: The complete mixability and convex minimization problems with monotone marginal densities. *J. Multivar. Anal.* **102**(10), 1344–1360 (2011)
34. Wang, R., Peng, L., Yang, J.: Bounds for the sum of dependent risks and worst Value-at-Risk with monotone marginal densities. *Finance Stochast.* **17**(2), 395–417 (2013)

A Beaufort Scale of Predictability

Mark H.A. Davis

Abstract This paper defines a Beaufort Predictability Index, analogous to the Beaufort Wind Force Scale, intended to assist analysts in sizing up statistical prediction problems, specifically in connection with financial risk management but having wider applicability. For background we outline the career of Francis Beaufort and the genesis of his Wind Force Scale. We then describe the various versions of probability that are relevant to prediction, from axiomatic assignment of probability to Knightian Uncertainty, and discuss the case of weather forecasting and attendant statistical methods. Mandelbrot's 'Seven States of Randomness' are also presented in brief outline. On the basis of this background we present a definition of an 11-point Predictability Index. Examples of application are given.

Keywords Prediction · Financial risk management · Knightian uncertainty · Mandelbrot's states of randomness · Prequential statistics

1 Introduction

While 'the quants' cannot really be held responsible for the financial crisis of 2008,¹ it nevertheless seems likely that a contributory factor was over-reliance on the output from models that, in fact, had very little credibility. A case in point is the CDO (collateralised debt obligation) market, see [3] for complete details. Conceptually, investors in (say) three classes contribute \$100m between them, which is used to buy a portfolio of corporate bonds. The subsequent interest payments are pooled and paid out to the investors in priority order. Class I investors are paid first and get a payment corresponding to a fixed interest rate on their investment. Then Class II investors are paid in the same way but at a higher rate. All residual receipts are paid to the Class III investors. Class I investors have the least risk, since any losses (i.e.

¹Alex Lipton points out that banks that did not fail were using the same models as those that did.

M.H.A. Davis (✉)
Imperial College London, London SW7 2AZ, UK
e-mail: mark.davis@imperial.ac.uk

failure of corporates to pay the interest on their bonds) will be absorbed by Class III investors first and then by Class II; the investors get higher reward for greater risk. The big problem in designing and valuing these structures is CORRELATION [15]. While it is possible to gauge the risk on individual bonds with reasonable accuracy, the only market data relating to dependence are the iTraxx and CDX indices relating to standardised portfolios, which may or may not be relevant in particular cases. In the absence of further information, analysts often use the empirical correlation of the corporates' stock returns as a proxy for 'default correlation', a highly questionable procedure.

All financial pricing and risk management is at base a prediction problem. In the case of option pricing it is mainly a question of predicting volatility, whereas risk management is concerned with 'real-world' forecasting, as are other branches of science and engineering such as weather forecasting, product planning, etc. Some time ago it occurred to me that it might be helpful to have some kind of classification of prediction problems which would help analysts to size up a problem before they dive in trying to do the impossible. The paradigm of classification is of course the Beaufort Wind Force Scale, which enables mariners to assess the 'state' they are in by simple, factual and unambiguous observations. The Beaufort scale and its history are outlined in Sect. 2. I floated the idea of a Beaufort scale of predictability in a talk at Imperial College in 2013, an endeavour which was encouraged at the time by Paul Embrechts. Ole Barndorff-Nielsen also expressed some enthusiasm, so it seems appropriate to enlarge on the idea in this Festschrift. Of course, the project stands or falls according to whether Ole is more or less favourably impressed having read this paper. Curiously, in his invited discussion [2] of Phil Dawid's paper on Prequential Statistics [5], Ole comments that '[a]n obvious reason why the literature on prediction is rather limited is that in statistical practice problems of forecasting are not very commonly met'. I do not believe he would make such a statement today.

A related effort appears in Benoît Mandelbrot's book *Fractals and Scaling in Finance* [16] in which he proposes a classification of randomness:

Instead of seeking immediately to specify a distribution or a process by an analytic formula with a panoply of parameters, one should perhaps first sort out the possible outcomes into a smaller number of distinct discrete categories.

Mandelbrot introduces what he calls the 'seven states of randomness' which we discuss below in Sect. 5. This is actually a classification of probability distributions and does not bring in the data/statistical side as we do here but, as always with this author, the ideas are thought-provoking.

In the next section we outline the career of Francis Beaufort and the genesis of his Wind Force Scale. The 1832 version of the Scale is given in Table 1. Section 3 describes the various versions of probability that are relevant to prediction, from axiomatic assignment of probability to Knightian Uncertainty. Section 4 covers weather forecasting and attendant statistical methods, particularly ideas based on Dawid's theory of Prequential Statistics. Mandelbrot's ideas are presented in brief outline in Sect. 5. With all this background material in hand we are ready in Sect. 6 to make a first attempt at defining a Beaufort Predictability Index.

Table 1 The Beaufort Wind Force Scale, 1832

No.	Description	Beaufort's criterion	
0	Calm	Calm	
1	Light Air	Just sufficient to give steerage way	
2	Light breeze	With which a well-conditioned man of war, under all sail, and 'clean full', would go in smooth water from ...	1–2 knots
3	Gentle breeze		3–4 knots
4	Moderate breeze		5–6 knots
5	Fresh breeze	In which a well-conditioned man of war, under all sail, 'clean full', and close-hauled could just carry ...	royals
6	Strong breeze		Single-reefs and top-gallant sails
7	Moderate gale		Double-reefs, jib, etc.
8	Fresh gale		Triple-reefs, courses, etc.
9	Strong gale		Close-reefs and courses
10	Whole gale	With which she could only bear close-reefed main top-sail and reefed fore-sail	
11	Storm	With which she would be reduced to storm staysails	
12	Hurricane	To which she could show no canvas	

In order to size up a prediction problem, we must consider all its aspects. First, the data. Is it generated by some well-understood physical mechanism? What are the ‘stylized facts’: the state of randomness, degree of non-stationarity, etc? How much data have we got? Secondly, the prediction problem: what exactly are we trying to predict, and what is the prediction horizon—is it long or short in relation to the time-scale of the data? Thirdly, verification: having chosen and applied a prediction algorithm, can we check *ex post* that the answers it produces are correct? To construct an index, these questions are formalised in five criteria $C_1, \dots, C_5$ specified in Table 2 of Sect. 6. A specific prediction problem will be assigned a score taking values $C_i = -1, 0, +1$ on each criterion, where $+1$ is the direction of easy prediction. The Beaufort Predictability Index (BPI) is then defined by

$$\text{BPI} = 5 - \sum_1^5 C_i,$$

so that BPI takes values in the range $0, \dots, 10$, where 0 corresponds to axiomatic assignment of probability and 10 to Knightian uncertainty. Examples of application are given in Table 3. Final remarks are given in the concluding Sect. 7.



Fig. 1 South-Westerly wind at 12 m/s. One whole ‘feather’ represents 5 m/s

2 Sir Francis Beaufort and the Wind Force Scale

Francis Beaufort was born in Ireland in 1774 and went to sea as a cabin boy in the British Navy at the age of 13, very soon acquiring the interest in meteorological observation and record-keeping that stayed with him for the rest of his career. He rapidly rose through the ranks and was given his first command, HMS Woolwich, in 1805; he was directed to conduct a hydrographic study of the Rio de la Plata region of South America. He developed the first versions of his Wind Force Scale, originally for his own use, in the same year, commenting that ‘[...] nothing gives a more indefinite representation of the wind [...] than the previously used expressions like *moderate wind* or *cloudy weather*’.

In 1812 Beaufort was seriously wounded in a skirmish with pirates on the Turkish coast. He never returned to active duty but remained an officer in the Royal Navy, being appointed Hydrographer to the Admiralty in 1829. Table 1 shows the Wind Force Scale in its 1832 incarnation. It will be seen that Wind Forces in the range 0–4 report the speed of the vessel under full sail, while from 5 upwards the criterion is the current sail configuration. The Scale was adopted for mandatory use on every ship of the Royal Navy in 1838. Beaufort was promoted to Rear Admiral in 1846 and was knighted two years later, becoming Sir Francis. He retired from the Navy in 1855 at the age of 81 and died two years later. Entirely self-educated, he had long been recognised as one of the 19th century’s leading hydrographers.

When steam replaced sail the scale had to be reformulated, and nowadays is simply a quantisation of wind speed, measured in metres/second. The relationship between wind speed W and scale number B is $W = 0.836B^{3/2}$. For example $B = 9.5$ when $W = 24.5$ m/s, the lower limit of ‘10 Beaufort’. The scale and the stylized verbal descriptions (Column 2 of Table 1) are used *verbatim* on many weather websites including the excellent Norwegian site² yr.no, which also uses a version of Beaufort’s ‘arrow’ notation to denote wind direction and speed, see Fig. 1.

3 Probabilistic Modelling in Statistics

The use and interpretation of probability has been a matter of debate for centuries. What is no longer a matter for debate is the mathematical formulation of probability in terms of Kolmogorov’s $(\Omega, \mathcal{F}, \mathbb{P})$. One should be aware however that specifying Ω is already a modelling step (is it $C(0, T)$, $D(0, T)$, ...?) while, in relation to

²Visitors to this site are invited to check the 5 languages in which it is available.

prediction, specifying $\mathbb{P}$ is committing oneself to some relationship between the past and the future, the credibility of which will depend on the problem. The connection between the Kolmogorov model and the real world is of course the frequentist interpretation based on the SLLN but this depends on a certain degree of stationarity. In the remainder of this section we summarise some of the many concepts of probability and randomness that play a role in the prediction problem.

3.1 Objective Probability

This refers to the frequentist interpretation and is what Mandelbrot would call *mild randomness*; see Sect. 5. There are three cases.

CASE 1. Probability is fixed a priori by symmetry and independence, examples being coin tossing, roulette etc. Here no statistical modelling is required, and no-one will dispute, for example, the statement that the distribution of heads in the next n coin tosses is $B(n, \frac{1}{2})$.

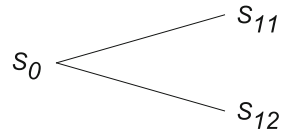
CASE 2. Random events are produced by a well-understood physical phenomenon. A case in point would be Geiger counter detection of radioactive emissions. The physics tells us that the emissions must follow a Poisson process with slowly varying rate, effectively constant over sufficiently short time scales. We need to estimate the rate from data, but having done so the prediction process is clear.

CASE 3. In the first two cases, we have a good theory and don't need much data. Here we have the opposite situation: lots of data but no theory. Many insurance problems fall under this heading, for example car insurance. Accidents are largely independent and occur at rates that do not vary much over time except in predictable ways due to improvements in safety, imposition of speed limits or other changes in the law, and changes in repair costs. In this situation accurate predictions can be made using time series models, which can be monitored and improved as the business develops.

3.2 Subjective Probability

This is probability as quantification of opinion. The decisive advantage is that no justification is required: I just assert that this is my view and act on it, while on the other hand there is no reason *a priori* why anybody else should buy into it. One industry that is entirely run on the basis of subjective probability is the asset allocation side of fund management. The situation here is perfectly explained by Harry Markowitz in his paper *Portfolio Selection* [17], incontestably the most influential paper in the history of investment management:

Fig. 2 Single-stage
Binomial model



The process of selecting a portfolio may be divided into two stages. The first stage starts with observation and experience and ends with beliefs about the future performances of available securities. The second stage starts with the relevant beliefs about future performances and ends with the choice of portfolio.

The second stage may be paraphrased ‘if this is what you believe, here’s what you should do’. Of course, while a firm’s investment managers may be subjectivists, the same freedom cannot be awarded to its risk managers who must take a more objective approach. What they should do is discussed in Sect. 4.

3.3 Risk-Neutral Probability

This is something of a digression, but should be mentioned as mathematical finance is the story of the interplay between ‘ $\mathbb{P}$ -measure’ and ‘ $\mathbb{Q}$ -measure’, the former being used for risk management and the latter for pricing. The key point here is that *$\mathbb{Q}$ -measure is not a probability at all. It is a set of normalised Arrow-Debreu prices.* For an explicit demonstration, consider the single-stage binomial model³ of Fig. 2. The asset price (‘stock’) starts at S_0 at time 0 and moves to one of two values $S_{11} > S_{12}$ at time 1. Each of these prices, but no other price, is regarded as possible. In addition there is a riskless account paying interest rate r , so \$1 deposited at time 0 accrues to $\$R = \$(1 + r)$ at time 1. The no-arbitrage condition is $S_{12} < RS_0 < S_{11}$; if this is violated non-negative, and in at least one state strictly positive, profit is realised by shorting one unit of the stock and investing $\$S_0$ in the riskless account or *vice versa*.

A trading strategy is defined by an investment at time 0 of N units of stock and $\$B$ in the riskless account; its value is $NS_0 + B$. It is easily checked that with $\Delta = S_{11} - S_{12}$ the strategies

$$(N_1, B_1) = \left(\frac{1}{\Delta}, -\frac{S_{12}}{R\Delta} \right), \quad (N_2, B_2) = \left(-\frac{1}{\Delta}, \frac{S_{11}}{R\Delta} \right)$$

realise payoffs (1, 0), (0, 1) respectively in the up and down states, so their values $p_1 = (RS_0 - S_{12})/R\Delta$, $p_2 = (S_{11} - RS_0)/R\Delta$ are the Arrow-Debreu prices of the two time-1 states. Since $p_1 + p_2 = 1/R$ we can define $q_i = Rp_i$, $i = 1, 2$ and then $\mathbb{Q} = (q_1, q_2)$ is a probability distribution on the two states, with the property that the

³The argument extends easily to the multi-stage binomial tree as discussed in all mathematical finance textbooks, for example [20].

value at time 0 of any time-1 payoff X is $\mathbb{E}_{\mathbb{Q}}[X]/R$. Note that this argument does not start with *any* probabilistic model for the prices, only a definition of ‘ Ω ’.

Choice of any corresponding $\mathbb{P}$ -measure is restricted by the Fundamental Theorem of Asset Pricing [6] which says, roughly speaking, that $\mathbb{P}$ and $\mathbb{Q}$ must have the same support, else there will be arbitrage in the model. Specialised to the binomial model here, this just means that any $\mathbb{P}$ -measure must assign strictly positive probabilities, summing to 1, to the two states S_{11} and S_{12} .

3.4 Knightian Uncertainty

This concept originated in Frank H. Knight’s⁴ book *Risk, Uncertainty and Profit* [13], still in print today. Knight set out to solve the following puzzle in economics: why is it that in a competitive market prices are not driven down to the cost of production, leaving no profit at all? The argument is clear: I produce a widget and sell it at \$5.00; my competitor steps in and markets the same thing at \$4.50, grabbing all the market share, so I have to reduce my price to, say, \$4.25. And so on. Knight’s answer is in three parts.

- (i) If we are in a static economy where there is no risk and no technological progress, the prices will indeed be driven down in this way.
- (ii) If there is mild randomness in the form of business risks and random but evolutionary technological progress then the same will be true. Entrepreneurs can hedge against these, leaving themselves in the same position as Case (i).
- (iii) The only way entrepreneurs can consistently make profits is by taking a series of unhedgeable risks—by launching new products, for example. The perfect example is provided by the Apple iPad. When launched, the iPad was not just a new product; it was a new *category* of product. There was nothing to compare it with and no significant amount of market research could be carried out without destroying the essential element of surprise. Apple just had to go for it.

In a word: entrepreneurs are rewarded for taking unquantifiable risks. The same distinction between risk and uncertainty was taken up later by Keynes [12, pp. 213–214], who explains it as follows.

By ‘uncertain’ knowledge, let me explain, I do not mean merely to distinguish what is known for certain from what is only probable. The game of roulette is not subject, in this sense, to uncertainty; nor is the prospect of a Victory bond being drawn. Or, again, the expectation of life is only slightly uncertain. Even the weather is only moderately uncertain. The sense in which I am using the term is that in which the prospect of a European war is uncertain, or the price of copper and the rate of interest twenty years hence, or the obsolescence of a new invention, or the position of private wealthowners in the social system in 1970. About these matters there is no scientific basis on which to form any calculable probability whatever. We simply do not know.

⁴Not to be confused with Frank B. Knight of the Ray-Knight Theorem (see [18, Sect. VI.52]). Frank B. was son of Frank H.

Keynes goes on to outline how we deal with uncertainty from a practical point of view, but his descriptions are straw-man caricatures, there to be demolished, which he proceeds to do. The best modern approach, certainly for long-term financial and investment problems such as pension fund management is *stochastic programming*; the reader will find a general overview plus a range of case studies in the collection [9]. This approach rightly emphasizes *hedging* as the primary means of guarding against uncertainty.

Recently, connections have been drawn between the idea of Knightian uncertainty and the theory of coherent or convex risk measures, see [8, Sect. 4.2]. Specifically, these authors prove the following *robust representation theorem*, in which $\mathcal{X}$ is the set of bounded measurable functions on a measurable space $(\Omega, \mathcal{F})$, and $\mathcal{M}_1^f$ denotes the set of finitely-additive set functions Q on $\mathcal{F}$ such that $Q(\Omega) = 1$.

Theorem [8, Proposition 4.15] A functional $\rho : \mathcal{X} \rightarrow \mathbb{R}$ is a coherent risk measure if and only if there exists $\mathcal{Q} \subset \mathcal{M}_1^f$ such that

$$\rho(X) = \sup_{Q \in \mathcal{Q}} \mathbb{E}_Q[-X], \quad X \in \mathcal{X}.$$

This result is called ‘robust’ because no fixed probability Q is specified; instead we take the ‘worst case’. This is an elegant result with many applications but to this author it is not Knightian uncertainty. The sample space Ω is fixed, meaning that we have already settled on the possible set of events, and then we are choosing some class of possible probability measures $\mathcal{Q}$, unrelated to the data, in a situation where, according to Keynes, ‘there is no scientific basis on which to form any calculable probability whatever’. We leave this as a subject for debate.

4 Weather Forecasting, Prequential Statistics and Consistency

We now turn to the third component in sizing up a prediction problem, namely how we might check that our predictions are correct. Much of the work in this direction originates in meteorology, a major discipline in its own right in which forecasting has always been recognised as one of the primary objectives. The techniques used, which are described in detail in the recent account [22], have developed in a rather different way to those of other applications such as finance, but there is much to learn from them. Weather and climate prediction is based on large-scale computer code solving complex systems of PDEs that model the atmosphere. A large part of the problem is estimating the sensitivity of the results to small errors in the parameters, an endeavour that is now known as Uncertainty Quantification [21]. Weather forecasting is based on the output from these models. One application is probability prediction in which the forecaster predicts the probability of, say, ‘rain tomorrow’. There is no way to say whether such a forecast is correct on any particular day, but we can check

correctness in the long run by monitoring the outputs over time, which is the approach advocated in Dawid's 'prequential' approach to statistics [5]. In particular, the *weak prequential principle* prescribes that evaluation should be based only on observed data and the numerical values of predictions, short-circuiting any consideration of 'model correctness'.

In [4] this process is formalised to give a definition of *consistency* of predictions. In summary, the idea is as follows. We observe a real-valued price series $X(1), \dots, X(n)$ and a vector series of other, auxiliary, data $H(1), \dots, H(n)$ and wish to compute some statistic relating to the conditional distribution of $X(n+1)$ given $\{X(k), H(k), k = 1, \dots, n\}$. A *statistic* of a distribution F is some functional of F such as a quantile or mean value. Let $\mathfrak{s}(F)$ denote the value of this statistic for a candidate distribution function F .

A *model* for the data is a discrete-time stochastic process $(\tilde{X}(k), \tilde{H}(k))$ defined on a canonical sequence space, so that different models correspond to different choices of probability measure $\mathbb{P}$ on this space. We will be considering families $\mathcal{P}$ of probability measures, and use the notation $\mathcal{P} = \{\mathbb{P}^m, m \in \mathfrak{M}\}$, where $\mathfrak{M}$ is an arbitrary index set, to identify different elements $\mathbb{P}^m$ of $\mathcal{P}$.

Now we can define the concept of *consistency* for a statistic $\mathfrak{s}$ relative to a class of models $\mathcal{P}$. Let $\mathcal{I}(\mathcal{P})$ denote the set of strictly increasing predictable processes (b_n) on $(\Omega, (\mathcal{F}_k))$ such that $\lim_{n \rightarrow \infty} b_n = \infty$ a.s. $\forall \mathbb{P}^m \in \mathcal{P}$. Often, b_k will actually be deterministic. A *calibration function* is a measurable function $\ell : \mathbb{R}^2 \rightarrow \mathbb{R}$, chosen so that

$$\mathbb{E}^m[\ell(\tilde{X}(k), \mathfrak{s}(F_k^m)) | \mathcal{F}_{k-1}] = 0 \quad (1)$$

for all $\mathbb{P}^m$ in some class $\mathcal{P}$. This property is equivalent to saying that $Y_k := \ell(\tilde{X}(k), \mathfrak{s}(F_k^m))$ is a martingale difference sequence, whatever the underlying measure $\mathbb{P}^m \in \mathcal{P}$.

Definition 1 A statistic $\mathfrak{s}$ is $(\ell, b, \mathcal{P})$ -consistent, where ℓ is a calibration function, $b \in \mathcal{I}(\mathcal{P})$ and $\mathcal{P}$ is a set of probability measures on $(\Omega, \mathcal{F})$, if

$$\lim_{n \rightarrow \infty} \frac{1}{b_n} \sum_{k=1}^n \ell(\tilde{X}(k), \mathfrak{s}(F_k^m)) = 0 \quad \mathbb{P}\text{-a.s. for all } \mathbb{P} \in \mathcal{P}. \quad (2)$$

As an example, consider the case of quantile forecasting, where $\mathfrak{s}(F) = q_\beta(F) = \{x : F(x-) \leq \beta \text{ and } F(x+) \geq \beta\}$, the set of β -quantiles. Possible choices for ℓ and b are $\ell(x, q) = \mathbf{1}_{(-\infty, q]}(x) - \beta$ and $b_n = n$. Then the n 'th term in (2) measures the difference between β and the average frequency of times the realised value $\tilde{X}(k)$ lies below the quantile q_β^k predicted at time $k-1$ over the time interval $1, \dots, n$. This is the obvious 'back-test' for a quantile estimator. The key point is that the criterion (2) satisfies the weak prequential principle: we are not using anything we do not directly observe. Theorem 5.2 of [4] shows that the β -quantile is $(\ell, b, \mathcal{P})$ -consistent with these choices of ℓ, b if $\mathcal{P}$ is the set of measures such that all conditional distributions F_k^m belong to the set $\mathfrak{F}_c$ of continuous distribution functions.

In a practical application we observe the data sequence $X(1), \dots, X(n-1)$ and produce an estimate $\pi(n)$, based on some algorithm, for what we claim to be $\mathfrak{s}(F_n)$. We evaluate the quality of this prediction by calculating

$$J_n(X, \pi) = \frac{1}{b_n} \sum_{k=1}^n \ell(X(k), \pi(n)).$$

Consistency is a ‘reality check’: it says that if X_i were actually a sample function of some process and we did use the correct predictor $\pi(i) = \mathfrak{s}(F_i)$ then the loss J_n will tend to 0 for large n , and this will be true whatever the model generating $X(i)$, within the class $\mathcal{P}$, so a small value of J_n is evidence that our prediction procedure is well-calibrated. The evidence is strongest when $\mathcal{P}$ is a huge class of distributions and b_n is the slowest-diverging sequence that guarantees convergence in (2) for all $\mathbb{P} \in \mathcal{P}$. Several cases are analysed in detail in [4].

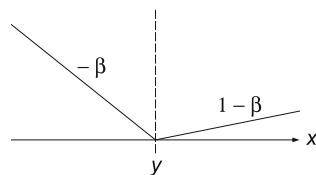
Consistency is closely related to *elicitability*, a concept in decision-theoretic statistics whose origins go back to Savage [19], recent contributors including [10, 14, 23]. It is a familiar fact that if Y is a finite-variance random variable the function $x \mapsto \mathbb{E}[(Y - x)^2]$ is minimised at $x = \mathbb{E}[Y]$, and this is true whatever the distribution of Y within the L_2 class. Elicitability is a generalisation of this idea: given a random variable Y with distribution function F chosen from some class $\mathfrak{F}$, and a statistic $\mathfrak{s}$, can we characterise $\mathfrak{s}(F)$ as a minimizing value for the function $x \mapsto \mathbb{E}[S(x, Y)]$, for all $F \in \mathfrak{F}$, for some *score function* $S(x, y)$? If the answer is yes, $\mathfrak{s}$ is said to be elicitable (within the class $\mathfrak{F}$). In general, a function $S(x, y)$ is a score function if it is measurable and for each y the function $x \mapsto S(x, y)$ is convex, equal to 0 if and only if $x = y$ and smooth at all $x \neq y$. It is possible to prove theorems stating that various statistics $\mathfrak{s}$ either are, or are not, elicitable. Besides the mean value, with score function $(x - y)^2$, another elicitable statistic is the β -quantile, for which a possible score function, shown in Fig. 3, is $S(x, y) = (\mathbf{1}_{x \geq y} - \beta)(x - y)$ with $\mathfrak{F} = L_1$.

When S is C^1 we can characterise minimality by solving

$$\frac{\partial}{\partial x} \mathbb{E}[S(x, Y)] = \mathbb{E} \left[\frac{\partial S}{\partial x}(x, Y) \right] = 0.$$

For the score function of Fig. 3, which is not C^1 , we have $\partial S / \partial x = \mathbf{1}_{x \geq y} - \beta$, so $\mathbb{E}[(\partial S / \partial x)(x, Y)] = 0$ when $F(x) = \beta$, i.e. x is any β -quantile, so long as $F \in \mathfrak{F}_c$. Note that $\mathbf{1}_{x \geq y} - \beta$ is exactly the calibration function ℓ specified in the quantile

Fig. 3 Score function for β -quantile



consistency test above, and this is the general situation: if we have an elicitable statistic $\mathfrak{s}$ with score function S then the corresponding calibration function is $\ell = \partial S / \partial x$. For the connection with martingales, note from (1) that at each time $k - 1$ we are taking the *conditional* expectation, with distribution F_k^m , different at each k , which characterises the martingale difference property while being in line with the elicibility idea that all F_k^m lie in some fixed class $\mathfrak{F}$.

It would be nice to say that there is a one-to-one correspondence between elicibility and consistency using the above relationship, but this is not completely straightforward. We saw above that for quantiles the appropriate class of distributions is L_1 for elicibility but $\mathfrak{F}_c$ for consistency. There is also the role of the normalizing sequence b_n in (2) to be considered. In some cases such as mean estimation b_n has to be a stochastic sequence and things become much less clear-cut. See [4] for further discussion of these points.

5 Mandelbrot's Seven States of Randomness

Mandelbrot introduces this idea in Chapter E5 of his book *Fractals and Scaling in Finance: Discontinuity, Concentration, Risk* [16]. We will only give the briefest of details here since, while the ideas are highly suggestive, we do not make use of Mandelbrot's technical results in this paper. The discussion in Chapter E5—and here—is centred around i.i.d. random variables $X_1, X_2, \dots$ although many of the results extend *mutatis mutandis* to stationary sequences and beyond.

In the i.i.d. case the classification of states refers to the distribution function of X_1 and specifically to its right-hand tail. First, there is a 3-way classification into *mild*, *slow* and *wild* states of randomness which, it is cogently explained, are analogous to the gas, liquid and solid states of matter (in that order). In Sect. 4.4 of the chapter these three states are refined into 2, 2 and 3 sub-categories respectively, giving the advertised seven states, but we will not consider this sub-division further.

The 3-way classification is based on the idea of *portioning*, the relative contributions of the addends X_i to the sum $S_n = \sum_i^n X_i$. Portioning is *even* if all addends play a similar role and this is mild randomness, whereas wildness is the case in which portioning is *concentrated*, meaning that the sum is dominated by a few addends. Slow randomness is an intermediate case in which portioning is concentrated in the short run but even in the long run. Paradigm distributions for the three categories mild, slow wild are the normal, log-normal and stable distributions respectively.

The question of portioning can be studied by examining the ratio of the maximum to the sum, i.e.

$$R_n = \frac{M_n}{S_n}$$

where $M_n = \max\{X_1, \dots, X_n\}$. If the X_i are non-negative then $R_n \in [0, 1]$ and one expects $R_n \sim 1/n$ for large n . Figures 4 and 5 show sample functions of nR_n . In Fig. 4 we have sample functions from the normal and log-normal distributions (both

Fig. 4 $n \times \max/\text{sum}$ for i.i.d. samples. At data point 1000, ordering is log-normal (*highest*), FTSE100 index returns (*middle*), normal (*lowest*)

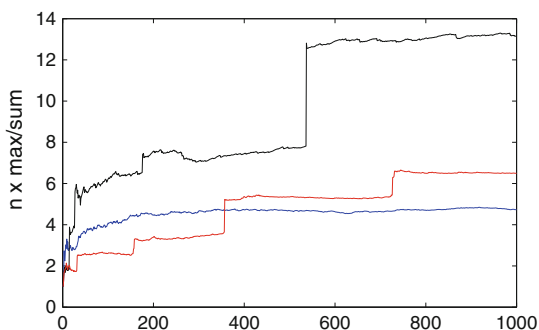
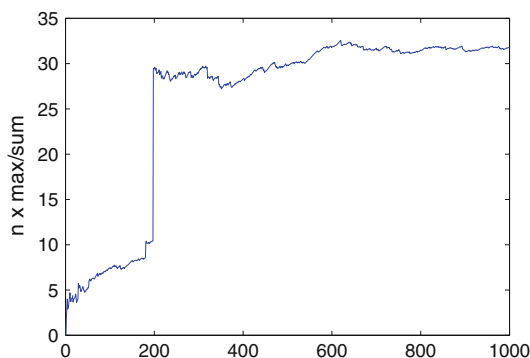


Fig. 5 $n \times \max/\text{sum}$ for i.i.d. samples maximum of 0 and symmetric α -stable with $\alpha = 1.8$



with mean and variance equal to 1) and the third curve is nR_n computed from the weekly returns of the FTSE100 stock index over 1000 weeks in 1992–2013 rescaled to mean=variance=1. The ‘mild’ behaviour of the normal case is evident, as is the fact that individual summands play a non-negligible role in the other two cases. From this admittedly anecdotal evidence it appears that the FTSE returns are no more wild than the log-normal. Figure 5 is based on a sample function of the symmetric α -stable distribution with $\alpha = 1.8$. Huge jumps play a dominating role.

There is a sizeable literature on the asymptotic properties of R_n . A brief summary for i.i.d. sequences will be found in [7, Sect. 6.2.6]. Further papers are [1, 11]. In [1] the authors show that, for stationary strong mixing X_n , if suitably scaled versions $\tilde{S}_n, \tilde{M}_n$ converge individually to $N(0, 1)$ and G respectively where G is an extreme value distribution then (under technical conditions)

$$\lim_{n \rightarrow \infty} \mathbb{E} \left[e^{it\tilde{S}_n} \mathbf{1}_{(\tilde{M}_n \leq x)} \right] = e^{-t^2/2} G(x),$$

so that asymptotically the two terms are independent.

A more difficult problem, as Mandelbrot acknowledges, is to delineate the boundaries of the ‘slow’ classification, as this by definition involves finite sample lengths, not just asymptotic behaviour.

6 The Beaufort Predictability Index (BPI)

Following the above discussion, we now propose a preliminary (‘1805’) formulation of the Predictability Index. It is based on the five criteria $C_1, \dots, C_5$ specified in Table 2; with each criterion is associated a score of +1, 0, −1 (denoted +, 0, − in the table) attained when that criterion is fully, partially, or not at all satisfied. The predictability index, taking values 0, 1, . . . , 10, is defined by

$$\text{BPI} = 5 - \sum_{i=1}^5 C_i.$$

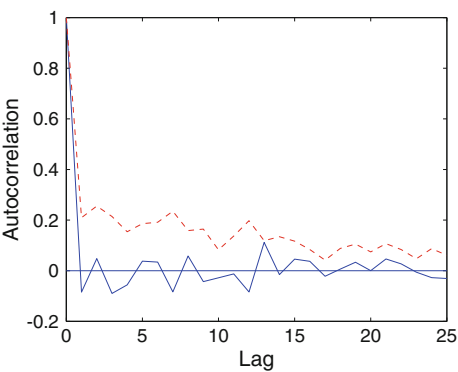
(3)

The criteria are interpreted as follows. C_1 specifies whether there is some physical model describing the production of the data. This may be a simple model, as in the Geiger counter, or a highly complex model, as in weather forecasting. C_2 classifies the state of randomness along the lines proposed by Mandelbrot but, crucially, taking into account what statistic it is that one is trying to predict. For example in the case of financial data, point (i.e. conditional expectation) prediction is Wild but quantile prediction is Intermediate. The conventional autocorrelation function plot, Fig. 6, showing that the ACF of absolute value of returns is much higher than that of the returns themselves, gives a hint why this is the case. C_3 assesses the amount of relevant data available. The score +1 includes the case where, as in roulette, data is not needed, while the score −1 is incurred when some critical piece of data is missing, as in the CDO example of Sect. 1. C_4 is concerned with the question of how far ahead

Table 2 Predictability Criteria

Criterion		Description	
C_1	Physical model	+	Simple model
		0	Complex model
		−	No physical model
C_2	State of Randomness	+	Mild
		0	Intermediate
		−	Wild
C_3	Relevant data	+	Plenty (or not needed)
		0	Limited amount
		−	Key data missing
C_4	Prediction horizon	+	Short (or irrelevant)
		0	Intermediate
		−	Long
C_5	Consistency test	+	Yes (or not needed)
		0	Not reliable
		−	No

Fig. 6 Empirical autocorrelation functions for the FTSE100 return data, for returns (*solid line*) and absolute value of returns (*dashed line*)



we are trying to predict *in relation to the randomness of the data*. For example, in weather forecasting a horizon of a few days is ‘Short’, while anything over two weeks is ‘Long’, whereas for insurance even a time horizon of some years could be Short. Finally, C_5 asks whether our prediction algorithm can survive consistency tests. As we pointed out in Sect. 4, it is quite possible to predict return quantiles in financial data even when that data is not Mild and we have no conventional statistical model. We score +1 in that situation. In assessing criteria $C_3 - C_5$ we must also award a score of +1 when the criterion is irrelevant or not needed (this covers the ‘axiomatic’ case of roulette etc.). Now let us evaluate the scores for the various prediction problems

Table 3 BPI examples

Problem	C_1	C_2	C_3	C_4	C_5	BPI
1. Roulette	+	+	+	+	+	0
2. Geiger Counter						
– Short horizon	+	+	+	+	+	0
– Long horizon	+	+	+	–	+	2
3. Weather forecasting						
– Short term	0	0	+	+	+	2
– Long term	0	–	0	–	–	8
4. Car insurance	–	+	+	0	+	3
5. Life insurance						
– Mortality	0	0	+	–	0	5
– Pensions	0	–	+	–	0	6
6. Finance						
– Point estimates	–	–	–	–	–	10
– Quantiles, short term	–	0	0	+	+	4
– Quantiles, long term	–	0	0	–	0	7
– Credit risk, CDOs	–	0	–	–	–	9

Table 4 Descriptions of BPI categories

BPI	Description
0	Axiomatic probability assignment
1	Accurate physical model with simple calibration
2	Physical model and/or abundant data
3	Reliable statistical or physical model
4	Less satisfactory model but survives consistency tests
5	Problems with data or validity of model
6	Model uncertainty must be explicitly allowed for
7	More serious model uncertainty
8	Severe model uncertainty
9	Borderline Knightian Uncertainty
10	Knightian Uncertainty

considered in this paper. The author's estimates are presented in Table 3. Readers may, and no doubt will, dispute the scores awarded in detail, but the table does point up dramatic differences in the ease of prediction occasioned by deteriorating data quality and increasing horizon in relation to the degree of randomness.

Finally, on the basis of these examples we can arrive at some preliminary classification of the prediction problem in the light of the achieved scores, which we present in Table 4.

7 Concluding Remarks

This is very much a first, or 1805, attempt at constructing a Predictability Index, and no doubt there is plenty of room for improvement. In particular it would be desirable to establish on a scientific basis the relationship between the quality and state of randomness of the data and a reasonable estimate of prediction horizon. A rough and ready method is the obvious one of constructing the empirical autocorrelation function and regarding the horizon as the time beyond which the autocorrelation is essentially zero, but there must be better ways for specific problems. On the other hand the five criteria of Table 2 do seem to be, at the least, factors that any analyst should take into account in sizing up a prediction problem.

Acknowledgments I would like to thank Paul Embrechts, Thomas Mikosch and Martijn Pistorius for helpful advice and comment. Special thanks to Sébastien Lleo for introducing me to Mandelbrot's seven states of randomness.

References

1. Anderson, C., Turkman, K.: The joint limiting distribution of sums and maxima of stationary sequences. *J. Appl. Prob.* **28**, 33–44 (1991). [Addendum, *ibid*, pp. 715–716.]
2. Barndorff-Nielsen, O.E.: Invited discussion to Dawid (1984). *J. R. Statist. Soc. A* **147**, 290–292 (1984)
3. Brigo, D., Pallavicini, A., Torresetti, R.: *Credit Models and the Crisis: A Journey into CDOs, Copulas, Correlations and Dynamic Models*. Wiley (2010)
4. Davis, M.H.A.: Consistency of risk measure estimates (2014). arxiv.org/abs/1410.4382
5. Dawid, A.P.: Present position and potential developments: some personal views. *Statistical theory: the prequential approach* (with discussion). *J. Roy. Statist. Soc. A* **147**, 278–292 (1984)
6. Delbaen, F., Schachermayer, W.: *The Mathematics of Arbitrage*. Springer (2006)
7. Embrechts, P., Klüppelberg, C., Mikosch, T.: *Modelling Extremal Events: for Insurance and Finance*. Springer (1997)
8. Föllmer, H., Schied, A.: *Stochastic Finance*, 3rd edn. De Gruyter, Berlin (2011)
9. Gassmann, H., Ziemba, W. (eds.): *Stochastic Programming: Applications in Finance, Energy, Planning and Logistics*, World Scientific Series in Finance, vol 4. World Scientific (2013)
10. Gneiting, T.: Making and evaluating point forecasts. *J. Amer. Statist. Assoc.* **106**, 746–762 (2011)
11. Ho, H.C., Hsing, T.: On the asymptotic joint distribution of the sum and maximum of stationary normal random variables. *J. Appl. Prob.* **33**(1), 138–145 (1996)
12. Keynes, J.M.: The general theory of employment. *Quart. J. Econ.* **51**(2), 209–223 (1937)
13. Knight, F.H.: *Risk, Uncertainty and Profit*. Houghton Mifflin Co., Boston (1921) [Final 1957 edition reprinted by Dover Publications, Mineola, NY, 2006]
14. Lambert, N., Pennock, D., Shoham, Y.: Eliciting properties of probability distributions. In: *Proceedings of the 9th ACM Conference on Electronic Commerce*, Chicago IL, 2008
15. Lipton, A., Rennie, A. (eds.): *Credit Correlation: Life after Copulas*. World Scientific (2008)
16. Mandelbrot, B.: *Fractals and Scaling in Finance: Discontinuity, Concentration, Risk*. Springer (1997)
17. Markowitz, H.: Portfolio selection. *J. Finance* **7**, 77–91 (1952)
18. Rogers, L.C.G., Williams, D.: *Diffusions, Markov Processes and Martingales*, vol. 2. Cambridge University Press (2000)
19. Savage, L.: Elicitation of personal probabilities and expectations. *J. Amer. Statist. Assoc.* **66**, 783–801 (1971)
20. Shreve, S.: *Stochastic Calculus for Finance I: The Binomial Asset Pricing Model*. Springer (2004)
21. Smith, R.C.: *Uncertainty Quantification: Theory, Implementation, and Applications*. SIAM Publications (2014)
22. Warner, T.T.: *Numerical Weather and Climate Prediction*. Cambridge University Press (2010)
23. Ziegel, J.F.: *Coherence and elicibility*. *Mathematical Finance* (2014) (Published online September 2014)

A Stochastic HJB Equation for Optimal Control of Forward-Backward SDEs

Bernt Øksendal, Agnès Sulem and Tusheng Zhang

Abstract We study optimal stochastic control problems of general coupled systems of forward-backward stochastic differential equations with jumps. By means of the Itô-Ventzell formula, the system is transformed into a controlled backward stochastic partial differential equation. Using a comparison principle for such equations we obtain a general stochastic Hamilton-Jacobi-Bellman (HJB) equation for the value function of the control problem. In the case of Markovian optimal control of jump diffusions, this equation reduces to the classical HJB equation. The results are applied to study risk minimization in financial markets.

Keywords Stochastic control · Forward-backward SDEs · Stochastic HJB equation · Comparison principle · Portfolio optimization · Risk minimization

This research was carried out with support of CAS—Centre for Advanced Study, at the Norwegian Academy of Science and Letters, within the research program SEFE.

The research leading to these results has received funding from the European Research Council under the European Community's Seventh Framework Programme (FP7/2007–2013)/ERC grant agreement no [228087].

B. Øksendal (✉) · A. Sulem

Department of Mathematics, University of Oslo, Blindern, 1053, N 0316 Oslo, Norway
e-mail: oksendal@math.uio.no

B. Øksendal

Norwegian School of Economics, Helleveien 30, N 5045 Bergen, Norway

A. Sulem

INRIA Paris-Rocquencourt, Domaine de Voluceau,
BP 105, 78153 Rocquencourt, Le Chesnay Cedex, France
e-mail: agnes.sulem@inria.fr

A. Sulem

Université Paris-Est, Marne-la-Vallée, 77455 Champs-sur-Marne, France

T. Zhang

School of Mathematics, University of Manchester,
Oxford Road, Manchester M139PL, UK
e-mail: Tusheng.zhang@manchester.ac.uk

1 Introduction

In classical theory of stochastic control of systems described by stochastic differential equations (SDEs) there are two main solution methods:

- (i) Dynamic programming, which leads to a Hamilton-Jacobi-Bellman (HJB) equation. This is a deterministic non-linear partial differential equation (PDE) in the (unknown) value function for the problem.
- (ii) The maximum principle, which involves the optimization of the Hamiltonian and an associated backward stochastic differential equation (BSDE) in the (unknown) adjoint processes, to obtain optimality conditions for the control.

Dynamic programming is a very efficient solution method for Markovian systems. The maximum principle, on the other hand, can also be applied in a non-Markovian setting, but leads to a coupled system of forward-backward SDEs (FBSDEs) with constraints, which is difficult to solve in general.

In view of this it is natural to ask if there is an extension of the HJB approach to non-Markovian systems. The answer has been known to be yes, at least in some cases; we refer in particular to the papers [1, 10] and the references therein where stochastic HJB equations associated to optimal control of non-Markovian SDEs driven by Brownian motions are studied. Note that in these papers the control does not enter the diffusion coefficient. We also mention the paper [2] where path-dependent stochastic control problems and associated path-dependent HJB equations are studied.

The purpose of the present paper is to show that the answer is yes also in the general context of optimal control of non-Markovian systems of *forward-backward* stochastic differential equations (FBSDEs) driven by Itô-Lévy processes. We obtain a *stochastic* HJB equation, which is a backward stochastic partial differential equation (BSPDE). This equation reduces to a classical deterministic HJB equation when the system is a Markovian SDE.

The relation between FBSDEs and BSPDEs has been known for several years. See in particular [5] and the references therein, and [4] for the Markovian case, which leads to backward deterministic PDEs. None of the above papers deal with jumps.

The contribution of our paper lies in the application of the results in [5] to *optimal control* of FBSDEs. However, we do not discuss existence and uniqueness issues for BSPDEs.

In the last part of the paper we illustrate our results by studying some applications to finance, in particular, a risk minimization portfolio problem.

2 Optimal Control of FBSDEs

Consider the following controlled coupled FBSDE: The forward equation in $X(t)$ has the form

$$\begin{cases} dX(t) = \alpha(t, X(t), Y(t), Z(t), K(t, \cdot), u(t, X(t)))dt \\ \quad + \beta(t, X(t), Y(t), Z(t), K(t, \cdot), u(t, X(t)))dB(t) \\ \quad + \int_{\mathbb{R}} \gamma(t, X(t), Y(t), Z(t), K(t, \cdot), u(t, X(t)), \zeta) \tilde{N}(dt, d\zeta); \quad t \in [0, T] \\ X(0) = x \in \mathbb{R}. \end{cases} \quad (1)$$

The backward equation in $Y(t), Z(t), K(t, \zeta)$ has the form

$$\begin{cases} dY(t) = -g(t, X(t), Y(t), Z(t), K(t, \cdot), u(t, X(t)))dt + Z(t)dB(t) \\ \quad + \int_{\mathbb{R}} K(t, \zeta) \tilde{N}(dt, d\zeta); \quad t \in [0, T]; \\ Y(T) = h(X(T)), \end{cases} \quad (2)$$

where $B(t) = B(t, \omega)$ and $\tilde{N}(dt, d\zeta) = N(dt, d\zeta) - \nu(d\zeta)dt$; $t \in [0, T], \omega \in \Omega$, $\zeta \in \mathbb{R}_0 := \mathbb{R} - \{0\}$ are a Brownian motion and an independent compensated Poisson random measure, respectively, on a given filtered probability space $(\Omega, \mathcal{F}, \mathbb{F} := \{\mathcal{F}_t\}_{t \geq 0}, P)$ satisfying the usual conditions. The measure ν is the Lévy measure of the Poisson random measure $N(\cdot, \cdot)$. The given functions

$$\begin{aligned} \alpha(t, x, y, z, k, u, \omega) &: [0, T] \times \mathbb{R}^3 \times \mathcal{R} \times W \times \Omega \rightarrow \mathbb{R} \\ \beta(t, x, y, z, k, u, \omega) &: [0, T] \times \mathbb{R}^3 \times \mathcal{R} \times W \times \Omega \rightarrow \mathbb{R} \\ \gamma(t, x, y, z, k, u, \zeta, \omega) &: [0, T] \times \mathbb{R}^3 \times \mathcal{R} \times W \times \mathbb{R}_0 \times \Omega \rightarrow \mathbb{R} \\ g(t, x, y, z, k, u, \omega) &: [0, T] \times \mathbb{R}^3 \times \mathcal{R} \times W \times \Omega \rightarrow \mathbb{R} \end{aligned}$$

are assumed to be $\mathbb{F}$ -predictable for each x, y, z, k, u .

Here, $\mathcal{R}$ denotes the set of functions $k(\zeta) : \mathbb{R}_0 \rightarrow \mathbb{R}$ and W is a given set of admissible control values $u(t, x, \omega)$, where $u(t) = u(t, X(t), \omega)$ is our control process. The function $h(x, \omega) : \mathbb{R} \times \Omega \rightarrow \mathbb{R}$ is assumed to be $\mathcal{F}_T$ -measurable for each x and C^1 with respect to x for each ω .

We let $\mathcal{A}$ denote a given set of controls, contained in the set of predictable control processes $u(t)$ such that the FBSDE system (1) and (2) has a unique solution $X = X^u \in S^2$, $(Y = Y^u, Z = Z^u, K = K^u) \in S^2 \times H^2 \times H_\nu^2$ where S^2 is the set of real-valued càdlàg adapted processes ϕ with $E(\sup_{0 \leq t \leq T} |\phi(t)|^2) < \infty$; H^2 is the set of real-valued predictable processes ϕ such that $E\left[\int_0^T \phi(t)^2 dt\right] < \infty$, and H_ν^2 is the set of predictable processes ℓ such that $E\left[\int_0^T \int_{\mathbb{R}} |\ell(t, \zeta)|^2 \nu(d\zeta) dt\right] < \infty$.

We refer to [7] for information about stochastic calculus and control for jump diffusions, and to [3] for existence and uniqueness results on FBSDEs.

If $u \in \mathcal{A}$, we call u *admissible*. We want to find $\hat{u} \in \mathcal{A}$ such that

$$\sup_{u \in \mathcal{A}} Y^u(0) = Y^{\hat{u}}(0). \quad (3)$$

First we try to write the solution $Y(t)$ above in the form

$$Y(t) = y(t, X(t)) \quad (4)$$

for some random field $y(t, x) = y(t, x, \omega)$ which, together with $z(t, x)$ and $k(t, x, \zeta)$, satisfies a BSPDE of the form

$$\begin{cases} dy(t, x) = A_u(y(\cdot), z(\cdot), k(\cdot))(t, x)dt + z(t, x)dB(t) \\ \quad + \int_{\mathbb{R}} k(t, x, \zeta) \tilde{N}(dt, d\zeta); \quad t \in [0, T] \\ y(T, x) = h(x), \end{cases} \quad (5)$$

for some partial integro-differential operator A_u acting on x .

Remark 1 We interpret the Eq. (5) in the weak (variational) sense, which means that $y \in L^2([0, T] \times \mathbb{R} \times \Omega)$ and $y(\cdot, \cdot, \omega) \in C([0, T]; L^2(\mathbb{R})) \cap L^2([0, T]; V)$ for all ω , and for all $\phi \in C_0^\infty(\mathbb{R})$; $t \geq 0$,

$$\begin{aligned} \langle y(t, \cdot), \phi \rangle &= \langle y_0(\cdot), \phi \rangle + \int_0^t \langle A_u y(s, \cdot), \phi \rangle ds \\ &+ \int_0^t \langle z(s, \cdot), \phi \rangle dB(s) + \int_0^t \int_{\mathbb{R}} \langle k(s, \cdot, \zeta), \phi \rangle \tilde{N}(ds, d\zeta), \end{aligned} \quad (6)$$

for some $y_0(\cdot) \in V$, where $\langle \cdot, \cdot \rangle$ denotes the dual pairing between the space V and its dual V^* , and $V := W_0^{1,2}(\mathbb{R})$ is the Sobolev space of order one with zero boundary condition at infinity. Note that with this framework the Itô calculus can be applied to (5). See [10, 11].

By the Itô-Ventzell formula (see [9] and the references therein),

$$\begin{aligned} dY(t) &= A_u(y(\cdot), z(\cdot), k(\cdot))(t, X(t))dt + z(t, X(t))dB(t) \\ &+ \int_{\mathbb{R}} k(t, X(t), \zeta) \tilde{N}(dt, d\zeta) + y'(t, X(t))[\alpha(t)dt + \beta(t)dB(t)] \\ &+ \frac{1}{2} y''(t, X(t)) \beta^2(t)dt \\ &+ \int_{\mathbb{R}} \{y(t, X(t) + \gamma(t, \zeta)) - y(t, X(t)) - y'(t, X(t))\gamma(t, \zeta)\} \nu(d\zeta)dt \\ &+ \int_{\mathbb{R}} \{y(t, X(t) + \gamma(t, \zeta)) - y(t, X(t))\} \tilde{N}(dt, d\zeta) + z'(t, X(t))\beta(t)dt \\ &+ \int_{\mathbb{R}} \{k(t, X(t) + \gamma(t, \zeta), \zeta) - k(t, X(t), \zeta)\} \nu(d\zeta)dt \\ &+ \int_{\mathbb{R}} k(t, X(t^-) + \gamma(t, \zeta), \zeta) \tilde{N}(dt, d\zeta), \end{aligned} \quad (7)$$

where $y'(t, x) = \frac{\partial y}{\partial x}(t, x)$ etc. and using the shorthand notation $\alpha(t) = \alpha(t, X(t), Y(t), Z(t), K(t, \cdot), u(t))$ etc. Rearranging the terms we see that

$$\begin{aligned}
 dY(t) = & [A_u(y(\cdot), z(\cdot), k(\cdot))(t, X(t)) + y'(t, X(t))\alpha(t) + \frac{1}{2}y''(t, X(t))\beta^2(t) \\
 & + \int_{\mathbb{R}} \{y(t, X(t) + \gamma(t, \zeta)) - y(t, X(t)) - y'(t, X(t))\gamma(t, \zeta)\}\nu(d\zeta) \\
 & + z'(t, X(t))\beta(t) + \int_{\mathbb{R}} \{k(t, X(t) + \gamma(t, \zeta), \zeta) - k(t, X(t), \zeta)\}\nu(d\zeta)]dt \\
 & + [z(t, X(t)) + y'(t, X(t))\beta(t)]dB(t) \\
 & + \int_{\mathbb{R}} \{y(t, X(t) + \gamma(t, \zeta)) - y(t, X(t)) + k(t, X(t) + \gamma(t, \zeta), \zeta)\}\tilde{N}(dt, d\zeta).
 \end{aligned} \tag{8}$$

Comparing (8) with the original Eq. (2) for Y we deduce the theorem

Theorem 1 Suppose that $(y(t, x), z(t, x), k(t, x, \cdot))$ satisfies the BSPDE

$$dy(t, x) = -A_u(t, x)dt + z(t, x)dB(t) + \int_{\mathbb{R}} k(t, x, \zeta)\tilde{N}(dt, d\zeta); \quad y(T, x) = h(x) \tag{9}$$

where

$$\begin{aligned}
 A_u(t, x) = & A_u(y(\cdot), z(\cdot), k(\cdot))(t, x) \\
 := & g(t, x, y(t, x), z(t, x) + y'(t, x)\beta(t), \\
 & y(t, x + \gamma(t, \cdot)) - y(t, x) + k(t, x + \gamma(t, \cdot), \cdot), u(t, x)) \\
 & + y'(t, x)\alpha(t) + \frac{1}{2}y''(t, x)\beta^2(t) + z'(t, x)\beta(t) \\
 & + \int_{\mathbb{R}} \{y(t, x + \gamma(t, \zeta)) - y(t, x) - y'(t, x)\gamma(t, \zeta)\}\nu(d\zeta) \\
 & + \int_{\mathbb{R}} \{k(t, x + \gamma(t, \zeta), \zeta) - k(t, x, \zeta)\}\nu(d\zeta).
 \end{aligned} \tag{10}$$

Then $(Y(t), Z(t), K(t, \zeta))$, given by

$$Y(t) := y(t, X(t)), \tag{11}$$

$$Z(t) := z(t, X(t)) + y'(t, X(t))\beta(t), \tag{12}$$

$$K(t, \zeta) := y(t, X(t) + \gamma(t, \zeta)) - y(t, X(t)) + k(t, X(t) + \gamma(t, \zeta), \zeta), \tag{13}$$

is a solution of the FBSDE system (1) and (2).

Definition 1 We say that the BSPDE (9) satisfies the comparison principle with respect to u if for all $u_1, u_2 \in \mathcal{A}$ and all $\mathcal{F}_T$ -measurable random variables h_1, h_2 with corresponding solutions $(y_i, z_i, k_i), i = 1, 2$, of (9) such that

$$A_{u_1}(t, x) \leq A_{u_2}(t, x) \text{ for all } t, x \in [0, T] \times \mathbb{R}$$

$$\text{and } h_1(x) \leq h_2(x) \text{ for all } x \in \mathbb{R},$$

we have

$$y_1(t, x) \leq y_2(t, x) \text{ for all } t, x \in [0, T] \times \mathbb{R}.$$

Sufficient conditions for the validity of comparison principles for BSPDEs with jumps is still an open question in this setting. For related results see [6]. However in the Brownian case, sufficient conditions for the validity of comparison principles for BSPDEs of the type (10) are given in Theorem 7.1 in [5]. Using this result we get

Theorem 2 Assume that the following holds:

- $N = K = 0$, i.e. there are no jumps
- The coefficients α, β , and g are $\mathbb{F}$ —progressively measurable for each fixed (x, y, z) and h is $\mathcal{F}_T$ —measurable for each fixed x
- α, β, g, h are uniformly Lipschitz-continuous in (x, y, z)
- α and β are bounded and $E[\int_0^T g^2(t, 0, 0, 0)dt + h^2(0)] < \infty$
- α and β do not depend on z .

Then the comparison principle holds for the BSPDE (9).

From the above we deduce the following result, which may be regarded as a stochastic HJB equation for optimal control of possibly non-Markovian FBSDEs.

Theorem 3 (Stochastic HJB equation.) Suppose the comparison principle holds for the BSPDE (9). Moreover, suppose that for all t, x, ω there exists a maximizer $u = \hat{u}(t, x) = \hat{u}(y, y', y'', z, z', k)(t, x, \omega)$ of the function $u \rightarrow A_u(t, x)$. Suppose the system (9) with $u = \hat{u}$ has a unique solution $(\hat{y}(t, x), \hat{z}(t, x), \hat{k}(t, x, \cdot))$ and that $\hat{u}(t, X(t)) \in \mathcal{A}$. Then $\hat{u}(t, X(t))$ is an optimal control for the problem (3), with optimal value

$$\sup_{u \in \mathcal{A}} Y^u(0) = Y^{\hat{u}}(0) = \hat{y}(0, x). \quad (14)$$

Note that in this general non-Markovian setting the classical value function from the dynamic programming is replaced by the solution $\hat{y}(t, x)$ of the BSPDE (9) for $u = \hat{u}$.

3 Applications

We now illustrate Theorem 3 by looking at some examples.

Example 1 (Maximizing expected utility from terminal wealth) First we consider the Merton problem, the solution of which is well known in the Markovian case with deterministic coefficients. Here we consider the general non-Markovian case, when the coefficients are stochastic processes. Consider a financial market consisting of a risk free investment, with unit price $S_0(t) := 1$; $t \in [0, T]$, and a risky investment, with unit price

$$dS_1(t) = S_1(t)[b(t)dt + \sigma(t)dB(t)]; \quad t \in [0, T]. \quad (15)$$

Here $b(t) = b(t, \omega)$ and $\sigma(t) = \sigma(t, \omega) > 0$ are given adapted processes. Let $u(t, X(t))$ be a *portfolio*, representing the *amount* invested in the risky asset at time t . If u is self-financing, then the corresponding wealth $X(t)$ at time t is given by the stochastic differential equation

$$dX(t) = dX_x^u(t) = u(t, X(t))[b(t)dt + \sigma(t)dB(t)], \quad t \in [0, T]; \quad X(0) = x > 0. \quad (16)$$

Let $(Y(t), Z(t)) = (Y_x^u(t), Z_x^u(t))$ be the solution of the BSDE

$$dY(t) = Z(t)dB(t), \quad t \in [0, T]; \quad Y(T) = U(X(T)), \quad (17)$$

where $U(X) = U(X, \omega)$ is a given utility function, possibly random. Then

$$Y_x^u(0) = E[U(X_x^u(T))].$$

Therefore, the classical portfolio optimization problem of Merton is to find $\hat{u} \in \mathcal{A}$ such that

$$\sup_{u \in \mathcal{A}} Y_x^u(0) = Y_x^{\hat{u}}(0). \quad (18)$$

In the following we assume that

$$\sup_{u \in \mathcal{A}} Y_x^u(0) < \infty. \quad (19)$$

In this general non-Markovian setting with stochastic coefficients $b(t) = b(t, \omega)$ and $\sigma(t) = \sigma(t, \omega) > 0$, an explicit expression for the optimal portfolio $\hat{u}$ is not known. We apply the theory from the previous sections to study this problem. In this case we get, from (10),

$$A_u(t, x) = y'(t, x)ub(t) + \frac{1}{2}y''(t, x)u^2\sigma^2(t, x) + z'(t, x)u\sigma(t) \quad (20)$$

which is maximal when

$$u = \hat{u}(t, x) = -\frac{y'(t, x)b(t) + z'(t, x)\sigma(t)}{y''(t, x)\sigma^2(t)}. \quad (21)$$

Substituting this into $A_{\hat{u}}(t, x)$ we obtain

$$A_{\hat{u}}(t, x) = -\frac{(y'(t, x)b(t) + z'(t, x)\sigma(t))^2}{2y''(t, x)\sigma^2(t)}. \quad (22)$$

Hence the BSPDE for $y(t, x)$ gets the form

$$\begin{cases} dy(t, x) = \frac{(y'(t, x)b(t) + z'(t, x)\sigma(t))^2}{2y''(t, x)\sigma^2(t)}dt + z(t, x)dB(t); & t \in [0, T] \\ y(T, x) = U(x). \end{cases} \quad (23)$$

We have proved:

Proposition 1 *Suppose there exists a solution $(y(t, x), z(t, x))$ of the BSPDE (23) with $y''(t, x) < 0$. Suppose that $\hat{u}$ defined in (21) is admissible. Then $\hat{u}$ is optimal for problem (18) and*

$$y(0, x) = \sup_{u \in \mathcal{A}} Y_x^u(0) = Y_x^{\hat{u}}(0). \quad (24)$$

Note that if b, σ and U are deterministic, we can choose $z(t, x) = 0$ in (23) and this leads to the following (deterministic) PDE for $y(t, x)$:

$$\frac{\partial y}{\partial t}(t, x) - \frac{y'(t, x)^2 b^2(t)}{2y''(t, x)\sigma^2(t)} = 0; \quad t \in [0, T]; \quad y(T, x) = U(x). \quad (25)$$

This is the classical Merton PDE for the value function, usually obtained by dynamic programming and the HJB equation. Hence we may regard (21)–(23) as a generalization of the Merton equation (25) to the non-Markovian case with stochastic $b(t)$, $\sigma(t)$ and $U(x)$. The Markovian case corresponds to the special case when $z(t, x) = 0$ in the BSDE (25). Therefore $\hat{y}(s, x)$ is a stochastic generalization of the value function

$$\varphi(s, x) := \sup_{u \in \mathcal{A}} E[U(X_{s,x}^u(T))] \quad (26)$$

where

$$dX_{s,x}^u(t) = u(t)[b(t)dt + \sigma(t)dB(t)]; \quad t \geq s; \quad X_{s,x}^u(s) = x. \quad (27)$$

Let us compare with the use of the classical HJB:

$$\begin{cases} \frac{\partial \varphi}{\partial s}(s, x) + \max_v \left\{ \frac{1}{2} v^2 \sigma_0^2(s) \varphi''(s, x) + v b_0(s) \varphi'(s, x) \right\} & = 0; \quad s < T \\ \varphi(T, x) & = U(x). \end{cases} \quad (28)$$

The maximum is attained at

$$v = \hat{u}(s, x) = -\frac{b_0(s) \varphi'(s, x)}{\varphi''(s, x) \sigma_0^2(s)}. \quad (29)$$

Substituted into (28) this gives the HJB equation

$$\frac{\partial \varphi}{\partial s}(s, x) - \frac{\varphi'(s, x)^2 b_0^2(s)}{\varphi''(s, x) \sigma_0^2(s)} = 0, \quad (30)$$

which is identical to (25).

Example 2 (Risk minimizing portfolios) Now suppose $X(t) = X_x^u(t)$ is as in (16), while $(Y(t), Z(t)) = (Y_x^u(t), Z_x^u(t))$ is given by the BSDE

$$dY(t) = -\left(-\frac{1}{2} Z^2(t)\right) dt + Z(t) dB(t); \quad Y(T) = X(T). \quad (31)$$

Note that the driver $g(z) := -\frac{1}{2} z^2$ is concave. We want to minimize the *risk* of the terminal financial standing $X(T)$, denoted by $\rho(X(T))$. If we interpret the risk in the sense of the *convex risk measure* defined in terms of the BSDE (31) we have

$$\rho(X(T)) = -Y(0).$$

See e.g. [12, 13] for more information about the representation of risk measures via BSDEs. The risk minimization problem is to find $\hat{u} \in \mathcal{A}$ such that

$$\inf_{u \in \mathcal{A}} -Y_x^u(0) = -Y_x^{\hat{u}}(0), \quad (32)$$

where $Y_x^u(t)$ is given by (31). By changing sign we can consider the supremum problem instead. In this case we get

$$\begin{aligned} A_u(t, x) = & -\frac{1}{2} (z(t, x) + y'(t, x) u \sigma(t))^2 + y'(t, x) u b(t) \\ & + \frac{1}{2} y''(t, x) u^2 \sigma^2(t) + z'(t, x) u \sigma(t), \end{aligned} \quad (33)$$

which is minimal when $u = \hat{u}(t, x)$ satisfies

$$\hat{u}(t, x) = -\frac{z(t, x)y'(t, x)\sigma(t) - y'(t, x)b(t) - z'(t, x)\sigma(t)}{((y'(t, x))^2 - y''(t, x))\sigma^2(t)}. \quad (34)$$

This gives

$$A_{\hat{u}}(t, x) = -\frac{1}{2}\hat{z}^2(t, x) + \frac{(\hat{z}(t, x)\hat{y}'(t, x)\sigma(t) - \hat{y}'(t, x)b(t) - \hat{z}'(t, x)\sigma(t))^2}{2((\hat{y}'(t, x))^2 - \hat{y}''(t, x))\sigma^2(t)}. \quad (35)$$

and hence $(\hat{y}(t, x), \hat{z}(t, x))$ solves the BSPDE

$$d\hat{y}(t, x) = -A_{\hat{u}}(t, x)dt + \hat{z}(t, x)dB(t), \quad 0 \leq t \leq T; \quad \hat{y}(T, x) = x. \quad (36)$$

We have proved:

Proposition 2 *Suppose there exists a solution $(\hat{y}(t, x), \hat{z}(t, x))$ of the BSPDE (36). Suppose $\hat{u}$ defined by (34) belongs to $\mathcal{A}$. Then $\hat{u}$ is optimal for the risk minimizing problem (32), and the minimal risk is*

$$\inf_{u \in \mathcal{A}} -Y_x^u(0) = -Y_x^{\hat{u}}(0) = -\hat{y}(0, x). \quad (37)$$

Next we look at the special case when $b(t)$ and $\sigma(t)$ are *deterministic*. Let us try to choose $\hat{z}(t, x) = 0$ in (36). Then this BSPDE reduces to the PDE

$$\begin{cases} \frac{\partial \hat{y}(t, x)}{\partial t} = -\frac{(\hat{y}'(t, x)b(t))^2}{2((\hat{y}'(t, x))^2 - \hat{y}''(t, x))\sigma^2(t)}; & 0 \leq t \leq T \\ \hat{y}(T, x) = x. \end{cases} \quad (38)$$

We try a solution of the form

$$\hat{y}(t, x) = x + a(t), \quad (39)$$

where $a(t)$ is deterministic. Substituted into (38) this gives

$$a'(t) = -\frac{1}{2} \left(\frac{b(t)}{\sigma(t)} \right)^2, \quad 0 \leq t \leq T; \quad a(T) = 0 \quad (40)$$

which gives

$$a(t) = \int_t^T \frac{1}{2} \left(\frac{b(s)}{\sigma(s)} \right)^2 ds; \quad 0 \leq t \leq T.$$

With this choice of $a(t)$, (38) is satisfied and the minimal risk is

$$\rho_{\min}(X(T)) = -Y^{(\hat{u})}(0) = -\hat{y}(0, x) = -x - \int_0^T \frac{1}{2} \left(\frac{b(s)}{\sigma(s)} \right)^2 ds \quad (41)$$

Hence by (34) the optimal (risk minimizing) portfolio is

$$\hat{u}(t, X(t)) = \frac{b(t)}{\sigma^2(t)}. \quad (42)$$

Remark 2 Note that (41) can be interpreted by means of *entropy* as follows: Recall that in general the *entropy* of a measure Q with respect to the measure P is defined by

$$H(Q \mid P) := E \left[\frac{dQ}{dP} \ln \frac{dQ}{dP} \right].$$

Define

$$\Gamma(t) = \exp \left(- \int_0^t \frac{b(s)}{\sigma(s)} dB(s) - \frac{1}{2} \int_0^t \left(\frac{b(s)}{\sigma(s)} \right)^2 ds \right). \quad (43)$$

By the Itô formula we have

$$\begin{aligned} d(\Gamma(t) \ln \Gamma(t)) &= \Gamma(t) \left[-\frac{b(t)}{\sigma(t)} dB(t) - \frac{1}{2} \left(\frac{b(t)}{\sigma(t)} \right)^2 dt \right] \\ &\quad + (\ln \Gamma(t)) \Gamma(t) \left(-\frac{b(t)}{\sigma(t)} dB(t) \right) + \Gamma(t) \left(-\frac{b(t)}{\sigma(t)} \right) \left(-\frac{b(t)}{\sigma(t)} \right) dt. \end{aligned}$$

Hence, if we define the measure $Q_\Gamma(\omega)$ by

$$dQ_\Gamma(\omega) := \Gamma(T) dP(\omega) \quad (44)$$

we get

$$\begin{aligned} E \left[\frac{dQ_\Gamma}{dP} \ln \frac{dQ_\Gamma}{dP} \right] &= E[\Gamma(T) \ln \Gamma(T)] \\ &= E \left[\int_0^T \Gamma(t) \frac{1}{2} \left(\frac{b(t)}{\sigma(t)} \right)^2 dt \right] = \frac{1}{2} \int_0^T \left(\frac{b(t)}{\sigma(t)} \right)^2 dt, \end{aligned}$$

which proves that (41) can be written

$$\rho_{\min}(X(T)) = -x - H(Q_\Gamma \mid P). \quad (45)$$

Note that Q_Γ is the *unique equivalent martingale measure* for the market (15), when deterministic coefficients are considered.

Thus we have proved that if the coefficients $b(t)$ and $\sigma(t)$ in (16) are deterministic and if the portfolio $\hat{u}(t, X(t)) := \frac{b(t)}{\sigma^2(t)}$ is admissible, then $\hat{u}$ is a risk minimizing portfolio for the problem (32) and the minimal risk is equal to minus the initial wealth x minus the entropy of the equivalent martingale measure. For alternative solution approaches to this problem based on (i) the maximum principle for optimal control of forward-backward SDEs, and on (ii) stochastic differential games, see the survey paper [8].

References

1. Buckdahn, R., Ma, J.: Pathwise stochastic control problems and stochastic HJB equations. *SIAM J. Control Optim.* **45**, 2224–2256 (2007)
2. Ekren, I., Touzi, N., Zhang, J.: Viscosity solutions of fully nonlinear parabolic path dependent PDEs: part I. [arXiv:1210.0006v2](https://arxiv.org/abs/1210.0006v2)
3. Hu, Y., Peng, S.: Solution of forward-backward stochastic differential equations. *Probab. Theory Relat. Fields* **103**, 273–283 (1995)
4. Ma, J., Protter, P., Yong, J.: Solving forward-backward stochastic differential equations explicitly—a four step scheme. *Probab. Theory Relat. Fields* **98**, 339–359 (1994)
5. Ma, J., Yin, H., Zhang, T.: On non-Markovian forward-backward SDEs and backward stochastic PDEs. *Stoch. Process. Appl.* **122**, 3980–4004 (2012)
6. Øksendal, B., Sulem, A., Zhang, T.: A comparison theorem for backward SPDEs with jumps (10 pages). In: Chen, Z.-Q., Jacob, N., Takeda, M., Uemura, T. (eds.) *Festschrift Masatoshi Fukushima*. World Scientific **2015**, 479–487 (2014). [arXiv:1402.4244](https://arxiv.org/abs/1402.4244)
7. Øksendal, B., Sulem, A.: *Applied Stochastic Control of Jump Diffusions*, 2nd edn. Springer (2007)
8. Øksendal, B., Sulem, A.: Risk minimization in financial markets modelled by Itô-Lévy processes. [arXiv:1402.3131](https://arxiv.org/abs/1402.3131) (February 2014). *Afr. Matematika* doi:[10.1007/s13370-014-0248-9](https://doi.org/10.1007/s13370-014-0248-9) (2014)
9. Øksendal, B., Zhang, T.: The Itô-Ventzell formula and forward stochastic differential equations driven by Poisson random measures. *Osaka J. Math.* **44**, 207–230 (2007)
10. Peng, S.: Stochastic Hamilton-Jacobi-Bellman equations. *SIAM J. Control Optim.* **30**, 284–304 (1992)
11. Prévôt, C.I., Röckner, M.: A concise course on stochastic partial differential equations. In: *Lecture Notes in Mathematics 1905*. Springer (2007)
12. Quenez, M.-C., Sulem, A.: BSDEs with jumps, optimization and applications to dynamic risk measures. *Stoch. Process. Appl.* **123**, 3328–3357 (2013)
13. Royer, M.: Backward stochastic differential equations with jumps and related non-linear expectations. *Stoch. Process. Appl.* **116**, 1358–1376 (2006)

CoCos with Extension Risk. A Structural Approach

José Manuel Corcuera, José Fajardo, Wim Schoutens
and Arturo Valdivia

Abstract In this paper we obtain some formulas for pricing contingent convertibles subject to what we call extension risk, i.e., the possibility that the bond issuer does not buy back the bond at pre-specified call dates. We follow a structural approach and we address the finite and infinite maturity cases.

Keywords Contingent capital · Coco bond · Extension risk · Callable bond

1 Introduction

Contingent Convertible (CoCo) Bonds are hybrid debt instruments that have been recently introduced in the light of the new regulatory reforms (CRD 4, Basel III, ...) to enhance the capital levels of financial institutions. The associated loss absorbing capacity of these bonds tries to minimize the future use of tax-payer money in a potential rescue or wind-down of the financial institution when it comes in a life-threatening situation. Upon the appearance of a trigger mechanism, related with the insolvency of the issuer, the *CoCo* is either converted into a predefined number of shares (conversion CoCo) or the face value of the bond is written down partially or in full (write-down CoCo). We focus here on conversion CoCos. Since the evolution of the stock price $(S_t)_{t \geq 0}$ is stochastic, the (recovery) value of the bond after conversion is unknown initially. We note that conversion is mandatory as opposed to convertible

J.M. Corcuera (✉) · A. Valdivia

University of Barcelona, Gran Via de Les Corts Catalanes, 585, 08007 Barcelona, Spain
e-mail: jmcrcuera@ub.edu

A. Valdivia

e-mail: arturo.valdivia@ub.edu

J. Fajardo

EBAPE, Getulio Vargas Foundation, Rio de Janeiro, Brazil
e-mail: jose.fajardo@fgv.br

W. Schoutens

Catholic University of Leuven, Leuven, Belgium
e-mail: wim.schoutens@wis.kuleuven.be

© Springer International Publishing Switzerland 2016

M. Podolskij et al. (eds.), *The Fascination of Probability, Statistics and their Applications*, DOI 10.1007/978-3-319-25826-3_21

bonds, where conversion is a choice that the investor has. In addition, in practice the bonds come often with a regulatory trigger, which gives the local regulator the authority to trigger conversion or write-down whenever he deems this necessary.

The definition of a *CoCo* requires the specification of the following parameters.

K	<u>Face value</u> of the <i>CoCo</i> .
C_p	<u>Conversion price</u> : the prefixed price of the share, for the investor, in case of conversion.
T	<u>Maturity of the CoCo</u> .
$(T_{c_i}, c_i)_{i=1}^m$	<u>Coupon structure</u> : defines the time T_{c_i} at which an amount c_i is paid as coupon, $i = 1, \dots, m$.
τ_c	<u>Conversion time</u> : the random time that defines when the <i>CoCo</i> conversion takes place. In other words, τ_c defines the trigger mechanism.
$T_1, T_2, \dots, T_n$	<u>Call dates</u> : at these dates the issuer of the contract can redeem the bond or he can opt to postpone the payment of the face value until the next call date.

In absence of any call dates and only having a fixed maturity T , the CoCo's final payoff is given by

$$K + \sum_{i=1}^m c_i \exp \left\{ \int_{T_{c_i}}^T r_u du \right\} \mathbf{1}_{\{T_{c_i} < \tau_c\}} + \left(\frac{K}{C_p} S_T - K \right) \mathbf{1}_{\{\tau_c \leq T\}},$$

where $(r_t)_{t \geq 0}$ indicates the short rate process and we implicitly assumed that when a conversion occurred shares then received are also held until maturity and coupon cashflows are invested into the risk-free account. Notice that the CoCo's payoff may be seen as a defaultable bond with face value K , coupons c_i , $i = 1, \dots, m$, and random recovery $\frac{K}{C_p} S_T - K$ in case of "default". These products are receiving an increasing attention in the literature, see [3, 4]. For a comprehensive presentation of these financial contracts see [12]. The differences among the contributions in the literature are how stock dynamics, interest rates and conversion times are modelled. For an intensity approach when modelling the conversion time, see for instance [5, 9], for a structural approach see [3, 4]. In this latter work the authors develop a capital structure model. See also [6], where they consider the reduced and the structural approaches in a very general context.

One of the criticism against CoCos is the possibility of a so-called *death spiral effect*. This latter effect happens when CoCo's holders, facing an undesirable conversion, try to cover their risk by shortselling shares and, as a result, add on the decrease in the share price, which brings the issuing company in even more distress. However in [8], it was indicated that by allowing the possibility of cancellations of the coupons associated to the CoCo bond the death spiral effect was mitigated; see also [10] for another treatment of the same problem.

Bonds and CoCo bonds in particular often come with a series of call dates at which the issuer can redeem earlier the bond. In some cases and in particular in old

style bonds the issuer has an incentive to redeem earlier. For example, if the issuer chooses not to redeem the bond at a call date, a step-up can apply: the original coupon payment is changed and an additional spread is added (the step-up). This step-up can be seen as a penalty to the issuer for not redeeming the bond at the call date. Investors do not get their principal back at the call date but have to wait at least until the next call date, but they are entitled to receive higher coupon payments. These incentives to redeem the bond make that, in practice, T_1 is in most cases assumed to be the actual maturity of the bond and therefore most models have been specified without call dates and just are in terms of a fixed maturity (taken equal to the first call date T_1).

However, under the new regulation (especially the Basel III requirements and the CRD IV directive, see [1]), in order that certain (hybrid) bonds account for Additional Tier 1 capital for the issuing financial institutions, no incentives to redeem are allowed anymore. Therefore, all the new AT1 (CoCo) bonds have no step-up anymore, making extension of the bond much more likely.

Furthermore, it is up to the issuer to decide to redeem the bond at a call date and the issuer will do so only if this is economically reasonable from his perspective. In theory, one could say that the issuer would basically compare the situation of issuing a new bond in line with the going market circumstances with the choice of keeping the existing bond alive. If issuing a new bond would be more expensive in coupon payments, compared with the coupons to be paid on the existing bond, he will not redeem.

In addition, in the past, often reputation risk was an additional reason for the issuing companies to redeem the bond. Issuing institutions did not want to upset their bondholders with an extension of a bond. However, in the last years, this has changed and major institutions like Deutsche Bank (Jan 2009) have made economically reasonable decisions not to redeem outstanding bonds and have not really witnessed reputational risk issues and have tapped the markets without much problem afterwards. In this respect see additional comments in [7, 11–13, 17].

In this perspective, investors look now much closer to the callability features of bonds and the likelihood of an extension has grown significantly. Indeed, on one hand, reputation risk is not that prominent any more, but more importantly for AT1s, since there is no incentive (step-up) any more to redeem, the probability that it makes economically sense not to redeem is much higher. In this paper we incorporate this risk of extension into the price of CoCo bonds. This has been already done by using an intensity approach in [11]; here we follow a structural approach as in [8].

The paper is organized as follows. In the next section, we explain the meaning of the extension risk and its effect on the price of bonds whose original prices, fixing a maturity time and without call dates, are an increasing function of a contingent stock. We elaborate how, when different call dates are introduced, one can view these bonds as a kind of equity structured product in the sense that the payoff of this product will depend just on the value of the stock in reference to some prefixed barriers. Then, in the third section we consider CoCo bonds with cancellable coupons and extension risk. We obtain an expression for their payoff and their price that depends on the value of the stock at the call dates and certain *credit events*. The pricing of CoCo bonds with extension risk is far from being straightforward. Therefore, in the fourth

section we assume that the credit event times are the times when the value of the stock hits some lower barriers and make the setting more specific by assuming the Black-Scholes model, see [2], for the stock price dynamics. Finally in section five, we consider the case of a perpetual CoCo bond.

2 Corporate Bonds with Extension Risk

Let $\pi(t; c, K, T)$ be the price, at time t , of a bond with maturity T and face value K that pays a coupon c at $T_c \leq T$ (we consider a single coupon for simplicity). Next, consider a contract that pays a coupon c_1 at $T_{c_1} \leq T_1$ and at T_1 can either pay back the face value K or postpone the payment of it until T_2 but pays a higher coupon c_2 at T_{c_2} between T_1 and T_2 . We denote its time t price with $\pi(t; c_1, c_2, K, T_1, T_2)$ (where we omit the times of the coupon payments). To postpone or not the payment of the face value depends on which is economically beneficial for the issuer of the contract. In other words at time T_1 the value of the contract is

$$\min\{K, \pi(T_1; c_2, K, T_2)\} = K - (K - \pi(T_1; c_2, K, T_2))_+.$$

T_1 is named the *call date* or *call time*, and the contract can be seen as consisting out of a long position in a corporate bond with a coupon c_1 , face value K and maturity time T_1 , and a short position in a put option with strike K , maturity time T_1 and as underlying a corporate bond with a coupon c_2 (in the period (T_1, T_2)), face value K and maturity time T_2 . We shall call this bond a *bond with extension risk* and we shall denote its price at time t by $\pi(t; c_1, c_2, K, T_1, T_2)$.

We can have a more complex contract with more call dates. For instance, consider now two call dates T_1, T_2 , and a time T_3 that indicates the maturity of the contract. In the period $(0, T_1)$ this bond, pays a coupon c_1 , and K at T_1 , or the latter payment is postponed, depending on whether the condition $K > \pi(T_1; c_2, c_3, K, T_2, T_3)$ is satisfied or not. Then if $K \leq \pi(T_1; c_2, c_3, K, T_2, T_3)$ the contract continues and it pays c_2 in (T_1, T_2) , and K at T_2 if $K > \pi(T_2; c_3, K, T_3)$ otherwise it pays the coupon c_3 and K at T_3 . Now again this contract can be seen as a long position in a bond with coupon c_1 a face value K and maturity time T_1 and a short position in a put option with strike K , maturity time T_1 and as underlying a *bond with extension risk* with call date T_2 and maturity T_3 , face value K and coupons c_2, c_3 . The main question this paper addresses is how to price these contracts.

Suppose that, for any fixed c_2, K and T_2 , $\pi(t; c_2, K, T_2) = f(S_t)$ where f is a continuous non-decreasing function and $(S_t)_{t \geq 0}$ is a positive process with *relative* independent increments, w.r.t. a risk neutral measure $\mathbb{P}^*$, then $\pi(t; c_1, c_2, K, T_1, T_2)$ will be an increasing function of S_t : in fact the payoff at T_1 is a non-decreasing function of S_{T_1} given by

$$\min\{K, \pi(T_1; c_2, K, T_2)\} = \min\{K, f(S_{T_1})\} := g(S_{T_1}),$$

and, using the tilde for indicating discounted prices,

$$\begin{aligned}
 \tilde{\pi}(t; c_1, c_2, K, T_1, T_2) &= c_1 \tilde{\pi}(t, T_{c_1}) + \mathbb{E}^* \left[\frac{g(S_{T_1})}{B_{T_1}} \middle| \mathcal{F}_t \right] \\
 &= c_1 \tilde{\pi}(t, T_{c_1}) + \mathbb{E}^* \left[\frac{g\left(\frac{S_{T_1}}{S_t} x\right)}{B_{T_1}} \right] \bigg|_{x=S_t} \\
 &= c_1 \tilde{\pi}(t, T_{c_1}) + h(t, S_t), \quad 0 \leq t \leq T_1
 \end{aligned}$$

where B_{T_1} is the value of the numéraire (usually the bank account) at T_1 and, to shorten notation, we use $\tilde{\pi}(t, T_{c_1})$ instead of $\tilde{\pi}(t, 0, 1, T_{c_1})$, T_{c_1} is the time where the coupon c_1 is paid and

$$h(t, S_t) := \mathbb{E}^* \left[\frac{g\left(\frac{S_{T_1}}{S_t} x\right)}{B_{T_1}} \right] \bigg|_{x=S_t},$$

that is a non decreasing function of S_t . Hereafter we write $\mathbb{E}^*$, to indicate the expectation with respect to $\mathbb{P}^*$. So, if we consider $N - 1$ call dates $T_1, T_2, \dots, T_{N-1}$ and a maturity time T_N we can write the payoff of this contract as

$$\begin{aligned}
 &c_i \text{ in } (T_{i-1}, T_i) \text{ if } S_{T_0} < M_0, \dots, S_{T_{i-1}} < M_{i-1}, i \geq 1 \\
 &K \text{ at } T_i \text{ if } S_{T_1} < M_1, \dots, S_{T_{i-1}} < M_{i-1}, S_{T_i} > M_i, i \geq 1
 \end{aligned}$$

for certain constants M_i that depend on K and $c_{i+1}, \dots, c_N$ and the price formula $\pi(T_i; c_{i+1}, \dots, c_N, K, T_{i+1}, \dots, T_N)$:

$$M_i = \max \{ S_{T_i} : \pi(T_i; c_{i+1}, \dots, c_N, K, T_{i+1}, \dots, T_N) < K \}, i = 1, \dots, N - 1.$$

and the convention $M_0 = +\infty, M_N = 0$.

Note the above is a backward procedure, in the sense that if we know how to obtain $\pi(T_{N-1}; c_N, K, T_N)$, a corporate bond without extension risk, we can calculate

$$M_{N-1} = \max \{ S_{T_{N-1}} : \pi(T_{N-1}; c_N, K, T_N) < K \}$$

and then can obtain

$$\pi(T_1; c_{N-1}, c_N, K, T_{N-1}, T_N),$$

and from here we can obtain M_{N-2} . Iterating this procedure, we finally obtain M_1 .

3 CoCos with Extension Risk

In this section we treat the case where the corporate bond is a CoCo bond with cancellable coupons, see [8] for the definition and pricing of such bonds. Set $T_1, T_2, \dots, T_N$ for the call times where the issuer can redeem the CoCo bond. Set τ_c for the conversion time and assume here that there are coupons c_{ij} at times T_{ij} : $T_{i-1} < T_{i1}, \dots \leq T_{im} = T_i, i = 1, \dots, N$, with $T_0 = 0$, provided they are not cancelled, in such a way that no credit event happens (this is indicated by $\tau_{ij} > T_{ij}$ where τ_{ij} are certain stopping times) and the running stock price at T_{ij} is above certain value M_{ij} , prefixed in advance, see [8] for the details. Then if we assume that the bond is going to be called at time T_l and we are at time T_n there will be payments at times T_{ij} :

$$c_{ij} \mathbf{1}_{\{\tau_c > T_{ij}, \tau_{ij} > T_{ij}, S_{T_{ij}} > M_{ij}\}}, \quad n+1 \leq i < l, 1 \leq j \leq m,$$

with

$$c_{ij} = c_{i-1,j} e^{s_i(T_i - T_{i-1})}, \quad i \geq 1$$

where $s_i, i \geq 1$ correspond to possible step-ups after extensions of the contract. Finally at T_l we will have the payment of the face value of the bond, K , if there is no conversion before:

$$K \mathbf{1}_{\{\tau_c > T_l\}}$$

and, in case of conversion, at τ_c :

$$\frac{K}{C_p} S_{\tau_c} \mathbf{1}_{\{\tau_c \leq T_l\}}$$

where C_p is the conversion price. Then, the issuer will call the bond, in the most economically favorable situation, in such a way that the discounted arbitrage price of the contract at time T_n will be

$$\tilde{V}_{T_n} = \inf_{\tau \in \mathcal{T}_n} \mathbb{E}^* \left[\tilde{Z}_{\tau,n} \middle| \mathcal{F}_{T_n} \right]$$

where $\mathcal{T}_n$ is the set of stopping times taking values in $\{T_n, T_{n+1}, \dots, T_N\}$ and

$$\tilde{Z}_{\tau,n} = \sum_{i=n+1}^{l:T_l=\tau} \sum_{j=1}^m \tilde{c}_{ij} \mathbf{1}_{\{\tau_{ij} > T_{ij}, S_{T_{ij}} > M_{ij}\}} + K e^{-r\tau} \mathbf{1}_{\{\tau_c > \tau\}} + \frac{K}{C_p} \tilde{S}_{\tau_c} \mathbf{1}_{\{\tau_c \leq \tau\}}.$$

It is known that then $\tilde{V}_{T_n}$ is given by the *initial* value of the *lower*¹ Snell envelope of

¹By a lower Snell envelope of a discrete time $\mathbb{F}$ -adapted process $X := (X_i)_{i=1}^N$, we understand the minus Snell envelope of $-X$. See, for instance, Sect. 2.2 in [14] for the definition of the Snell envelope of a process and its relationship with the problem of optimal stopping.

$$\tilde{X}_{l,n} := \tilde{Z}_{T_l,n} = \sum_{i=n+1}^l \sum_{j=1}^m \tilde{c}_{ij} \mathbf{1}_{\{\tau_{ij} > T_{ij}, S_{T_{ij}} > M_{ij}\}} + K e^{-r T_l} \mathbf{1}_{\{\tau_c > T_l\}} + \frac{K}{C_p} \tilde{S}_{T_l} \mathbf{1}_{\{\tau_c \leq T_l\}},$$

$l = n, \dots, N$ and where we assume that $\tau_c > \tau_{ij}$ on $[0, T_{ij}]$.

Then

$$\tilde{V}_{T_n} = \tilde{Y}_{n,n} = \min \left\{ \tilde{X}_{n,n}, \mathbb{E}^* \left[\tilde{Y}_{n+1,n} \middle| \mathcal{F}_{T_n} \right] \right\},$$

with

$$\tilde{Y}_{N,n} = \tilde{X}_{N,n} = \sum_{i=n+1}^N \sum_{j=1}^m \tilde{c}_{ij} \mathbf{1}_{\{\tau_{ij} > T_{ij}, S_{T_{ij}} > M_{ij}\}} + K e^{-r T_N} \mathbf{1}_{\{\tau_c > T_N\}} + \frac{K}{C_p} \tilde{S}_{T_N} \mathbf{1}_{\{\tau_c \leq T_N\}}.$$

In particular

$$\begin{aligned} \tilde{Y}_{N-1,n} &= \sum_{i=n+1}^{N-1} \sum_{j=1}^m \tilde{c}_{ij} \mathbf{1}_{\{\tau_{ij} > T_{ij}, S_{T_{ij}} > M_{ij}\}} + \frac{K}{C_p} \tilde{S}_{T_{N-1}} \mathbf{1}_{\{\tau_c \leq T_{N-1}\}} \\ &\quad + \mathbf{1}_{\{\tau_c > T_{N-1}\}} \min \left\{ K e^{-r T_{N-1}}, \right. \\ &\quad \left. \mathbb{E}^* \left[\sum_{j=1}^m \tilde{c}_{Nj} \mathbf{1}_{\{\tau_{Nj} > T_{Nj}, S_{T_{Nj}} > M_{Nj}\}} + K e^{-r T_N} \mathbf{1}_{\{\tau_c > T_N\}} + \frac{K}{C_p} \tilde{S}_{T_N} \mathbf{1}_{\{\tau_c \leq T_N\}} \middle| \mathcal{F}_{T_{N-1}} \right] \right\}. \end{aligned}$$

If we write

$$\begin{aligned} &\tilde{\pi}_{N-1,N}(S_{T_{N-1}}) \\ &:= \mathbb{E}^* \left[\sum_{j=1}^m \tilde{c}_{Nj} \mathbf{1}_{\{\tau_{Nj} > T_{Nj}, S_{T_{Nj}} > M_{Nj}\}} + K e^{-r T_N} \mathbf{1}_{\{\tau_c > T_N\}} + \frac{K}{C_p} \tilde{S}_{T_N} \mathbf{1}_{\{\tau_c \leq T_N\}} \middle| \mathcal{F}_{T_{N-1}} \right], \end{aligned}$$

then $\pi_{N-1,N}(S_{T_{N-1}})$ is the price of a CoCo bond with cancellable coupons issued at time T_{N-1} with maturity time T_N , face value K and conversion price C_p . Now, we assume that this is an increasing function of $S_{T_{N-1}}$ (in the cases considered in [8], at that we shall treat later, this is satisfied). Now, if we write

$$M_{N-1} := \inf \{x > 0 : \pi_{N-1,N}(x) \geq K\},$$

we have that

$$\begin{aligned}
\tilde{Y}_{N-1,n} &= \sum_{i=n+1}^{N-1} \sum_{j=1}^m \tilde{c}_{ij} \mathbf{1}_{\{\tau_{ij} > T_{ij}, S_{T_{ij}} > M_{ij}\}} + \frac{K}{C_p} \tilde{S}_{\tau_c} \mathbf{1}_{\{\tau_c \leq T_{N-1}\}} \\
&\quad + K e^{-rT_{N-1}} \mathbf{1}_{\{\tau_c > T_{N-1}, S_{T_{N-1}} \geq M_{N-1}\}} + \tilde{\pi}_{N-1,N}(S_{T_{N-1}}) \mathbf{1}_{\{\tau_c > T_{N-1}, S_{T_{N-1}} < M_{N-1}\}} \\
&= \sum_{i=n+1}^{N-1} \sum_{j=1}^m \tilde{c}_{ij} \mathbf{1}_{\{\tau_{ij} > T_{ij}, S_{T_{ij}} > M_{ij}\}} + \frac{K}{C_p} \tilde{S}_{\tau_c} \mathbf{1}_{\{\tau_c \leq T_{N-1}\}} + K e^{-rT_{N-1}} \mathbf{1}_{\{\tau_c > T_{N-1}, S_{T_{N-1}} \geq M_{N-1}\}} \\
&\quad + \mathbb{E}^* \left[\sum_{j=1}^m \tilde{c}_{Nj} \mathbf{1}_{\{\tau_{Nj} > T_{Nj}, S_{T_{Nj}} > M_{Nj}, S_{T_{N-1}} < M_{N-1}\}} + K e^{-rT_N} \mathbf{1}_{\{\tau_c > T_N, S_{T_{N-1}} < M_{N-1}\}} \right. \\
&\quad \left. + \frac{K}{C_p} \tilde{S}_{\tau_c} \mathbf{1}_{\{S_{T_{N-1}} < M_{N-1}, T_{N-1} < \tau_c \leq T_N\}} \middle| \mathcal{F}_{T_{N-1}} \right].
\end{aligned}$$

We know, by the induction argument in the previous section, that $\tilde{Y}_{N-1,n}$ is an increasing function of $S_{T_{N-1}}$ and we can proceed backwards to obtain M_{N-2} , in fact

$$\begin{aligned}
\tilde{Y}_{N-2,n} &= \min \left\{ \sum_{i=n+1}^{N-2} \sum_{j=1}^m \tilde{c}_{ij} \mathbf{1}_{\{\tau_{ij} > T_{ij}, S_{T_{ij}} > M_{ij}\}} + K e^{-rT_{N-2}} + \frac{K}{C_p} \tilde{S}_{\tau_c} \mathbf{1}_{\{\tau_c \leq T_{N-2}\}}, \mathbb{E}^* \left[\tilde{Y}_{N-1,n} \middle| \mathcal{F}_{T_{N-2}} \right] \right\} \\
&= \sum_{i=n+1}^{N-2} \sum_{j=1}^m \tilde{c}_{ij} \mathbf{1}_{\{\tau_{ij} > T_{ij}, S_{T_{ij}} > M_{ij}\}} + \frac{K}{C_p} \tilde{S}_{\tau_c} \mathbf{1}_{\{\tau_c \leq T_{N-2}\}} \\
&\quad + \mathbf{1}_{\{\tau_c > T_{N-2}\}} \min \left\{ K e^{-rT_{N-2}}, \mathbb{E}^* \left[\tilde{A}_{N-2,N} \middle| \mathcal{F}_{T_{N-2}} \right] \right\}
\end{aligned}$$

where

$$\begin{aligned}
\tilde{A}_{N-2,N} &= \sum_{j=1}^m \tilde{c}_{N-1,j} \mathbf{1}_{\{\tau_{N-1,j} > T_{N-1,j}, S_{T_{N-1,j}} > M_{N-1,j}\}} + \frac{K}{C_p} \tilde{S}_{\tau_c} \mathbf{1}_{\{T_{N-2} < \tau_c \leq T_{N-1}\}} \\
&\quad + K e^{-rT_{N-1}} \mathbf{1}_{\{\tau_c > T_{N-1}, S_{T_{N-1}} \geq M_{N-1}\}} + \sum_{j=1}^m \tilde{c}_{Nj} \mathbf{1}_{\{\tau_{Nj} > T_{Nj}, S_{T_{Nj}} > M_{Nj}, S_{T_{N-1}} < M_{N-1}\}} \\
&\quad + K e^{-rT_N} \mathbf{1}_{\{\tau_c > T_N, S_{T_{N-1}} < M_N\}} + \frac{K}{C_p} \tilde{S}_{\tau_c} \mathbf{1}_{\{S_{T_{N-1}} < M_{N-1}, T_{N-1} < \tau_c \leq T_N\}}.
\end{aligned}$$

Then we have that the payoff of this contract can be written as

$$\begin{aligned}
&c_{ij} \mathbf{1}_{\{\tau_{ij} > T_{ij}, S_{T_{ij}} > M_{ij}, S_{T_0} < M_0, \dots, S_{T_{i-1}} < M_{i-1}\}} \text{ at times } T_{ij}, i \geq 1, \\
&K \mathbf{1}_{\{\tau_c > T_i, S_{T_0} < M_0, \dots, S_{T_{i-1}} < M_{i-1}, S_{T_i} > M_i\}} \text{ at } T_i, i \geq 1, \\
&\frac{K}{C_p} S_{\tau_c} \mathbf{1}_{\{\lceil \tau_c \rceil \leq T_N, S_{T_0} < M_0, \dots, S_{\lceil \tau_c \rceil - 1} < M_{\lceil \tau_c \rceil - 1}\}} \text{ at } \tau_c.
\end{aligned}$$

where $\lceil \tau_c \rceil$ is the element of $\{T_1, T_2, \dots, T_N\}$ such that $\lceil \tau_c \rceil - 1 < \tau_c \leq \lceil \tau_c \rceil$, and, by convention, $M_0 = \infty$ and $M_N = 0$.

So, on $\{t < \tau_c \wedge \tau\}$, where τ is the *call time*,

$$\begin{aligned}
\tilde{V}_t = & \sum_{i,j:T_{ij}>t} \tilde{c}_{ij} \mathbb{P}^* \left(\tau_{ij} > T_{ij}, S_{T_{ij}} > M_{ij}, S_{T_0} < M_0, \dots, S_{T_{i-1}} < M_{i-1} \mid \mathcal{F}_t \right) \\
& + \sum_{i:T_i>t} K e^{-r(T_i-t)} \mathbb{P}^* \left(\tau_c > T_i, S_{T_0} < M_0, \dots, S_{T_{i-1}} < M_{i-1}, S_{T_i} > M_i \mid \mathcal{F}_t \right) \\
& + \frac{K}{C_p} \mathbb{E}^* \left[\tilde{S}_{\tau_c} \mathbf{1}_{[\tau_c] \leq T_N, S_{T_0} < M_0, \dots, S_{[\tau_c]-1} < M_{[\tau_c]-1}} \mid \mathcal{F}_t \right]. \tag{1}
\end{aligned}$$

4 Structural Approach to Pricing

Let us assume that all credit events are defined by the movements of $(S_t)_{t \geq 0}$ with respect to some critical barriers, in such a way that

$$\tau_c = \inf\{t \geq 0 : S_t \leq S^*\} \quad \text{and} \quad \tau_{ij} = \inf\{t \geq 0 : S_t \leq S_{ij}^*\}, \tag{2}$$

for a series of given constants S^* and $\{S_{ij}^*, i = 1, \dots, N, j = 1, \dots, m\}$, and with the usual convention $\inf \emptyset = \infty$.

The following simple proposition give us a useful characterization of the discounted price of a CoCo with extension risk, in terms of the *share measure* $\mathbb{P}^{(S)}$. Recall that $\mathbb{P}^{(S)}$ is the probability measure, equivalent to $\mathbb{P}^*$, obtained by taking the share price as numéraire; its Radon-Nikodým derivative is thus given by

$$\frac{d\mathbb{P}^{(S)}}{d\mathbb{P}^*} = \frac{e^{-\int_0^T r_t dt} S_T}{S_0}.$$

Proposition 1 *The discounted price of a CoCo with extension risk is given by*

$$\begin{aligned}
\tilde{V}_t = & \sum_{i,j:T_{ij}>t}^{N,m} \tilde{c}_{ij} \\
& \times \mathbb{P}^* \left(\inf_{0 \leq s \leq T_{ij}} U_s > \log \frac{S_{ij}^*}{S^*}, U_{T_{ij}} > \log \frac{M_{ij}}{S^*}, U_{T_0} < \log \frac{M_0}{S^*}, \dots, U_{T_{i-1}} < \log \frac{M_{i-1}}{S^*} \mid \mathcal{F}_t \right) \\
& + \sum_{i:T_i>t}^{N,m} K e^{-r(T_i-t)} \\
& \times \mathbb{P}^* \left(\inf_{0 \leq s \leq T_i} U_s > 0, U_{T_0} < \log \frac{M_0}{S^*}, \dots, U_{T_{i-1}} < \log \frac{M_{i-1}}{S^*}, U_{T_i} > \log \frac{M_i}{S^*} \mid \mathcal{F}_t \right) \\
& + \frac{K \tilde{S}_t}{C_p} \sum_{i=1}^N \mathbb{P}^{(S)} \left(\inf_{0 \leq s \leq T_i} U_s > 0, U_{T_0} < \log \frac{M_0}{S^*}, \dots, U_{T_{i-1}} < \log \frac{M_{i-1}}{S^*}, U_{T_i} \geq \log \frac{M_i}{S^*} \mid \mathcal{F}_t \right), \tag{3}
\end{aligned}$$

where $(U_t := \log \frac{S_t}{S^*})_{t \geq 0}$.

Proof Taking into account (1), it suffices to rewrite its third term by means of the identity

$$\begin{aligned} & \{\lceil \tau_c \rceil \leq T_N, S_{T_0} < M_0, \dots, S_{\lceil \tau_c \rceil - 1} < M_{\lceil \tau_c \rceil - 1}\} \\ &= \uplus_{i=1}^N \{\tau_c \leq T_i, S_{T_0} < M_0, \dots, S_{T_{i-1}} < M_{i-1}, S_{T_i} \geq M_i\}, \end{aligned}$$

along with the share measure $\mathbb{P}^{(S)}$.

This result shows that the computation of the price in (3) boils down to computing the joint distribution of $(\inf_{0 \leq s \leq T_n} U_s, U_{T_1}, \dots, U_{T_n})$. In what follows we work out a simple model.

4.1 A Simple Model

Let us assume that the share price $(S_t)_{t \geq 0}$ is given as in the Black-Scholes model (see [2]), i.e.,

$$dS_t = S_t(rdt + \sigma dW_t^*),$$

where $(W_t^*)_{t \geq 0}$ follows a $\mathbb{P}^*$ -Brownian motion. Further, let us assume that the critical barriers appearing in (2) are equal, that is, $S_{ij}^* = S^*$. Under these assumptions, the process $(U_t)_{t \geq 0}$ defined in Proposition 1 satisfies

$$U_t = \log \frac{S_t}{S^*} = \log \frac{S_0}{S^*} + \left(r - \frac{1}{2}\sigma^2\right)t + \sigma W_t^*, \quad t \geq 0.$$

Further, an application of the Girsanov theorem tells us that the dynamics of $(U_t)_{t \geq 0}$ under $\mathbb{P}^{(S)}$ is given by

$$U_t = \log \frac{S_0}{S^*} + \left(r + \frac{1}{2}\sigma^2\right)t + \sigma W_t^{(S)}, \quad t \geq 0.$$

4.1.1 Joint Distribution of $(\inf_{0 \leq s \leq T_n} U_s, U_{T_1}, \dots, U_{T_n})$ in a Black-Scholes Setting

Let $(W_t)_{t \geq 0}$ be a Brownian motion with respect to a probability measure $\mathbb{P}$, and denote its natural filtration by $\mathbb{F} := (\mathcal{F}_t)_{t \geq 0}$. Let $(U_t := u_0 + \mu t + \sigma W_t)_{t \geq 0}$, $u_0 > 0$, and consider its first-passage time to the level zero, i.e.,

$$\tau := \inf\{t \geq 0 : U_t = 0\},$$

with the usual convention $\inf \emptyset = \infty$.

The joint distribution of (τ, U_{T_1}) is well-known (see, for instance, Corollary B.3.3 in [15]) and, in particular, it allows to get, for any $a_1 \geq 0$, $T_1 > t$ and on the set $\{\tau > t\}$,

$$\begin{aligned}
& \mathbb{P}(\tau \geq T_1, U_{T_1} \geq a_1 \mid \mathcal{F}_t) \\
&= \mathbb{P}(U_{T_1} \geq a_1 \mid \mathcal{F}_t) - e^{-2\mu\sigma^{-2}U_t} \mathbb{P}(U_{T_1} \leq -a_1 + 2\mu(T_1 - t) \mid \mathcal{F}_t) \\
&= \Phi\left(\frac{-a_1 + U_t + \mu(T_1 - t)}{\sigma\sqrt{T_1 - t}}\right) - e^{-2\mu\sigma^{-2}U_t} \Phi\left(\frac{-a_1 - U_t + \mu(T_1 - t)}{\sigma\sqrt{T_1 - t}}\right).
\end{aligned} \tag{4}$$

where Φ stands for the standard normal distribution. Below we extend this result in order to obtain an explicit expression for the joint finite-dimensional distributions of a drifted Brownian motion in different times and its infimum. The results are based on Eq. (4), an iterative procedure and the following lemma.

Lemma 1 Consider the processes $(U_t)_{t \geq 0}$ and $(\eta_t^j)_{t \geq 0}$, $j = 1, \dots, n$, given by

$$\begin{aligned}
U_t &:= u_0 + \int_0^t \mu(s) ds + \int_0^t \sigma(s) dW_s, \quad t \geq 0, \\
\eta_t^j &:= \exp \left\{ \int_0^{t \wedge T_j} \theta(s) dW_s - \frac{1}{2} \int_0^{t \wedge T_j} \theta^2(s) ds \right\}, \quad t \geq 0,
\end{aligned}$$

where $\mu(s)$ and $\theta(s)$ are deterministic and càdlàg functions. For every function F (satisfying suitable integrability conditions) and every $0 \leq t \leq T_1 \leq T_n$ the following equivalence holds true

$$\begin{aligned}
& \mathbb{E} \left[e^{\int_0^{T_j} \theta(s) dW_s - \frac{1}{2} \int_0^{T_j} \theta^2(s) ds} F(U_{T_1}, \dots, U_{T_n}) \mid \mathcal{F}_t \right] \\
&= \eta_t^j \mathbb{E} \left[F \left(U_{T_1} + \int_t^{T_1 \wedge T_j} \sigma(s) \theta(s) ds, \dots, U_{T_n} + \int_t^{T_n \wedge T_j} \sigma(s) \theta(s) ds \right) \mid \mathcal{F}_t \right].
\end{aligned}$$

In particular, for constant coefficients μ, σ we have

$$\begin{aligned}
& \mathbb{E} \left[e^{-2\mu\sigma^{-2}U_{T_j}} F(U_{T_1}, \dots, U_{T_n}) \mid \mathcal{F}_t \right] \\
&= e^{-2\mu\sigma^{-2}U_t} \mathbb{E} \left[F(U_{T_1} - 2\mu(T_1 \wedge T_j - t), \dots, U_{T_n} - 2\mu(T_n \wedge T_j - t)) \mid \mathcal{F}_t \right].
\end{aligned}$$

Proof Consider the change of measure given by

$$\left. \frac{d\mathbb{P}^{\theta, j}}{d\mathbb{P}} \right|_{\mathcal{F}_t} = \exp \left\{ \int_0^{t \wedge T_j} \theta(s) dW_s - \frac{1}{2} \int_0^{t \wedge T_j} \theta^2(s) ds \right\} = \eta_t^j.$$

Applying the abstract Bayes' rule we get

$$\begin{aligned} \mathbb{E} \left[\exp \left\{ \int_0^{T_j} \theta(s) dW_s - \frac{1}{2} \int_0^{T_j} \theta^2(s) ds \right\} F(U_{T_1}, \dots, U_{T_n}) \middle| \mathcal{F}_t \right] \\ = \eta_t^j \mathbb{E}_{\mathbb{P}^{\theta,j}} [F(U_{T_1}, \dots, U_{T_n}) | \mathcal{F}_t]. \end{aligned}$$

Now, let $(W_t^{\theta,j})_{t \geq 0}$ be the correspondent $\mathbb{P}^{\theta,j}$ -Brownian motion given by the Girsanov theorem, i.e.,

$$W_t^{\theta,j} := W_t - \int_0^{t \wedge T_j} \theta(s) ds, \quad t \geq 0,$$

and notice that

$$\begin{aligned} U_t &= u_0 + \int_0^t \mu(s) ds + \int_0^t \sigma(s) dW_s \\ &= u_0 + \int_0^t \mu(s) ds + \int_0^t \sigma(s) dW_s^{\theta,j} + \int_0^{t \wedge T_j} \sigma(s) \theta(s) ds \\ &:= \tilde{U}_t + \int_0^{t \wedge T_j} \sigma(s) \theta(s) ds. \end{aligned}$$

Thus

$$\begin{aligned} \mathbb{E}_{\mathbb{P}^{\theta,j}} [F(U_{T_1}, \dots, U_{T_n}) | \mathcal{F}_t] \\ &= \mathbb{E}_{\mathbb{P}^{\theta,j}} [F(U_{T_1} - U_t + x, \dots, U_{T_n} - U_t + x) | x=U_t] \\ &= \mathbb{E}_{\mathbb{P}^{\theta,j}} \left[F(\tilde{U}_{T_1} - \tilde{U}_t + \int_t^{T_1 \wedge T_j} \sigma(s) \theta(s) ds + x, \dots, \tilde{U}_{T_n} - \tilde{U}_t + \int_t^{T_n \wedge T_j} \sigma(s) \theta(s) ds + x) \middle| x=U_t \right] \\ &= \mathbb{E} \left[F(U_{T_1} - U_t + \int_t^{T_1 \wedge T_j} \sigma(s) \theta(s) ds + x, \dots, U_{T_n} - U_t + \int_t^{T_n \wedge T_j} \sigma(s) \theta(s) ds + x) \middle| x=U_t \right] \\ &= \mathbb{E} \left[F \left(U_{T_1} + \int_t^{T_1 \wedge T_j} \sigma(s) \theta(s) ds, \dots, U_{T_n} + \int_t^{T_n \wedge T_j} \sigma(s) \theta(s) ds \right) \middle| \mathcal{F}_t \right]. \end{aligned}$$

For the particular case take simply $\theta(t) := -2\mu\sigma^{-1}$.

Proposition 2 On $\{\tau > t\}$, for $a_1 \geq 0$ and $B_i \in \mathcal{B}(\mathbb{R})$, $i = 2, \dots, n$,

$$\begin{aligned} \mathbb{P}(\tau \geq T_1, U_{T_1} < a_1, U_{T_2} \in B_2, \dots, U_{T_n} \in B_n | \mathcal{F}_t) \\ &= \mathbb{P}(0 < U_{T_1} < a_1, U_{T_2} \in B_2, \dots, U_{T_n} \in B_n | \mathcal{F}_t) \\ &\quad - e^{-2\mu\sigma^{-2}U_t} \mathbb{P}(-a_1 < \bar{U}_{T_1} < 0, \bar{U}_{T_2} \in -B_2, \dots, \bar{U}_{T_n} \in -B_n | \mathcal{F}_t), \end{aligned}$$

with $\bar{U}_s = U_s - 2\mu(s - t)$, $s \geq t$.

Proof Take $x \geq 0$ and set

$$\begin{aligned}\vartheta_t(x) &:= \frac{d\mathbb{P}(\tau \geq T_1, U_{T_1} \leq x | \mathcal{F}_t)}{dx} \\ &= c\phi\left(\frac{-x + U_t + \mu(T_1 - t)}{\sigma\sqrt{T_1 - t}}\right) - ce^{-2\mu\sigma^{-2}U_t}\phi\left(\frac{-x - U_t + \mu(T_1 - t)}{\sigma\sqrt{T_1 - t}}\right),\end{aligned}$$

with $c = \frac{1}{\sigma\sqrt{T_1 - t}}$, where we apply (4). Let f_n denote the density of $(U_{T_2} - U_{T_1}, \dots, U_{T_n} - U_{T_1})$; this Gaussian vector is independent of U_{T_1} and so we have

$$\begin{aligned}\mathbb{P}(\tau \geq T_1, U_{T_1} < a_1, U_{T_2} \in B_2, \dots, U_{T_n} \in B_n | \mathcal{F}_t) \\ = \int_0^{a_1} \int_{B_2} \dots \int_{B_n} \vartheta_t(x) f_n(y_1 - x, \dots, y_n - x) dy_1 \dots dy_n dx \\ - e^{-2\mu\sigma^{-2}U_t} \int_0^{a_1} \int_{B_2} \dots \int_{B_n} c\phi\left(\frac{-x - U_t + \mu(T_1 - t)}{\sigma\sqrt{T_1 - t}}\right) f_n(y_1 - x, \dots, y_n - x) dy_1 \dots dy_n dx.\end{aligned}$$

The first term above is clearly equal to $\mathbb{P}(0 < U_{T_1} < a_1, U_{T_2} \in B_2, \dots, U_{T_n} \in B_n | \mathcal{F}_t)$ since, conditioned on $\mathcal{F}_t$, the random variable U_{T_1} has density

$$c\phi\left(\frac{-x + U_t + \mu(T_1 - t)}{\sigma\sqrt{T_1 - t}}\right) = c\phi\left(\frac{x - U_t - \mu(T_1 - t)}{\sigma\sqrt{T_1 - t}}\right).$$

Further, for the second term, we can see that

$$\begin{aligned}\int_0^{a_1} \int_{B_2} \dots \int_{B_n} c\phi\left(\frac{-x - U_t + \mu(T_1 - t)}{\sigma\sqrt{T_1 - t}}\right) f_n(y_1 - x, \dots, y_n - x) dy_1 \dots dy_n dx \\ = \int_{-a_1 + 2\mu(T_1 - t)}^{2\mu(T_1 - t)} \int_{B_2} \dots \int_{B_n} c\phi\left(\frac{z - U_t - \mu(T_1 - t)}{\sigma\sqrt{T_1 - t}}\right) \\ \times f_n(y_1 + z - 2\mu(T_1 - t), \dots, y_n + z - 2\mu(T_1 - t)) dy_1 \dots dy_n dz \\ = \mathbb{P}(-a'_1 < U_{T_1} < 2\mu(T_1 - t), U_{T_2} \in B_2 + 2U_{T_1} - 2\mu(T_1 - t), \dots, \\ U_{T_n} \in B_n + 2U_{T_1} - 2\mu(T_1 - t) | \mathcal{F}_t),\end{aligned}$$

Finally, a straightforward algebraic manipulation yields

$$\begin{aligned}\mathbb{P}(U_{T_2} \in B_2 + 2U_{T_1} - 2\mu(T_1 - t), \dots, U_{T_n} \in B_n + 2U_{T_1} - 2\mu(T_1 - t) | \mathcal{F}_{T_1}) \\ = \mathbb{P}(U_{T_2} - U_{T_1} \in B_2 + U_{T_1} - 2\mu(T_1 - t), \dots, U_{T_n} - U_{T_1} \in B_n + U_{T_1} - 2\mu(T_1 - t) | \mathcal{F}_{T_1}) \\ = \mathbb{P}(U_{T_1} - U_{T_2} \in B_2 + U_{T_1} - 2\mu(T_2 - t), \dots, U_{T_1} - U_{T_n} \in B_n + U_{T_1} - 2\mu(T_n - t) | \mathcal{F}_{T_1}) \\ = \mathbb{P}(-U_{T_2} + 2\mu(T_2 - t) \in B_2, \dots, -U_{T_n} + 2\mu(T_n - t) \in B_n | \mathcal{F}_{T_1}) \\ = \mathbb{P}(\tilde{U}_{T_2} \in -B_2, \dots, \tilde{U}_{T_n} \in -B_n | \mathcal{F}_{T_1}).\end{aligned}$$

From this result we have the following lemma.

Lemma 2 *Let $a_1, \dots, a_n$ be non-negative constants. On $\{\tau > t\}$, the following equation holds true*

$$\begin{aligned}
& \mathbb{P}(\tau \geq T_n, U_{T_1} < a_1, \dots, U_{T_{n-1}} < a_{n-1}, U_{T_n} > a_n | \mathcal{F}_t) \\
&= \mathbb{P}(-a_1 < U_{T_1} < a_1, \dots, -a_{n-1} < U_{T_{n-1}} < a_{n-1}, U_{T_n} > a_n | \mathcal{F}_t) \\
&\quad - e^{-2\mu\sigma^{-1}U_t} \mathbb{P}(-a_1 < \bar{U}_{T_1} < a_1, \dots, -a_{n-1} < \bar{U}_{T_{n-1}} < a_{n-1}, \bar{U}_{T_n} < -a_n | \mathcal{F}_t),
\end{aligned}$$

where $\bar{U}_{T_j} = U_{T_j} - 2\mu(T_j - t)$, $j = 1, \dots, n$.

Proof The result is obtained by induction. In order to prove the inductive basis, we start by preconditioning to $\mathcal{F}_{T_1}$ in order to get

$$\begin{aligned}
& \mathbb{P}(\tau \geq T_2, U_{T_1} < a_1, U_{T_2} > a_2 | \mathcal{F}_t) \\
&= \mathbb{E}[\mathbf{1}_{\{U_{T_1} < a_1\}} \mathbb{P}(\tau \geq T_2, U_{T_2} > a_2 | \mathcal{F}_{T_1}) | \mathcal{F}_t] \\
&= \mathbb{E}[\mathbf{1}_{\{U_{T_1} < a_1\}} \mathbf{1}_{\{\tau > T_1\}} \mathbb{P}(U_{T_2} > a_2 | \mathcal{F}_{T_1}) | \mathcal{F}_t] \\
&\quad - \mathbb{E}[\mathbf{1}_{\{U_{T_1} < a_1\}} e^{-2\mu\sigma^{-1}U_{T_1}} \mathbf{1}_{\{\tau > T_1\}} \mathbb{P}(U_{T_2} - 2\mu(T_2 - T_1) < -a_2 | \mathcal{F}_{T_1}) | \mathcal{F}_t] \\
&= \mathbb{P}(\tau \geq T_1, U_{T_1} < a_1, U_{T_2} > a_2 | \mathcal{F}_t) \\
&\quad - e^{-2\mu\sigma^{-1}U_t} \mathbb{P}(\bar{\tau} \geq T_1, \bar{U}_{T_1} < a_1, \bar{U}_{T_2} < -a_2 | \mathcal{F}_t).
\end{aligned}$$

Note that for the second identity we have used (4), whereas for the last identity we have used Lemma 1. The first term above can be obtained by means of Proposition 2 as

$$\begin{aligned}
& \mathbb{P}(\tau \geq T_1, U_{T_1} < a_1, U_{T_2} > a_2 | \mathcal{F}_t) \\
&= \mathbb{P}(0 < U_{T_1} < a_1, U_{T_2} > a_2 | \mathcal{F}_t) - e^{-2\mu\sigma^{-1}U_t} \mathbb{P}(-a_1 < \bar{U}_{T_1} < 0, \bar{U}_{T_2} < -a_2 | \mathcal{F}_t) \\
&= \mathbb{P}(U_{T_1} \in B_1, U_{T_2} \in B_2 | \mathcal{F}_t) \\
&\quad - e^{-2\mu\sigma^{-1}U_t} \mathbb{P}(U_{T_1} - 2\mu(T_1 - t) \in -B_1, U_{T_2} - 2\mu(T_2 - t) \in -B_2 | \mathcal{F}_t),
\end{aligned}$$

where we have set $B_1 := (0, a_1)$ and $B_2 := (a_2, \infty)$. Similarly, but applying now Proposition 2 to

$$\bar{U}_s = U_s - 2\mu(s - t) = \bar{u}_0 + \bar{\mu}s + \sigma W_s,$$

with $\bar{u}_0 := u_0 + 2\mu t$ and $\bar{\mu} := -\mu$, we get

$$\begin{aligned}
& \mathbb{P}(\bar{\tau} \geq T_1, \bar{U}_{T_1} < a_1, \bar{U}_{T_2} \in -B_2 | \mathcal{F}_t) \\
&= \mathbb{P}(\bar{U}_{T_1} \in B_1, \bar{U}_{T_2} \in -B_2 | \mathcal{F}_t) \\
&\quad - e^{-2\bar{\mu}\sigma^{-1}\bar{U}_t} \mathbb{P}(\bar{U}_{T_1} - 2\bar{\mu}(T_1 - t) \in -B_1, \bar{U}_{T_2} - 2\bar{\mu}(T_2 - t) \in B_2 | \mathcal{F}_t) \\
&= \mathbb{P}(U_{T_1} - 2\mu(T_1 - t) \in B_1, U_{T_2} - 2\mu(T_2 - t) \in -B_2 | \mathcal{F}_t) \\
&\quad - e^{2\mu\sigma^{-1}U_t} \mathbb{P}(U_{T_1} \in -B_1, U_{T_2} \in B_2 | \mathcal{F}_t).
\end{aligned}$$

Putting these expressions together we get

$$\begin{aligned}
& \mathbb{P}(\tau \geq T_2, U_{T_1} < a_1, U_{T_2} > a_2 \mid \mathcal{F}_t) \\
&= \mathbb{P}(U_{T_1} \in B_1, U_{T_2} \in B_2 \mid \mathcal{F}_t) + \mathbb{P}(U_{T_1} \in -B_1, U_{T_2} \in B_2 \mid \mathcal{F}_t) \\
&\quad - e^{-2\mu\sigma^{-1}U_t} \left\{ \mathbb{P}(U_{T_1} - 2\mu(T_1 - t) \in -B_1, U_{T_2} - 2\mu(T_2 - t) \in -B_2 \mid \mathcal{F}_t) \right. \\
&\quad \left. + \mathbb{P}(U_{T_1} - 2\mu(T_1 - t) \in B_1, U_{T_2} - 2\mu(T_2 - t) \in -B_2 \mid \mathcal{F}_t) \right\} \\
&= \mathbb{P}(-a_1 < U_{T_1} < a_1, U_{T_2} \in B_2 \mid \mathcal{F}_t) \\
&\quad - e^{-2\mu\sigma^{-1}U_t} \mathbb{P}(-a_1 < U_{T_1} - 2\mu(T_1 - t) < a_1, U_{T_2} - 2\mu(T_2 - t) \in -B_2 \mid \mathcal{F}_t)
\end{aligned}$$

This proves the inductive basis. The inductive step is straightforward.

Remark 1 Notice that by taking $a_1 = \dots = a_{n-1} = \infty$ we recover our base formula (4).

We conclude this section by noting that, under the Black-Scholes model, a direct combination of Proposition 1 and Lemma 2 allows us to obtain an almost (except for the constants $M_1, M_2, \dots, M_{N-1}$ that need to be computed in advance) closed-form formula, for the price of a CoCo with extension risk. The computations related with such formula are connected to multivariate Gaussian distributions as follows: in order to calculate the conditional joint distribution of $(U_{T_1}, U_{T_2}, \dots, U_{T_n})$, fixed $\mathcal{F}_t$, notice that

$$U_{T_j} = u_0 + \mu T_j + \sigma W_{T_j} = U_t + \mu(T_j - t) + \sigma(W_{T_j} - W_t),$$

and so

$$U_{T_j} \leq a_j \iff X_j := \frac{W_{T_j} - W_t}{\sqrt{T_j - t}} \leq \frac{a_j - U_t - \mu(T_j - t)}{\sigma \sqrt{T_j - t}} = -d_j + \frac{a_j}{\sigma \sqrt{T_j - t}}.$$

Clearly, conditioned to $\mathcal{F}_t$, $X_j \sim N(0, 1)$. Moreover, for $i \leq k$, we have

$$\mathbb{E}[X_i X_k \mid \mathcal{F}_t] = \frac{\mathbb{E}[W_{T_i-t} W_{T_k-t}]}{\sqrt{(T_i - t)(T_k - t)}} = \sqrt{\frac{T_i - t}{T_k - t}} =: \Sigma(i, k) =: \Sigma(k, i).$$

Therefore $(X_1, \dots, X_n)$ has a n -dimensional normal distribution with zero mean and covariance matrix $\Sigma_2 := (\Sigma(i, k))_{0 \leq i, k \leq n}$. Consequently,

$$\begin{aligned}
& \mathbb{P}(U_{T_1} \leq a_1, \dots, U_{T_n} \leq a_n \mid \mathcal{F}_t) \\
&= \mathbb{P}\left(X_1 \leq -d_1 + \frac{a_1}{\sigma \sqrt{T_1 - t}}, \dots, X_n \leq -d_n + \frac{a_n}{\sigma \sqrt{T_n - t}} \mid \mathcal{F}_t\right).
\end{aligned}$$

5 The Case of Infinite Horizon

The recursive method used in Sect. 3 cannot be employed if the horizon is infinite. Obviously one alternative is to take a sufficiently large horizon as an approximation. Nevertheless, one can pose the question whether there is an alternative to the backward procedure with a big enough horizon. This is the purpose of this section.

Consider, a d -dimensional time-inhomogeneous Markov chain $(X_n)_{n \geq 0}$ taking values in $\mathbb{R}^d$ and consider a gain or reward function G

$$V(x, n) := \inf_{\tau \in \mathcal{M}_n} \mathbb{E}_{(x, n)} [G(X_\tau, \tau)],$$

where, $\mathcal{M}_n$ is the set of stopping times greater or equal than n , $\mathbb{E}_{(x, n)}$ indicates the conditional expectation assuming that at time n we are at x , and introduce the operator T , such that

$$TF(x, n) := \mathbb{E}_{(x, n)}[F(X_{n+1}, n+1)],$$

where $F : \mathbb{R}^d \times \mathbb{Z}_+ \rightarrow \mathbb{R}$ is a measurable function, such that $F(X_{n+1}, n+1) \in \mathbb{L}^1(\mathbb{P}_{(x, n)})$ for all $n \geq 0$. We assume that

$$\mathbb{E}_{(x, n)} \left[\sup_{n \leq k} |G(X_k, k)| \right] < \infty,$$

with $G(X_\infty, \infty) = 0$. Define now another operator Q , such that

$$QF(x, n) := \min\{G(x, n), TF(x, n)\}.$$

Now, by Theorem 1.11 and Corollary 1.12 in [16],

$$V(x, n) = \lim_{k \rightarrow \infty} Q^k G(x, n)$$

Next, we apply this result to our case.

Define, using the same notation as in the previous section,

$$G(X_n, n) := K e^{-rT_n} \mathbf{1}_{\{\tau_c > T_n\}} + \frac{K}{C_p} \tilde{S}_{\tau_c} \mathbf{1}_{\{\tau_c \leq T_n\}} + \sum_{i=1}^n \sum_{j=1}^m \tilde{c}_{ij} \mathbf{1}_{\{\tau_{i,j} > T_{i,j}, S_{T_{i,j}} > M_{i,j}\}},$$

with

$$X_n = (S_{T_n}, \underline{S}_{T_n}),$$

where $\underline{S}_t := \inf_{0 \leq u \leq t} S_u$, and as above,

$$TG(X_n, n) := \mathbb{E}_{(X_n, n)}^*[G(X_{n+1}, n+1)],$$

and

$$\begin{aligned} QG(X_n, n) &:= \min\{G(X_n, n), TG(X_n, n)\} \\ &= \min \left\{ G(X_n, n), \sum_{i=1}^n \sum_{j=1}^m \tilde{c}_{ij} \mathbf{1}_{\{\tau_{i,j} > T_{i,j}, S_{T_{i,j}} > M_{i,j}\}} \right. \\ &\quad \left. + \mathbb{E}_{(X,n)}^* \left[K e^{-rT_{n+1}} \mathbf{1}_{\{\tau_c > T_{n+1}\}} + \frac{K}{C_p} \tilde{S}_{\tau_c} \mathbf{1}_{\{\tau_c \leq T_{n+1}\}} + \sum_{j=1}^m \tilde{c}_{n+1,j} \mathbf{1}_{\{\tau_{n+1,j} > T_{n+1,j}, S_{T_{n+1,j}} > M_{n+1,j}\}} \right] \right\} \\ &= K e^{-rT_n} \mathbf{1}_{\{\tau_c > T_n, S_{T_n} > M_1\}} + \frac{K}{C_p} \tilde{S}_{\tau_c} \mathbf{1}_{\{\tau_c \leq T_n\}} \\ &\quad + \sum_{i=1}^n \sum_{j=1}^m \tilde{c}_{i,j} \mathbf{1}_{\{\tau_{i,j} > T_{i,j}, S_{T_{i,j}} > M_{i,j}\}} + V_1(S_{T_n}, T_n) \mathbf{1}_{\{\tau_c > T_n, S_{T_n} \leq M_1\}}, \end{aligned}$$

where M_1 is a certain constant, since

$$\mathbb{E}_{(X_n, n)}^* \left[K e^{-rT_{n+1}} \mathbf{1}_{\{\tau_c > T_{n+1}\}} + \frac{K}{C_p} \tilde{S}_{\tau_c} \mathbf{1}_{\{\tau_c \leq T_{n+1}\}} + \sum_{j=1}^m \tilde{c}_{n+1,j} \mathbf{1}_{\{\tau_{n+1,j} > T_{n+1,j}, S_{T_{n+1,j}} > M_{n+1,j}\}} \right]$$

is assumed to be an increasing function of S_{T_n} , and where $V_1(S_{T_n}, T_n)$ (*the continuation value*) is defined in an obvious way. Iterating we obtain

$$\begin{aligned} Q^k G(X_n, n) &= K e^{-rT_n} \mathbf{1}_{\{\tau_c > T_n, S_{T_n} > M_k\}} + \frac{K}{C_p} \tilde{S}_{\tau_c} \mathbf{1}_{\{\tau_c \leq T_n\}} \\ &\quad + \sum_{i=1}^n \sum_{j=1}^m \tilde{c}_{i,j} \mathbf{1}_{\{\tau_{i,j} > T_{i,j}, S_{T_{i,j}} > M_{i,j}\}} + V_k(S_{T_n}, T_n) \mathbf{1}_{\{\tau_c > T_n, S_{T_n} \leq M_k\}}, \end{aligned}$$

for certain constants M_k and, according with the results mentioned above, the discounted price at time T_n is

$$\begin{aligned} V(X_n, n) &= \lim_{k \rightarrow \infty} \left(K e^{-rT_n} \mathbf{1}_{\{\tau_c > T_n, S_{T_n} > M_k\}} + \frac{K}{C_p} \tilde{S}_{\tau_c} \mathbf{1}_{\{\tau_c \leq T_n\}} + V_k(S_{T_n}, T_n) \mathbf{1}_{\{\tau_c > T_n, S_{T_n} \leq M_k\}} \right). \end{aligned}$$

A study about the computational behavior of this procedure can be matter of further research.

Acknowledgments The work of J.M. Corcuera is supported by the Grant of the Spanish MCINN MTM2013-40782. This research was partially carried out at CAS—Centre for Advanced Study at the Norwegian Academy of Science and Letters, Research group SEFE. J. Fajardo thanks financial support from CNPQ, grant 471131/2013-0.

References

1. Basel Committee on Banking Supervision. Basel III: A global regulatory framework for more resilient banks and banking systems. Bank for International Settlements (2010)
2. Black, F., Scholes, M.: The pricing of options and corporate liabilities. *J. Polit. Econ.* **81**, 637–654 (1973)
3. Brigo, D., Garcia, J., Pedraza, N.: Coco bonds valuation with equity and credit-calibrated first passage structural models. Preprint (2013)
4. Chen, N., Glasserman, P., Nouri, B., Pelger, M.: CoCos, bail-in, and tail risk. OFR Working paper. U.S. Department of the Treasury (2013)
5. Cheridito, P., Xu, Z.: A reduced form CoCo model with deterministic conversion intensity. Preprint (2013)
6. Cheridito, P., Xu, Z.: Pricing and hedging CoCos. Preprint (2013)
7. Commission of Experts.: Final report of the commission of experts for limiting the economic risks posed by large companies. Technical report, Swiss National Bank (2010)
8. Corcuera, J.M., Fajardo, J., Schoutens, W., Jonsson, H., Spiegeleer, J., Valdivia, A.: Close form pricing formulas for coupon cancellable CoCos. *J. Bank. Finance* **42**, 339–351 (2014)
9. De Spiegeleer, J., Schoutens, W.: Pricing contingent convertibles: a derivatives approach. *J. Deriv.* **20**(2), 27–36 (2012)
10. De Spiegeleer, J., Schoutens, W.: Multiple trigger CoCos: contingent debt without death spiral risk. *Financ. Markets Inst. Instrum. J.* **22**(2), 129–141 (2013)
11. De Spiegeleer, J., Schoutens, W.: Coco bonds with extension risk. *Wilmott* **71**, 78–91 (2014)
12. De Spiegeleer, J., Van Hulle, C., Schoutens, W.: The Handbook of Hybrid Securities: Convertible Bonds, CoCo Bonds and Bail-In. The Wiley Finance Series (2014)
13. European Banking Authority.: Buffer convertible capital securities—final termsheet. Technical report (2011)
14. Lamberton, D., Lapeyre, B.: Introduction to Stochastic Calculus Applied to Finance. Financial Mathematics Series. Chapman & Hall/CRC (2007)
15. Musiela, M., Rutkowski, M.: Martingale Methods in Financial Modelling. Applications of Mathematics (New York), vol. 36. Springer, Berlin (1997)
16. Peskir, G., Shiryaev, A.N.: Optimal Stopping and Free-Boundary Problems. Lectures in Mathematics. ETH Zurich. Birkhauser, Basel (2006)
17. Yu, P., Van Luu, B.: Lessons from the collapse in hybrid bank capital securities. *Int. J. Manage. Pract.* **5**(2), 125–148 (2012)

Hedging Under Worst-Case-Scenario in a Market Driven by Time-Changed Lévy Noises

Giulia Di Nunno and Erik Hove Karlsen

Abstract In an incomplete market driven by time-changed Lévy noises we consider the problem of hedging a financial position coupled with the underlying risk of model uncertainty. Then we study hedging under worst-case-scenario. The proposed strategies are not necessarily self-financing and include the interplay of a cost process to achieve the perfect hedge at the end of the time horizon. The hedging problem is tackled in the framework of stochastic differential games and it is treated via backward stochastic differential equations. Two different information flows are considered and the solutions compared.

Keywords Model uncertainty · Hedging · BSDEs · Stochastic differential games · Time-change · Martingale random fields

JEL Classification G11 · C70

1 Introduction

In a stylized incomplete market on the time horizon $[0, T]$ ($T > 0$), we consider the problem of hedging a contingent claim coupled with the underlying risk of an uncertain model description. This is referred to as model ambiguity in the literature, see e.g. [12]. We choose to consider a conservative evaluation of such exposure to

G. Di Nunno (✉)

CMA and Department of Mathematics, University of Oslo, P.O. Box 1053 Blindern,
N-0316 Oslo, Norway
e-mail: giulian@math.uio.no

G. Di Nunno

Department of Business and Management Science, NHH, Norwegian School
of Economics, Helleveien 30, N-5045 Bergen, Norway

E.H. Karlsen

Department of Mathematics, University of Oslo, P.O. Box 1053 Blindern,
N-0316 Oslo, Norway
e-mail: erikhk@math.uio.no

model ambiguity by measuring the robustness of the strategy in terms of a worst-case-scenario risk measure. In fact, on the complete filtered probability space $(\Omega, \mathcal{F}, \mathbb{P})$, $\mathbb{M} = \{\cdot|_{\mathcal{M}_t}, t \in [0, T]\}$, we fix the dynamic coherent risk measure

$$\rho_t(\xi) := \operatorname{ess\,sup}_{\mathbb{Q} \in \mathcal{Q}_M} \mathbb{E}_{\mathbb{Q}}[-\xi | \cdot |_{\mathcal{M}_t}], \quad t \in [0, T], \quad (1)$$

where $\mathcal{Q}_M$ is the set of all scenarios considered, which are given by the probability measures $\mathbb{Q}$ equivalent to $\mathbb{P}$ on the future outcomes $(\Omega, \mathcal{M}_T)$. The risk measure ρ is well-defined and finite for all financial positions ξ integrable with respect to all $\mathbb{Q} \in \mathcal{Q}_M$. This risk measure is naturally connected to the concept of worst-case-scenario, associating the risk evaluation with the highest expected value on the downside of the position over all $\mathbb{Q} \in \mathcal{Q}_M$. For this reason it is also a conservative evaluation of the risk exposure.

The financial market presents two investment possibilities: a saving account with price dynamics given by

$$dS_t^{(0)} = r_t S_t^{(0)} dt, \quad S_0^{(0)} = 1, \quad (2)$$

which is used as numéraire, and a stock with price dynamics of the type

$$dS_t^{(1)} = \alpha_t S_t^{(1)} dt + \sigma_t S_t^{(1)} dB_t + \int_{\mathbb{R}_0} \gamma_t(z) S_t^{(1)-} \tilde{H}(dt, dz), \quad S_0^{(1)} > 1. \quad (3)$$

The driving noises B and $\tilde{H}$ are a doubly stochastic Gaussian and Poisson type measures, respectively related to a time-changed Brownian and Poisson process. The parameters $r, \alpha, \sigma, \gamma$ are càglàd adapted stochastic processes and fields. To ensure the existence of a square integrable positive solution and to allow further analysis, we assume that $|r_t| < C$ (for some $C > 0$) $\mathbb{P} \times dt$ -a.e., $\gamma_t(z) > -1$ $\mathbb{P} \times \Lambda$ -a.e., and

$$\mathbb{E} \left[\int_0^T \left\{ |\alpha_t| + \sigma_t^2 \lambda_t^B + \int_{\mathbb{R}_0} |\ln(1 + \gamma_t(z)) - \gamma_t(z)| \nu(dz) \lambda_t^H \right\} dt \right] < \infty.$$

The elements $\nu, (\lambda^B, \lambda^H)$, and Λ are associated to the behaviour of the jumps in these dynamics and the intensity of the (stochastic) time distortion applied, see the next section for details.

Price dynamics of this type include various well-known stochastic volatility price models such as [10], where time-change of a pure jump Lévy process is used to take the erratic behaviour of volatility into account by mimicking the transition between a real-time clock to a transactions-time clock. See also e.g. [4, 21, 22, 39] where the stochastic volatility models lead to dynamics driven by time-changed Brownian motions, which are the doubly stochastic Gaussian noises in this paper. Also in the credit risk literature we can find examples of price dynamics of the type above. See e.g. [29], where doubly stochastic Poisson processes, also called the Cox processes, are largely used in the modelling of prices subject to default risk.

In the market above, we consider a financial claim $F \in L^2(\Omega, \mathcal{M}_T, \mathbb{P})$ with payoff at $T > 0$ and the $\mathbb{M}$ -predictable hedging strategies represented by the triple (π, V^π, C^π) , i.e. π_t is the wealth invested in the stock at t representing the market portfolio, V_t^π the value of the strategy on the market, and C_t^π is the cost process. We denote $V_0^\pi = v > 0$. We assume that

$$\begin{aligned} dV_t^\pi &= \frac{\pi_t}{S_t^{(1)}} dS_t^{(1)} + \frac{V_t^\pi - \pi_t}{S_t^{(0)}} dS_t^{(0)} \\ &= \left(V_t^\pi r_t + \pi_t (\alpha_t - r_t) \right) dt + \pi_t \sigma_t dB_t + \int_{\mathbb{R}_0} \pi_t \gamma_t(z) \tilde{H}(dt, dz) \end{aligned} \quad (4)$$

and that $C_t^\pi = \rho_t(e^{-\int_t^T r_s ds} (V_T^\pi - F))$. We assume that (4) admits a unique strong solution for any admissible π (see e.g. [24] for conditions). Moreover, we require that the solution is square integrable, and for this we assume

$$\mathbb{E} \left[\int_0^T \left\{ |\alpha_t - r_t| |\pi_t| + |\alpha_t \sigma_t|^2 \lambda_t^B + \int_{\mathbb{R}_0} |\pi_t \gamma_t(z)|^2 v(dz) \lambda_t^H \right\} dt \right] < \infty. \quad (5)$$

Moreover, the process C^π is also assumed square integrable.

The process $Y_t^\pi := V_t^\pi + C_t^\pi$, $t \geq 0$, is called the (total) price of the strategy. The hedging of F is obtained for a strategy $(\hat{\pi}, V^{\hat{\pi}}, C^{\hat{\pi}})$ that yields $Y_T^{\hat{\pi}} = F$. Note that $Y_0^\pi = v + C_0^\pi$ is $\mathcal{M}_0$ -measurable and, if the hedging strategy has $C^\pi \equiv 0$, then the market investments are enough to self-finance the hedge. It is only in a complete market that it is possible to hedge all claims with these self-financing strategies.

We observe that for any hedging strategy (π, V^π, C^π) with $Y_T^\pi = F$ we have

$$\rho_t(e^{-\int_t^T r_s ds} (Y_T^\pi - F)) = 0, \quad t \in [0, T].$$

This means that the risk given by the spread between the discounted final strategy price and the actual claim is zero according to the risk measure given, and actually all coherent risk measures.

Hereafter, we consider the problem of finding a hedging strategy $(\hat{\pi}, V^{\hat{\pi}}, C^{\hat{\pi}})$ for F such that $Y_t^{\hat{\pi}} = Y_t$, $t \in [0, T]$, where

$$Y_t := \operatorname{ess\,inf}_{\pi \in \Pi_{\mathbb{M}}} \rho_t \left(e^{-\int_t^T r_s ds} (V_T^\pi - F) - V_t^\pi \right), \quad t \in [0, T]. \quad (6)$$

Hence the strategy $\hat{\pi}$ minimizes the risk associated to the total price. Clearly $Y_T^{\hat{\pi}} = Y_T = F$. Here above, the set $\Pi_{\mathbb{M}}$ denotes the admissible strategies.

Considering the risk measure (1), we can write the problem (6) in the following way:

$$\begin{aligned}
Y_t^{\hat{\pi}} &= Y_t = \operatorname{ess\,inf}_{\pi \in \Pi_{\mathbb{M}}} \operatorname{ess\,sup}_{\mathbb{Q} \in \mathcal{Q}_{\mathbb{M}}} E_{\mathbb{Q}} \left[- \left(e^{-\int_t^T r_s ds} V_T^{\pi} - V_t^{\pi} - e^{-\int_t^T r_s ds} F \right) | \mathcal{M}_t \right] \\
&= \operatorname{ess\,inf}_{\pi \in \Pi_{\mathbb{M}}} \operatorname{ess\,sup}_{\mathbb{Q} \in \mathcal{Q}_{\mathbb{M}}} E_{\mathbb{Q}} \left[e^{-\int_t^T r_s ds} F - \int_t^T e^{-\int_t^s r_u du} \pi_s (\alpha_s - r_s) ds \right. \\
&\quad \left. - \int_t^T e^{-\int_t^s r_u du} \pi_s \sigma_s dB_s \right. \\
&\quad \left. - \int_t^T \int_{\mathbb{R}_0} e^{-\int_t^s r_u du} \pi_s \gamma_s(z) \tilde{H}(ds, dz) \middle| \mathcal{M}_t \right]. \quad (7)
\end{aligned}$$

A solution to the problem (7) corresponds to finding $(\hat{\pi}, \hat{\mathbb{Q}}) \in \Pi_{\mathbb{M}} \times \mathcal{Q}_{\mathbb{M}}$ such that

$$\begin{aligned}
Y_t &= E_{\hat{\mathbb{Q}}} \left[e^{-\int_t^T r_s ds} F - \int_t^T e^{-\int_t^s r_u du} \hat{\pi}_s (\alpha_s - r_s) ds \right. \\
&\quad \left. - \int_t^T e^{-\int_t^s r_u du} \hat{\pi}_s \sigma_s dB_s \right. \\
&\quad \left. - \int_t^T \int_{\mathbb{R}_0} e^{-\int_t^s r_u du} \hat{\pi}_s \gamma_s(z) \tilde{H}(ds, dz) \middle| \mathcal{M}_t \right], \quad t \in [0, T]. \quad (8)
\end{aligned}$$

The cost of the hedge is then $C_t^{\hat{\pi}} = E_{\hat{\mathbb{Q}}} [e^{-\int_t^T r_s ds} (F - V_T^{\hat{\pi}}) | \mathcal{M}_t]$. To have a unique description of the optimal strategy, we set $E_{\hat{\mathbb{Q}}} [C_0^{\hat{\pi}}] = 0$ and $v = E_{\hat{\mathbb{Q}}} [Y_0^{\hat{\pi}}]$.

This kind of approach to hedging is treated in [14] (see also [15]) in the context of a financial market driven by a Brownian motion and an insurance payment process driven by a Poisson process. See also [2] for a study of a similar problem within an insurance perspective, but Brownian driven dynamics. We also refer to [28], where a first study of this problem is given in the context of a market driven by a Brownian motion and a doubly stochastic Poisson noise. Comparatively, in the present paper we consider a more general market model and a substantially different structure of admissible strategies.

The admissible scenarios are described by a measure change via shift transformation. With respect to this, we suggest a version of the Girsanov theorem that explicitly deals with time-change. We note that, in this context, the measure change is not structure preserving in general.

The hedging problem (7) and (8) is tackled using backward stochastic differential equations (BSDEs) and stochastic differential games. Our study is carried through in the context of two different filtrations: $\mathbb{M} = \mathbb{F}$, which is substantially the information flow generated by the noises, and $\mathbb{M} = \mathbb{G}$, which is the filtration that, additionally to $\mathbb{F}$, includes initial knowledge of the time-change process. These two settings lead to different BSDEs depending on their measurability properties. We treat the solutions exploiting the martingale random field properties of the driving noises and, in the case of $\mathbb{G}$, we also rely on the better explicit structure of the noise (which allows for a more explicit stochastic representation theorem).

In the case of information flow $\mathbb{G}$, BSDEs driven by doubly stochastic Lévy noises are treated in [19]. We also mention that these integral representation theorems are studied in [18] with different approaches: via chaos expansions and via the non-anticipating derivative (see also [17] for a review on stochastic derivation). As for filtration $\mathbb{F}$, we rely on the general results of [9], which we adapt to the random field set-up.

Even though we can regard the information flow $\mathbb{F}$ as partial with respect to $\mathbb{G}$, the problems presented here are not the same as in the study on BSDEs with partial information, see e.g. [11] in the case of mean-variance hedging.

Finally, we remark that the hedging criteria we consider differs from mean-variance hedging in the objective function to minimize: mean-variance hedging identifies the strategy by minimizing the quadratic cost, see e.g. [26] in the context of prices modeled by general semimartingales, and [31] for the case of dynamics driven by a Brownian motion and doubly stochastic Poisson noises.

The paper is structured as follows: the next section provides details about the framework and the BSDEs considered. In Sect. 3, we study shift transformations, while Sect. 4 is dedicated to the actual solution of the hedging problem in the two information flows considered. Section 5 concludes with comments on the results obtained.

2 The Framework and Preliminary Results

Hereafter, we give full detail of the noises considered in (3) and the stochastic structures used. For this we refer to [18, 19]. In particular, we apply stochastic integration with respect to martingale random fields, see [8, 16].

2.1 The Random Measures and Their Properties

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a complete probability space and $\mathbb{X} := [0, T] \times \mathbb{R}$. We will consider $\mathbb{X} = ([0, T] \cup \{0\}) \cup ([0, T] \times \mathbb{R}_0)$, where $\mathbb{R}_0 := \mathbb{R} \setminus \{0\}$ and $T > 0$. Clearly, $[0, T] \cup \{0\} \simeq [0, T]$. Denote $\mathcal{B}_{\mathbb{X}}$ the Borel σ -algebra on $\mathbb{X}$. Whenever we write $\Delta \subset \mathbb{X}$, we intend a set Δ in $\mathcal{B}_{\mathbb{X}}$.

The two dimensional stochastic process $\lambda := (\lambda^B, \lambda^H)$ represents the intensity of the stochastic time distortion applied in the noise. Each component λ^l for $l = B, H$, satisfies

- (i) $\lambda_t^l \geq 0$ $\mathbb{P}$ -a.s. for all $t \in [0, T]$,
- (ii) $\lim_{h \rightarrow 0} \mathbb{P}(|\lambda_{t+h}^l - \lambda_t^l| \geq \varepsilon) = 0$ for all $\varepsilon > 0$ and almost all $t \in [0, T]$,
- (iii) $\mathbb{E}[\int_0^T \lambda_t^l dt] < \infty$.

Denote the space of all processes $\lambda := (\lambda^B, \lambda^H)$ satisfying (i), (ii) and (iii) by $\mathcal{L}$.

Correspondingly, we define the random measure Λ on $\mathbb{X}$ by

$$\Lambda(\Delta) := \int_0^T \mathbf{1}_\Delta(t, 0) \lambda_t^B dt + \int_0^T \int_{\mathbb{R}_0} \mathbf{1}_\Delta(t, z) \nu(dz) \lambda_t^H dt, \quad \Delta \in \mathbb{X}. \quad (9)$$

Here ν is a (deterministic) σ -finite measure on the Borel sets of $\mathbb{R}_0$ satisfying

$$\int_{\mathbb{R}_0} z^2 \nu(dz) < \infty.$$

We denote the σ -algebra generated by the values of Λ by $\mathcal{F}^\Lambda$. Furthermore, Λ^H denotes the restriction of Λ to $[0, T] \times \mathbb{R}_0$ and Λ^B the restriction of Λ to $[0, T] \times \{0\}$. Hence $\Lambda(\Delta) = \Lambda^B(\Delta \cap [0, T] \times \{0\}) + \Lambda^H(\Delta \cap [0, T] \times \mathbb{R}_0)$, $\Delta \subseteq \mathbb{X}$. Here below we introduce the noises driving the stochastic dynamics.

Definition 2.1 B is a signed random measure on the Borel sets of $[0, T] \times \{0\}$, satisfying

$$(A1) \quad \mathbb{P}(B(\Delta) \leq x \mid \mathcal{F}^\Lambda) = \mathbb{P}(B(\Delta) \leq x \mid \Lambda^B(\Delta)) = \Phi\left(\frac{x}{\sqrt{\Lambda^B(\Delta)}}\right), \quad x \in \mathbb{R},$$

$$\Delta \subseteq [0, T] \times \{0\},$$

(A2) $B(\Delta_1)$ and $B(\Delta_2)$ are conditionally independent given $\mathcal{F}^\Lambda$ whenever Δ_1 and Δ_2 are disjoint sets.

Here Φ stands for the cumulative probability distribution function of a standard normal random variable.

H is a random measure on the Borel sets of $[0, T] \times \mathbb{R}_0$, satisfying

$$(A3) \quad \mathbb{P}(H(\Delta) = k \mid \mathcal{F}^\Lambda) = \mathbb{P}(H(\Delta) = k \mid \Lambda^H(\Delta)) = \frac{\Lambda^H(\Delta)^k}{k!} e^{-\Lambda^H(\Delta)}, \quad k \in \mathbb{N},$$

$$\Delta \subseteq [0, T] \times \mathbb{R}_0,$$

(A4) $H(\Delta_1)$ and $H(\Delta_2)$ are conditionally independent given $\mathcal{F}^\Lambda$ whenever Δ_1 and Δ_2 are disjoint sets.

Furthermore, we assume that

(A5) B and H are conditionally independent given $\mathcal{F}^\Lambda$.

Substantially, conditional on Λ , we have that B is a Gaussian random measure and H is a Poisson random measure. We refer to [20] or [27] for the existence of the above conditional distributions.

Let $\tilde{H} := H - \Lambda^H$ be the signed random measure given by

$$\tilde{H}(\Delta) = H(\Delta) - \Lambda^H(\Delta), \quad \Delta \subseteq [0, T] \times \mathbb{R}_0.$$

Definition 2.2 We define the signed random measure μ on the Borel subsets of $\mathbb{X}$ by

$$\mu(\Delta) := B(\Delta \cap [0, T] \times \{0\}) + \tilde{H}(\Delta \cap [0, T] \times \mathbb{R}_0), \quad \Delta \subseteq \mathbb{X}. \quad (10)$$

The random measures B and H are related to a specific form of time-change for Brownian motion and pure jump Lévy process. More specifically define $B_t := B([0, t] \times \{0\})$, $\Lambda_t^B := \int_0^t \lambda_s^B ds$, $\eta_t := \int_0^t \int_{\mathbb{R}_0} z \tilde{H}(ds, dz)$ and $\hat{\Lambda}_t^H := \int_0^t \lambda_s^H ds$, for $t \in [0, T]$.

We can immediately see the role that the time-change processes Λ^B and $\hat{\Lambda}^H$ play by studying the characteristic function of B and η . In fact, from (A1) and (A3) we see that the conditional characteristic functions of B_t and η_t are given by

$$\mathbb{E}[e^{icB_t} | \mathcal{F}^\Lambda] = \exp \left\{ \int_0^t \frac{1}{2} c^2 \lambda_s^B ds \right\} = \exp \left\{ \frac{1}{2} c^2 \Lambda_t^B \right\}, \quad c \in \mathbb{R}, \quad (11)$$

$$\begin{aligned} \mathbb{E}[e^{ic\eta_t} | \mathcal{F}^\Lambda] &= \exp \left\{ \int_0^t \int_{\mathbb{R}_0} [e^{icz} - 1 - icz] v(dz) \lambda_s^H ds \right\} \\ &= \exp \left\{ \left(\int_{\mathbb{R}_0} [e^{icz} - 1 - icz] v(dz) \right) \hat{\Lambda}_t^H \right\}, \quad c \in \mathbb{R}. \end{aligned} \quad (12)$$

Indeed, there is a strong connection between the distributions of B and the Brownian motion, and between η and a centered pure jump Lévy process with the same jump behavior. The relationship is based on a random distortion of the time scale. The following characterization is due to [38, Theorem 3.1] (see also [20]).

Theorem 2.1 Let W_t , $t \in [0, T]$, be a Brownian motion and N_t , $t \in [0, T]$, be a centered pure jump Lévy process with Levy measure v . Assume that both W and N are independent of Λ . Then B satisfies (A1)–(11) and (A2) if and only if, for any $t \geq 0$,

$$B_t \stackrel{d}{=} W_{\Lambda_t^B},$$

and η satisfies (A3)–(12) and (A4) if and only if, for any $t \geq 0$,

$$\eta_t \stackrel{d}{=} N_{\hat{\Lambda}_t^H}.$$

In addition, B is infinitely divisible if Λ^B is infinitely divisible, and η is infinitely divisible if $\hat{\Lambda}^H$ is infinitely divisible, see [3, Theorem 7.1].

2.2 Stochastic Non-anticipating Integration and Representation Theorems

Let us define $\mathbb{F}^\mu = \{\mathcal{F}_t^\mu, t \in [0, T]\}$ as the filtration generated by $\mu(\Delta)$, $\Delta \subseteq [0, t] \times \mathbb{R}$, $t \in [0, T]$. In view of (10), (A1) and (A3), we can see that, for any $t \in [0, T]$,

$$\mathcal{F}_t^\mu = \mathcal{F}_t^B \vee \mathcal{F}_t^H \vee \mathcal{F}_t^\Lambda,$$

where $\mathcal{F}_t^B$ is generated by $B(\Delta \cap [0, T] \times \{0\})$, $\mathcal{F}_t^H$ by $H(\Delta \cap [0, T] \times \mathbb{R}_0)$, and $\mathcal{F}_t^\Lambda$ by $\Lambda(\Delta)$, $\Delta \in [0, t] \times \mathbb{R}$. This is an application of [40, Theorem 1] and [18, Theorem 2.8]. Set $\mathbb{F} = \{\mathcal{F}_t, t \in [0, T]\}$, where

$$\mathcal{F}_t := \bigcap_{r>t} \mathcal{F}_r^\mu.$$

Furthermore, we set $\mathbb{G} = \{\mathcal{G}_t, t \in [0, T]\}$ where $\mathcal{G}_t := \mathcal{F}_t \vee \mathcal{F}^\Lambda$. Remark that $\mathcal{G}_T = \mathcal{F}_T$, $\mathcal{G}_0 = \mathcal{F}^\Lambda$, while $\mathcal{F}_0$ is trivial. From now on we set $\mathcal{F} = \mathcal{F}_T$.

Lemma 2.1 *The filtration $\mathbb{G}$ is right-continuous.*

Proof This can be shown adapting classical arguments for the Lévy case as in e.g. [1, Theorem 2.1.9].

For $\Delta \subset (t, T] \times \mathbb{R}$, the conditional independence (A2) and (A4) means that

$$\mathbb{E}[\mu(\Delta) \mid \mathcal{G}_t] = \mathbb{E}[\mu(\Delta) \mid \mathcal{F}_t \vee \mathcal{F}^\Lambda] = \mathbb{E}[\mu(\Delta) \mid \mathcal{F}^\Lambda] = 0. \quad (13)$$

Hence, μ is a *martingale random field with conditional orthogonal values* in $L^2(\Omega, \mathcal{F}, \mathbb{P})$ with respect to $\mathbb{G}$ in the sense of [16] (see Definition 2.1), since

- μ has a σ -finite variance measure

$$m(\Delta) := \mathbb{E}[\mu(\Delta)^2] = \mathbb{E}[\Lambda(\Delta)], \quad \Delta \subseteq \mathbb{X},$$

with $m(\{0\} \times \mathbb{R}) = 0$,

- it is additive on pairwise disjoint sets in $\mathcal{B}_{\mathbb{X}}$ and σ -additive with convergence in L^2 ,
- μ is $\mathbb{G}$ -adapted,
- it has the martingale property (13),
- μ has conditionally orthogonal values, if $\Delta_1, \Delta_2 \subset (t, T] \times \mathbb{R}$ such that $\Delta_1 \cap \Delta_2 = \emptyset$ then, combining (A2), (A4), (A5) and (13),

$$\mathbb{E}[\mu(\Delta_1)\mu(\Delta_2) \mid \mathcal{G}_t] = \mathbb{E}[\mu(\Delta_1) \mid \mathcal{F}^\Lambda] \mathbb{E}[\mu(\Delta_2) \mid \mathcal{F}^\Lambda] = 0.$$

In [8] there is a discussion about martingale (difference) random fields and the role of ordering associated with the information flow. In their terminology the martingale random fields here treated is both a “strong” and a “weak” martingale.

Denote $\mathcal{J}_{\mathbb{G}}$ as the subspace of $L^2([0, T] \times \mathbb{R} \times \Omega, \mathcal{B}_{\mathbb{X}} \times \mathcal{F}, \Lambda \times \mathbb{P})$ of the random fields admitting a $\mathbb{G}$ -predictable modification, in particular

$$\|\phi\|_{\mathcal{J}_{\mathbb{G}}} := \left(\mathbb{E} \left[\int_0^T \phi_s(0)^2 \lambda_s^B ds + \int_0^T \int_{\mathbb{R}_0} \phi_s(z)^2 \nu(dz) \lambda_s^H ds \right] \right)^{\frac{1}{2}} < \infty. \quad (14)$$

For any $\phi \in \mathcal{J}_{\mathbb{G}}$, we define the (Itô type) non-anticipative stochastic integral $I : \mathcal{J}_{\mathbb{G}} \Rightarrow L^2(\Omega, \mathcal{F}, \mathbb{P})$ by

$$I(\phi) := \int_0^T \phi_s(0) dB_s + \int_0^T \int_{\mathbb{R}_0} \phi_s(z) \tilde{H}(ds, dz).$$

We refer to [16] for the details on the integration with respect to martingale random fields of the type discussed here. Recall that I is a linear isometric operator:

$$\sqrt{E[I(\phi)^2]} = \|I(\phi)\|_{L^2(\Omega, \mathcal{F}, \mathbb{P})} = \|\phi\|_{\mathcal{J}_{\mathbb{G}}}.$$

Because of the structure of the filtration considered, we have the following result (see [19]):

Lemma 2.2 Consider $\xi \in L^2(\Omega, \mathcal{F}^{\Lambda}, \mathbb{P})$ and $\phi \in \mathcal{J}_{\mathbb{G}}$. Then

$$\xi I(\phi) = I(\xi \phi),$$

whenever either side of the equality exists as an element in $L^2(\Omega, \mathcal{F}, \mathbb{P})$.

Remark 2.1 It is easy to see that the random field μ is also a martingale random field with respect to $\mathbb{F}$ and the non-anticipating integration can be done also with respect to $\mathbb{F}$ as for $\mathbb{G}$. We denote $\mathcal{J}_{\mathbb{F}}$ the corresponding set of integrands. However, results such as Lemma 2.2 and the forthcoming representation would not hold. See also [18, Remark 4.4].

We remark that $\mathcal{F}_t^{\mu} := \sigma\{\mu(\Delta), \Delta \subseteq [0, t] \times \mathbb{R}\} = \sigma\{I(\phi 1_{\Delta}), \Delta \subseteq [0, t] \times \mathbb{R}\}$, $\phi \in \mathcal{J}_{\mathbb{F}}$ (indeed $\mu(\Delta) = I(1_{\Delta})$) and $\mathcal{G}_t := \sigma\{\mu(\Delta), \Delta \subseteq [0, t] \times \mathbb{R}; \Lambda(\Delta), \Delta \subseteq [0, T] \times \mathbb{R}\} = \sigma\{I(\phi 1_{\Delta}), \Delta \subseteq [0, t] \times \mathbb{R}\}$, $\phi \in \mathcal{J}_{\mathbb{G}}$

The following representation theorems are given in [19].

Theorem 2.2 (Integral representation theorem) *Assume $\xi \in L^2(\Omega, \mathcal{F}, \mathbb{P})$. Then there exists a unique $\phi \in \mathcal{J}_{\mathbb{G}}$ such that*

$$\xi = \mathbb{E}[\xi \mid \mathcal{F}^\Lambda] + \int_0^T \int_{\mathbb{R}} \phi_s(z) \mu(ds, dz). \quad (15)$$

Note that the two summands in (15) are orthogonal. Here $\mathbb{E}[\xi \mid \mathcal{F}^\Lambda]$ represents the stochastic component of ξ that cannot be recovered by integration on μ .

Remark 2.2 The existence of such a representation is treated in [25, Chap. 3], where the result is obtained after a discussion on the solution of *the martingale problem*. In [19], the existence and uniqueness of the above representation is proved by classical density arguments inspired by [33, Sect. 4] and [32]. In [18], the representation is given with respect to $\tilde{H}$ using orthogonal polynomials. There, an *explicit formula* for the integrand ϕ is derived by means of the non-anticipating derivative with respect to $\mathbb{G}$, see [18, Theorem 5.1]. This result holds for more general choices of Λ^H , but with an assumption on the moments. The non-anticipating derivative is well-defined with respect to any martingale random-field with orthogonal values and is an operator on the whole $L^2(\Omega, \mathcal{F}, \mathbb{P})$. The random variable $\xi^0 = E[\xi \mid \mathcal{F}_T^\Lambda]$ is characterized by having non-anticipating derivative identically null.

There are other related results in the literature, e.g. in [41, Proposition 41] the same representation is proved for a class of Malliavin differentiable random variables (Clark-Ocone type results).

If an $\mathcal{F}_T^H$ -measurable ξ is considered, then an integral representation is given in the general context of (marked) point processes, see for instance [7, Theorems 4.12 and 8.8] or [6, 13, 23]. Theorem 2.2 differs in the choice of filtration, which also leads to different integrals. In [6, 7, 13, 23], the integrator in the representation theorem is given by $H - \vartheta$, where ϑ is $\mathbb{F}^H$ -predictable compensator of H . Our Λ^H is not $\mathbb{F}^H$ -predictable.

Theorem 2.3 (Martingale representation theorem) *Assume M_t , $t \in [0, T]$, is a $\mathbb{G}$ -martingale. Then there exists a unique $\phi \in \mathcal{J}_{\mathbb{G}}$ such that*

$$M_t = \mathbb{E}[M_T \mid \mathcal{F}^\Lambda] + \int_0^t \int_{\mathbb{R}} \phi_s(z) \mu(ds, dz), \quad t \in [0, T].$$

We observe that, in the case we consider μ to be a martingale random field with respect to $\mathbb{F}$, the corresponding results take a different form. See [16]. In particular, we have:

Theorem 2.4 (Integral representation theorem) *Assume $\xi \in L^2(\Omega, \mathcal{F}, \mathbb{P})$. Then there exists a unique $\phi \in \mathcal{J}_{\mathbb{F}}$ such that*

$$\xi = \xi^0 + \int_0^T \int_{\mathbb{R}} \phi_s(z) \mu(ds, dz), \quad (16)$$

where ξ^0 is a random variable in $L^2(\Omega, \mathcal{F}, \mathbb{P})$ orthogonal to the integral part.

In terms of the non-anticipating derivative with respect to $\mathbb{F}$ as studied in [16], the random variable ξ^0 is characterised by having derivative identically null.

2.3 Backward Stochastic Differential Equations Driven by μ

The problem of hedging considered in this paper leads to different types of BSDEs depending on the information considered. Hereafter, we give an overview of the results needed in the sequel related to both types. In particular, the comparison theorems will play a central role in the solution of the optimisation problem (7). Our references are [9, 19].

2.3.1 Information Flow $\mathbb{G}$

In the case of information flow $\mathbb{G}$, the BSDE of reference is of the form:

$$Y_t = \xi + \int_t^T g_s(\lambda_s, Y_s, \phi_s) ds - \int_t^T \int_{\mathbb{R}} \phi_s(z) \mu(ds, dz), \quad t \in [0, T]. \quad (17)$$

Given a terminal condition ξ and a driver (or generator) g , a solution is given by the couple of $\mathbb{G}$ -adapted processes (Y, ϕ) on $(\Omega, \mathcal{F}, \mathbb{P})$ satisfying the equation above. Hereafter, we characterise explicitly the functional spaces in use and the elements of the BSDE to obtain a solution.

Let $S_{\mathbb{G}}$ be the space of $\mathbb{G}$ -adapted stochastic processes $Y_t(\omega)$, $t \in [0, T]$, $\omega \in \Omega$, such that

$$\|Y\|_{S_{\mathbb{G}}} := \sqrt{\mathbb{E} \left[\sup_{t \in [0, T]} |Y_t|^2 \right]} < \infty,$$

and let $\mathcal{H}_2^{\mathbb{G}}$ be the space of $\mathbb{G}$ -predictable stochastic processes $f_t(\omega)$, $t \in [0, T]$, $\omega \in \Omega$, such that

$$\mathbb{E} \left[\int_0^T f_s^2 ds \right] < \infty.$$

Denote Φ the space of functions $\phi : \mathbb{R} \rightarrow \mathbb{R}$ such that

$$|\phi(0)|^2 + \int_{\mathbb{R}_0} \phi(z)^2 \nu(dz) < \infty,$$

where ν is the jump measure of the market dynamics.

Definition 2.5 We say that (ξ, g) are *standard parameters* when $\xi \in L^2(\Omega, \mathcal{F}, \mathbb{P})$ and $g : [0, T] \times [0, \infty)^2 \times \mathbb{R} \times \Phi \times \Omega \rightarrow \mathbb{R}$ such that g satisfies the following conditions:

- $g(\lambda, Y, \phi, \cdot)$ is $\mathbb{G}$ -adapted for all $\lambda \in \mathcal{L}$, $Y \in S_{\mathbb{G}}$, $\phi \in \mathcal{I}_{\mathbb{G}}$,
- $g(\lambda, \cdot, 0, 0, \cdot) \in \mathcal{H}_2^g$, for all $\lambda \in \mathcal{L}$
- there exists $K_g > 0$ for which

$$\begin{aligned} & |g_t((\lambda^B, \lambda^H), y^{(1)}, \phi^{(1)}) - g_t((\lambda^B, \lambda^H), y^{(2)}, \phi^{(2)})| \leq K_g \left(|y^{(1)} - y^{(2)}| \right. \\ & \left. + |\phi^{(1)}(0) - \phi^{(2)}(0)| \sqrt{\lambda^B} + \sqrt{\int_{\mathbb{R}_0} |\phi^{(1)}(z) - \phi^{(2)}(z)|^2 \nu(dz) \sqrt{\lambda^H}} \right), \end{aligned}$$

- for all $(\lambda^B, \lambda^H) \in [0, \infty)^2$, $y^{(1)}, y^{(2)} \in \mathbb{R}$, and $\phi^{(1)}, \phi^{(2)} \in \Phi$ $dt \times d\mathbb{P}$ a.e.

Theorem 2.6 Let (g, ξ) be standard parameters. Then there exists a unique couple $(Y, \phi) \in S_{\mathbb{G}} \times \mathcal{I}_{\mathbb{G}}$ such that

$$\begin{aligned} Y_t &= \xi + \int_t^T g_s(\lambda_s, Y_s, \phi_s) ds - \int_t^T \int_{\mathbb{R}} \phi_s(z) \mu(ds, dz) \\ &= \xi + \int_t^T g_s(\lambda_s, Y_s, \phi_s) ds - \int_t^T \phi_s(0) dB_s - \int_t^T \int_{\mathbb{R}_0} \phi_s(z) \tilde{H}(ds, dz). \end{aligned} \quad (18)$$

Remark 2.3 The initial point Y_0 of the solution Y is *not* necessarily a (deterministic) constant. From the definition of $\mathbb{G}$ and (18), we see that Y_0 is a square integrable $\mathcal{F}^\Lambda$ -measurable random variable. To be specific, we have:

$$\begin{aligned} Y_0 &= \mathbb{E} \left[\xi + \int_0^T g_s(\lambda_s, Y_s, \phi_s) ds - \int_0^T \phi_s(0) dB_s - \int_0^T \int_{\mathbb{R}_0} \phi_s(z) \tilde{H}(ds, dz) \mid \mathcal{F}^\Lambda \right] \\ &= \mathbb{E} \left[\xi + \int_0^T g_s(\lambda_s, Y_s, \phi_s) ds \mid \mathcal{F}^\Lambda \right]. \end{aligned}$$

For a linear BSDE of the form (19), there exists an explicit representation of the solution.

Theorem 2.7 Assume we have the following BSDE:

$$\begin{aligned} -dY_t = & \left[A_t Y_t + C_t + E_t(0) \phi_t(0) \sqrt{\lambda_t^B} + \int_{\mathbb{R}_0} E_t(z) \phi_t(z) \nu(dz) \sqrt{\lambda_t^H} \right] dt \\ & - \phi_t(0) dB_t - \int_{\mathbb{R}_0} \phi_t(z) \tilde{H}(dt, dz), \quad Y_T = \xi, \end{aligned} \quad (19)$$

where the coefficients satisfy

- (i) A is a bounded stochastic process, there exists $K_A > 0$ such that $|A_t| \leq K_A$ for all $t \in [0, T]$ $\mathbb{P}$ -a.s.,
- (ii) $C \in \mathcal{H}_2^{\mathcal{G}}$,
- (iii) $E \in \mathcal{I}_{\mathbb{G}}$,
- (iv) There exists a $K_E > 0$ such that $0 \leq E_t(z) < K_E z$ for $z \in \mathbb{R}_0$, and $|E_t(0)| < K_E$ $dt \times d\mathbb{P}$ -a.e.

Then (19) has a unique solution (Y, ϕ) in $S_{\mathbb{G}} \times \mathcal{I}_{\mathbb{G}}$ and Y has representation

$$Y_t = \mathbb{E} \left[\xi \Gamma_T^t + \int_t^T \Gamma_t^s C_s ds \middle| \mathcal{G}_t \right], \quad t \in [0, T],$$

where

$$\begin{aligned} \Gamma_s^t := & \exp \left\{ \int_t^s A_u - \frac{1}{2} E_u(0)^2 \mathbf{1}_{\{\lambda_u^B \neq 0\}} du + \int_t^s E_u(0) \frac{\mathbf{1}_{\{\lambda_u^B \neq 0\}}}{\sqrt{\lambda_u^B}} dB_u \right. \\ & + \int_t^s \int_{\mathbb{R}_0} \left[\ln \left(1 + E_u(z) \frac{\mathbf{1}_{\{\lambda_u^H \neq 0\}}}{\sqrt{\lambda_u^H}} \right) - E_u(z) \frac{\mathbf{1}_{\{\lambda_u^H \neq 0\}}}{\sqrt{\lambda_u^H}} \right] \nu(dz) \lambda_u^H du \\ & \left. + \int_t^s \int_{\mathbb{R}_0} \ln \left(1 + E_u(z) \frac{\mathbf{1}_{\{\lambda_u^H \neq 0\}}}{\sqrt{\lambda_u^H}} \right) \tilde{H}(du, dz) \right\}, \quad s \geq t. \end{aligned}$$

Note that $\Gamma_t^s = \frac{\Gamma_s^0}{\Gamma_t^0}$.

The next result is a comparison theorem. This result is crucial in the solution of the optimisation problem (7).

Theorem 2.8 (Comparison theorem) *Let $(g^{(1)}, \xi^{(1)})$ and $(g^{(2)}, \xi^{(2)})$ be two sets of standard parameters for the BSDE's with solutions $(Y^{(1)}, \phi^{(1)})$, $(Y^{(2)}, \phi^{(2)}) \in S_{\mathbb{G}} \times \mathcal{J}_{\mathbb{G}}$. Assume that*

$$g_t^{(2)}(\lambda, y, \phi, \omega) = f_t\left(y, \phi(0)\kappa(0)\sqrt{\lambda^B}, \int_{\mathbb{R}_0} \phi(z)\kappa(z)v(dz)\sqrt{\lambda^H}, \omega\right),$$

where $\kappa \in \mathcal{J}_{\mathbb{G}}$ satisfies condition (iv) from Theorem 2.7 and f is a function $f : [0, T] \times \mathbb{R} \times \mathbb{R} \times \mathbb{R} \times \Omega \rightarrow \mathbb{R}$ which satisfies, for some $K_f > 0$,

$$|f_t(y, b, h) - f_t(y', b', h')| \leq K_f(|y - y'| + |b - b'| + |h - h'|),$$

$dt \times d\mathbb{P}$ a.e. and

$$\mathbb{E}\left[\int_0^T |f_t(0, 0, 0)|^2 dt\right] < \infty.$$

If $\xi^{(1)} \leq \xi^{(2)}$ $\mathbb{P}$ -a.s. and $g_s^{(1)}(\lambda_s, Y_s^{(1)}, \phi_s^{(1)}) \leq g_s^{(2)}(\lambda_s, Y_s^{(1)}, \phi_s^{(1)})$ $dt \times d\mathbb{P}$ -a.e., then

$$Y_t^{(1)} \leq Y_t^{(2)} \quad dt \times d\mathbb{P}\text{-a.e.}$$

2.3.2 Information Flow $\mathbb{F}$

In the case of information flow $\mathbb{F}$ the BSDE of reference takes the form:

$$Y_t = \xi + \int_t^T \int_{\mathbb{R}} f_s(Y_s, \phi_s(z)) \langle \mu \rangle(ds, dz) - \int_t^T \int_{\mathbb{R}} \phi_s(z) \mu(ds, dz) - N_T + N_t, \quad (20)$$

where

- (i) $\mu(dt, dz)$, $t \in [0, T]$, $z \in \mathbb{R}$, is the $(\mathbb{F}, \mathbb{P})$ -martingale random field (10),
- (ii) $\langle \mu \rangle(dt, dz)$ is its conditional variance measure, see [16] Theorem 2.1, which is in fact the correspondent to the conditional quadratic variation for martingales, see [37], and
- (iii) N , with $N_0 = 0$, is a square integrable $(\mathbb{F}, \mathbb{P})$ -martingale orthogonal to μ , i.e. for every set $A \in \mathcal{B}_{\mathbb{R}}$, for $\mu_t(A) := \mu((0, t] \times A)$, $t \in [0, T]$, the quadratic variation $[N, \mu(A)]$ is a uniformly integrable martingale.

Moreover, we have

$$\langle \mu(A) \rangle_t = \int_0^t \int_A \lambda_s^B \delta_{\{0\}}(dz) ds + \int_0^t \int_A \mathbf{1}_{\mathbb{R}_0}(z) v(dz) \lambda_s^H ds.$$

The existence and uniqueness of the solution of (20) is treated adapting Proposition 2.1 and Lemma 2.2 of [9] to the martingale random field case. Here we present the variation of these results in the form used later. The definition of standard parameters is analogous to Definition 2.5, but referred to $\mathbb{F}$. The same is intended for the spaces involved. BSDEs of the type (20) with standard parameters admit a unique solution, which is characterized by the triple (Y, ϕ, N) .

Lemma 2.9 *Let a, b, c be $\mathbb{F}$ -predictable random fields with a bounded and*

$$\mathbb{E}\left[\int_0^T \int_{\mathbb{R}} b_s^2(z) \langle \mu \rangle(ds, dz)\right] < \infty.$$

Let $\mathcal{E}$ be the Doléans exponential of the martingale random field $\int_0^t \int_{\mathbb{R}} b_s(z) \mu(ds, dz)$, $t \in [0, T]$, and define

$$\psi_t := \exp\left(\int_0^t \int_{\mathbb{R}} a_s(z) \langle \mu \rangle(ds, dz)\right), \quad \Psi_t = \psi_t \mathcal{E}_t, \quad t \in [0, T].$$

Suppose that

- (i) $\mathcal{E}$ is a positive uniformly integrable martingale;
- (ii) $\mathbb{E}[(\sup_{t \in [0, T]} \psi_t)^2 \mathcal{E}_T^2] < \infty$,
- (iii) $\mathbb{E}\left[\int_0^T \int_{\mathbb{R}} \Psi_s(z) c_s(z) \langle \mu \rangle(ds, dz)\right] < \infty$.

If the linear backward equation

$$\begin{aligned} dY_t &= - \int_{\mathbb{R}} \left(a_t(z) Y_t + b_t(z) \phi_t(z) + c_t(z) \right) \langle \mu \rangle(ds, dz) + \int_{\mathbb{R}} \phi_t(z) \mu(dt, dz) + dN_t \\ Y_T &= \xi, \end{aligned} \tag{21}$$

has solution (Y, ϕ, N) in $S_{\mathbb{F}} \times \mathcal{J}_{\mathbb{F}} \times \mathcal{L}_{\mathbb{F}}^{2,b}$, where $\mathcal{L}_{\mathbb{F}}^{2,b}$ is the space of L^2 -bounded $(\mathbb{F}, \mathbb{P})$ -martingales, then Y is given by

$$Y_t = \mathbb{E}\left[\xi \frac{\Psi_T}{\Psi_t} + \int_t^T \int_{\mathbb{R}} \frac{\Psi_s}{\Psi_t} c_s(z) \langle \mu \rangle(ds, dz) \middle| \mathcal{F}_t\right], \quad 0 \leq t \leq T. \tag{22}$$

Proof The result presents weaker assumptions on the coefficients compared to Lemma 2.2 in [9] which serve better the applications to come. However, the proof follows substantially the same arguments and it will not be detailed here.

Recall that the Doléan exponential is positive if the martingale has jumps grater than -1 . Conditions for uniform integrability can be found in e.g. [30].

Remark 2.4 Whenever expression (22) makes sense, it is a solution of the linear equation (21).

The next result is a comparison theorem for equations of the type (20).

Theorem 2.10 (Comparison theorem) *Consider two linear BSDEs of the form (20):*

$$Y_t^{(i)} = \xi^{(i)} + \int_t^T \int_{\mathbb{R}} f_s^{(i)}(Y_s^{(i)}, \phi_s^{(i)}) \langle \mu \rangle (ds, dz) - \int_t^T \int_{\mathbb{R}} \phi_s^{(i)}(z) \mu(ds, dz) - N_T^{(i)} + N_t^{(i)},$$

for $i = 1, 2$. Define $\delta Y_t := Y_t^{(2)} - Y_t^{(1)}$, $\delta \phi_t(z) := \phi_t^{(2)}(z) - \phi_t^{(1)}(z)$, and

$$\begin{aligned} a_t(z) &:= \frac{f_t^{(2)}(Y_t^{(2)}, \phi_t^{(2)}(z)) - f_t^{(2)}(Y_t^{(1)}, \phi_t^{(2)}(z))}{\delta Y_t} 1_{\delta Y_t \neq 0}, \\ b_t(z) &:= \frac{f_t^{(2)}(Y_t^{(1)}, \phi_t^{(2)}(z)) - f_t^{(2)}(Y_t^{(1)}, \phi_t^{(1)}(z))}{\delta \phi_t(z)} 1_{\delta \phi_t(z) \neq 0}, \\ c_t(z) &:= f_t^{(2)}(Y_t^{(1)}, \phi_t^{(1)}(z)) - f_t^{(1)}(Y_t^{(1)}, \phi_t^{(1)}(z)). \end{aligned}$$

Then the process δY verifies the linear BSDE:

$$\begin{aligned} -d\delta Y_t &= \int_{\mathbb{R}} [a_t(z)\delta Y_t + b_t(z)\delta \phi_t(z) + c_t(z)] \langle \mu \rangle (dt, dz) + \int_{\mathbb{R}} \delta \phi_t(z) \mu(dt, dz) - d\delta N_t, \\ \delta Y_T &= \xi^{(1)} - \xi^{(2)}. \end{aligned}$$

Assume that a and b verify condition (i) and (ii) in Lemma 2.9. Assume also that $\xi^{(2)} \geq \xi^{(1)}$ and, for any t , $c_t(z) \geq 0$ $\mathbb{P}$ -a.s. Then, for any t , $Y_t^{(2)} \geq Y_t^{(1)}$ $\mathbb{P}$ -a.s.

3 Change of Measure: Shift Transformations

The various possible scenarios considered are given by all probability measures $\mathbb{Q}$ equivalent to $\mathbb{P}$ obtained by shift transformation, see (1). This is chosen as a feasible set of transformations that allow for an explicit evaluation of the Radon-Nikodym density. Hereafter, we study such transformations. We observe that, in the presence of time-change, such transformations do not lead to a self-preserving structure. As illustration we can see that the doubly stochastic Poisson random measure will not be of such structure after the measure change. We also recall that such shift transformations, when applied to Lévy noises, are actually structure preserving.

We introduce the Radon-Nikodym density process $Z_t = Z_t^\theta$, $t \in [0, T]$, by

$$\begin{aligned} dZ_t^\theta &= Z_t^\theta \left(\theta_t^B dB_t + \int_{\mathbb{R}_0} \theta_t^H(z) \tilde{H}(dt, dz) \right), \\ Z_0^\theta &= 1, \end{aligned} \tag{23}$$

for $\theta \in \mathcal{J}_{\mathbb{G}}$ with $\theta_t(0) = \theta_t^B$ and $\theta_t(z) = \theta_t^H(z)$, $z \in \mathbb{R}_0$, and $\theta_t^H(z) > -1$ $\mathbb{P} \times \Lambda$ - a.e.

An explicit strong solution of (23) is obtained by application of the Itô formula:

$$\begin{aligned} Z_t^\theta = & \exp \left(\int_0^t \theta_s^B dB_s - \int_0^t \frac{1}{2} (\theta_s^B)^2 \lambda_s^B ds \right. \\ & + \int_0^t \int_{\mathbb{R}_0} \left[\ln \left(1 + \theta_s^H(z) \right) - \theta_s^H(z) \right] v(dz) \lambda_s^H ds \\ & \left. + \int_0^t \int_{\mathbb{R}_0} \ln \left(1 + \theta_s^H(z) \right) \tilde{H}(ds, dz) \right), \quad t \in [0, T]. \end{aligned}$$

Since we have assumed $\theta_t^H(z) > -1$ $\mathbb{P} \times \Lambda$ -a.e., we know that $\ln \left(1 + \theta_s^H(z) \right)$ and the stochastic integration are well-defined. A generalized version of the Novikov condition, see [30], ensures uniform integrability of Z .

Theorem 3.1 *Let B and $\tilde{H}$ be as in Definition 2.1 with respect to $\mathbb{P}$. Assume that $Z_t = Z_t^\theta$, $t \in [0, T]$, with $\theta \in \mathcal{I}_{\mathbb{G}}$, is a positive uniformly integrable $(\mathbb{G}, \mathbb{P})$ -martingale with $\mathbb{E}[Z_T^2] < \infty$, and define the probability measure $\mathbb{Q}$, equivalent to $\mathbb{P}$, by*

$$\frac{d\mathbb{Q}}{d\mathbb{P}} = Z_T.$$

Define B^θ and $\tilde{H}^\theta$ by the dynamics

$$\begin{aligned} dB_t^\theta &:= dB_t - \theta_t^B d\Lambda_t^B, \\ \tilde{H}^\theta(dt, dz) &:= \tilde{H}(dt, dz) - \theta_t^H(z) \Lambda^H(dt, dz), \end{aligned}$$

where we recall that $\Lambda^B(dt, \{0\}) = d\Lambda_t^B = \lambda_t^B dt$ and $\Lambda^H(dt, dz) = v(dz) \lambda_t^H dt$. Moreover, for any bounded predictable ψ such that $\int_0^T \int_{\mathbb{R}_0} \psi_t(z) \Lambda^H(dt, dz) < \infty$, $\mathbb{P}$ -a.s., define the process

$$M_t^\theta(\psi) := \int_0^t \int_{\mathbb{R}_0} \psi_s(z) \tilde{H}^\theta(ds, dz), \quad 0 \leq t \leq T.$$

Then B^θ is a continuous $(\mathbb{G}, \mathbb{Q})$ -martingale and a time-changed $(\mathbb{G}, \mathbb{Q})$ -Brownian motion. Also, $M^\theta(\psi)$ is a $(\mathbb{G}, \mathbb{Q})$ -martingale, where $\tilde{H}^\theta$ is a $(\mathbb{G}, \mathbb{Q})$ -martingale random field.

Moreover, if

$$\mathbb{E}_{\mathbb{Q}} \left[\sup_{t \in [0, T]} |[B^\theta, M^\theta(\psi)]_t| \right] < \infty, \quad (24)$$

for $\psi_t(z) = 1_\Delta(t, z)$, $\Delta \in \mathcal{B}_{[0, T] \times \mathbb{R}_0}$: $m(\Delta) = E[\Lambda^H(\Delta)] < \infty$, then B^θ and $M^\theta(\psi)$ are strongly orthogonal under $\mathbb{Q}$.

We recall that two $\mathbb{G}$ -martingales with values in $L^2(\mathbb{Q})$ are strongly orthogonal if their product is a uniformly integrable $(\mathbb{G}, \mathbb{Q})$ -martingale or, equivalently, if their quadratic variation process is a uniformly integrable $(\mathbb{G}, \mathbb{Q})$ -martingale.

Proof With $\varepsilon \in [0, 1]$, define

$$X_t^\varepsilon := \varepsilon B_t^\theta + M_t^\theta(\psi).$$

We have

$$dX_t^\varepsilon = -\alpha_t^\varepsilon dt + \varepsilon dB_t + \int_{\mathbb{R}_0} \psi_t(z) \tilde{H}(dt, dz),$$

where

$$\alpha_t^\varepsilon = \varepsilon \theta_t^B \lambda_t^B + \int_{\mathbb{R}_0} \psi_t(z) \theta_t^H(z) \nu(dz) \lambda_t^H.$$

From Lemma 1.27 in [34], we know that if $Z_t X_t^\varepsilon$ is a local $(\mathbb{G}, \mathbb{P})$ -martingale, then X_t^ε is a local $(\mathbb{G}, \mathbb{Q})$ -martingale. From Definition 1.28 and Example 1.29 in [34], and recalling that

$$dZ_t = Z_{t-} \left(\theta_t^B dB_t + \int_{\mathbb{R}_0} \theta_t^H(z) \tilde{H}(dt, dz) \right),$$

we get the following:

$$\begin{aligned} d(Z_t X_t^\varepsilon) &= Z_{t-} dX_t^\varepsilon + X_{t-}^\varepsilon dZ_t + dZ_t dX_t^\varepsilon \\ &= Z_{t-} \left(\varepsilon + X_{t-}^\varepsilon \theta_t^B \right) dB_t \\ &\quad + Z_{t-} \int_{\mathbb{R}_0} \left(X_{t-}^\varepsilon \theta_t^H(z) + \psi_t(z) + \psi_t(z) \theta_t^H(z) \right) \tilde{H}(dt, dz). \end{aligned}$$

Thus, X_t^ε is a local $(\mathbb{G}, \mathbb{Q})$ -martingale for all $\varepsilon \in [0, 1]$. In particular, $X_t^0 = M_t^\theta(\psi)$ is a local $(\mathbb{G}, \mathbb{Q})$ -martingale. Moreover, $B_t^\theta = X_t^1 - M_t^\theta(\psi)$ is also a local $(\mathbb{G}, \mathbb{Q})$ -martingale.

Since B^θ is a continuous local martingale ($B_0^\theta = 0$), with quadratic variation $[B^\theta, B^\theta]_t = [B, B]_t = \Lambda_t^B$ (the quadratic variation is invariant under equivalent measure change), then B^θ is a time-changed $(\mathbb{G}, \mathbb{Q})$ -Brownian motion, see Theorem 16.4 in [27]. Hence, it is also a doubly stochastic Gaussian measure as in Definition 2.1.

As for $M^\theta(\psi)$, we can see that its quadratic variation is

$$[M^\theta(\psi), M^\theta(\psi)]_t = \int_0^t \int_{\mathbb{R}_0} \psi_s^2(z) H(ds, dz).$$

Now, let $\psi_t(z) = 1_{(0,t] \times B}(t, z)$, for $t \in [0, T]$ and $B \in \mathcal{B}_{\mathbb{R}_0}$. Then,

$$\begin{aligned} \mathbb{E}_{\mathbb{Q}} \left[[M^\theta(\psi), M^\theta(\psi)]_T \right] &= \mathbb{E}_{\mathbb{Q}} \left[\int_0^T \int_{\mathbb{R}_0} 1_{(0,t] \times B}(t, z) H(dt, dz) \right] \\ &= \mathbb{E} \left[Z_T \int_0^T \int_{\mathbb{R}_0} 1_{(0,t] \times B}(t, z) H(dt, dz) \right]. \end{aligned}$$

By Hölder's inequality, we have that

$$\mathbb{E}_{\mathbb{Q}} \left[[M^\theta(\psi), M^\theta(\psi)]_T \right] \leq \left(\mathbb{E}[Z_T^2] \right)^{1/2} \cdot \left(\mathbb{E} \left[H((0, T] \times B)^2 \right] \right)^{1/2} < \infty. \quad (25)$$

Hence $M^\theta(\psi)$ is a $(\mathbb{G}, \mathbb{Q})$ -martingale, see e.g. Corollary to Theorem 27.II in [37]. Denote $\mathcal{B}$ a semi-ring generating $\mathcal{B}_{\mathbb{R}_0}$. We can regard the σ -algebra $\mathcal{B}_{(0,T]}$ as generated by the semi-ring of intervals of the form $(s, t]$, where $0 \leq s < t \leq T$. The σ -algebra $\mathcal{B}_{(0,T] \times \mathbb{R}_0}$ is generated by the semi-ring of sets $(s, t] \times A$, where $A \in \mathcal{B}$. For an element $(s, t] \times A$ in the semi-ring, let $\psi = 1_{(0,t] \times A}$. Then we have

$$\tilde{H}^\theta((s, t] \times A) = M_t^\theta(\psi) - M_s^\theta(\psi). \quad (26)$$

By (25) and (26) $\tilde{H}^\theta$ is σ -finite on the semi-ring $\mathbb{P}$ -a.s. (equivalently $\mathbb{Q}$ -a.s.), hence we can uniquely extend (26) to the σ -algebra $\mathcal{B}_{[0,T] \times \mathbb{R}_0}$, see Theorems 11.3 and 10.3 in [5]. Hence, $\tilde{H}^\theta$ has the $(\mathbb{G}, \mathbb{Q})$ -martingale property, conditionally orthogonal values with respect to $(\mathbb{G}, \mathbb{Q})$, and its variance measure is σ -finite. $\tilde{H}^\theta$ is clearly $\mathbb{G}$ -adapted by its definition, and $\tilde{H}^\theta$ is additive and σ -additive in $L^2(\mathbb{Q})$ by its integral form and the condition on θ^H . In conclusion, $\tilde{H}^\theta$ is a $(\mathbb{G}, \mathbb{Q})$ -martingale random field with conditionally orthogonal values.

Finally, we show that B^θ and $M^\theta(\psi)$ are strongly orthogonal under $\mathbb{Q}$, for $\psi = 1_\Delta$ with $\Delta \in \mathcal{B}_{[0,T] \times \mathbb{R}_0}$: $m(\Delta) = \mathbb{E}[\Lambda^H(\Delta)] < \infty$. In fact, observe that

$$\begin{aligned} B_t^\theta &:= B_t - \left\langle B, \int_0^\cdot \theta_s^B dB_s \right\rangle_t, \\ M_t^\theta(\psi) &= \int_0^t \int_{\mathbb{R}_0} 1_\Delta(s, z) \tilde{H}^\theta(dt, dz) \\ &:= \int_0^t \int_{\mathbb{R}_0} 1_\Delta(s, z) \tilde{H}(dt, dz) - \left\langle \int_0^\cdot \int_{\mathbb{R}_0} 1_\Delta(s, z) \tilde{H}(ds, dz), \int_0^\cdot \int_{\mathbb{R}_0} \theta_s^H(z) \tilde{H}(ds, dz) \right\rangle_t, \end{aligned}$$

Then

$$\langle B^\theta, M^\theta(\psi) \rangle_t = \left\langle B, \int_0^\cdot \int_{\mathbb{R}_0} 1_\Delta(s, z) \tilde{H}(dt, dz) \right\rangle_t = 0,$$

as a consequence of (A5) in Definition 2.1. From this we know that $[B^\theta, M^\theta(\psi)]_t$ is a $\mathbb{Q}$ -local martingale and, by (24) and Theorem 51.I in [37], we get that $[B^\theta, M^\theta(\psi)]$

is a uniformly integrable $\mathbb{Q}$ -martingale. Then B_t^θ and $M_t^\theta(\psi)$ are strongly orthogonal square integrable martingales.

Remark 3.1 If (23) is defined with $\theta \in \mathcal{J}_{\mathbb{F}}$, then Z^θ is an $\mathbb{F}$ -adapted process. In this case, the fields B^θ and $\tilde{H}^\theta$ would be strongly orthogonal $(\mathbb{F}, \mathbb{Q})$ -martingale random fields in the sense discussed above.

Note that $\tilde{H}^\theta$ is not a doubly stochastic Poisson random field under $\mathbb{Q}$, in general.

Corollary 1 *Let $\tilde{H}^\theta$ and Z be defined as in Theorem 3.1. If the stochastic field θ^H is deterministic, then $\tilde{H}^\theta$ is a $(\mathbb{G}, \mathbb{Q})$ -centered doubly stochastic Poisson random field. Moreover, if $\theta_t^H(z) = \theta_t^H$, then the new jump measure and the new time distortion process are given by*

$$v^\theta(dz) = v(dz), \quad \lambda_t^\theta(\omega) = \{1 + \theta_t^H\} \lambda_t(\omega).$$

If $\theta_t^H(z) = \theta^H(z)$, then the new jump measure and the new time distortion process are given by

$$v^\theta(dz) = \{1 + \theta^H(z)\} v(dz), \quad \lambda_t^\theta(\omega) = \lambda_t(\omega).$$

Proof This can be shown by studying the characteristic function (under $\mathbb{Q}$) of $\tilde{H}^\theta(\Delta)$ for $\Delta \in \mathcal{B}_{[0, T] \times \mathbb{R}_0}$.

4 Hedging Under Worst Case Scenario

In this section we define explicitly the set of scenarios $\mathcal{Q}_{\mathbb{M}}$ considered in the definition of the risk measure in (1):

$$\rho_t(\xi) := \operatorname{ess\,sup}_{\mathbb{Q} \in \mathcal{Q}_{\mathbb{M}}} \mathbb{E}_{\mathbb{Q}}[-\xi | \mathcal{M}_t], \quad t \in [0, T].$$

We consider the cases of information flow given by $\mathbb{M} = \mathbb{G}, \mathbb{F}$. Considering the agent's perspective, it is natural to choose the filtration $\mathbb{F}$ as model for the information flow. In fact $\mathbb{G}$ carries the information of the whole process of stochastic time-change, which would be a form of anticipating information embedded in the information flow not reasonably available to an agent. We study the optimisation problem (7) in both cases on the market (2) and (3) with the $\mathbb{F}$ -adapted coefficients $r, \alpha, \sigma, \gamma$.

Definition 4.1 Let the process $Z_t, t \in [0, T]$, be a $(\mathbb{M}, \mathbb{P})$ -martingale defined by

$$\begin{aligned} dZ_t &= Z_{t-} \left(\theta_t^B dB_t + \int_{\mathbb{R}_0} \theta_t^H(z) \tilde{H}(dt, dz) \right), \\ Z_0 &= 1, \end{aligned}$$

for $\theta \in \mathcal{J}_{\mathbb{G}}$ (with the notation $\theta_t(0) = \theta_t^B$ and $\theta_t(z) = \theta_t^H(z)$ for $z \in \mathbb{R}_0$) and $\theta_t^H(z) > -1$ $\mathbb{P} \times \Lambda - a.e.$ Consider the cases such that $Z_T \in L^2(\mathbb{P})$ and (24) is satisfied. Moreover, for $K > 0$,

$$|\theta_t^B \lambda_t^B| < K, \quad 0 \leq \theta_t^H(z) \sqrt{\lambda_t^H} < K \cdot z, \quad z \in \mathbb{R}_0, \quad \mathbb{P} \times dt - a.e. \quad (27)$$

Then the set of admissible scenarios is given by:

$$\mathcal{Q}_{\mathbb{M}} := \{\mathbb{Q} \sim \mathbb{P} \mid \frac{d\mathbb{Q}}{d\mathbb{P}} = Z_T^\theta, \quad \theta \in \mathcal{J}_{\mathbb{M}}\},$$

where the Radon-Nikodym derivative are of the type above.

We remark that $\mathcal{Q}_{\mathbb{M}}$ is a convex set. Moreover, $\mathcal{Q}_{\mathbb{F}} \subseteq \mathcal{Q}_{\mathbb{G}}$.

We recall that the hedging problem considered (6), with the chosen risk measure (1), translates to the problem (7):

$$Y_t = \operatorname{ess\,inf}_{\pi \in \Pi_{\mathbb{M}}} \operatorname{ess\,sup}_{\mathbb{Q} \in \mathcal{Q}_{\mathbb{M}}} Y_t^{\pi, \mathbb{Q}} \quad (28)$$

with

$$\begin{aligned} Y_t^{\pi, \mathbb{Q}} := & E_{\mathbb{Q}} \left[e^{-\int_t^T r_s ds} F - \int_t^T e^{-\int_t^s r_u du} \pi_s (\alpha_s - r_s) ds \right. \\ & - \int_t^T e^{-\int_t^s r_u du} \pi_s \sigma_s dB_s \\ & \left. - \int_t^T \int_{\mathbb{R}_0} e^{-\int_t^s r_u du} \pi_s \gamma_s(z) \tilde{H}(ds, dz) \middle| \mathcal{M}_t \right]. \end{aligned}$$

Hence, a solution to (7) is given by $(\hat{\pi}, \hat{\mathbb{Q}}) \in \Pi_{\mathbb{M}} \times \mathcal{Q}_{\mathbb{M}}$ such that

$$\begin{aligned} Y_t = Y_t^{\hat{\pi}, \hat{\mathbb{Q}}} = & E_{\hat{\mathbb{Q}}} \left[e^{-\int_t^T r_s ds} F - \int_t^T e^{-\int_t^s r_u du} \hat{\pi}_s (\alpha_s - r_s) ds \right. \\ & - \int_t^T e^{-\int_t^s r_u du} \hat{\pi}_s \sigma_s dB_s \\ & \left. - \int_t^T \int_{\mathbb{R}_0} e^{-\int_t^s r_u du} \hat{\pi}_s \gamma_s(z) \tilde{H}(ds, dz) \middle| \mathcal{M}_t \right], \quad t \in [0, T]. \quad (29) \end{aligned}$$

Analogously, we define the set of admissible portfolio $\Pi_{\mathbb{M}}$ with respect to the filtrations $\mathbb{M} = \mathbb{F}, \mathbb{G}$.

Definition 4.2 The portfolio $\pi : [0, T] \times \Omega \rightarrow \mathbb{R}$ is admissible if

- (i) $\pi_t \gamma_t(z) > -1$ $\mathbb{P}$ -a.s.,
- (ii) π is $\mathbb{M}$ -predictable such that there exists a unique strong càdlàg $\mathbb{M}$ -adapted solution V^π to the dynamics (4) on $[0, T]$,
- (iii) for all $\mathbb{Q} \in \mathcal{Q}_{\mathbb{M}}$

$$\mathbb{E}_{\mathbb{Q}} \left[\int_0^T \left[|\alpha_t - r_t| |\pi_t| + |\pi_t \sigma_t|^2 \lambda_t^B + \int_{\mathbb{R}_0} |\pi_t \gamma_t(z)|^2 \nu(dz) \lambda_t^H \right] dt \right] < \infty.$$

Note that $\Pi_{\mathbb{M}}$ is a convex set.

The solution to the problem (28) is studied via BSDEs and the comparison theorem. The two filtrations lead to different types of BSDEs.

4.1 Flow of Information $\mathbb{G}$

First of all we consider the filtration $\mathbb{G}$ and the corresponding stochastic process:

$$Y_t^{\pi, \theta} = \mathbb{E}_{\mathbb{Q}} \left[e^{-\int_t^T r_s ds} F - \int_t^T e^{-\int_t^s r_u du} \pi_s (\alpha_s - r_s) ds - \int_t^T e^{-\int_t^s r_u du} \pi_s \sigma_s dB_s - \int_t^T \int_{\mathbb{R}_0} e^{-\int_t^s r_u du} \pi_s \gamma_s(z) \tilde{H}(ds, dz) \middle| \mathcal{G}_t \right],$$

where $\pi \in \Pi_{\mathbb{G}}$ and $\mathbb{Q} = \mathbb{Q}^\theta \in \mathcal{Q}_{\mathbb{G}}$. By Theorem 3.1, define the $\mathbb{G}$ -martingale random fields:

$$dB_t^\theta := dB_t - \theta_t^B \lambda_t^B dt,$$

and

$$\tilde{H}^\theta(dt, dz) := \tilde{H}(dt, dz) - \theta_t^H(z) \nu(dz) \lambda_t^H dt.$$

So when $\mathbb{M} = \mathbb{G}$, $Y^{\pi, \mathbb{Q}}$ takes the form:

$$\begin{aligned} Y_t^{\pi, \theta} = \mathbb{E}_{\mathbb{Q}} \left[e^{-\int_t^T r_s ds} F - \int_t^T e^{-\int_t^s r_u du} \pi_s \left((\alpha_s - r_s) + \sigma_s \theta_s^B \lambda_s^B \right. \right. \\ \left. \left. + \int_{\mathbb{R}_0} \gamma_s(z) \theta_s^H(z) \nu(dz) \lambda_s^H \right) ds - \int_t^T e^{-\int_t^s r_u du} \pi_s \sigma_s dB_s^\theta \right. \\ \left. - \int_t^T \int_{\mathbb{R}_0} e^{-\int_t^s r_u du} \pi_s \gamma_s(z) \tilde{H}^\theta(ds, dz) \middle| \mathcal{G}_t \right] \end{aligned}$$

$$\begin{aligned}
&= \mathbb{E}_{\mathbb{Q}} \left[e^{-\int_t^T r_s ds} F - \int_t^T e^{-\int_t^s r_u du} \pi_s \left((\alpha_s - r_s) + \sigma_s \theta_s^B \lambda_s^B \right. \right. \\
&\quad \left. \left. + \int_{\mathbb{R}_0} \gamma_s(z) \theta_s^H(z) v(dz) \lambda_s^H \right) ds \middle| \mathcal{G}_t \right]. \quad (30)
\end{aligned}$$

Hence,

$$\begin{aligned}
e^{-\int_0^t r_u du} Y_t^{\pi, \theta} &= \mathbb{E}_{\mathbb{Q}} \left[e^{-\int_0^T r_s ds} F - \int_0^T e^{-\int_0^s r_u du} \pi_s \right. \\
&\quad \times \left((\alpha_s - r_s) + \sigma_s \theta_s^B \lambda_s^B + \int_{\mathbb{R}_0} \gamma_s(z) \theta_s^H(z) v(dz) \lambda_s^H \right) ds \middle| \mathcal{G}_t \right] \quad (31) \\
&\quad + \int_0^t e^{-\int_0^s r_u du} \pi_s \left((\alpha_s - r_s) + \sigma_s \theta_s^B \lambda_s^B + \int_{\mathbb{R}_0} \gamma_s(z) \theta_s^H(z) v(dz) \lambda_s^H \right) ds.
\end{aligned}$$

The martingale representation Theorem 2.3 applied to

$$\xi := e^{-\int_0^T r_s ds} F - \int_0^T e^{-\int_0^s r_u du} \pi_s \left((\alpha_s - r_s) + \sigma_s \theta_s^B \lambda_s^B + \int_{\mathbb{R}_0} \gamma_s(z) \theta_s^H(z) v(dz) \lambda_s^H \right) ds \quad (32)$$

gives us the existence of the $\mathbb{G}$ -predictable integrands $Z^{\pi, \theta}$ and $U^{\pi, \theta}$ for the two corresponding stochastic integrals, so we have

$$\begin{aligned}
e^{-\int_0^t r_u du} Y_t^{\pi, \theta} &= \mathbb{E}_{\mathbb{Q}}[\xi | \mathcal{F}_t^{\Lambda}] + \int_0^t Z_s^{\pi, \theta} dB_s^{\theta} + \int_0^t \int_{\mathbb{R}_0} U_s^{\pi, \theta}(z) \tilde{H}^{\theta}(ds, dz) \\
&\quad + \int_0^t e^{-\int_0^s r_u du} \pi_s \left((\alpha_s - r_s) + \sigma_s \theta_s^B \lambda_s^B + \int_{\mathbb{R}_0} \gamma_s(z) \theta_s^H(z) v(dz) \lambda_s^H \right) ds.
\end{aligned}$$

The Itô formula allows to obtain the linear BSDE

$$\begin{aligned}
dY_t^{\pi, \theta} &= \left(r_t Y_t^{\pi, \theta} + \pi_t (\alpha_t - r_t) + \left(\pi_t \sigma_t - e^{\int_0^t r_u du} Z_t^{\pi, \theta} \right) \theta_t^B \lambda_t^B \right. \\
&\quad \left. + \int_{\mathbb{R}_0} \left(\pi_t \gamma_t(z) - e^{\int_0^t r_u du} U_t^{\pi, \theta}(z) \right) \theta_t^H(z) v(dz) \lambda_t^H \right) dt \\
&\quad + e^{\int_0^t r_u du} Z_t^{\pi, \theta} dB_t + \int_{\mathbb{R}_0} e^{\int_0^t r_u du} U_t^{\pi, \theta}(z) \tilde{H}(dt, dz), \quad (33) \\
Y_T^{\pi, \theta} &= F,
\end{aligned}$$

the solution of which is guaranteed by Theorem 2.7 thanks to (27). The generator of this BSDE is:

$$\begin{aligned}
g(\lambda, y, z, u(\cdot), \pi, \theta) = & -yr - (\mu - r)\pi - (\pi\sigma - e^{\int_0^t r_s ds} z)\theta^B \lambda^B \\
& - \int_{\mathbb{R}_0} (\pi \gamma(x) - e^{\int_0^t r_s ds} u(x))\theta^H(x) \nu(dx) \lambda^H. \quad (34)
\end{aligned}$$

The min-max type of problem corresponding to (28) arises in stochastic differential games. With the comparison Theorem 2.8 in hands, we can justify the proof of the following result due to [35] in our setting for the driving noises considered in this paper.

As short hand notation, denote $g(\pi_t, \theta_t) = g(\lambda, y, z, u(\cdot), \pi_t, \theta_t)$. The solution of a BSDE with standard parameters (ξ, g) is denote by (Y, Z, U) , for an optimal $\hat{\theta}$ the solution is denoted by (Y^π, Z^π, U^π) , and for an optimal $\hat{\pi}$ the solution is then $(Y^\theta, Z^\theta, U^\theta)$. The solution given in the case $g(\hat{\pi}_t, \hat{\theta}_t)$ is denoted $(\hat{Y}, \hat{Z}, \hat{U})$.

Theorem 4.1 *Let (ξ, g) be standard parameters. Suppose that for all $(\omega, t, \lambda, y, z, u)$ there exist $\hat{\pi}_t = \hat{\pi}(\omega, t, \lambda, y, z, u)$ and $\hat{\theta}_t = \hat{\theta}(\omega, t, \lambda, y, z, u)$ such that for all admissible portfolios $\pi \in \Pi_{\mathbb{G}}$ and all admissible probability measures $\mathbb{Q} = \mathbb{Q}^\theta \in \mathcal{Q}_{\mathbb{G}}$, we have:*

$$g(\hat{\pi}_t, \theta_t) \leq g(\pi_t, \theta_t) \leq g(\pi_t, \hat{\theta}_t), \quad (35)$$

for a.a. (ω, t) . Assume that the conditions of Theorem 2.8 hold. Suppose $\hat{\pi}$ and $\hat{\theta}$ are admissible, and suppose that for all admissible (π, θ) there exists a unique solution to the BSDE with $(\xi, g(\pi_t, \theta_t))$ as terminal condition and generator, respectively. Then

$$\hat{Y}_t = Y_t^{\hat{\pi}} = \operatorname{ess\,inf}_{\pi \in \Pi_{\mathbb{G}}} Y_t^\pi =: Y_t = \operatorname{ess\,sup}_{\mathbb{Q} \in \mathcal{Q}_{\mathbb{G}}} \left\{ \operatorname{ess\,inf}_{\pi \in \Pi_{\mathbb{G}}} Y_t^{\pi, \theta} \right\} = \operatorname{ess\,sup}_{\mathbb{Q} \in \mathcal{Q}_{\mathbb{G}}} Y_t^\theta.$$

Proof The proof is due to [35]. By applying the comparison theorem to the solutions of the BSDEs of the couples of standard parameters $(F, g(\hat{\pi}_t, \theta_t))$, $(F, g(\pi_t, \theta_t))$, $(F, g(\pi_t, \hat{\theta}_t))$, by (35) we get that $Y_t^\theta \leq Y_t^{\pi, \theta} \leq Y_t^\pi$, for all admissible (π, θ) , thus:

$$\text{For all } \theta : Y_t^\theta \leq \operatorname{ess\,inf}_{\pi \in \Pi} Y_t^{\pi, \theta} \quad \mathbb{P} \times dt - a.e. \quad (36)$$

$$\text{For all } \pi : \operatorname{ess\,sup}_{\theta \in \mathcal{Q}} Y_t^{\pi, \theta} \leq Y_t^\pi \quad \mathbb{P} \times dt - a.e. \quad (37)$$

From definition of essential supremum and (36), we get

$$\hat{Y}_t \leq \operatorname{ess\,sup}_{\mathbb{Q} \in \mathcal{Q}_{\mathbb{G}}} Y_t^\theta = \operatorname{ess\,sup}_{\mathbb{Q} \in \mathcal{Q}_{\mathbb{G}}} \left(\operatorname{ess\,inf}_{\pi \in \Pi_{\mathbb{G}}} Y_t^{\pi, \theta} \right).$$

From (37) and definition of essential infimum, we get

$$Y_t = \operatorname{ess\,inf}_{\pi \in \Pi_{\mathbb{G}}} \left\{ \operatorname{ess\,sup}_{\mathbb{Q} \in \mathcal{Q}_{\mathbb{G}}} Y_t^{\pi, \theta} \right\} \leq \operatorname{ess\,inf}_{\pi \in \Pi_{\mathbb{G}}} Y_t^{\pi} \leq \hat{Y}_t.$$

Hence, we obtain the following chain of inequalities:

$$\begin{aligned} Y_t &= \operatorname{ess\,inf}_{\pi \in \Pi_{\mathbb{G}}} \left\{ \operatorname{ess\,sup}_{\mathbb{Q} \in \mathcal{Q}_{\mathbb{G}}} Y_t^{\pi, \theta} \right\} \leq \operatorname{ess\,inf}_{\pi \in \Pi_{\mathbb{G}}} Y_t^{\pi} \leq \hat{Y}_t \\ &\leq \operatorname{ess\,sup}_{\mathbb{Q} \in \mathcal{Q}_{\mathbb{G}}} Y_t^{\theta} \leq \operatorname{ess\,sup}_{\mathbb{Q} \in \mathcal{Q}_{\mathbb{G}}} \left(\operatorname{ess\,inf}_{\pi \in \Pi_{\mathbb{G}}} Y_t^{\pi, \theta} \right). \end{aligned}$$

Since $\sup(\inf) \leq \inf(\sup)$ we get equality between all terms.

We shall apply this result. The generator (34) satisfies the conditions of $g^{(2)}$ in Theorem 2.8. In fact for an admissible probability measure $\mathbb{Q}^{\theta}$ and an admissible π , we have:

$$\begin{aligned} g_t(\lambda, y, \zeta, u(\cdot), \pi_t, \theta_t) \\ &= -yr_t - \pi_t \left[\alpha_t - r_t + \sigma_t \theta_t^B \lambda^B + \int_{\mathbb{R}_0} \gamma_t(z) \theta_t^H(z) v(dz) \lambda^H \right] \\ &\quad + e^{\int_0^t r_s ds} \zeta \theta_t^B \sqrt{\lambda^B} \sqrt{\lambda^B} + \int_{\mathbb{R}_0} e^{\int_0^t r_s ds} u(z) \sqrt{\lambda^H} \theta_t^H(z) v(dz) \sqrt{\lambda^H}. \end{aligned}$$

Moreover, condition (35) leads to the study of the equations

$$\frac{\partial g}{\partial \theta^B}(\hat{\pi}, \hat{\theta}) = 0, \quad \frac{\partial g}{\partial \theta^H}(\hat{\pi}, \hat{\theta}) = 0, \quad \frac{\partial g}{\partial \pi}(\hat{\pi}, \hat{\theta}) = 0.$$

The determinant of the Hessian is null, and these equations correspond to a critical point. Recall that $\mathcal{Q}_{\mathbb{G}}$ and $\Pi_{\mathbb{G}}$ are convex. These yield to the characterising equations for the optimal solution.

Summarising, we have the following result.

Theorem 4.2 *Let the reference filtration be $\mathbb{G}$. If $(\hat{\pi}, \hat{\mathbb{Q}}) \in \Pi_{\mathbb{G}} \times \mathcal{Q}_{\mathbb{G}}$ satisfy the equations:*

$$\left(e^{\int_0^t r_s ds} \hat{Z}_t - \hat{\pi}_t \sigma_t \right) \lambda_t^B = 0, \quad (38)$$

$$\int_{\mathbb{R}_0} \left(e^{\int_0^t r_s ds} \hat{U}_t(z) - \hat{\pi}_t \gamma_t(z) \right) v(dz) \lambda_t^H = 0, \quad (39)$$

$$(\alpha_t - r_t) + \sigma_t \hat{\theta}_t^B \lambda_t^B + \int_{\mathbb{R}_0} \gamma_t(z) \hat{\theta}_t^H(z) v(dz) \lambda_t^H = 0, \quad (40)$$

where $(\hat{Z}, \hat{U}) \in \mathcal{I}_{\mathbb{G}}$ are the integrands in the integral representation (Theorem 2.2):

$$e^{-\int_0^T r_t dt} F = \mathbb{E}_{\hat{\mathbb{Q}}}[e^{-\int_0^T r_t dt} F | \mathcal{F}_T^{\Lambda}] + \int_0^T \hat{Z}_s dB_s^{\hat{\theta}} + \int_0^T \int_{\mathbb{R}_0} \hat{U}_s(z) \tilde{H}^{\hat{\theta}}(ds, dz), \quad (41)$$

then $(\hat{\pi}, \hat{\mathbb{Q}})$ is the optimal solution of the problem (8). The optimal price process $\hat{Y}_t = Y_t^{\hat{\pi}, \hat{\mathbb{Q}}} = Y_t$, $t \in [0, T]$, is given by:

$$\hat{Y}_t = \mathbb{E}_{\hat{\mathbb{Q}}}\left[e^{-\int_t^T r_s ds} F - \int_t^T e^{\int_0^s r_u du} \hat{Z}_s dB_s^{\hat{\theta}} - \int_t^T \int_{\mathbb{R}_0} e^{\int_0^s r_u du} \hat{U}_s(z) \tilde{H}^{\hat{\theta}}(ds, dz) | \mathcal{G}_t\right] \quad (42)$$

where

$$dB_t^{\hat{\theta}} := dB_t - \hat{\theta}_t^B \lambda_t^B dt$$

is a $(\mathbb{G}, \hat{\mathbb{Q}})$ -martingale and a time-changed $(\mathbb{G}, \hat{\mathbb{Q}})$ -Brownian motion, and

$$\tilde{H}^{\hat{\theta}}(dt, dz) := \tilde{H}(dt, dz) - \hat{\theta}_t^H(z) \nu(dz) \lambda_t^H dt.$$

is a $(\mathbb{G}, \hat{\mathbb{Q}})$ -martingale random field orthogonal to $B^{\hat{\theta}}$. Under probability measure $\mathbb{P}$, the optimal price process follows the following dynamics:

$$\begin{aligned} d\hat{Y}_t &= \left(\hat{Y}_t r_t + \hat{\pi}_t (\alpha_t - r_t) \right) dt + \hat{\pi}_t \sigma_t dB_t + \int_{\mathbb{R}_0} \hat{\pi}_t \gamma_t(z) \tilde{H}(dt dz) \\ \hat{Y}_0 &= E_{\hat{\mathbb{Q}}}[e^{-\int_0^T r_t dt} F | \mathcal{F}_T^{\Lambda}]. \end{aligned} \quad (43)$$

On the other side, if $(\hat{\pi}, \hat{\mathbb{Q}}) \in \Pi_{\mathbb{G}} \times \mathcal{Q}_{\mathbb{G}}$ is an optimal solution, then the equations (38)–(40) are satisfied.

Remark 4.1 We observe that the optimal strategy $(\hat{\pi}, V^{\hat{\pi}}, C^{\hat{\pi}})$ is then given by the process $\hat{\pi}$ as characterised above, the value process $V^{\hat{\pi}}$ from (4) has the initial value $v = V_0^{\hat{\pi}} = E_{\hat{\mathbb{Q}}}[e^{-\int_0^T r_t dt} F]$, and the initial cost $C_0^{\hat{\pi}} = E_{\hat{\mathbb{Q}}}[e^{-\int_0^T r_t dt} F | \mathcal{F}_T^{\Lambda}] - E_{\hat{\mathbb{Q}}}[e^{-\int_0^T r_t dt} F]$ and is $C_t^{\hat{\pi}} = 0$ for $t \in (0, T]$.

Remark 4.2 Observe that the optimal $\hat{\mathbb{Q}}$ is a martingale measure for the optimal price process $\hat{Y}$. In fact from (42), we have that

$$e^{-\int_0^t r_s ds} \hat{Y}_t = \mathbb{E}_{\hat{\mathbb{Q}}}[e^{-\int_0^T r_s ds} \hat{Y}_T | \mathcal{G}_t].$$

4.2 Flow of Information $\mathbb{F}$

In the case of flow of information $\mathbb{F}$, we see that the problem (7) leads to a different type of BSDE than the one considered so far.

Recall that $\mu(dt, dz) = 1_{\{0\}} B(dt, dz) + 1_{\mathbb{R}_0}(z) \tilde{H}(dt, dz)$, is an $(\mathbb{F}, \mathbb{P})$ -martingale random field and that its conditional variance measure is:

$$\langle \mu \rangle(dt, dz) = \delta_{\{0\}}(dz) \lambda_s^B ds + 1_{\mathbb{R}_0}(z) v(dz) \lambda_s^H ds.$$

Moreover, by Theorem 3.1 we define the $(\mathbb{F}, \mathbb{Q})$ -martingale random field, μ^θ , by

$$\mu^\theta(dt, dz) = \mu(dt, dz) - \beta_t(z) \langle \mu \rangle(dt, dz), \quad (44)$$

where β is given by:

$$\beta_t(z) := \theta_t^B 1_{\{0\}}(z) + \theta_t^H(z) 1_{\mathbb{R}_0}(z).$$

Note that $\langle \mu \rangle = \langle \mu^\theta \rangle$.

Let us consider the process $Y^{\pi, \mathbb{Q}}$ under filtration $\mathbb{F}$ for $\pi \in \Pi_{\mathbb{F}}$ and $\mathbb{Q} \in \mathcal{Q}_{\mathbb{F}}$:

$$\begin{aligned} Y_t^{\pi, \theta} = \mathbb{E}_{\mathbb{Q}} \Bigg[& e^{-\int_t^T r_s ds} F - \int_t^T e^{-\int_t^s r_u du} \pi_s (\alpha_s - r_s) ds \\ & - \int_t^T e^{-\int_t^s r_u du} \pi_s \sigma_s dB_s \\ & - \int_t^T \int_{\mathbb{R}_0} e^{-\int_t^s r_u du} \pi_s \gamma_s(z) \tilde{H}(ds, dz) \Big| \mathcal{F}_t \Bigg]. \end{aligned} \quad (45)$$

Then, by direct computation and using Theorem 2.4, we obtain the following equalities:

$$\begin{aligned} e^{-\int_0^t r_s ds} Y_t^{\pi, \theta} &= \mathbb{E}_{\mathbb{Q}} \Bigg[e^{-\int_0^T r_s ds} F - \int_0^T \int_{\mathbb{R}} e^{-\int_0^s r_u du} \pi_s \left[\left((\alpha_s - r_s) + \sigma_s \theta_s^B \right) 1_{\{0\}}(z) \right. \\ &\quad \left. + \gamma_s(z) \theta_s^H(z) 1_{\mathbb{R}_0}(z) \right] \langle \mu^\theta \rangle(ds, dz) \Big| \mathcal{F}_t \Bigg] \\ &\quad + \int_0^t \int_{\mathbb{R}} e^{-\int_0^s r_u du} \pi_s \left[\left((\alpha_s - r_s) + \sigma_s \theta_s^B \right) 1_{\{0\}}(z) \right. \\ &\quad \left. + \gamma_s(z) \theta_s^H(z) 1_{\mathbb{R}_0}(z) \right] \langle \mu^\theta \rangle(ds, dz) \\ &= \mathbb{E}_{\mathbb{Q}} \Bigg[\xi_0 + \int_0^T \int_{\mathbb{R}} Z_t 1_{\{0\}}(z) + U_t(z) 1_{\mathbb{R}_0}(z) \mu^\theta(dt, dz) \Big| \mathcal{F}_t \Bigg] \\ &\quad + \int_0^t \int_{\mathbb{R}} e^{-\int_0^s r_u du} \pi_s \left[\left((\alpha_s - r_s) + \sigma_s \theta_s^B \right) 1_{\{0\}}(z) \right. \\ &\quad \left. + \gamma_s(z) \theta_s^H(z) 1_{\mathbb{R}_0}(z) \right] \langle \mu^\theta \rangle(ds, dz). \end{aligned}$$

Equivalently,

$$\begin{aligned} e^{-\int_0^t r_s ds} Y_t^{\pi, \theta} &= \mathbb{E}_{\mathbb{Q}}[\xi_0 | \mathcal{F}_t] + \int_0^t \int_{\mathbb{R}} Z_s 1_{\{0\}}(z) + U_s(z) 1_{\mathbb{R}_0}(z) \mu^\theta(ds, dz) \\ &\quad + \int_0^t \int_{\mathbb{R}} e^{-\int_0^s r_u du} \pi_s \left[\left((\alpha_s - r_s) + \sigma_s \theta_s^B \right) 1_{\{0\}}(z) \right. \\ &\quad \left. + \gamma_s(z) \theta_s^H(z) 1_{\mathbb{R}_0}(z) \right] \langle \mu^\theta \rangle(ds, dz). \end{aligned}$$

Recall that the random variable $\xi_0 \in L^2(\Omega, \mathcal{F}, \mathbb{Q})$ is orthogonal to the stochastic integrals. The equation above, together with (44), yields the BSDE:

$$\begin{aligned} dY_t^{\pi, \theta} &= - \int_{\mathbb{R}} \left[\left\{ -r_t Y_t - \pi_t(\alpha_t - r_t) + \theta_t^B \{ e^{\int_0^t r_s ds} Z_t - \pi_t \sigma_t \} \right\} 1_{\{0\}}(z) \right. \\ &\quad \left. + \left\{ e^{\int_0^t r_s ds} U_t(z) - \pi_t \gamma_t(z) \right\} \theta_t^H(z) 1_{\mathbb{R}_0}(z) \right] \langle \mu \rangle(dt, dz) \\ &\quad + \int_{\mathbb{R}} \left[e^{\int_0^t r_s ds} Z_t 1_{\{0\}}(z) + e^{\int_0^t r_s ds} U_t(z) 1_{\mathbb{R}_0}(z) \right] \mu(dt, dz) \\ &\quad + e^{\int_0^t r_s ds} d\mathbb{E}_{\mathbb{Q}}[\xi_0 | \mathcal{F}_t] \\ Y_T^{\pi, \theta} &= F. \end{aligned} \tag{46}$$

We remark that $\mathbb{E}_{\mathbb{Q}}[\xi_0 | \mathcal{F}_t]$, $t \in [0, T]$, is an $(\mathbb{F}, \mathbb{Q})$ -martingale orthogonal to μ^θ . Thus the process $\int_0^t e^{\int_0^s r_u du} d\mathbb{E}_{\mathbb{Q}}[\xi_0 | \mathcal{F}_s]$, $t \in [0, T]$, is an $(\mathbb{F}, \mathbb{Q})$ -martingale orthogonal to μ^θ . By direct computation that this process is also an $(\mathbb{F}, \mathbb{P})$ -martingale orthogonal to μ . In fact, for $A \in \mathcal{B}(\mathbb{R})$, we have

$$\begin{aligned} \left\langle \int_0^\cdot e^{\int_0^s r_u du} d\mathbb{E}_{\mathbb{Q}}[\xi_0 | \mathcal{F}_s], \mu(A) \right\rangle &= \left\langle \int_0^\cdot e^{\int_0^s r_u du} d\mathbb{E}_{\mathbb{Q}}[\xi_0 | \mathcal{F}_s], \mu^\theta(A) \right\rangle \\ &\quad + \left\langle \int_0^\cdot e^{\int_0^s r_u du} d\mathbb{E}_{\mathbb{Q}}[\xi_0 | \mathcal{F}_s], \int_0^\cdot \int_{\mathbb{R}} \beta_t(z) d\langle \mu \rangle(dt, dz) \right\rangle = 0. \end{aligned}$$

Another way to look at (45) is by application of Theorem 3.1. In fact, setting $\psi_t := e^{-\int_0^t r_s ds}$, $t \in [0, T]$, we have:

$$\begin{aligned} \psi_t Y_t^{\pi, \theta} &= \mathbb{E}_{\mathbb{Q}} \left[\psi_T F - \int_t^T \psi_s \pi_s \left[(\alpha_s - r_s) ds - \sigma_s dB_s - \int_{\mathbb{R}_0} \gamma_s(z) \tilde{H}(ds, dz) \right] \middle| \mathcal{F}_t \right] \\ &= \mathbb{E}_{\mathbb{Q}} \left[\psi_T F - \int_t^T \psi_s \pi_s \left[(\alpha_s - r_s) ds + \sigma_s \left(dB_s^\theta + \theta_s^B \lambda_s^B ds \right) \right. \right. \\ &\quad \left. \left. + \int_{\mathbb{R}_0} \gamma_s(z) \left(\tilde{H}^\theta(ds, dz) + \theta_s^H(z) v(dz) \lambda_s^H ds \right) \right] \middle| \mathcal{F}_t \right] \end{aligned}$$

By use of the martingale property, we get

$$\begin{aligned} \psi_t Y_t^{\pi, \theta} = \mathbb{E}_{\mathbb{Q}} \left[\psi_T F - \int_t^T \int_{\mathbb{R}} \psi_s \pi_s \left[\left((\alpha_s - r_s) + \sigma_s \theta_s^B \right) 1_{\{0\}}(z) \right. \right. \\ \left. \left. + \gamma_s(z) \theta_s^H(z) 1_{\mathbb{R}_0}(z) \right] \langle M \rangle(ds, dz) \middle| \mathcal{F}_t \right]. \end{aligned} \quad (47)$$

With $d\mathbb{Q} = Z_T d\mathbb{P}$ and $Z_T = \mathcal{E}_T \left(\int_0^T \int_{\mathbb{R}} \beta_s(z) \mu(ds, dz) \right)$, we recognize (47) as the solution of the linear BSDE of type, cf. (22):

$$\begin{aligned} Y_t^{\pi, \theta} = \mathbb{E} \left[\frac{\Psi_T}{\Psi_t} F - \int_t^T \int_{\mathbb{R}} \frac{\Psi_s}{\Psi_t} \pi_s \left[\left((\alpha_s - r_s) + \sigma_s \theta_s^B \right) 1_{\{0\}}(z) \right. \right. \\ \left. \left. + \gamma_s(z) \theta_s^H(s, z) 1_{\mathbb{R}_0}(z) \right] \langle \mu \rangle(ds, dz) \middle| \mathcal{F}_t \right]. \end{aligned}$$

Here we recall that r is bounded, $\theta \in \mathcal{J}_{\mathbb{F}}$ so that $\mathbb{Q} \in \mathcal{Q}_{\mathbb{F}}$, and (5) holds. By Lemma 2.9 and Remark 2.4, the $\mathbb{P}$ -dynamics correspond to

$$\begin{aligned} dY_t^{\pi, \theta} = & - \int_{\mathbb{R}} \left\{ -r_t 1_{\{0\}}(z) Y_t + \theta_t^B 1_{\{0\}}(z) \bar{Z}_t + \theta_t^H(z) 1_{\mathbb{R}_0}(z) \bar{U}_t(z) \right. \\ & \left. - \pi_t \left[\left((\alpha_t - r_t) + \sigma_t \theta_t^B \right) 1_{\{0\}}(z) + \gamma_t(z) \theta_t^H(z) 1_{\mathbb{R}_0}(z) \right] \right\} \langle \mu \rangle(dt, dz) \\ & + \int_{\mathbb{R}} \bar{Z}_t 1_{\{0\}}(z) + \bar{U}_t(z) 1_{\mathbb{R}_0}(z) \mu(dt, dz) + dN_t \\ Y_T^{\pi, \theta} = & F, \end{aligned} \quad (48)$$

where N is an $(\mathbb{F}, \mathbb{P})$ -martingale orthogonal to μ . Comparing (48) and (46). We see that

$$\begin{aligned} \bar{Z}_t &= e^{\int_0^t r_s ds} Z_t \\ \bar{U}_t(z) &= e^{\int_0^t r_s ds} U_t(z), \\ N_t &= \int_0^t e^{\int_0^s r_u du} d\mathbb{E}_{\mathbb{Q}}[\xi_0 | \mathcal{F}_s], \quad (N_0 = 0). \end{aligned}$$

We can state the corresponding result to Theorem 4.1 for the case of information flow $\mathbb{F}$. Set $f(\pi_t, \theta_t) := f_t(\lambda, y, z, u, \pi_t, \theta_t)$ as short-hand notation.

Theorem 4.3 *Let (ξ, f) be standard parameters. Suppose that for all $(\omega, t, \lambda, y, z, u)$ there exist $\hat{\pi}_t = \hat{\pi}(\omega, t, \lambda, y, z, u)$ and $\hat{\theta}_t = \hat{\theta}(\omega, t, \lambda, y, z, u)$ such that, for all admissible portfolios $\pi \in \Pi_{\mathbb{F}}$ and all admissible probability measures $\mathbb{Q} = \mathbb{Q}^{\theta} \in \mathcal{Q}_{\mathbb{F}}$, we have:*

$$f(\hat{\pi}_t, \theta_t) \leq f(\pi_t, \theta_t) \leq f(\pi_t, \hat{\theta}_t), \quad (49)$$

for a.a. (ω, t) . Assume that the conditions of Theorem 2.10 hold, and that $\hat{\pi}$ and $\hat{\theta}$ are admissible. Suppose that for all admissible (π, θ) there exists a unique solution to the BSDE with $(\xi, f(\pi_t, \theta_t))$ as terminal condition and generator, respectively. Then

$$\hat{Y}_t = \operatorname{ess\,inf}_{\pi \in \Pi_{\mathbb{F}}} Y_t^\pi =: Y_t = \operatorname{ess\,sup}_{\mathbb{Q} \in \mathcal{Q}_{\mathbb{F}}} \left\{ \operatorname{ess\,inf}_{\pi \in \Pi_{\mathbb{F}}} Y_t^{\pi, \theta} \right\} = \operatorname{ess\,sup}_{\mathbb{Q} \in \mathcal{Q}_{\mathbb{F}}} Y_t^\theta.$$

Proof The argument is the same as for Theorem 4.1, but rely on the comparison theorem in the case $\mathbb{F}$.

From the BSDE in (48), we can see that

$$\begin{aligned} f_t(y, \zeta, u(\cdot), \pi, \theta) = & - \left\{ -r_t y - \pi_t(\alpha_t - r_t) + \theta^B(\zeta - \pi_t \sigma_t) \right\} 1_{\{0\}}(z) \\ & + \left\{ u(z) - \pi_t \gamma_t(z) \right\} \theta^H(z) 1_{\mathbb{R}_0}(z), \end{aligned} \quad (50)$$

where $\langle \mu \rangle(dt, dz) = \delta_{\{0\}}(dz) \lambda_s^B ds + 1_{\mathbb{R}_0}(z) v(dz) \lambda_s^H ds$. So we obtain

$$\begin{aligned} g.(\lambda, y, \zeta, u(\cdot), \pi, \theta) = & \left\{ -ry - \pi(\alpha - r) - \theta^B \pi \sigma \right\} \lambda^B - \int_{\mathbb{R}_0} \pi \gamma(z) \theta^H(z) v(dz) \lambda^H \\ & + \theta^B \zeta \lambda^B + \int_{\mathbb{R}_0} \theta^H(z) u(z) v(dz) \lambda^H. \end{aligned}$$

We observe that BSDEs of the type (48) with (49) satisfy the conditions of Theorem 2.10. In the same way as for Theorem 4.2, condition (49) yields to the study of saddle points. Hence we obtain the following result:

Theorem 4.4 *Let the reference filtration be $\mathbb{F}$. If $(\hat{\pi}, \hat{\mathbb{Q}}) \in \Pi_{\mathbb{F}} \times \mathcal{Q}_{\mathbb{F}}$ satisfies the equations:*

$$\left(e^{\int_0^t r_s ds} \hat{Z}_t - \hat{\pi}_t \sigma_t \right) \lambda_t^B = 0, \quad (51)$$

$$\int_{\mathbb{R}_0} \left(e^{\int_0^t r_s ds} \hat{U}_t(z) - \hat{\pi}_t \gamma_t(z) \right) v(dz) \lambda_t^H = 0, \quad (52)$$

$$(\alpha_t - r_t) + \sigma_t \hat{\theta}_t^B \lambda_t^B + \int_{\mathbb{R}_0} \gamma_t(z) \hat{\theta}_t^H(z) v(dz) \lambda_t^H = 0, \quad (53)$$

where $(\hat{Z}, \hat{U}) \in \mathcal{I}_{\mathbb{F}}$ are the integrands in the integral representation:

$$e^{-\int_0^T r_t dt} F = \hat{\xi}_0 + \int_0^T \hat{Z}_s d\mathbf{B}_s^{\hat{\theta}} + \int_0^T \int_{\mathbb{R}_0} \hat{U}_s(z) \tilde{H}^{\hat{\theta}}(ds, dz), \quad (54)$$

and $\hat{\xi}_0 \in L^2(\Omega, \mathcal{F}, \hat{\mathbb{Q}})$ is a random variable orthogonal to the stochastic integrals (cf. Theorem 2.4), then $(\hat{\pi}, \hat{\mathbb{Q}})$ is the solution of the optimisation problem (28). The optimal price process $\hat{Y}_t = Y_t^{\hat{\pi}, \hat{\mathbb{Q}}} = Y_t$, $t \in [0, T]$, is given by:

$$\hat{Y}_t = \mathbb{E}_{\hat{\mathbb{Q}}} \left[e^{-\int_t^T r_s ds} F - \int_t^T e^{\int_0^s r_u du} \hat{Z}_s d\hat{B}_s - \int_t^T \int_{\mathbb{R}_0} e^{\int_0^s r_u du} \hat{U}_s(z) \tilde{H}^{\hat{\theta}}(ds, dz) | \mathcal{F}_t \right], \quad (55)$$

where

$$d\hat{B}_t^{\hat{\theta}} := dB_t - \hat{\theta}_t^B \lambda_t^B dt,$$

and

$$\tilde{H}^{\hat{\theta}}(dt, dz) := \tilde{H}(dt, dz) - \hat{\theta}_t^H(z) v(dz) \lambda_t^H dt.$$

These are orthogonal $(\mathbb{F}, \hat{\mathbb{Q}})$ -martingale random fields. Under probability measure $\mathbb{P}$, the optimal price process has the following dynamics:

$$\begin{aligned} d\hat{Y}_t &= \left(\hat{Y}_t r_t + \hat{\pi}_t (\alpha_t - r_t) \right) dt + \hat{\pi}_t \sigma_t dB_t + \int_{\mathbb{R}_0} \hat{\pi}_t \gamma_t(z) \tilde{H}(dt, dz) + e^{\int_0^t r_s ds} d\mathbb{E}_{\hat{\mathbb{Q}}}[\hat{\xi}_0 | \mathcal{F}_t] \\ \hat{Y}_0 &= E_{\hat{\mathbb{Q}}} [e^{-\int_0^T r_t dt} F]. \end{aligned} \quad (56)$$

Remark 4.3 We observe that the optimal strategy $(\hat{\pi}, V^{\hat{\pi}}, C^{\hat{\pi}})$ is then given by the process $\hat{\pi}$ as characterised above. The wealth $V^{\hat{\pi}}$ on the market has the initial value $V_0^{\hat{\pi}} = v = E_{\hat{\mathbb{Q}}} [e^{-\int_0^T r_t dt} F]$, and the cost is a process $C_t^{\hat{\pi}}$, $t \in [0, T]$, given as follows.

Observe that

$$d\hat{Y}_t = dV_t^{\hat{\pi}} + (\hat{Y}_t - V_t^{\hat{\pi}}) r_t dt + e^{\int_0^t r_s ds} d\mathbb{E}_{\hat{\mathbb{Q}}}[\hat{\xi}_0 | \mathcal{F}_t],$$

then

$$dC_t^{\hat{\pi}} = C_t^{\hat{\pi}} r_t dt + e^{\int_0^t r_s ds} d\mathbb{E}_{\hat{\mathbb{Q}}}[\hat{\xi}_0 | \mathcal{F}_t]; \quad C_0^{\hat{\pi}} = 0.$$

which results in $C_t^{\hat{\pi}} = e^{\int_0^t r_s ds} (\mathbb{E}_{\hat{\mathbb{Q}}}[\hat{\xi}_0 | \mathcal{F}_t] - \mathbb{E}_{\hat{\mathbb{Q}}}[\hat{\xi}_0])$.

Remark 4.4 Observe that the optimal $\hat{\mathbb{Q}}$ is a martingale measure for the optimal price process $\hat{Y}$. In fact, the price process is given by:

$$e^{-\int_0^t r_s ds} \hat{Y}_t = \mathbb{E}_{\hat{\mathbb{Q}}} [e^{-\int_0^T r_s ds} \hat{Y}_T | \mathcal{F}_t].$$

Remark 4.5 The characterizing Eqs. (38)–(40) and (51)–(53) are formally the same, being the difference on the measurability properties of the processes involved. Denote $\hat{\mathbb{Q}}_{\mathbb{G}} = \hat{\mathbb{Q}}^{\hat{\pi}} \in \mathcal{Q}_{\mathbb{G}}$ in Theorem 4.2 and $\hat{\mathbb{Q}}_{\mathbb{F}} = \hat{\mathbb{Q}}^{\hat{\pi}} \in \mathcal{Q}_{\mathbb{F}}$ in Theorem 4.4. From Eqs. (40) and (53) combined with Theorem 3.1, we can see that $\hat{\mathbb{Q}}_{\mathbb{G}|\mathbb{F}} = \hat{\mathbb{Q}}_{\mathbb{F}}$.

5 Conclusions and Example

With the intent of finding a hedging strategy in the incomplete market (2) and (3) we have studied the optimization problem (6). We have developed the solution using BSDEs and their comparison theorems. For this we refer to [9, 19], of which we adapt the results. Our approach relies on results developed for stochastic differential games as given in [35]. Indeed we have transformed the hedging problem under model uncertainty in a min-max type problem (7) by exploiting the explicit representation of the risk-measure considered (1).

The noises considered are naturally linked to two different filtrations. The filtration $\mathbb{G}$ captures all the statistical properties of the noises, allowing to exploit the underlying Gaussian and Poisson structure, see Definition 2.1. The filtration $\mathbb{F}$ is substantially the filtration generated by the noises. We study the problem (7) with respect to both situations and we observe that $\mathcal{F}_T = \mathcal{G}_T$. Correspondingly, we have proposed two BSDEs related to the two filtered probability spaces. The terminal condition is the same. The results obtained show differences in terms of adaptability of the solutions and the structure of the solution itself, cf. (43) and (56).

Both set-ups lead to hedging strategies $(\hat{\pi}, V^{\hat{\pi}}, C^{\hat{\pi}})$ with presence of cost process. In the case of $\mathbb{G}$, the cost process accounts for the anticipated knowledge of the time-change. In the case of $\mathbb{F}$, the cost process represents the spread between the perfect hedge and the best self-financing strategy. Here the distances are evaluated in terms of the risk-measure (1).

From the methodological point of view, we remark that the BSDEs presented for the study in the case of filtration $\mathbb{F}$ are based on the properties of the martingale random fields and we recall that our noises are martingale random fields with respect to both filtrations. In fact in this framework of general martingales we can see the correspondence between the two set-ups generated by the two filtrations.

As explained earlier, from a financial modeling perspective it is better suited to consider the information flow given by $\mathbb{F}$.

Example 5.1 Toy example. Let $e^{-\int_0^T r_t dt} F$ is $\mathcal{F}_T^\Lambda$ -measurable (r deterministic). In the information flow $\mathbb{G}$, the integral representation (41) shows that the optimal integrands $(\hat{Z}, \hat{U}) \in \mathcal{J}_G$ are null and that the optimal strategy presents $\hat{\pi} = 0$, and $v = \mathbb{E}_{\hat{\mathbb{Q}}_G}[e^{-\int_0^T r_t dt} F]$. Here the values $(\hat{\theta}^B, \hat{\theta}^H)$ for the probability measure $\hat{\mathbb{Q}}_G = \mathbb{Q}^{\hat{\theta}} \in \mathcal{Q}_G$ are given by (40). See Theorem 4.2. The cost process is $C_0^{\hat{\pi}} = e^{-\int_0^T r_t dt} F - \mathbb{E}_{\hat{\mathbb{Q}}_G}[e^{-\int_0^T r_t dt} F]$, $C_t^{\hat{\pi}} = 0$, $t \in (0, T]$.

Consider the case of information flow $\mathbb{F}$. Denote $\mathbb{H}_F \subset L^2(\Omega, \mathcal{F}, \hat{\mathbb{Q}}_F)$ and $\mathbb{H}_G \subset L^2(\Omega, \mathcal{F}, \hat{\mathbb{Q}}_G)$ the spaces generated by the integrals $\int_0^T \int_{\mathbb{R}} \phi_t(z) \mu(dt, dz)$ for all $\phi \in \mathcal{F}_F$ and $\phi \in \mathcal{J}_G$, respectively. Being $\hat{\mathbb{Q}}_F = \hat{\mathbb{Q}}_{G|F}$ (see Remark 4.5), then $\mathbb{H}_F \subset \mathbb{H}_G$. Hence,

$$L^2(\Omega, \mathcal{F}, \hat{\mathbb{Q}}_G) \ominus \mathbb{H}_F \supset L^2(\Omega, \mathcal{F}, \hat{\mathbb{Q}}_G) \ominus \mathbb{H}_G \ni e^{-\int_0^T r_t dt} F.$$

Hence $\hat{\xi}_0 = e^{-\int_0^T r_t dt} F$ and $(\hat{Z}, \hat{U}) \equiv 0$ in the representation (54). The optimal strategy is then $\hat{\pi} = 0$, $v = \mathbb{E}_{\hat{\mathbb{Q}}_F}[e^{-\int_0^T r_t dt} F]$, and $C_t^{\hat{\pi}} = \mathbb{E}_{\hat{\mathbb{Q}}_F}[e^{-\int_0^T r_t dt} F | \mathcal{F}_t] - \mathbb{E}_{\hat{\mathbb{Q}}_F}[e^{-\int_0^T r_t dt} F]$, $t \in [0, T]$.

In line with Remarks 4.2 and 4.4, we see that the optimal measures $\mathbb{Q}_F$ and $\mathbb{Q}_G$ are risk-neutral in the given market. In fact, applying (40) or (53), i.e.

$$(\alpha_t - r_t) + \sigma_t \hat{\theta}_t^B + \int_{\mathbb{R}_0} \gamma_t(z) \hat{\theta}_t^H(z) v(dz) \lambda_t = 0,$$

we have

$$\begin{aligned} d \left(e^{-\int_0^t r_s ds} S_t^{(1)} \right) &= e^{-\int_0^t r_s ds} S_t^{(1)} \left[(\alpha_t - r_t) dt + \sigma_t dB_t + \int_{\mathbb{R}_0} \gamma_t(z) \tilde{H}(dt, dz) \right] \\ &= e^{-\int_0^t r_s ds} S_t^{(1)} \left[\sigma_t dB_t^{\hat{\theta}} + \int_{\mathbb{R}_0} \gamma_t(z) \tilde{H}^{\hat{\theta}}(dt, dz) \right]. \end{aligned}$$

This result is consistent with the observations of [36] in the context of dynamics given by a jump diffusion and in the literature related to risk-minimizing strategies.

Acknowledgments We acknowledge the support of the Centre of Advanced Study (CAS) at the Norwegian Academy of Science and Letters that has hosted and funded the research project Stochastics in Environmental and Financial Economics (SEFE) during the academic year 2014/15.

References

1. Applebaum, D.: Lévy Processes and Stochastic Calculus. Cambridge University press (2004)
2. Balter, A., Pelsser, A.: Pricing and hedging in incomplete markets with model ambiguity. SSRN 2459405 (2015)
3. Barndorff-Nielsen, O., Maejima, M., Sato, K.i.: Infinite divisibility for stochastic processes and time change. J. Theor. Probab. **19**(2), 411–446 (2006). <http://dx.doi.org/10.1007/s10959-006-0020-7>
4. Barndorff-Nielsen, O.E., Nicolato, E., Shephard, N.: Some recent developments in stochastic volatility modelling. Quant. Financ. **2**(1), 11–23 (2002). doi:[10.1088/1469-7688/2/1/301](https://doi.org/10.1088/1469-7688/2/1/301)
5. Billingsley, P.: Probability and Measure, 3rd edn. Wiley Series in Probability and Mathematical Statistics. Wiley, New York (1995). A Wiley-Interscience Publication
6. Boel, R., Varaiya, P., Wong, E.: Martingales on jump processes I. Representation results. SIAM J. Control **13**(5) (1975)
7. Brémaud, P.: Point Processes and Queues—Martingale Dynamics. Springer, New York (1981)
8. Cairoli, R., Walsh, J.B.: Stochastic integrals in the plane. Acta Math. **134**, 111–183 (1975)
9. Carbone, R., Ferrario, B., Santacroce, M.: Backward stochastic differential equations driven by càdlàg martingales. Theor. Probab. Appl. **52**(2), 304–314 (2008)
10. Carr, P., German, H., Madan, D., Yor, M.: Stochastic volatility for Lévy processes. Math. Finan. **13**(3), 345–382 (2003)
11. Ceci, C., Cretarola, A., Russo, F.: BSDEs under partial information and financial applications. Stochastic Process. Appl. **124**(8), 2628–2653 (2014)
12. Chen, Z., Epstein, L.: Ambiguity, risk, and asset returns in continuous time. Econometrica **70**(4), 1403–1443 (2002)

13. Davis, M.H.A.: The representation of martingales of jump processes. *SIAM J. Control Optim.* **14**(4), 623–638 (1976)
14. Delong, L.: No-good-deal, local mean variance, ambiguity pricing and hedging of an insurance payment process. *ASTIN Bull.* **42**, 203–232 (2012)
15. Delong, L.: *Backward Stochastic Differential Equations with Jumps and their Actuarial and Financial Applications*. Springer, European Actuarial Academy Series (2013)
16. Di Nunno, G., Eide, I.B.: Minimal-variance hedging in large financial markets: random fields approach. *Stoch. Anal. Appl.* **28**(1), 54–85 (2010)
17. Di Nunno, G., Rozanov, Y.: Stochastic integrals and adjoint derivatives. In: Benth, F., Di Nunno, G., Lindstrøm, T., Øksendal, B., Zhang, T. (eds.) *Stochastic Analysis and Applications*, Abel Symposia, vol. 2, pp. 265–307. Springer, Berlin (2007)
18. Di Nunno, G., Sjursen, S.: On chaos representation and orthogonal polynomials for the doubly stochastic Poisson process. In: Dalang, R.C., Dozzi, M., Russo, F. (eds.) *Seminar on Stochastic Analysis, Random Fields and Applications VII, Progress in Probability*, vol. 67, pp. 23–54. Springer, Basel (2013). doi:[10.1007/978-3-0348-0545-2_2](https://doi.org/10.1007/978-3-0348-0545-2_2)
19. Di Nunno, G., Sjursen, S.: BSDEs driven by time-changed Lévy noises and optimal control. *Stoch. Processes Appl.* **124**, 1679–1709 (2014)
20. Grigelionis, B.: Characterization of stochastic processes with conditionally independent increments. *Lith. Math. J.* **15**(4), 562–567 (1975). <http://dx.doi.org/10.1007/BF00969787>
21. Heston, S.L.: A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Rev. Financ. Stud.* **6**(2), 327–343 (1993)
22. Hull, J., White, A.: The pricing of options on assets with stochastic volatilities. *J. Finance* **42**(2), 281–300 (1987)
23. Jacod, J.: Multivariate point processes: predictable projection, Radon-Nikodym derivatives, representation of martingales. *Probab. Theory Relat. Fields* **31**(3), 235–253 (1975)
24. Jacod, J.: Existence and uniqueness for stochastic differential equations. In: Kohlmann, M., Vogel, W. (eds.) *Stochastic Control Theory and Stochastic Differential Systems, Lecture Notes in Control and Information Sciences*, vol. 16, pp. 435–446. Springer, Berlin (1979). <http://dx.doi.org/10.1007/BFb0009402>
25. Jacod, J., Shiryaev, A.N.: *Limit Theorems for Stochastic Processes*. Springer (2003)
26. Jeanblanc, M., Mania, M., Santacroce, M., Schweizer, M.: Mean-variance hedging via stochastic control and BSDEs for general semimartingales. *Ann. Appl. Probab.* **22**(6), 2388–2428 (2012)
27. Kallenberg, O.: *Foundations of Modern Probability*. Springer (1997)
28. Karlsen, E.H.: *Optimal portfolio problems under model ambiguity*. Thesis, University of Oslo (May 2014)
29. Lando, D.: On Cox processes and credit risky securities. *Rev. Deriv. Res.* **2**(2–3), 99–120 (1998)
30. Lépine, D., Mémin, J.: Sur l'intégrabilité uniforme des martingales exponentielles. *Z. Wahrsch. Verw. Gebiete* **42**(3), 175–203 (1978)
31. Lim, A.: Mean-variance hedging when there are jumps. *Siam J. Control Optim.* **44**(5), 1893–1922 (2005). <http://dx.doi.org/10.1137/040610933>
32. Løkka, A.: Martingale representation of Lévy processes. *Stoch. Anal. Appl.* **22**(4), 867–892 (2005)
33. Øksendal, B.: *Stochastic Differential Equations*. Springer (2005)
34. Øksendal, B., Sulem, A.: *Applied Stochastic Control of Jump Diffusions*, 2nd edn. Universitext. Springer, Berlin (2007)
35. Øksendal, B., Sulem, A.: Portfolio optimization under model uncertainty and BSDE games. *Quant. Finance* **11**(11), 1665–1674 (2011)
36. Øksendal, B., Sulem, A.: Robust stochastic control and equivalent martingale measures. In: *Stochastic Analysis with Financial Applications, Progress in Probability*, vol. 65, pp. 179–189. Birkhäuser/Springer Basel AG, Basel (2011)
37. Protter, P.: *Stochastic Integration and Differential Equations*. Springer (2005). Version 2.1
38. Serfozo, R.F.: Processes with conditional stationary independent increments. *J. Appl. Probab.* **9**(2), (1972)

39. Stein, E.M., Stein, J.C.: Stock price distributions with stochastic volatility: an analytic approach. *Rev. Financ. Stud.* **4**(4), 727–752 (1991)
40. Winkel, M.: The recovery problem for time-changed Lévy processes. Research Report MaPhySto 2001-37 (2001)
41. Yablonski, A.L.: The Malliavin calculus for processes with conditionally independent increments. In: Benth, F., Di Nunno, G., Lindstrøm, T., Øksendal, B., Zhang, T. (eds.) *Stochastic Analysis and Applications*, Abel Symposia, vol. 2, pp. 641–678. Springer, Berlin (2007)

Markov Renewal Methods in Restart Problems in Complex Systems

Søren Asmussen, Lester Lipsky and Stephen Thompson

Abstract A task with ideal execution time L such as the execution of a computer program or the transmission of a file on a data link may fail, and the task then needs to be restarted. The task is handled by a complex system with features similar to the ones in classical reliability: failures may be mitigated by using server redundancy in parallel or k -out-of- n arrangements, standbys may be cold or warm, one or more repairmen may take care of failed components, etc. The total task time X (including restarts and pauses in failed states) is investigated with particular emphasis on the tail $\mathbb{P}(X > x)$. A general alternating Markov renewal model is proposed and an asymptotic exponential form $\mathbb{P}(X > x) \sim Ce^{-\gamma x}$ identified for the case of a deterministic task time $L \equiv \ell$. The rate γ is given by equating the spectral radius of a certain matrix to 1, and the asymptotic form of $\gamma = \gamma(\ell)$ as $\ell \rightarrow \infty$ is derived, leading to the asymptotics of $\mathbb{P}(X > x)$ for random task times L . A main finding is that X is always heavy-tailed if L has unbounded support. The case where the Markov renewal model is derived by lumping in a continuous-time finite Markov process with exponential holding times is given special attention, and the study includes analysis of the effect of processing rates that differ with state or time.

Keywords Alternating renewal process · Computer reliability · Data transmission · Failure rate · Fault-tolerant computing · Heavy tails · Markov renewal equation · Matrix perturbation · Phase-type distribution · RESTART · Tail asymptotics · Perron-Frobenius theory · Spectral radius

S. Asmussen (✉)

Department of Mathematics, Aarhus University, Aarhus, Denmark
e-mail: asmus@math.au.dk; asmus@imf.au.dk

L. Lipsky · S. Thompson

Department of Computer Science and Engineering,
University of Connecticut, Storrs, CT, USA
e-mail: lester@engr.uconn.edu

S. Thompson

e-mail: stevet@alexzilla.com

1 Introduction

This paper studies some systems of a similar flavour as those of interest in Reliability and Availability Theory [5, 6, 23]. The system consists of a large number of components and at a given time it is in a state depending on the characteristics of the components (operating, under repair, in cold or hot standby, rebooting etc.). In brief, a finite state space $\mathcal{E}$ is partitioned into two subspaces, $\mathcal{D}$ and $\mathcal{U}$ (Down/Up, or Operating/Failed, etc.). Individual components may be failed and under repair also in an Up state, but the system as a whole can only operate in the Up states and is paused upon exit from there to remain in the Down states until all necessary repairs have been undertaken.

In classical reliability theory one is interested in the length of time the system is in some Up state (the longer without failure, the better). Repair time is of less importance. Availability is concerned with the fraction of time that the system is in an Up state, taking into account the multiplicity of processors (thus the availability fraction could be greater than 1). In this paper we focus on the processing of a task of length L (the *task time*) which can only finish processing after the system has been in the set $\mathcal{U}$ of Up states for a time greater than L (L may be random or constant, $L = \ell$). If during processing the system fails and goes to a Down state, the performed work is lost and when entering an Up state again, the processing has to be restarted from scratch. That is, each time the system leaves $\mathcal{U}$ and goes to $\mathcal{D}$, it must start over when reentering $\mathcal{U}$. The quantity of interest is the *total task time* X .

The study of total task times in problems of this type has a long tradition in many specific and distinct failure recovery schemes. In particular in the RESUME (also referred to as preemptive resume) scenario, if there is a processor failure while a job is being executed, after repair is implemented the job can continue where it left off. In the REPLACE situation (also referred to as preemptive repeat different), if a job fails, it is replaced by a different job from the same distribution. Here, no details concerning the previous job are necessary in order to continue. For these two schemes, see Kulkarni et al. [16, 17]. Further models and methods of failure recovery are in Chlebus et al. [11] for restartable processors and in DePrisco et al. [12] (stage checkpointing).

As indicated above, we are here concerned with RESTART (also referred to as preemptive repeat identical). There are many examples where this is relevant. The obvious one alluded to above involves execution of a program on some computer. If the computer fails, and the intermediate results are not saved externally (e.g., by *checkpointing*), then the job must restart from the beginning. As another example, one might wish to copy a file from a remote system using some standard protocol as FTP or HTTP. The time it takes to copy a file is proportional to its length. A transmission error immediately aborts the copy and discards the partially received data, forcing the user to restart the copy from the beginning. Yet another example would be receiving ‘customer service’ by telephone. Often, while dealing with a particular service agent, the connection is broken. Then the customer must redial the

service center, and invariably (after waiting in a queue) ends up talking to a different agent, and having to explain everything from the beginning.

Computing expected values and transforms is usually easy in any of the models mentioned above. Also the important problem of tail behaviour, that is, the probability of long delays, has been successfully attacked in a number of the above references. For example for the REPLACE model the delay $X - L$ is just a geometric sum of terms distributed as the failure time U conditioned on $U < L$, even if L is random, which assuming phase-type forms of L, U makes matrix-analytic methods available, cf. Bobbio and Trivedi [8]. However, for the RESTART policy the tail behaviour resisted detailed analysis until the recent work of Sheahan et al. [20] and Asmussen et al. [2] where the tail of L was found in a variety of combinations of tail behavior of X and the failure time U . A main and surprising finding was that the tail of X is always heavy-tailed if L has unbounded support. The consequence is that delays can be very severe in the RESTART setting. For further recent work in this direction, see Jelenković et al. [13, 14].

2 Statement of Main Results

We now proceed to a more precise description of the model. Without the task being processed (equivalently, with task time $T \equiv 0$), the system is assumed to develop according to an alternating Markov renewal process with state space $\mathcal{E} = \mathcal{U} \cup \mathcal{D}$ and imbedded Markov chain at jump times $\xi_0, \xi_1, \dots$. If $\xi_0 = u \in \mathcal{U}$, the first sojourn in state ξ_0 terminates at time T_u and then state $\xi_1 \in \mathcal{D}$ is entered, and if $\xi_0 = d \in \mathcal{D}$, the first sojourn in state ξ_0 terminates at time T_d where state $\xi_1 \in \mathcal{U}$ is entered. When it is unimportant to specify whether states are in $\mathcal{U}$ or in $\mathcal{D}$, they are just denoted by $i, j, \dots$; when a state in $\mathcal{E}$ is denoted by u , it is silently understood that $u \in \mathcal{U}$, and similarly for $d \in \mathcal{D}$.

The transition rules are thus specified by the set of probability measures

$$\begin{aligned} F_{du}(t) &= \mathbb{P}_d(T_d \leq t, \xi_1 = u), & d \in \mathcal{D}, u \in \mathcal{U}, \\ F_{ud}(t) &= \mathbb{P}_u(T_u \leq t, \xi_1 = d), & u \in \mathcal{U}, d \in \mathcal{D}, \end{aligned}$$

$F_{ij} = 0$ when i, j are either both in $\mathcal{U}$ or both in $\mathcal{D}$. In particular, the transition probabilities p_{ij} of the Markov chain ξ are given by $p_{du} = F_{du}(\infty)$, $p_{ud} = F_{ud}(\infty)$, $p_{ij} = 0$ when i, j are both in either $\mathcal{U}$ or in $\mathcal{D}$, and we have

$$F_d(t) = \mathbb{P}_d(T_d \leq t) = \sum_{u \in \mathcal{U}} F_{du}(t), \quad F_u(t) = \mathbb{P}_u(T_u \leq t) = \sum_{d \in \mathcal{D}} F_{ud}(t).$$

The term *alternating* comes from only transitions from $\mathcal{U}$ to $\mathcal{D}$ and vice versa being possible, not ones within $\mathcal{U}$ or $\mathcal{D}$. At the time T_u or T_d of the first state space change,

the process repeats itself with new starting value ξ_1 . See [1, pp. 206–7] for more detail on the basics of Markov renewal processes.

The model for how the task is handled is that processing only occurs in an Up state u of the Markov renewal process and then the processing rate is $\rho_u(t)$ at time t for a suitable stochastic process ρ_u . Thus, if the task length is $L \equiv \ell$ and $\xi_0 = u \in \mathcal{U}$, the task finishes processing in the first Markov renewal interval $[0, T_u)$ at time

$$\tau_u(\ell) = \inf \left\{ t > 0 : \int_0^t \rho_u(s) ds = \ell \right\}$$

if $\tau_u(\ell) < T_u$. Otherwise, no processing takes place in the down state $\xi_1 = d \in \mathcal{D}$ entered after T_u , and at the time where the next state ξ_2 in $\mathcal{U}$ is entered, the whole process is restarted with the same processing time ℓ but in Markov state ξ_2 . No specific assumptions are made at the moment on the dependence between T_u , ρ_u , ξ_1 , but to avoid trivialities, we need $\mathbb{P}(\tau_u(\ell) < T_u) > 0$ for at least one $u \in \mathcal{U}$, ensuring that the task will eventually be finished.

The main example of processing rate modeling is of course $\rho_u(t) \equiv 1$ for all u and t so that $\tau_u(\ell) = \ell$, and this may safely be used as basis for intuition for quite a while. We look at more complicated time-varying and random rates in Sect. 8.

Our main example of the Markov renewal set-up comes from continuous-time finite Markov processes modeling the time evolution of models in classical reliability theory [5, 6, 23]. We present this in Sect. 3. However, there are others, e.g. vanilla RESTART with repair. Here we can take $\mathcal{U}$, $\mathcal{D}$ as one-point sets $\mathcal{U} = \{u\}$, $\mathcal{D} = \{d\}$, and $\rho_u(t) \equiv 1$. Then F_u can be seen as the distribution of either the operating time or the failure time (denoted G in [2]), F_d the distribution of the repair time, and the Markov renewal model allows for incorporating general distributions of F_u , F_d , not just exponential ones.

Our main result on the general set-up with deterministic task length is the following:

Theorem 2.1 *Consider a deterministic task length $L = \ell$ and denote by $\mathbf{R}(\alpha)$ the $\mathcal{E} \times \mathcal{E}$ matrix with entries*

$$\begin{aligned} r_{du}(\alpha) &= \mathbb{E}_d[e^{\alpha T_d}; \xi_1 = u], & d \in \mathcal{D}, u \in \mathcal{U}, \\ r_{ud}(\alpha) &= \mathbb{E}_u[e^{\alpha T_u}; \tau_u(\ell) \geq T_u, \xi_1 = d], & u \in \mathcal{U}, d \in \mathcal{D}, \end{aligned}$$

all other $r_{ij}(\alpha) = 0$. Assume there exists $\gamma = \gamma(\ell)$ such that $\mathbf{R}(\gamma)$ is irreducible with spectral radius 1. Then $\mathbb{P}_i(X > x) \sim C_i e^{-\gamma x}$ as $x \rightarrow \infty$, with $C_i = C_i(\ell)$ specified in Corollary 4.1 below.

The assumption that γ exists is essentially a condition on light tails of the T_d , for example that each T_d is gamma-like. By this we mean that the density exists and has asymptotic form

$$by^{a-1}e^{-\delta y} \tag{1}$$

We discuss this in more detail in Sect. 4. The case of heavy tails of some T_d is not included by Theorem 2.1 but can be treated as well. For example, we have:

Theorem 2.2 *Consider a deterministic task length $L = \ell$ and assume that each T_d has a subexponential density, more precisely that there exists a subexponential density $\varphi(y)$ such that, for all d , $\mathbb{P}(T_d \in \cdot)$ is absolutely continuous for large y with a density of asymptotic form $c_d \varphi(y)$. Then $\mathbb{P}_i(X > x) \sim C_i \bar{\Phi}(x)$ as $x \rightarrow \infty$, with $C_i = C_i(\ell)$ specified in (20) below and $\bar{\Phi}(x) = \int_x^\infty \varphi(y) dy$.*

[See (18) below for the definition of a subexponential density. The constants C_i are obviously not the same as in Theorem 2.1; similar slight abuse of notation occurs throughout the paper].

For a random task length L , conditioning on $L = \ell$ gives

$$\mathbb{P}_i(X > x) = \int_0^\infty \mathbb{P}_i(X > x | L = \ell) \mathbb{P}(L \in d\ell). \quad (2)$$

Since now ℓ is a variable (and not constant, as previously assumed), we write $\gamma(\ell)$ rather than ℓ , $\mathbf{R}(\alpha, \ell)$ rather than $\mathbf{R}(\alpha)$ to stress the dependence on both variables. In (2), $\mathbb{P}_i(X > x | L = \ell)$ has an asymptotic exponential form as $x \rightarrow \infty$ by Theorem 2.1. Further, the rate $\gamma(\ell)$ goes to 0 as $\ell \rightarrow \infty$ since $\mathbf{R}(0, \infty)$ equals the spectral radius one matrix $\mathbf{P}$. We thus obtain the following easy but notable result:

Corollary 2.3 *If the task length L has unbounded support, the distribution of the total task time X is heavy-tailed in the sense that $e^{\delta x} \mathbb{P}(X > x) \rightarrow \infty$ for all $\delta > 0$.*

The key step for obtaining more precise results turns out to be to identify the asymptotics of $\gamma(\ell)$ as $\ell \rightarrow \infty$. Our result is:

Theorem 2.4 *Assume that the assumptions of Theorem 2.1 on $\mathbf{R}(\alpha, \ell)$ hold for all large ℓ , and that for some function $\varphi(\ell)$ it holds that*

$$k_{ud}^*(\ell) = \mathbb{P}(T_u > \tau_u(\ell), \xi_1 = d) \sim k_{ud} \varphi(\ell) \quad (3)$$

as $\ell \rightarrow \infty$ for some set of constants such that $k_{ud} > 0$ for at least one pair $u \in \mathcal{U}$, $d \in \mathcal{D}$. Then

$$\gamma(\ell) \sim \mu \varphi(\ell) \quad \text{as } \ell \rightarrow \infty, \quad \text{where } \mu = \frac{\sum_{u \in \mathcal{U}, d \in \mathcal{D}} \pi_u k_{ud}}{\sum_{i \in \mathcal{E}} \pi_i \mathbb{E}_i T_i} \quad (4)$$

and $\boldsymbol{\pi} = (\pi_i)_{i \in \mathcal{E}} = (\pi_i)_{i \in \mathcal{U} \cup \mathcal{D}}$ is the stationary distribution of the Markov chain ξ , that is, the invariant probability vector for the matrix $\mathbf{P} = \mathbf{R}(0, \infty)$.

Remark 2.5 In practice, $\varphi(\ell)$ will be found by first determining the growth rate of the $h_{ud}^*(\ell) = \mathbb{P}(T_u > \tau_u(\ell), \xi_1 = d)$ and letting $\varphi(\ell) = \min_{d \in \mathcal{D}} h_{ud}^*(\ell)$. $\square$

Given Theorem 2.4, it is easy to adapt the calculations of [2] to get the tail of X , in more or less sharp forms depending on the form of the distribution of L . For example:

Corollary 2.6 *Assume that L is Gamma-like as in (1), that $\mathbb{P}_u(\tau_u(\ell) \rightarrow \infty) = 1$ for all u and that the assumptions of Theorem 2.4 hold with $\varphi(\ell) = e^{-q\ell}$ for some q . Then*

$$\mathbb{P}_i(X > x) \sim C \frac{\log^{a-1} x}{x^{\delta/q}}$$

with C given in (16) below.

The paper is organised as follows. Section 3 presents some main examples and a general class of models is described where the Markov renewal model is derived by lumping in a continuous-time finite Markov process with exponential holding times. This case provided our main motivation, but the Markov renewal conversion is necessary because of non-Markovian features of the RESTART mechanism. Section 4 contains the proof of Theorem 2.1, where the main ingredient is a suitable set of Markov renewal equations. Also the forms of the knowns and unknowns are specialized to the Markov model. Section 5 contains the proof of Theorem 2.4. Perron-Frobenius theory for nonnegative matrices plays a main role (see [1, I.6] for a short introduction and [7, 19] for more extensive treatments). However, the proof is by bare-hand arguments rather than general perturbation theory. Section 6 relates these results to rare event theory for regenerative processes. In particular, the asymptotic form of $\mathbb{E}[X \mid L = \ell]$ is shown to be $1/\gamma(\ell)$. The proof of Theorem 2.2 is given in Sect. 7; it involves discussion of Markov renewal equations with heavy tails, a less established topic than in the light tailed case. Non-constant processing rates $\rho_u(t)$ are studied in Sect. 8 and lead into matrix formalism and differential equations somewhat similar to the theory of fluid reward models and multivariate phase-type distributions. Finally, Sect. 9 sketches via an example how non-exponential distributions may be incorporated in the Markov model, and some preliminaries and technical steps are deferred to the Appendix.

The paper only contains one numerical example (a more extensive set is presented in the companion paper [21]). Nevertheless, it should be stressed that our aim is computational. In particular, we have chosen a finite state set-up leading to explicit and numerically implementable matrix formulas rather than a general state one where one would need to impose many not easily verifiable technical conditions, operators would take the role of matrices, etc.

3 From Markov to Semi-Markov Models

In many examples, the Markov renewal structure may be derived from embedding into a larger Markov process $J(t)$ with state space $\mathcal{E}^* = \mathcal{U}^* \cup \mathcal{D}^*$. In a state $u \in \mathcal{U}^*$, the task is processed at rate r_u and no processing can take place in a state $d \in \mathcal{D}^*$. We then take $\mathcal{U} \subseteq \mathcal{U}^*$ as the subset of states that can be entered from $\mathcal{D}$ and $\mathcal{D} \subseteq \mathcal{D}^*$ as

the subset of states that can be entered from $\mathcal{U}$. If $\xi_0 = J(0) = u \in \mathcal{U}$, we further let $T_u = \inf\{T > 0 : J(t) \in \mathcal{D}^*\}$ (and similarly for T_d) and $\rho_u(t) = r_{J(t)}$.

It should be noted that in this Markovian scheme, the F_u, F_d become phase-type (PH) and the F_{ud}, F_{du} defective PH (recall that a PH distribution is defined as the time to absorption in a finite continuous time Markov process with one absorbing state, cf. [1, III.4]). The implication is that standard matrix-analytic formulas apply to rewrite the expressions in the general set-up in terms of matrix expressions (matrix inverses and matrix-exponentials).

The Markovian imbedding scheme is illustrated in the following examples, where the states in $\mathcal{U}$ are dark green, the ones in $\mathcal{U}^* \setminus \mathcal{U}$ light green, and similar dark and light red coloring is used for the states in $\mathcal{D}$ and $\mathcal{D}^* \setminus \mathcal{D}$.

Example 3.1 Consider a system with two exponential parallel servers and two repairmen. The failure rate of a server is β , the repair rate is λ , and a task can be (re)started if at least one server is up. We can take $\mathcal{E}^* = \{2, 1, 0\}$, with $J(t) = i$ meaning that i servers are up (and thus $2 - i$ down, i.e. under repair).

We have $\mathcal{U}^* = \{1, 2\}$, $\mathcal{U} = \{1\}$, $\mathcal{D}^* = \mathcal{D} = \{0\}$ (Fig. 1); note that $2 \notin \mathcal{U}$ since even if the first service may be started in state 2, none of the following ones will be so because a failed period (a sojourn in state 0) cannot be terminated by a jump to 2. The rates are $r_1 = 1$, $r_2 = 2$ if both servers can work on the task and $r_1 = r_2 = 1$ if only one can (hot back-up).

With only one repairman, 2λ should be changed to λ , and with cold back-up, 2β should be changed to β . Everything else remains unchanged. $\square$

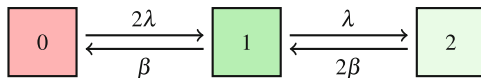
Example 3.2 Consider the system in Example 3.1 with the modification that 2 servers have to be up before a service can start. We then take $\mathcal{E}^* = \{2, 1+, 1-, 0\}$, with 0, 2 having the same meaning as before, and $1-$ meaning that there is one repaired server and one under repair, $1+$ that there is one operating and one under repair. Assuming that a server that is repaired but waiting for the other to be repaired before going into operation cannot fail, we have the transition diagram in Fig. 2.

We have $\mathcal{U}^* = \{2, 1+\}$, $\mathcal{U} = \{2\}$, $\mathcal{D}^* = \{0, 1-\}$, $\mathcal{D} = \{0\}$. Again depending on the interpretation, we may take $r_{1+} = r_2 = 1$ or $r_{1+} = 1$, $r_2 = 2$. $\square$

Example 3.3 Consider again the system in Example 3.1, this time modified such that the two servers a, b are heterogeneous, i.e. with failure rates β_a, β_b and repair rates λ_a, λ_b . We take $\mathcal{E}^* = \{0, 1a, 1b, 2\}$, with 0, 2 having the same meaning as before and $1a$ meaning that server a is up and server b down, and vice versa for $1b$. Cf. Fig. 3a for the case of homogeneous rates. We have $\mathcal{U}^* = \{2, 1a, 1b\}$, $\mathcal{U} = \{1a, 1b\}$, $\mathcal{D}^* = \mathcal{D} = \{0\}$.

Non-preemptive service and heterogeneous service rates $r_a \neq r_b$ for the servers can be handled by splitting state 2 into two states $2a, 2b$, cf. Fig. 3b. Then $r_{2a} = r_{1a} = r_a$, $r_{2b} = r_{1b} = r_b$. $\square$

Fig. 1 Parallel servers



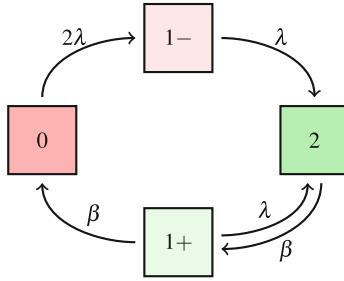


Fig. 2 Start requires both servers up

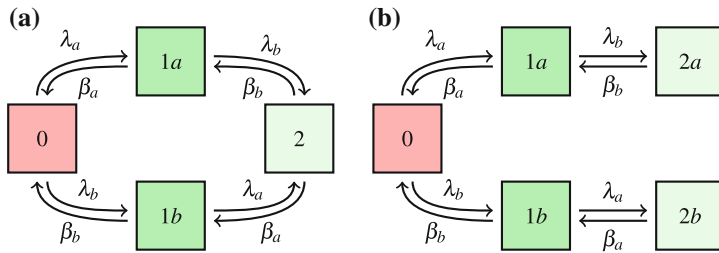


Fig. 3 Heterogeneous parallel servers. **a** Homogeneous service rates, **b** heterogeneous service rates

4 Theorem 2.1: Proof and Amendments

Proof of Theorem 2.1 For $i \in \mathcal{E}$, let $Z_i(x) = \mathbb{P}_i(X > x)$. If $i = d \in \mathcal{D}$, then service starts at the time T_d of exit from d , so that the whole of T_d contributes to the total task time X and for $X > x$ to occur, only delay $(x - T_d)^+$ needs to be accumulated after T_d . Conditioning on $y = T_d$, considering the cases $y > x$ and $y \leq x$ separately, and partitioning according to the possible values $u = \xi_1 \in \mathcal{U}$ gives

$$Z_d(x) = z_d(x) + \sum_{u \in \mathcal{U}} \int_0^x Z_u(x - y) F_{du}(dy)$$

where $z_d(x) = \mathbb{P}_d(T_d > x)$. If $i = u \in \mathcal{U}$, then the task will be terminated before $y = T_u$ if $y > \tau_u(\ell)$ so that $X > x$ if and only if $\tau_u(\ell) > x$. This gives

$$Z_u(x) = z_u(x) + \sum_{d \in \mathcal{D}} \int_0^x Z_d(x - y) \mathbb{P}(T_u \in dy, \tau_u(\ell) > y, \xi_1 = d)$$

where $z_u(x) = \mathbb{P}(T_u > \tau_u(\ell) > x)$. Recalling that transitions within $\mathcal{U}$ or within $\mathcal{D}$ need not be taken into account and letting $\tilde{F}_{du} = F_{du}$,

$$\tilde{F}_{ud}(t) = \mathbb{P}(T_u \leq t, \tau_u(\ell) > T_u, \xi_1 = d),$$

this can be summarised as the set

$$Z_i(x) = z_i(x) + \sum_{j \in \mathcal{E}} \int_0^x Z_j(x-y) \tilde{F}_{ij}(\mathrm{d}y), \quad i \in \mathcal{E}, \quad (5)$$

of Markov renewal equations where $\tilde{F}_{ij} \equiv 0$ if i, j are both in $\mathcal{U}$ or both in $\mathcal{D}$. That the $Z_i(x)$ decay exponentially at rate γ then follows immediately from the assumption $\text{spr}(\mathbf{R}(\gamma)) = 1$ via the generalisation of the key renewal theorem stated in Lemma 9.1. In more detail, that result also gives the form of the C_i which we next state separately. $\square$

Corollary 4.1 *Let $\mathbf{v} = (v_i)_{i \in \mathcal{E}}$, $\mathbf{h} = (h_i)_{i \in \mathcal{E}}$ be the left, resp. right, eigenvectors of $\mathbf{R}(\gamma)$ corresponding to the eigenvalue γ . Then $C_i = h_i C_1 / C_2$ where*

$$C_1 = \frac{1}{\gamma} \sum_{u \in \mathcal{U}} v_u \int_0^\infty \mathbb{E}[(e^{\gamma \tau_u(\ell)} - 1) \mathbf{1}\{\tau_u(\ell) \leq y\} \mid T_u = y] F_u(\mathrm{d}y) \quad (6)$$

$$+ \frac{1}{\gamma} \sum_{d \in \mathcal{D}} v_d \int_0^\infty [e^{\gamma y} - 1] F_d(\mathrm{d}y), \quad (7)$$

$$C_2 = \sum_{u \in \mathcal{U}, d \in \mathcal{D}} v_u h_d \int_0^\infty y e^{\gamma y} \mathbb{P}(\tau_u(\ell) \geq y \mid T_u = y) F_{ud}(\mathrm{d}y) \quad (8)$$

$$+ \sum_{d \in \mathcal{D}, u \in \mathcal{U}} v_h h_u \int_0^\infty y e^{\gamma y} F_{du}(\mathrm{d}y). \quad (9)$$

Proof For $d \in \mathcal{D}$, integration by parts gives that

$$\gamma \int_0^\infty e^{\gamma x} z_d(x) \mathrm{d}x = \gamma \int_0^\infty e^{\gamma x} \mathbb{P}(T_d > x) \mathrm{d}x$$

reduces to the integral in (7). For $u \in \mathcal{U}$,

$$\begin{aligned} \gamma \int_0^\infty e^{\gamma x} z_u(x) \mathrm{d}x &= \gamma \int_0^\infty e^{\gamma x} \mathbb{P}(T_u \geq \tau_u(\ell) > x) \mathrm{d}x \\ &= \int_0^\infty F_u(\mathrm{d}y) \int_0^\infty \gamma e^{\gamma x} \mathbb{P}(\tau_u(\ell) \in (x, y] \mid T_u = y) \mathrm{d}x \\ &= \int_0^\infty F_u(\mathrm{d}y) \int_0^\infty \gamma e^{\gamma x} \mathbb{P}(\tau_u(\ell) \mathbf{1}\{\tau_u(\ell) \leq y\} > x \mid T_u = y) \mathrm{d}x \\ &= \int_0^\infty F_u(\mathrm{d}y) \mathbb{E}[e^{\gamma \tau_u(\ell)} \mathbf{1}\{\tau_u(\ell) \leq y\} - 1 \mid T_u = y]. \end{aligned}$$

Considering the cases $\tau_u(\ell) \leq y$ and $\tau_u(\ell) > y$ separately gives the integral in (6). Inserting in (22) gives the expression for C_1 . The one for C_2 follows from (23) by similar manipulations. $\square$

Remark 4.2 In the simplest case $\rho_u(t) \equiv 1$, the integral in (6) reduces to

$$(e^{\gamma\ell} - 1) \int_{\ell}^{\infty} F_u(dy) = (e^{\gamma\ell} - 1) \mathbb{P}(T_u > \ell)$$

and the one in (8) to

$$\int_0^{\ell} ye^{\gamma y} F_{du}(dy) = \mathbb{E}[T_u e^{\gamma T_u}; T_u \leq \ell, \xi_1 = d]. \quad \square$$

Remark 4.3 The expressions for γ are implicit even for simple RESTART [2], so in the present generality, numerical evaluation seems inevitable. This has two steps. The first is computing the elements of $\mathbf{R}(\alpha)$. How difficult this is depends on the specific model parameters (but see Sect. 8). The next step is then evaluating eigenvalues of $\mathbf{R}(\alpha)$ and finding the roots of the equation $1 = \text{spr}(\mathbf{R}(\alpha))$ which can be done using standard software.

The dimension of the matrices can be reduced from $\mathcal{E} \times \mathcal{E}$ to $\mathcal{U} \times \mathcal{U}$ or $\mathcal{D} \times \mathcal{D}$ by noting that $[\text{spr}(\mathbf{R}(\alpha))]^2 = \text{spr}(\mathbf{R}(\alpha)^2)$, and hence the equation determining γ is (in obvious block notation)

$$1 = \text{spr}(\mathbf{R}_{\mathcal{U}\mathcal{D}}(\gamma)\mathbf{R}_{\mathcal{D}\mathcal{U}}(\gamma)) \quad \text{or} \quad 1 = \text{spr}(\mathbf{R}_{\mathcal{D}\mathcal{D}}(\gamma)\mathbf{R}_{\mathcal{U}\mathcal{U}}(\gamma)). \quad \square$$

We proceed to discuss Theorem 2.1 in the setting of the Markov model of Sect. 3. In the following, we will use the partitioning

$$\mathbf{Q} = \begin{pmatrix} \mathbf{Q}_{\mathcal{U}^*\mathcal{U}^*} & \mathbf{Q}_{\mathcal{U}^*\mathcal{D}^*} \\ \mathbf{Q}_{\mathcal{D}^*\mathcal{U}^*} & \mathbf{Q}_{\mathcal{D}^*\mathcal{D}^*} \end{pmatrix}$$

of the generator $\mathbf{Q}$ of the Markov process J on $\mathcal{E}^* = \mathcal{U}^* \cup \mathcal{D}^*$. Recall that the parameters of a PH distribution is the rate matrix giving transition rates of the non-absorbing states of the underlying Markov process and the initial vector; the exit vector is the vector of rates of transitions to the absorbing state. Cf. [1, III.4]. Let further $\mathbf{e}_u$ denote the $\mathcal{U}^*$ -column vector with 1 at entry $u \in \mathcal{U}^*$ and 0 otherwise (and similarly for $\mathbf{e}_d$), and $\mathbf{e}$ the column vector with 1 at all entries (of dimension $\mathcal{U}^*$ or $\mathcal{D}^*$ depending on the context).

Proposition 4.4 *In the Markov model,*

- F_{ud} is defective PH with rate matrix $\mathbf{Q}_{\mathcal{U}^*\mathcal{U}^*}$, initial vector $\mathbf{e}_u^{\top}$ and exit vector $\mathbf{Q}_{\mathcal{U}^*\mathcal{D}^*}\mathbf{e}_d$;
- F_{du} is defective PH with rate matrix $\mathbf{Q}_{\mathcal{D}^*\mathcal{D}^*}$, initial vector $\mathbf{e}_d^{\top}$ and exit vector $\mathbf{Q}_{\mathcal{D}^*\mathcal{U}^*}\mathbf{e}_u$;
- F_u is PH with rate matrix $\mathbf{Q}_{\mathcal{U}^*\mathcal{U}^*}$, initial vector $\mathbf{e}_u^{\top}$ and exit vector $\mathbf{Q}_{\mathcal{U}^*\mathcal{D}^*}\mathbf{e}$;
- F_d is PH with rate matrix $\mathbf{Q}_{\mathcal{D}^*\mathcal{D}^*}$, initial vector $\mathbf{e}_d^{\top}$ and exit vector $\mathbf{Q}_{\mathcal{D}^*\mathcal{U}^*}\mathbf{e}$.

Remark 4.5 The implication of Proposition 4.4 is that standard matrix-analytic machinery becomes available to rewrite many expressions considered so far in terms of matrices (facilitating computer implementation). For example,

$$z_d(x) = \mathbb{P}(T_d > x) = \mathbf{e}_d^\top \mathbf{e}^{\mathcal{Q}^* \mathcal{D}^* x} \mathbf{e}, \quad r_{du}(\alpha) = \mathbf{e}_d^\top (-\alpha \mathbf{I} - \mathcal{Q}^* \mathcal{D}^*)^{-1} \mathcal{Q}^* \mathcal{U}^* \mathbf{e}_u.$$

Similarly, with constant processing rate $\rho_u(t) \equiv 1$, we have $\tau_u(\ell) = \ell$ and so

$$\begin{aligned} r_{ud}(\alpha) &= \mathbb{E}[\mathbf{e}^{\alpha T_u}; T_u < \ell, \xi_1 = d] = \int_0^\ell \mathbf{e}^{\alpha y} \mathbf{e}_u^\top \mathbf{e}^{\mathcal{Q}^* \mathcal{U}^* y} \mathcal{Q}^* \mathcal{U}^* \mathbf{e}_d \, dy \\ &= \mathbf{e}_u^\top (-\alpha \mathbf{I} - \mathcal{Q}^* \mathcal{U}^*)^{-1} (\mathbf{e}^{\alpha \ell} \mathbf{e}^{\mathcal{Q}^* \mathcal{U}^* \ell} - \mathbf{I}) \mathcal{Q}^* \mathcal{U}^* \mathbf{e}_d. \end{aligned}$$

We return in Sect. 8 to the question of how to evaluate such quantities in more general settings than $\rho_u(t) \equiv 1$. $\square$

We also have

Proposition 4.6 *In the Markov model γ always exists and is unique.*

Proof From general results on phase-type distributions based on Perron-Frobenius theory, the tail of F_{du} has asymptotic form $c_{du} x^{n_{ud}} e^{-\alpha^* x}$ for suitable constants α^* , n_{ud} , c_{du} (if $\mathcal{Q}^* \mathcal{D}^*$ is irreducible, $-\alpha^*$ is the eigenvalue of maximal real part and all $n_{ud} = 0$). Thus the $r_{du}(\alpha)$ are defined for $\alpha < \alpha^*$ and some have limit ∞ as $\alpha \uparrow \alpha^*$. Further, letting $\rho^* = \max_{u \in \mathcal{U}^*} \rho_u$, $r_{ud}(\alpha)$ can be bounded by $\mathbb{E}[\mathbf{e}^{\alpha T_u}; T_u \leq \ell/\rho^*]$ which is finite for all α . Thus $\text{spr}(\mathbf{R}(\alpha))$ is defined for all $\alpha \in [0, \alpha^*)$ and by dominated convergence, it is continuous in that domain. The limit at $\alpha = 0$ is < 0 because $\mathbf{R}(0)$ is a proper subtransition matrix (some row sums are < 1 because of the condition $T_u < \tau_u(\ell)$ imposed), and the limit as $\alpha \uparrow \alpha^*$ is ∞ by Lemma 9.2. Also that lemma gives that $\text{spr}(\mathbf{R}(\alpha))$ is strictly increasing in α . Putting these facts together completes the proof. $\square$

5 Theorem 2.4: Proof and Amendments

Let $\mathbf{h}(\ell)$ be the right Perron-Frobenius eigenvector of $\mathbf{R}(\gamma(\ell), \ell)$ corresponding to the eigenvalue 1 and normalized by $\boldsymbol{\pi} \mathbf{h}(\ell) = 1$. Since $\gamma(\ell) \rightarrow 0$ and hence $\mathbf{R}(\gamma(\ell), \ell) \rightarrow \mathbf{P}$ as $\ell \rightarrow \infty$, one expects the following lemma to hold in view of $\mathbf{P} \mathbf{e} = \mathbf{e}$, $\boldsymbol{\pi} \mathbf{e} = 1$; we include the proof since it is short.

Lemma 5.1 $\mathbf{h}(\ell) \rightarrow \mathbf{e}$ as $\ell \rightarrow \infty$.

Proof Assume that $\mathbf{h}(\ell_n) \rightarrow \mathbf{e}$ fails for some sequence $\{\ell_n\}$. The assumption $\boldsymbol{\pi} \mathbf{h}(\ell) = 1$ and $\pi_i > 0$ for all i ensures that the sequence $\{\mathbf{h}(\ell_n)\}$ is relatively compact, so passing to a subsequence, if necessary, we may assume that $\mathbf{h}(\ell_n) \rightarrow \mathbf{e} + \mathbf{f}$ with $\mathbf{f} \neq 0$. From $\mathbf{R}(\gamma(\ell), \ell) \mathbf{h}(\ell_n) = \mathbf{h}(\ell_n)$ and $\mathbf{R}(\gamma(\ell), \ell) \rightarrow \mathbf{P}$ we then get

$\mathbf{P}(\mathbf{e} + \mathbf{f}) = \mathbf{e} + \mathbf{f}$. Since $\mathbf{P}$ is a transition matrix, we have $\mathbf{P}\mathbf{e} = \mathbf{e}$ and so $\mathbf{f}$ is an eigenvector of $\mathbf{P}$ corresponding to the eigenvalue 1. Thus $\mathbf{f} = c\mathbf{e}$ for some c by the Perron-Frobenius theorem, and

$$1 = \boldsymbol{\pi}\mathbf{h}(\ell_n) \rightarrow \boldsymbol{\pi}(\mathbf{e} + \mathbf{f}) = 1 + c\boldsymbol{\pi}\mathbf{e} = 1 + c$$

then gives $c = 0$, a contradiction. $\square$

Proof of Theorem 2.4 Let $\mathbf{K}^*(\ell)$ denote the $\mathcal{E} \times \mathcal{E}$ matrix with ud th element $k_{ud}^*(\ell)$ and the $\mathcal{U} \times \mathcal{U}$ -, $\mathcal{D} \times \mathcal{U}$ - and $\mathcal{D} \times \mathcal{D}$ blocks identically 0. Let further $\mathbf{A}(\alpha, \ell)$ be the matrix with elements

$$a_{ud}(\alpha, \ell) = \int_0^\ell (e^{\alpha y} - 1 - \alpha y) F_{ud}(\mathrm{d}y), \quad a_{du}(\alpha, \ell) = \int_0^\infty (e^{\alpha y} - 1 - \alpha y) F_{du}(\mathrm{d}y)$$

for $u \in \mathcal{U}$, $d \in \mathcal{D}$ and all other $a_{ij}(\alpha, \ell) = 0$, let $\mathbf{M}^*(\ell)$ be the matrix with elements

$$m_{ud}^*(\ell) = \int_0^\ell y F_{ud}(\mathrm{d}y), \quad m_{du}^*(\ell) = \int_0^\infty y F_{du}(\mathrm{d}y)$$

for $u \in \mathcal{U}$, $d \in \mathcal{D}$ and all other $m_{ij}(\alpha, \ell; 1) = 0$, and let $\mathbf{M} = \lim_{\ell \rightarrow \infty} \mathbf{M}(\ell)$, i.e. $\mathbf{M}$ is the matrix with elements

$$m_{ud} = \int_0^\infty y F_{ud}(\mathrm{d}y) = \mathbb{E}[T_u; \xi_1 = d], \quad m_{du} = \int_0^\infty y F_{du}(\mathrm{d}y) = \mathbb{E}[T_d; \xi_1 = u]$$

and all other $m_{ij} = 0$. We then have the identity

$$\mathbf{A}(\alpha, \ell) = \mathbf{R}(\alpha, \ell) - \mathbf{P} + \mathbf{K}^*(\ell) - \alpha \mathbf{M}^*(\ell). \quad (10)$$

Write for convenience $\mathbf{h}(\ell) = \mathbf{e} + \gamma(\ell)\mathbf{n}(\ell)$. Taking $\alpha = \gamma(\ell)$ in (10) and multiplying by $\mathbf{h}(\ell)$ to the right, we obtain

$$\begin{aligned} \mathbf{O}(\gamma(\ell)^2) &= \mathbf{e} + \gamma(\ell)\mathbf{n}(\ell) - \mathbf{P}\mathbf{e} - \gamma(\ell)\mathbf{P}\mathbf{n}(\ell) \\ &\quad + \mathbf{K}^*(\ell)\mathbf{e} + \gamma(\ell)\mathbf{K}^*(\ell)\mathbf{n}(\ell) - \gamma(\ell)\mathbf{M}^*(\ell)\mathbf{e} - \gamma(\ell)^2\mathbf{M}^*(\ell)\mathbf{n}(\ell) \end{aligned}$$

Noting that $\mathbf{P}\mathbf{e} = \mathbf{e}$ and $\boldsymbol{\pi}\mathbf{P} = \boldsymbol{\pi}$, that $\gamma(\ell)\mathbf{n}(\ell) \rightarrow \mathbf{0}$ by Lemma 5.1 and that $\boldsymbol{\pi}\mathbf{n}(\ell) = \mathbf{0}$ because of $\boldsymbol{\pi}\mathbf{h}(\ell) = 1$, it follows by multiplication by $\boldsymbol{\pi}$ to the left that

$$\begin{aligned} \mathbf{o}(\gamma(\ell)) &= \boldsymbol{\pi}\mathbf{K}^*(\ell)\mathbf{e} + \gamma(\ell)\boldsymbol{\pi}\mathbf{K}^*(\ell)\mathbf{e} - \gamma(\ell)\boldsymbol{\pi}\mathbf{M}^*(\ell)\mathbf{e}, \\ \mathbf{o}(1) &= \frac{1}{\gamma(\ell)}\boldsymbol{\pi}\mathbf{K}^*(\ell)\mathbf{e} + \boldsymbol{\pi}\mathbf{K}^*(\ell)\mathbf{n}(\ell) - \boldsymbol{\pi}\mathbf{M}^*(\ell)\mathbf{e} \end{aligned}$$

Since $\pi \mathbf{M}^*(\ell) \mathbf{e} \rightarrow \pi \mathbf{M} \mathbf{e}$ and $\mathbf{K}^*(\ell) = o(1)$ by (3), this gives

$$\pi \mathbf{M} \mathbf{e} = \lim_{\ell \rightarrow \infty} \left[\frac{\varphi(\ell)}{\gamma(\ell)} \frac{\pi \mathbf{K}^*(\ell) \mathbf{e}}{\varphi(\ell)} + o(1) \right] = \lim_{\ell \rightarrow \infty} \frac{\varphi(\ell)}{\gamma(\ell)} \pi \mathbf{K} \mathbf{e}$$

where $\mathbf{K} = \lim \mathbf{K}^*(\ell)/\varphi(\ell)$ is given by (3). Hence

$$\lim_{\ell \rightarrow \infty} \frac{\gamma(\ell)}{\varphi(\ell)} = \frac{\pi \mathbf{K} \mathbf{e}}{\pi \mathbf{M} \mathbf{e}} = \mu. \quad \square$$

Corollary 5.2 Consider the Markov model with $\rho_u(t) \equiv 1$, assume that $\mathbf{Q}_{\mathcal{U}^* \mathcal{U}^*}$ is irreducible, let $-\delta$ be the eigenvalue with maximal real part and $\mathbf{h}_{\mathcal{U}^*}^*, \mathbf{v}_{\mathcal{U}^*}^*$ the corresponding right and left eigenvectors normalized by $\mathbf{v}_{\mathcal{U}^*}^* \mathbf{h}_{\mathcal{U}^*}^* = 1$. Then $\gamma(\ell) \sim \mu e^{-\delta \ell}$ as $\ell \rightarrow \infty$, where

$$\mu = \frac{(\pi_{\mathcal{U}^*}^* \mathbf{h}_{\mathcal{U}^*}^*) \cdot (\mathbf{v}_{\mathcal{U}^*}^* \mathbf{e})}{\pi_{\mathcal{U}^*}^* (-\mathbf{Q}_{\mathcal{U}^* \mathcal{U}^*})^{-1} \mathbf{e} + \pi_{\mathcal{D}^*}^* (-\mathbf{Q}_{\mathcal{D}^* \mathcal{D}^*})^{-1} \mathbf{e}}$$

with $\pi^* = (\pi_{\mathcal{U}^*}^*, \pi_{\mathcal{D}^*}^*)$ the stationary distribution of $\mathbf{Q}$.

Proof By Perron-Frobenius theory, $-\delta$ is a simple eigenvalue and

$$\begin{aligned} k_{ud}^*(\ell) &= \mathbb{P}(T_u > \tau_u(\ell), \xi_1 = d) = \mathbb{P}(T_u > \ell, \xi_1 = d) \\ &= \mathbf{e}_u \exp\{\mathbf{Q}_{\mathcal{U}^* \mathcal{U}^*} \ell\} \mathbf{e}_d \sim \mathbf{e}_u (\mathbf{h}_{\mathcal{U}^*}^* \mathbf{v}_{\mathcal{U}^*}^*) \mathbf{e}_d \cdot e^{-\delta \ell} \end{aligned}$$

Thus we may take $\varphi(\ell) = e^{-\delta \ell}$ and the expression for μ then easily comes out. $\square$

Remark 5.3 The assumption of $\mathbf{Q}_{\mathcal{U}^* \mathcal{U}^*}$ being irreducible holds in all the examples we present. For an example where it fails, assume

$$\mathcal{U}^* = \{1, 2\}, \quad \mathcal{D}^* = \{3\}, \quad \mathbf{Q} = \begin{pmatrix} -a & a & 0 \\ 0 & -a & a \\ b & c & -b-c \end{pmatrix}$$

with at least $a, b > 0$. Here the eigenvalue $-a$ of $\mathbf{Q}_{\mathcal{U}^* \mathcal{U}^*}$ is not simple and $\mathbf{Q}_{\mathcal{U}^* \mathcal{U}^*}$ is reducible (1 cannot be reached from 2). Nevertheless, the assumptions of Theorem 2.4 hold since $\mathbf{Q}_{\mathcal{U}^* \mathcal{U}^*}$ is the rate matrix of an Erlang(2, a) distribution and so

$$\begin{aligned} \mathbb{P}_1(T_1 > \ell, \xi_1 = 3) &= \mathbb{P}_1(T_1 > \ell) = a\ell e^{-a\ell} + e^{-a\ell}, \\ \mathbb{P}_2(T_2 > \ell, \xi_1 = 3) &= \mathbb{P}_2(T_2 > \ell) = e^{-a\ell}. \end{aligned}$$

I.e., we may take $\varphi(\ell) = \ell e^{-a\ell}$ and get $k_{13}^* = a, k_{23}^* = 0$. In general, dealing with reducibility and eigenvalues that are not simple lead into the Jordan canonical form; we omit the details. $\square$

Proof of Corollary 5.2 We use once more formula (2) for a random task length L , stating that

$$\mathbb{P}_i(X > x) = \int_0^\infty \mathbb{P}_i(X > x \mid L = \ell) \mathbb{P}(L \in d\ell). \quad (11)$$

It is shown in [2] that here it is permissible to insert the approximations in Theorem 2.4 and Corollary 5.2, leading to

$$\mathbb{P}_i(X > x) \sim \int_{\ell_0}^\infty h_i C_1 / C_2 \exp\{-\mu e^{-q\ell}\} a e^{-\delta\ell} d\ell. \quad (12)$$

Here h_i, C_1, C_2 depend on ℓ . Now $h_i \rightarrow 1$ by Lemma 5.1 and similarly $v_i \rightarrow \pi_i$ as $\ell \rightarrow \infty$. It then follows from the assumption $\mathbb{P}_u(\tau(\ell) \rightarrow \infty) = 1$ that C_1, C_2 have limits

$$D_1 = \sum_{u \in \mathcal{U}} \pi_u \cdot 1 \cdot 0 + \sum_{d \in \mathcal{D}} \pi_d \mathbb{E} T_d = \sum_{d \in \mathcal{D}} \pi_d \mathbb{E} T_d \quad (13)$$

$$D_2 = \sum_{u \in \mathcal{U}, d \in \mathcal{D}} \pi_u \cdot 1 \int_0^\infty y \cdot 1 \cdot 1 F_{ud}(dy) + \sum_{d \in \mathcal{D}, u \in \mathcal{U}} \pi_d \cdot 1 \int_0^\infty y \cdot 1 F_{du}(dy) \quad (14)$$

$$= \sum_{i \in \mathcal{E}} \pi_i \mathbb{E}_i T_i \quad (15)$$

The asymptotics of the integral in (12) is determined in Lemma 9.3, and the result follows with

$$C = \frac{\Gamma(\delta/q) D_1}{\mu^{\delta/q} q^a D_2}. \quad (16)$$

□

Remark 5.4 Given Theorem 2.4, the asymptotics of $\mathbb{P}(X > x)$ can easily be found in a variety of combinations of the forms of $\varphi(\ell)$ and the distribution of L . To this end, simply insert in the integral estimates of [2] (note that in many cases only logarithmic asymptotics comes out). We omit the lengthy statement of all of the available results, but remark that the case where $\varphi(\ell)$ and the distribution of L are somewhat alike is particularly nice. More precisely, with the assumption

$$\mathbb{P}(L \in d\ell) = \mu \varphi'(\ell) (\mu \varphi(\ell))^{\beta-1} L_0(\varphi(\ell))$$

on the density of L where $\beta > 0$ and L_0 is slowly varying at 0, a Tauberian argument gives

$$\mathbb{P}(X > x) \sim \frac{\Gamma(\beta) D_1}{\mu^\beta D_2} \frac{L_0(1/x)}{x^\beta}, \quad x \rightarrow \infty. \quad \square$$

Example 5.5 The asymptotic parameter, γ , depends heavily on the type of system under examination. But it is still of interest to get some idea of how it behaves. A detailed study of this set of problems is presented in our paper [21], but we consider here the simplest system, a single server (On state) that has an exponential failure rate of β , and a single repairman (Down state) with exponential repair rate, λ .

In this case, the equation $\text{spr}(\mathbf{R}(\alpha)) = 1$ reduces to the root finding formula,

$$f(\alpha = \gamma) = \frac{\lambda}{(\lambda - \alpha)} \frac{\beta[1 - e^{-\ell(\beta - \alpha)}]}{(\beta - \alpha)} - 1 = 0.$$

From Fig. 4 we see that γ depends heavily on both ℓ and λ . The property that $\gamma(0|\lambda, \beta) = \lambda$ is an artifact of the fact that the repair time distribution is exponential(λ), and for $\delta = \ell\beta \ll 1$ the probability that the system will fail one more time is of order δ . One expects in general that the behavior of γ near $\ell = 0$ should be dominated by the asymptotic failure distribution, namely $\pi_d \exp(\mathbf{Q}_{\mathcal{D}\mathcal{D}} t) \mathbf{e}$. For large ℓ , the exponential decay of γ given by Theorem 2.4 is confirmed by the figure, as seen by the straight lines, all with asymptotic slope $= -\beta = -1$, at least for the simple system examples given. $\square$

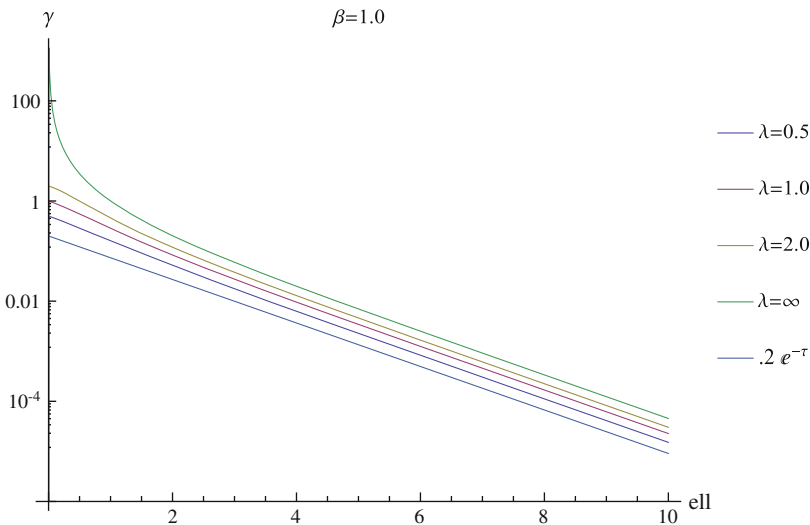


Fig. 4 $\log[\gamma(\ell | \lambda, \beta)]$ as a function of ℓ for $\lambda \in \{0.5, 1.0, 2.0, \infty\}$ and $\beta = 1$, for the simplest system of one *Up* and one *Down* state. The function, $\log[.2 \exp(-\ell)]$ is also included, showing all the curves have asymptotic slope $= -\beta = -1$. For this system, $\gamma(0|\lambda, \beta) = \lambda$

6 The Rare Events Approach. The Mean Task Time

We now take a different approach to the study of $X = X(\ell)$, the theory of rare events in regenerative processes. The key fact is that for such a process the time until a rare event (when properly normalized) has an approximate exponential distribution, and that the corresponding convergence in distribution is accompanied by convergence of expected values, cf. [1, VI.4]. For simplicity, we will assume that all $\rho_u(t) \equiv 1$.

Theorem 6.1 *Under the assumptions of Theorem 2.4, it holds as $\ell \rightarrow \infty$ that $\gamma(\ell)X(\ell) \xrightarrow{\mathcal{D}} V$, $\gamma(\ell)\mathbb{E}X(\ell) \xrightarrow{\mathcal{D}} 1$ where V is standard exponential.*

Proof Fix a reference state $u^* \in \mathcal{U}$ and define

$$\sigma = \inf\{n > 0 : \xi_n = u^* \mid \xi_0 = u^*\}, \quad \tau = \sum_{n=0}^{\sigma-1} T_{\xi_n}^n$$

where $T_i^n \stackrel{\mathcal{D}}{=} T_i$ and obvious independence assumptions apply. Thus σ is the number of Markov steps needed to return to u^* and τ the time the Markov renewal process takes for that, which is our choice of the regenerative cycle. By [1] Proposition 4.2, p. 207, slightly rewritten, we have

$$\mathbb{E}\tau = \frac{1}{\pi_{u^*}} \sum_{i \in \mathcal{C}} \pi_i \mathbb{E}_i T_i.$$

Let $M = \max_{n < \sigma, \xi_n \in \mathcal{U}} T_{\xi_n}^n$. Then the task is handled in the cycle if and only if $M > \ell$, and we let $a(\ell)$ denote the corresponding probability. Using Proposition 10.1 of [4] gives

$$a(\ell) \sim \frac{1}{\pi_{u^*}} \sum_{u \in \mathcal{U}} \pi_u \mathbb{P}(T_u > \ell).$$

Further Propositions V.4.1 and V.4.2 of [1, pp. 180–181] give

$$\frac{a(\ell)}{\mathbb{E}\tau} X(\ell) \xrightarrow{\mathcal{D}} V, \quad \frac{a(\ell)}{\mathbb{E}\tau} \mathbb{E}X(\ell) \xrightarrow{\mathcal{D}} 1, \quad \ell \rightarrow \infty$$

(note that these results squeeze $X(\ell)$ in between the start and the of the cycle where the task is terminated). Now just note that $a(\ell)/\mathbb{E}\tau \sim \gamma(\ell)$ by Theorem 2.4. $\square$

Remark 6.2 It is tempting to write the conclusion $\gamma(\ell)X(\ell) \xrightarrow{\mathcal{D}} V$ of Theorem 6.1 as $\mathbb{P}(X > x) \sim e^{-\gamma x}$ which is similar to Theorem 2.1, but the crucial difference is that the limit there is $x \rightarrow \infty$ with ℓ fixed whereas Theorem 6.1 requires that x is of

order $1/\gamma(\ell)$ and that $\ell \rightarrow \infty$. The implication is that whereas none of the results imply the other, they provide together support for the interchangeability of the limits $x \rightarrow \infty$ and $\ell \rightarrow \infty$. Similar remarks apply to $\mathbb{E}X$. $\square$

7 Theorem 2.2: Proof and Amendments

For Theorem 2.2, one needs a Markov renewal version of the key renewal theorem with defective heavy tails. This appears, however, not to be in the literature. One may note in this connection that already for an ordinary renewal equation

$$Z(x) = z(x) + \int_0^\infty Z(x-y) F(dy) \quad (17)$$

(with only one unknown function Z), such a result is relatively new and was only given fairly recently in Asmussen et al. [3]. The details of this analysis are quite technical and one needs to go somewhat beyond the ordinary subexponential setting by imposing conditions not only on the tail of F but also on the local behavior. This involves the definition of a distribution F to have a subexponential density, namely that there exists $\hat{x} < \infty$ such that F has a density $f(x)$ on $[\hat{x}, \infty)$ and that

$$f^{*2}(x) = 2 \int_0^{\hat{x}} f(x-y) F(dy) + \int_{\hat{x}}^{x-\hat{x}} f(x-y) f(y) dy \sim 2f(x), \quad x \rightarrow \infty. \quad (18)$$

A complete and rigorous treatment of the relevant version of the Markov key renewal theorem will be presented elsewhere, with Lemma 7.1 stated below without proof being the main extension of [3], but we present here the basic intuition. If $\theta = \int_0^\infty F(dx) < 1$ in the simple case of (17), the idea in [3] is to use the convolution structure of the renewal equation to view $Z(x)$ as the density of the sum of a geometric sum of r.v.'s with distribution F/θ and a r.v. W with density $z(x)/\int z$. Three cases arise according to the balance between the heaviness of z and the tail of F . The one corresponding to the present case is $z(x)$ being heavier than the density f of F , and here the contribution from the geometric sum vanishes, giving $Z(x) \sim z(x)/\int z$.

In the setting of Theorem 2.2, the representation in [1, Proposition 4.4 p. 209, of the solution $(Z_i(x))_{i \in \mathcal{E}}$ to the set of Markov renewal equations (5) can be written as

$$Z_i(x) = \sum_{n=0}^{\infty} \sum_{j \in \mathcal{E}} \int_0^x z_j(x-y) \mathbb{P}_i(\xi_n(\ell) = j, S_n(\ell) \in dy) \quad (19)$$

where $\xi_n(\ell)$, $S_n(\ell)$ are defined as follows. Let N be the Markov renewal epoch at which the task is processed, Δ some extra absorbing state Δ , and $\xi(\ell)$ the Markov chain with state space $\mathcal{E} \cup \{\Delta\}$ such that $\xi_n(\ell) = \xi_n$ for $n < N$ and $\xi_n(\ell) = \Delta$ for $n \geq N$. Further, $S_n(\ell) = T_0^* + \dots + T_n^*$ for $n < N$, $S_n(\ell) = \infty$ for $n \geq N$ where

given the Markov chain $\xi(\ell)$, the T_k^* are independent with T_k^* distributed as T_i on the event $\xi_k(\ell) = i$ (the definition of $S_n(\ell)$ for $n \geq N$ is redundant but one may take, e.g., $S_n(\ell) = \infty$). We need the following extension of Propositions 7, 8 of [3]:

Lemma 7.1 *Let $G_1, \dots, G_A$ be a finite set of distributions such that each G_a admits a subexponential density $g_a(x)$ with $g_a(x) \sim c_a \varphi(x)$ for some subexponential density $\varphi(x)$. Then for any $n_1, \dots, n_A$*

$$(g_1^{n_1} * \dots * g_A^{n_A})(x) \sim (n_1 c_1 + \dots + n_A c_A) \varphi(x),$$

and for any $\varepsilon > 0$, there exists C_ε such that

$$(g_1^{n_1} * \dots * g_A^{n_A})(x) \leq C_\varepsilon (1 + \varepsilon)^{n_1 + \dots + n_A} \varphi(x),$$

Now the transition probabilities $p_{ij}(\ell)$ of $\xi(\ell)$ are given by $p_{\Delta\Delta}(\ell)(\ell) = 1$, $p_{du}(\ell) = p_{du}$,

$$p_{ud}(\ell) = \mathbb{P}_u(T_d < \tau_u(\ell), \xi_1 = d), \quad p_{u\Delta} = 1 - \sum_{d \in \mathcal{D}} p_{ud}(\ell) = \mathbb{P}_u(T_d \geq \tau_u(\ell))$$

all other $p_{ij}(\ell) = 0$. Since Δ is absorbing, $\text{spr}(\mathbf{P}(\ell)) < 1$ and so Perron-Frobenius theory gives the existence of $b < 1$ such that the n -step transition probabilities $p_{ij}^n(\ell)$ decay at rate at most b^n . For $j \neq \Delta$ one has $\mathbb{P}_i(\xi_n(\ell) = j) = p_{ij}^n(\ell)$ and letting W_d be a r.v. with density $z_j^*(y) = z_j(y)/\bar{z}_j$ where $\bar{z}_j = \int_0^\infty z_j(y) dy$, we can rewrite (19) as

$$Z_i(x) = \sum_{n=0}^{\infty} \sum_{j \in \mathcal{E}} p_{ij}^n(\ell) \bar{z}_j \mathbb{E}[(g_{n,j} * z_j^*)(x)]$$

where $g_{n,j}$ is the conditional density of $S_n(\ell)$ given $\xi_0 = i, \xi_1, \dots, \xi_{n-1}$. By the first part of Lemma 7.1, $g_{n,j}(x) \sim c_n \varphi(x)$ for some (random) c_n . Further, the $z_d(x)$ are heavier than $\varphi(x)$ (that $\bar{\Phi}(x)/\varphi(x) \rightarrow \infty$ is a standard estimate in the heavy-tailed area), and so $(g_{n,j} * z_j^*)(x) \sim z_j^*(x)$ by Proposition 7 of [3]. Choosing ε with $(1 + \varepsilon)b < 1$, we can then use the second part of Lemma 7.1 and a dominated convergence argument to conclude that

$$Z_i(x) \sim \sum_{n=0}^{\infty} \sum_{j \in \mathcal{E}} p_{ij}^n(\ell) \bar{z}_j z_j^*(x) \sim \sum_{n=0}^{\infty} \sum_{j \in \mathcal{E}} p_{ij}^n(\ell) c_d \Phi(x).$$

This completes the proof of Theorem 2.2, with the expression

$$C_i = \mathbf{e}_i (\mathbf{I} - \mathbf{P}(\ell))^{-1} \mathbf{c} \quad (20)$$

for the C_i where $\mathbf{c}$ is the vector with j -th entry c_d for $i = d \in \mathcal{D}$ and 0 for $i = u \in \mathcal{U}$. $\square$

Example 7.2 Consider as in Sect. 2 vanilla RESTART with repair, $\mathcal{U} = \{u\}$, $\mathcal{D} = \{d\}$, and $\rho_u(t) \equiv 1$. Then F_u is the distribution of the failure time, F_d the distribution of the repair time, and with heavy-tailed F_d , Theorem 2.2 immediately gives that the total task time X has tail proportional to $\overline{F_d}(x)$.

A particular case of a heavy-tailed F_d could arise if repair means rebooting of the processor, such that the ideal time for rebooting is B but that rebooting may fail at Poisson(β) times and then itself needs to be restarted. This means that given $B = b$, T_d is the RESTART total task time with exponential(β) failure and task length b , so that by [2] $\overline{F_d} \sim C(b)e^{-\eta(b)x}$, with the values of $C(b)$, $\eta(b)$ given there. For a random B with unbounded support, T_d becomes heavy-tailed. For example, for a Gamma-like B we have that (again by reference to [2])

$$\overline{F_d}(x) \sim c_{\alpha,\delta} \frac{\log^{\alpha-1} x}{x^{\eta_{\alpha,\delta}}}$$

with $c_{\alpha,\delta}$, $\eta_{\alpha,\delta}$ specified in [2]. Thus F_d is regularly varying and Theorem 2.2 applies. $\square$

8 Time-Varying Processing Rates

As is clear from the formulas presented so far, a main problem for numerical implementation beyond the simplest case $\rho_u(t) \equiv \rho$ is evaluation of quantities like

$$g_{ud}(\ell) = \mathbb{P}(T_u > \tau_u(\ell), \xi_1 = d) = \mathbb{P}\left(\int_0^{T_u} \rho_u(t) dt > \ell, \xi_1 = d\right).$$

Example 8.1 A simple example is obtained by letting the server speed depend on time. E.g. the server slowing down with time could be modeled by the service rate being proportional to $t^{-\alpha}$ with $0 < \alpha < 1$, and then $\tau_u(\ell)$ is proportional to $\ell^{1-\alpha}$. $\square$

Another tractable and more interesting case is what we call the *independent Markov rate model* where the server speed is determined by an independent external environment. This is defined in terms of a set of Markov process $I = \{I(t)\}_{t \geq 0}$ with state space $\{1, \dots, m\}$ and a set of rates $r_1, \dots, r_m$, such that $\rho_u(t) = r_{I(t)}$. The basic assumption is that I is independent of the Markov renewal process. We denote the rate matrix by $\mathbf{A}$ and the initial distribution by $\boldsymbol{\alpha}$. A natural choice of $\boldsymbol{\alpha}$ is the stationary distribution for I , and dependence of the rates on u can be obtained by taking I complicated enough. Write $\mathbf{\Delta}_r$ for the diagonal matrix with the r_i on the diagonal.

Proposition 8.2 Consider the independent Markov model with all $r_i > 0$ and assume that T_u is independent of ξ_1 and exponential (δ). Then

$$g_{ud}(\ell) = p_{ud} \boldsymbol{\alpha} \exp\{\mathbf{B}(\delta)\ell\} \mathbf{e}$$

$$\text{where } \mathbf{B}(\delta) = \Delta_r^{-1} \mathbf{A} - \frac{\delta}{2} (\Delta_r^{-1} \mathbf{e} \mathbf{e}^\top + \mathbf{e} \mathbf{e}^\top \Delta_r^{-1}).$$

Proof Fix u, d and define $\mathbf{f}(x)$ as the row vector with elements

$$f_j(x) = \mathbb{P}_\alpha \left(\int_0^{T_u} \rho_u(t) dt > x, I(\tau_u(x)) = j \right).$$

We shall identify $f_j(x+h)$ up to $o(h)$ terms and thereby derive a differential equation for $\mathbf{f}(x)$, giving $\mathbf{f}(x) = \alpha e^{B(\delta)x}$ from which the result follows by letting $x = \ell$, summing over j (corresponding to multiplication by $\mathbf{e}$) and using the assumed independence of ξ_1 to extract the transition probability p_{ud} .

For $\int_0^{T_u} \rho_u(t) dt > x+h$ to occur, we must have $\tau_u(x) < T_u$. The contribution to $f_j(x+h)$ from the event $I(\tau_u(x+h)) = I(\tau_u(x)) = j$ is therefore

$$\begin{aligned} f_j(x)(1 + a_{jj}h/r_j) \mathbb{P}(T_u > \tau_u(x) + h/r_j \mid T_u > \tau_u(x)) + o(h) \\ = f_j(x)(1 + a_{jj}h/r_j - \delta h/r_j) + o(1), \end{aligned}$$

where the $o(h)$ terms take care of, e.g., the possibility of jumps out of j and back in the time interval $(\tau_u(x), \tau_u(x+h)]$. Consider next the contribution from the event $I(\tau_u(x)) = k, I(\tau_u(x+h)) = j$ and a single jump from k to j in the time interval $(\tau_u(x), \tau_u(x+h)]$. This jump must occur before $\tau_u(x) + h/r_k$, and its position is then approximately at $\tau_u(x) + Uh/r_k$ where U is an independent uniform(0, 1) r.v. Passage to $x+h$ must therefore occur at time $\tau_u(x) + Uh/r_k + (1-U)h/r_j$, and the probability that T_u survives from $\tau_u(x)$ to there is approximately

$$\mathbb{E}\{-\delta[Uh/r_k + (1-U)h/r_j]\} = 1 - h\frac{\delta}{2}(1/r_k + 1/r_j).$$

The asked for contribution is thus

$$\sum_{k \neq j} f_k(x) \left(a_{kj}h/r_k - h\frac{\delta}{2}(1/r_k + 1/r_j) \right). \quad (21)$$

Alltogether,

$$\begin{aligned} f_j(x+h) &= f_j(x) + \sum_{k=1}^m f_k(x) \left(a_{kj}h/r_k - h\frac{\delta}{2}(1/r_k + 1/r_j) \right) + o(h), \\ f'_j(x) &= \sum_{k=1}^m f_k(x) \left(a_{kj}/r_k - \frac{\delta}{2}(1/r_k + 1/r_j) \right). \end{aligned}$$

In matrix notation, this means $\mathbf{f}'(x) = \mathbf{f}(x)\mathbf{B}(\delta)$ which together with $\mathbf{f}(0) = \alpha$ gives the desired conclusions. $\square$

Remark 8.3 The assumption of T_u being exponential is easily generalized to a PH distribution at the expense of the differential equations and the form of $\mathbf{B}$ become somewhat more complicated. We omit the details. $\square$

We next turn to the Markov model of Sect. 3, where the rate generating process is internal rather than external. That is, the role of I is taken by the J -process of Sect. 3.

Proposition 8.4 *Consider the Markov model with $r_v > 0$ for all $v \in \mathcal{U}$. Then*

$$g_{ud}(\ell) = \mathbf{e}_u^\top \exp\{\Delta_r^{-1} \mathbf{Q}_{\mathcal{U}^* \mathcal{U}^*} \ell\} (-\mathbf{Q}_{\mathcal{U}^* \mathcal{U}^*})^{-1} \mathbf{Q}_{\mathcal{U}^* \mathcal{D}^*} \mathbf{e}_d.$$

Proof Define $f(x)$ as the row vector with elements

$$f_v(x) = \mathbb{P}_u \left(\int_0^{T_u} r_{J(t)} dt > x, J(\tau_u(x-)) = v \right), \quad v \in \mathcal{U}.$$

Then $f'(x) = f(x) \Delta_r^{-1} \mathbf{Q}_{\mathcal{U}^* \mathcal{U}^*}$ and thus $f(x) = \mathbf{e}_u^\top \exp\{\Delta_r^{-1} \mathbf{Q}_{\mathcal{U}^* \mathcal{U}^*} x\}$. Indeed, the only difference from the proof of Proposition 8.2 is that the $\delta/2$ term in (21) does not enter because the possibility $T_u \in (\tau_u(x), \tau_u(x+h)]$ is taken care of by the rows of $\mathbf{Q}_{\mathcal{U}^* \mathcal{U}^*}$ not summing to 0 (what is missing is the rates of entering $\mathcal{D}^*$). Now just note that

$$\begin{aligned} \mathbb{P}_u \left(\int_0^{T_u} \rho_u(t) dt > \ell, \xi_1 = d \mid J(\tau_u(\ell-)) = v \right) \\ = \int_{\tau_u(\ell)}^\infty \mathbf{e}_v^\top \exp\{\mathbf{Q}_{\mathcal{U}^* \mathcal{U}^*} (t - \tau_u(\ell))\} \mathbf{Q}_{\mathcal{U}^* \mathcal{D}^*} \mathbf{e}_d dt = \mathbf{e}_v^\top (-\mathbf{Q}_{\mathcal{U}^* \mathcal{U}^*})^{-1} \mathbf{Q}_{\mathcal{U}^* \mathcal{D}^*} \mathbf{e}_d. \end{aligned}$$

$\square$

Proposition 8.5 *Consider the Markov model with $r_v > 0$ for all $v \in \mathcal{U}$. Then*

$$r_{ud}(\alpha, \ell) = \mathbf{e}_u^\top \left(\mathbf{I} - \exp\{\Delta_r^{-1} (\mathbf{Q}_{\mathcal{U}^* \mathcal{U}^*} + \alpha \mathbf{I}) \ell\} \right) (-\mathbf{Q}_{\mathcal{U}^* \mathcal{U}^*} - \alpha \mathbf{I})^{-1} \mathbf{Q}_{\mathcal{U}^* \mathcal{D}^*} \mathbf{e}_d.$$

Proof We have

$$\begin{aligned} r_{ud}(\alpha, \ell) &= \mathbb{E}[e^{\alpha T_u}; \xi_1 = d] - \mathbb{E}[e^{\alpha T_u}; \int_0^{T_u} r_{J(t)} dt > \ell, \xi_1 = d] \\ &= \mathbf{e}_u^\top (-\mathbf{Q}_{\mathcal{U}^* \mathcal{U}^*} - \alpha \mathbf{I})^{-1} \mathbf{Q}_{\mathcal{U}^* \mathcal{D}^*} \mathbf{e}_d - \mathbb{E}[e^{\alpha T_u}; \int_0^{T_u} r_{J(t)} dt > \ell, \xi_1 = d]. \end{aligned}$$

The idea is now to note the identity

$$\mathbf{e}^{\alpha T_u} = \exp\{\alpha \tau_u(\ell)\} \cdot \exp\{\alpha (T_u - \tau_u(\ell))\}$$

(valid on the set $\{T_u > \tau_u(\ell)\}$), take care of the first factor by the differential equation approach and the second by the conditioning argument in the proof of Proposition 8.4.

So, define this time $\mathbf{f}(x)$ as the row vector with elements

$$f_v(x) = \mathbb{E}_u[\exp\{\alpha\tau_u(x)\}; \int_0^{T_u} r_{J(t)} dt > x, J(\tau_u(x-)) = v], \quad v \in \mathcal{U}.$$

Now

$$\exp\{\alpha\tau_u(x+h)\} = \exp\{\alpha\tau_u(x)\} \cdot \exp\{\alpha h/r_k\} \approx \exp\{\alpha\tau_u(x)\} \cdot (1 + \alpha h/r_k)$$

on the set $\{T_u > \tau_u(x+h), J(\tau_u(x)) = k\}$. This gives

$$\mathbf{f}'(x) = \mathbf{f}(x)\mathbf{\Delta}_r^{-1}(\mathbf{Q}_{\mathcal{U}^*\mathcal{U}^*} + \alpha\mathbf{I}), \quad \mathbf{f}(x) = \mathbf{e}_u^\top \exp\{\mathbf{\Delta}_r^{-1}(\mathbf{Q}_{\mathcal{U}^*\mathcal{U}^*} + \alpha\mathbf{I})x\}.$$

Further,

$$\begin{aligned} & \mathbb{E}_u[\exp\{\alpha(T_u - \tau_u(\ell))\}, \int_0^{T_u} \rho_u(t) dt > \ell, \xi_1 = d \mid J(\tau_u(\ell-)) = v] \\ &= \int_{\tau_u(\ell)}^{\infty} \exp\{\alpha(t - \tau_u(\ell))\} \mathbf{e}_v^\top \exp\{\mathbf{Q}_{\mathcal{U}^*\mathcal{U}^*}(t - \tau_u(\ell))\} \mathbf{Q}_{\mathcal{U}^*\mathcal{U}^*} \mathbf{e}_d dt \\ &= \mathbf{e}_v^\top (-\mathbf{Q}_{\mathcal{U}^*\mathcal{U}^*} - \alpha\mathbf{I})^{-1} \mathbf{Q}_{\mathcal{U}^*\mathcal{U}^*} \mathbf{e}_d. \end{aligned}$$

The rest of the proof is easy manipulations. $\square$

Remark 8.6 In Propositions 8.2–8.5, the assumption of strictly positive rates can be dispensed with by working with a reduced state space formed only by the states having rates > 0 . Consider for example Proposition 8.2 and assume $r_i > 0$ for $i = 1, \dots, n$, $r_i = 0$ for $i = n+1, \dots, m$ with $1 \leq n < m$. The $f_i(x)$ then only need to be defined for $i = 1, \dots, n$. Dividing $\mathbf{A}$ as

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix},$$

where $\mathbf{A}_{11}$ is $n \times n$, $\mathbf{A}_{12}$ $n \times (m-n)$ etc., the rate of starting an excursion in $\{n+1, \dots, m\}$ from i and ending it by a jump to j is

$$\sum_{k=n+1}^n a_{jk} \int_0^{\infty} \mathbf{e}_k e^{A_{22}t} \mathbf{e}_j dt = a_{jk} \mathbf{e}_k (-\mathbf{A}_{22})^{-1} \mathbf{e}_j,$$

and so Proposition 8.2 holds if $\mathbf{B}(\delta)$ is modified to

$$\mathbf{\Delta}_{r;11}^{-1} (\mathbf{A}_{11} - \mathbf{A}_{11}(\mathbf{A}_{22})^{-1} \mathbf{A}_{21}) - \frac{\delta}{2} (\mathbf{\Delta}_{r;11}^{-1} \mathbf{e} \mathbf{e}^\top + \mathbf{e} \mathbf{e}^\top \mathbf{\Delta}_{r;11}^{-1}),$$

in obvious block notation for $\mathbf{\Delta}_{r;11}^{-1}$.

If in the Markov model $\mathcal{D}$ consists of a single state, Proposition 8.4 gives the tail probability of $\int_0^{T_u} r_{J(t)} dt$. This is a ‘fluid reward’, and with the extension sketched in Remark 8.6, Proposition 8.4 reduces to a well-known expression for the marginal distributions in the Kulkarni class of multivariate PH distributions, [15]. $\square$

9 Phase-Type Modeling

Standard Markov models assume exponential distributions of failure times, repair times etc. This may seem intrinsically inherent in the Markovian set-up because of the memoryless property of the exponential distribution, but there is in fact a simple approach going far beyond this by using PH distributions as introduced in Sect. 3. We shall not go into the general formulation but only consider a basic example, the *Erlang distribution* $E_p = E_p(\lambda)$ with $p = 1, 2, \dots$ stages defined as the sum of p i.i.d. exponential(μ) r.v.’s (the stages). The density is $\lambda^p t^{p-1} e^{-\lambda t} / \Gamma(p)$ where $\Gamma(p) = (p-1)!$. The Erlang is therefore just a Gamma with integer shape parameter. However, the important fact is the probabilistic interpretation in terms of the stages which we model as a Markovian movement between stages 1, 2 and the final one, completion, with transition rate λ for each of the two transitions. In some examples the stages may have a physical interpretation; e.g., stage 1 may be the real repair and stage 2 checking or warm-up. Usually, the Erlang(p) distribution is, however, mainly used in a descriptive way to allow for including non-exponential distributions in a Markovian set-up.

We will consider an elaborate example of the use of the Erlang distribution which from the general view of the paper also illustrates how identical Markov processes may lead to different Markov renewal schemes. We look at a system with 3 identical components, each with exponential(λ) operating times, and two repairmen with $E_p(\mu)$ repair times. A Markov state of the whole system is specified with the number $i = 0, 1, 2, 3$ of failed components and the stage(s) in which the servers are currently operating. In Fig. 6, state 2:12 indicates that 2 components are failed and one server working in stage 1, the other in stage 2. In state 1:2 only one component is failed, i.e. under repair, and the repairman working on it is in stage 2. In state 3:11, all 3 components are failed, one waiting for repair and the servers working on the other two are both in stage 1; etc. The system may be parallel as in Fig. 5, in which case $\mathcal{U}^* = \{0, 1:1, 1:2, 2:11, 2:12, 2:22\}$ or, e.g., 2-out-of-3 as in Fig. 6 so that $\mathcal{U}^* = \{0, 1:1, 1:2\}$.

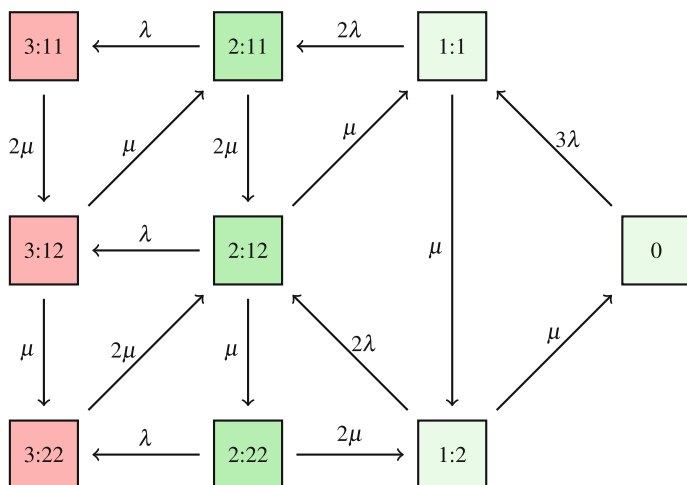


Fig. 5 E_2 repair times, parallel

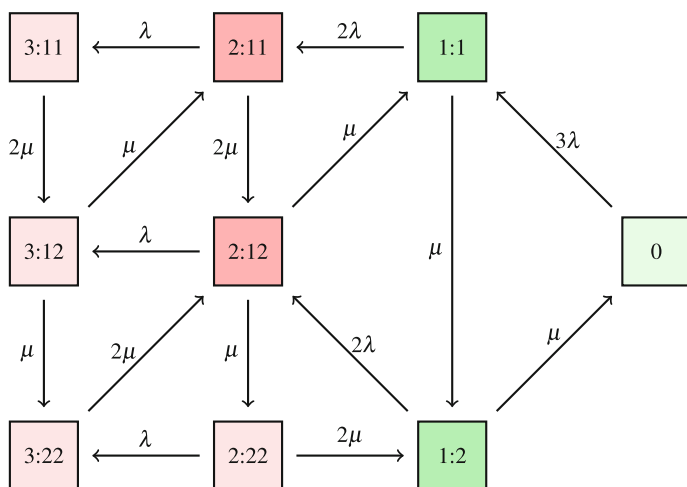


Fig. 6 E_2 repair times, 2-out-of-3

Appendix

The following result is given in [1, Th. VII.4.6]:

Lemma 9.1 *Consider a Markov renewal equation*

$$Z_i(x) = z_i(x) + \sum_{j \in \mathcal{E}} \int_0^x Z_j(x-y) \tilde{F}_{ij}(dy), \quad i \in \mathcal{E},$$

with $\mathcal{E}$ finite and the matrix $(\|\tilde{F}_{ij}\|)$ irreducible. Let $r_{ij} = \int_0^\infty e^{\gamma u} \tilde{F}_{ij}(du)$, suppose that for some real γ the matrix $\mathbf{R} = (r_{ij})$ has spectral radius 1, and choose $\mathbf{v}, \mathbf{h}$ with $\mathbf{vA} = \mathbf{v}$, $\mathbf{Ah} = \mathbf{h}$, $v_i > 0$, $h_i > 0$, $i \in \mathcal{E}$. Then $Z_i(x) \sim h_i C_1 / C_2$ where

$$C_1 = \sum_{j=1}^p v_j \int_0^\infty e^{\gamma x} z_j(x) dx, \quad (22)$$

$$C_2 = \sum_{k,j=1}^p v_k h_j \int_0^\infty x e^{\gamma x} \tilde{F}_{kj}(dx). \quad (23)$$

The following lemma is unsurprising given the Perron-Frobenius theory of non-negative matrices, but included here in this precise form for the sake of easy reference:

Lemma 9.2 *Let for $k = 1, 2$, $\mathbf{A}(k) = (a_{ij}(k))$ be irreducible non-negative matrices and let $\gamma(k) = \text{spr}(\mathbf{A}(k))$. Assume that $\mathbf{A}(1) \geq \mathbf{A}(2)$ and that $a_{i'j'}(1) > a_{i'j'}(2) + b > 0$ for at least one pair $i'j'$ and some $b > 0$. Then $\gamma(1) \geq c(b)\gamma(2)$ for some constant $c(b) > 1$ depending only on $\mathbf{A}(2)$ and satisfying $c(b) \rightarrow \infty$ as $b \rightarrow \infty$.*

Proof Irreducibility of $\mathbf{A}(2)$ ensures that for some n and all ij there exists a path $i_0 i_1 \cdots i_{n-1} i_n$ with $i_0 = i$, $i_n = j$, all $a_{i_{k-1}i_k}(2) > 0$ and $i_{k-1}i_k = i'j'$ for some k' . Hence

$$a_{ij}(1) > a_{ij}(2) + b \prod_{k=1, k \neq k'}^n a_{i_{k-1}i_k}(2) > a_{ij}(2)(1 + bc_1(i, j))$$

for some $c_1(i, j) > 0$. Thus $\mathbf{A}(1)^n \geq c(b)^n \mathbf{A}(2)^n$ where $c(b) = (1 + \inf_{i,j} c_1(i, j))$. But with $\boldsymbol{\pi}(1)$ the positive left (row) eigenvector of $\mathbf{A}(1)$ corresponding to $\gamma(1)$ and $\mathbf{h}(1)$ the positive right (column) eigenvector of $\mathbf{A}(2)$ corresponding to $\gamma(2)$, we have

$$\boldsymbol{\pi}(2)\mathbf{A}(1)^{nm}\mathbf{h}(1) \sim \gamma(1)^{nm}\boldsymbol{\pi}(2)\mathbf{h}(1), \quad \boldsymbol{\pi}(2)\mathbf{A}(2)^{nm}\mathbf{h}(1) \sim \gamma(2)^{nm}\boldsymbol{\pi}(2)\mathbf{h}(1)$$

which in view of $\mathbf{A}(1)^n \geq c(b)^n \mathbf{A}(2)^n$ and $\boldsymbol{\pi}(2)\mathbf{h}(1) > 0$ is only possible if $\gamma(1) \geq c(b)\gamma(2)$. $\square$

Lemma 9.3 *Let $I = \int_0^\infty \exp\{-\mu e^{-q\ell} x\} \ell^{a-1} e^{-\delta\ell} d\ell$. Then $I \sim \frac{\Gamma(\delta/q) \log^{a-1} x}{\mu^{\delta/q} q^a x^{\delta/q}}$ as $x \rightarrow \infty$.*

Proof Using the substitution $y = \mu e^{-q\ell} x$, we can rewrite I as

$$\int_0^{\mu x} e^{-y} \left(\frac{\log x + \log \mu - \log y}{q} \right)^{a-1} \left(\frac{y}{\mu x} \right)^{\delta/q} \frac{dy}{qy} = \frac{1}{\mu^{\delta/q} q^a} \frac{\log x}{x^{\delta/q}} I_1$$

where

$$I_1 = \int_0^\infty \omega(x, y)^{\alpha-1} y^{\delta/q-1} e^{-y} dy, \quad \omega(x, y) = 1 + \log \mu / \log x - \log y / \log x.$$

Now for $x \geq e$ and $\varepsilon > 0$, we have $|\omega(x, y)| \leq \omega^*(y)$ where $\omega^*(y) = d \max(y^\varepsilon, y^{-\varepsilon})$ for some $d = d_\varepsilon$. Choosing ε such that $\varepsilon|\alpha - 1| < \alpha$, $\omega^*(y)^{\alpha-1} y^{\delta/q-1} e^{-y}$ is integrable, and since $\omega(x, y) \rightarrow 1$ as $x \rightarrow \infty$ with y fixed, the dominated convergence theorem implies that $I_1 \rightarrow \Gamma(\delta/q)$, concluding the proof. $\square$

References

1. Asmussen, S.: Applied Probability and Queues, 2nd ed. Springer (2003)
2. Asmussen, S., Fiorini, P., Lipsky, L., Rolski, T., Sheahan, R.: On the distribution of total task times for tasks that must restart from the beginning if failure occurs. *Math. Oper. Res.* **33**, 932–944 (2007)
3. Asmussen, S., Foss, S., Korshunov, D.: Asymptotics for sums of random variables with local subexponential behaviour. *J. Theoret. Probab.* **16**, 489–518 (2003)
4. Asmussen, S., Perry, D.: On cycle maxima, first passage problems and extreme value theory for queues. *Stoch. Models* **8**, 421–458 (1992)
5. Aven, T., Jensen, U.: Stochastic Models in Reliability. Springer (1999)
6. Barlow, R.E., Proschan, F.: Mathematical Theory of Reliability. Wiley; reprinted 1996 in SIAM Classics in Applied Mathematics (1965)
7. Berman, A., Plemmons, R.J.: Nonnegative Matrices in the Mathematical Sciences. SIAM (1994)
8. Bobbio, A., Trivedi, K.: Computation of the distribution of the completion time when the work requirement is a PH random variable. *Stoch. Models* **6**, 133–150 (1990)
9. Castillo, X., Siewiorek, D.P.: A performance-reliability model for computing systems. In: Proceedings of FTCS-10, IEEE Computer Society, pp. 187–192. Silver Spring, MD (1980)
10. Chimento, P.F. Jr., Trivedi, K.S.: The completion time of programs on processors subject to failure and repair. *IEEE Trans. Comput.* **42**(1) (1993)
11. Chlebus, B.S., De Prisco, R., Shvartsman, A.A.: Performing tasks on synchronous restartable message-passing processors. *Distrib. Comput.* **14**, 49–64 (2001)
12. DePrisco, R., Mayer, A., Yung, M.: Time-optimal message-efficient work performance in the presence of faults. In: Proceedings of 13th ACM PODC, pp. 161–172 (1994)
13. Jelenković, P., Tan, J.: Characterizing heavy-tailed distributions induced by retransmissions. *Adv. Appl. Probab.* **45**, 106–138 (2013)
14. Jelenković, P., Skiani, E.: Distribution of the number of retransmissions of bounded documents. *Adv. Appl. Probab.* **47**, 425–449 (2015)
15. Kulkarni, V.G.: A new class of multivariate phase type distributions. *Oper. Res.* **37**, 151–158 (1989)
16. Kulkarni, V., Nicola, V., Trivedi, K.: On modeling the performance and reliability of multimode systems. *J. Syst. Softw.* **6**, 175–183 (1986)
17. Kulkarni, V., Nicola, V., Trivedi, K.: The completion time of a job on a multimode system. *Adv. Appl. Probab.* **19**, 932–954 (1987)
18. Nicola, V.F., Martini, R., Chimento, P.F.: The completion time of a job in a failure environment and partial loss of work. In: Proceedings of the 2nd International Conference on Mathematical Methods in Reliability (MMR'2000), pp. 813–816. Bordeaux (2000)
19. Seneta, E.: Non-negative Matrices and Markov Chains. Springer (1994)

20. Sheahan, R., Lipsky, L., Fiorini, P., Asmussen, S.: On the distribution of task completion times for tasks that must restart from the beginning if failure occurs. *SIGMETRICS Perform. Eval. Rev.* **34**, 24–26 (2006)
21. Thompson, S., Lipsky, L., Asmussen, S.: Linear algebraic methods in RESTART problems in Markovian systems. In: Fiondella, L., Puliafito, A. (eds.) *Principles of Performance and Reliability Modeling and Evaluation: Essays in Honor of Kishor Trivedi on his 70th Birthday*, Springer Series in Reliability Engineering (2015)
22. Thøgersen, J., Asmussen, S.: On Markovian systems of Wiener-Hopf and renewal integral equations. Work in progress (2015)
23. Trivedi, K.S. *Probability and Statistics with Reliability, Queuing and Computer Science Applications*, 2nd ed. Wiley (2002)