PhD Dissertation

Continuous-Time Stationary Processes And Wind Power

Infinitely divisible distributions, stochastic delay differential equations, and applcations to wind power production

Victor Rohde
Department of Mathematics
Aarhus University

2019

Continuous-time stationary processes and wind power

PhD dissertation by
Victor Rohde

Department of Mathematics, Aarhus University
Ny Munkegade 118, 8000 Aarhus C, Denmark

Supervised by
Associate Professor Andreas Basse-O'Connor
Associate Professor Jan Pedersen

Submitted to Graduate School of Science and Technology, Aarhus, December 30, 2019

Dissertation was typesat in kpfonts with
pdfLaTeX and the memoir class

AARHUS
UNIVERSITY
DEPARTMENT OF MATHEMATICS

Contents

Preface

This dissertation is a product of my PhD studies at the Department of Mathematics, Aarhus Univeristy, from August 2015 to December 2019. The studies where done under the supervision of Andreas Basse-O'Connor (main supervisor) and Jan Pedersen (co-supervisor), and fully funded by Andreas' grant (DFF-4002-00003) from the Danish Council for Independent Research.

This dissertation is based around the following nine self-contained papers:

Paper A On infinite divisibility of a class of two-dimensional vectors in the second Wiener chaos. *Submitted*.

Paper B The polylogarithmic distribution. Working paper.

Paper C Stochastic delay differential equations and related autoregressive models. *Stochastics* (forthcoming), 24 pages.

Paper D Multivariate stochastic delay differential equations and CAR representations of CARMA processes. *Stochastic Processes and their Applications* 129.10, 4119-4143.

Paper E Stochastic differential equations with a fractionally filtered delay: a semimartingale model for long-range dependent processes. *Bernoulli* (forthcoming), 30 pages.

Paper F Recovering the background noise of a Lévy-driven CARMA process using an SDDE approach. *Proceedings ITISE 2017* 2, 707–718.

Paper G On non-negative modeling with CARMA processes. *Journal of Mathematical Analysis and Applications* 476.1, 196-214.

Paper H Multivariate continuous-time modeling of wind indexes and hedging of wind risk. *Submitted*.

Paper I A statistical view on a surrogate model for estimating extreme events with an application to wind turbines. Working paper.

Besides layout, all published or submitted papers correspond to their published or submitted versions. Each paper introduces notational conventions that may not be consistent with the remaining papers, and the notation is therefore only to be understood in the context of the corresponding paper.

I have made major contributions to the research and writing of Papers A-E and G-H. Paper F and I are written jointly with Mikkel Slot Nielsen, and we have made equal contribution to these. Papers A, C, and F where largely written during the first half of my PhD studies and are therefore included to a varying extent in my progress report written as part of the qualifying examination after which I obtained a master's degree in Mathematics-Economics.

The first chapter is an introduction to the studied topics. A small introduction to each paper is also included. The introduction is meant to tie the papers together under three main themes: infinitely divisible distributions, stochastic delay differential equations, and wind power production. While these themes cover a variety of very different topics, it is an aim of the introduction to unveil some common ground among them. The papers appear ordered in the continuum introduced by the three main themes, ranging from infinitely divisible distributions over stochastic delay differential equations to wind power production. In this way, Paper A is the paper most concerned with infinitely divisible distributions and Paper I deals most exclusively with wind power production. This progression of the papers through the main themes is not chronological.

Together with the conclusion of my PhD studies belongs a sincere thanks to a lot of people. I wish to express my gratitude to my supervisor Andreas Basse-O'Connor for many insightful comments, considerate attitude, and cheerful mood. My other supervisor Jan Pedersen also deserve a big praise for always having an open door, asking the good question, and having an abundance of helpful comments both in and outside the academic setup. I am truly grateful to both my supervisors for showing patience with me, for the guidance through the years, and the many meetings with advise and laughter.

I would like to thank Fred Espen Benth for being the central figure in a very pleasant visit to the University of Oslo. I enjoyed my time there immensely, and the conversations in Oslo and the following correspondence via email have always been insightful and enjoyable. Our collaborations have inspired me to pursue different areas of mathematics, and have therefore sparked a lot of passion in me. I would also like to express my gratitude to Mikkel Slot Nielsen for the many interesting collaborations and discussions. Furthermore, a warm thank you goes out to Troels Sønderby Christensen for a very pleasant collaboration and visits, both the visit to Aarhus and when I went to Aalborg, and for the good times in the office and on the skis in Oslo. James Nichols and Vestas Wind Systems also deserve a big thank you for a great collaboration and interesting meetings.

Many fellow PhD student at Aarhus University also deserve a big praise for the many joyful experiences over the years, in particular, Julie, Thorbjørn, Mikkel, Jeanett, Patrick, Mads, Claudio, and Mathias.

Finally, my most heartfelt thanks goes to my family. I am deeply grateful to my significant other, Line, who is always loving, considerate, takes an interest in me, and forgives my flaws. My greatest affection goes to my kids Theodor and Albert; your laughter and cheer are the best part of any day. Spending time with you and seeing your development gives a profound meaning to it all. To my whole family, I look forward to our future together, and I am truly grateful that you have been with me during my PhD studies.

Victor Rohde
Aarhus, December 2019

Summary

Infinitely divisible distributions play a pivotal role in an abundance of statistical models. From models of the movement of stock prices to models of wind speed, very often, an infinitely divisible distribution is used to capture the noisy behavior. This dissertation consider infinitely divisible distributions and models based on them. A particular focus is on applications to wind power production. The dissertation is divided into three smaller parts: infinitely divisible distributions, stochastic delay differential equations (SDDEs), and wind power production.

The topic of the first paper is the infinte divisibility of two-dimensional sums of Gaussian squares. This is related to the infinite divisibility of a two-dimensional Wiener chaos. The second paper introduces the polylogarithmic distribution which is a class of infinitely divisible distributions that has applications to models based on non-negative Ornstein-Uhlenbeck (OU) processes and continuous-time autoregressive moving average (CARMA) processes. These two papers make up the first part of the dissertation.

The next part consists of four papers devoted to SDDEs. These equations will be connected to the popular CARMA processes. Exploring this connection leads to a new understanding of and to theoretical results concerning CARMA processes. Both univariate and multivariate settings are considered, and a model that builds on SDDEs to introduce a semi-martingale process with a long memory property is examined. Finally, a simulation study of SDDEs ability to recover the driving Lévy process is presented.

Statistical models with application to wind power production are studied in the last three papers. Here, both theoretical and applied considerations are done. First, a class of processes with a potential application to short-term wind forecasting is introduced. This class of processes combines ideas from two popular processes: the CARMA process and the Cox-Ingersoll-Ross process. Next, two multivariate continuous-time models for the wind power utilization in Germany based on OU processes are proposed. The models are applied in a minimum variance hedging setup and the risk premiums are investigated. Finally, a non-parametric regression technique is examined for use as a surrogate model for the distribution of extreme loads on wind turbines. Using the surrogate model has the potential to relieve a heavy computational burden when analyzing extreme loads on wind turbines.

Resumé

Uendeligt delelige fordelinger spiller en central rolle i mange statistiske modeller. Fra modeller for udviklingen af aktiekurser til modeller for vindhastigheder gælder at der ofte bruges en uendeligt delelig fordeling til at fange støjens opførsel. Denne afhandling betragter uendeligt delelige fordelinger og modeller baseret på dem. Et særligt fokus er på anvendelser indenfor produktionen af vindkraft. Afhandlingen er opdelt i tre mindre dele: uendeligt delelige fordelinger, stokastiske forsinket differential ligninger (SDDE'er) og produktion af vindkraft.

Emnet for den første artikel er uendelig deleligheden af to-dimensionelle summer af Gaussiske kvadrater. Dette er relateret til den uendelige delelighed af et to-dimensionelt Wiener-kaos. Den anden artikel introducerer den polylogaritmiske fordeling, som er en klasse af uendeligt delelige fordelinger, der har anvendelse i modeller baseret på ikke-negative Ornstein-Uhlenbeck (OU) processer og kontinuert-tids autoregressive glidende gennemsnit (CARMA) processer. Disse to artikler udgør den første del af afhandlingen.

Den næste del består af fire artikler, der beskæftiger sig med SDDE'er. Disse ligninger bliver forbundet til de populære CARMA processer. Undersøgelsen af denne forbindelse fører til en ny forståelse af og til teoretiske resultater vedrørende CARMA processer. Både endimensionelle og flerdimensionelle udgaver bliver udforsket, og brugen af SDDE'er til at introducere en semi-martingale proces med en lang hukommelsesegenskab undersøges. Endelig præsenteres en simuleringsundersøgelse af SDDE'ers evne til at gendanne den drivende Lévy proces.

Statistiske modeller med anvendelse indenfor produktion af vindkraft studeres i de sidste tre artikler. Her gøres både teoretiske og anvendte overvejelser. Først introduceres en klasse af processer med en potentiel anvendelse indenfor kortvarig vind forudsigelse. Denne klasse af processer kombinerer ideer fra de to populære processer: CARMA processen og Cox-Ingersoll-Ross processen. Dernæst foreslås to flerdimensionelle kontinuerttidsmodeller til udnyttelsen af vindkraft i Tyskland baseret på OU processer. Modellerne anvendes til at danne et minimumsvarians hedge, og risikopræmierne undersøges. Endelig undersøges en ikke-parametrisk regressionsteknik med henblik på at konstruere en surrogat-model for fordelingen af ekstreme belastninger på vindmøller. Brugen af surrogat-modellen har potentiale til at aflaste en tung beregningsbyrde, når man analyserer ekstreme belastninger på vindmøller.

Introduction

1 Infinitely divisible distributions

A random variable U is said to be infinitely divisible if, for any $n \in \mathbb{N}$, there exist independent random variables $U_1, \dots, U_n$ such that $U_1 + \dots + U_n$ has the same distribution as U. Infinitely divisible distributions can also be characterized through the so-called Lévy-Khinchin representation of the characteristic function. In particular, a random variable U is infinitely divisible if and only if

$$\log \mathbb{E}[\exp\{ixU\}] = ibx - \frac{1}{2}ax^2 + \int_{\mathbb{R}\setminus\{0\}} \left(e^{-xy} - 1 - ixy\mathbb{1}_{\|y\|<1}(y)\right)\nu(\mathrm{d}y) \tag{1.1}$$

for $b \in \mathbb{R}$, $a \geq 0$, and a Borel measure ν on $\mathbb{R} \setminus \{0\}$ such that $\int_{\mathbb{R}\setminus\{0\}}(\|y\|^2 \wedge 1)\nu(\mathrm{d}y) < \infty$ (see [17]). It is a desirable property of the infinitely divisible random variables to have the expression in (1.1) for the characteristic function, especially if the integral with respect to the measure ν can be calculated analytically. Characteristic functions play a crucial role in many applications, for example when calculating the moments (which can be found by differentiation of the characteristic function) or in Fourier pricing of financial derivatives (see [9]). Often, it may also be reasonable to consider an infinitely divisible distribution from an applied perspective, where the behaviour of many phenomena can be seen as the accumulation of many independent small effects. For example a pollen subdued in water which famously led Robert Brown to describe the Brownian motion or the noisy, short term behaviour of the return of a financial asset.

Every infinitely divisible distribution can be associated with a Lévy process. Lévy processes are continuous-time versions of random walks, and they are a very valuable modeling tool; offering a rich class of distributions and tractability. Lévy processes play an important part in a variety of models, for example, models of financial assets (see [1, 8]) and, more recently, in models related to wind energy, temperature, and electricity prices (see [3, 4, 5, 11]). Lévy processes are also at the core of most of the models considered in this dissertation. In particular, Paper B-H will to a varying extend consider Lévy driven models.

The class of infinitely divisible distributions covers many popular distributions, including the Gaussian, lognormal, gamma, Weibull, Gumbel and Pareto distribution (see [14] and references therein). These distributions will play a role in this dissertation. For example, in Paper H where we consider models where the stationary distributions are described by the gamma and lognormal distribution. While the Weibull, Gumbel and Pareto distribution are not central to the models considered in Paper I, application of the models would often rely on these distributions to estimate so-called 50-year return loads as in [16]. The distribution of the solutions to the stochastic delay differential equations considered in Paper C, D, E, and F are also infinitely divisible when the noise is a Lévy process. This is a consequence of the solutions being moving averages with respect to the noise process. The Lévy process constitute the primary example of a noise process and is the most frequently considered example in practise.

In the following, Paper A and B are introduced which studies infinitely divisible distributions. An introduction of the remaining papers is deferred to the next sections.

Paper A

For a four-dimensional mean zero multivariate Gaussian vector (X_1, X_2, X_3, X_4), Paper A investigates under which conditions

$$(X_1^2 + X_2^2, X_3^2 + X_4^2) \tag{1.2}$$

is infinitely divisible. This problem is motivated by an interest in understanding the infinite divisibility of the two-dimensional Wiener chaos which consists of limits of a two-dimensional sum of Gaussian squares multiplied by either 1 or -1.

We give sufficient conditions for (1.2) being infinitely divisible that goes beyond the results in the literature. It is still an open question whether (1.2) is always infinitely divisible.

Paper B

In Paper B the polylogarithmic distribution, a family of infinitely divisible distributions, is introduced. A random variable is said to have a polylogarithmic distribution of order $n \in \mathbb{N}_0$ with parameter $\alpha > 0$ and $\beta > 0$ if the cumulant-generating function of U, ψ_U, is given by

$$\psi_U(z) = \log \mathbb{E}[e^{zu}] = \alpha \operatorname{Li}_n(z/\beta), \quad z \in \mathbb{C} \setminus ([\beta, \infty) \times i\mathbb{R}).$$

where Li_n is the polylogarithm of order n (see Paper B for a definition of the polylogarithm). The polylogarithm has a long history in mathematics dating back to 1696 and has been studied by some of the great mathematicians of the past such as Euler, Abel, and Ramanujan.

The distribution of a compound Poisson process with exponential jumps is a polylogarithmic distribution of order 0 and the gamma distribution is a polylogarithmic distribution of order 1.

If consider a Lévy process $(L(t))_{t\in\mathbb{R}}$ where $L(1)$ has a polylogarithmic distribution of order $n \in \mathbb{N}_0$ and with parameter $\alpha > 0$ and $\beta > 0$, then the Ornstein-Uhlenbeck process with mean reversion $\lambda > 0$ driven by $(L(t))_{t\in\mathbb{R}}$ has a polylogarithmic distribution of order $n+1$ with parameter α/λ and β as its stationary distribution. This makes it possible to do fast calculation of the cumulant-generating function, and therefore also the characteristic function and moment-generating function, for a polylogarithmic driven Ornstein-Uhlenbeck process. We also investigate the stationary distribution of certain CARMA processes driven by a polylogarithmic Lévy process. The results obtained have been used in Paper H to improve the computation time of one of the employed models.

2 Stochastic delay differential equations

This section explores a connection between stochastic delay differential equations (SDDEs) and CARMA processes. This connection has motivated the introduction of a class of stationary processes that have a long-memory property and are semimartingales. Also, writing a CARMA process as an SDDE gives a straightforward way to recover the driving noise process which we also investigate.

2.1 Ornstein-Uhlenbeck processes

Continuous-time stationary processes are a key ingredient in a vast amount of probabilistic models. The class of all continuous-time stationary processes is enormous, and subclasses need to be considered in applications. Such a subclass is for example the class of Ornstein-Uhlenbeck processes, which are stationary processes that are widely applicable. Ornstein-Uhlenbeck processes are analytically tractable and giving rise to a good description of a diverse range of data (such as in [1, 4]). They also serve as a powerful building block for more complicated models.

An example of an Ornstein-Uhlenbeck process severing as a building block for a more complicated probabilistic model is the CARMA process, which is a linear transformation of a multivariate Ornstein-Uhlenbeck process. CARMA processes are extensively studied and have found numerous applications, for example in modelling wind speed, electricity prices, stochastic volatility, and temperature (see [3, 5, 7, 11, 18]).

Another stationary process that can be formed by considering Ornstein-Uhlenbeck processes is the Cox-Ingersoll-Ross (CIR) process (see [10]). The sum of squared independent and identically distributed zero mean, Gaussian Ornstein-Uhlenbeck processes gives a CIR process. Not all CIR processes can be obtained this way, and a slight generalization is needed to extend to the full class of CIR processes (see Paper G). The CIR process is, amongst other things, well-known as a model for interest rate (see [10]) and for its role in the Heston model (see [13]) where it models the evolution of the volatility of a financial asset. More recently, the CIR process has also been proposed as a model for wind speed with a particular focus on short term forecasting in [2].

2.2 Continuous-time autoregressive moving average processes

While the Ornstein-Uhlenbeck process is widely applicable it lacks flexibility in the auto-covariance function (since it is just an exponential function). Many applications call for models with a more rich auto-correlation structure. CARMA processes are an example of a class of stationary processes that have a more rich auto-correlation structure and can be seen as a natural generalization of Ornstein-Uhlenbeck processes (see [6] for a thorough treatment of CARMA processes). Intuitively, an Ornstein-Uhlenbeck process, $(X(t))_{t\in\mathbb{R}}$, solves

$$DX(t) = -\lambda X(t) + DL(t), \quad \lambda > 0. \tag{2.1}$$

Here, D denotes the differential operator with respect to t, and $(L(t))_{t\in\mathbb{R}}$ is a Lévy process. Of course, the differential operator is not defined for $(L(t))_{t\in\mathbb{R}}$, and (2.1) is therefore only understood heuristically. The CARMA process generalizes this by instead considering the solution to

$$P(D)X(t) = Q(D)DL(t) \tag{2.2}$$

where $P(z) = z^p + a_1 z^{p-1} + \cdots + a_p$ and $Q(z) = b_0 + b_1 z + \cdots + b_q z^q$, $p > q$. Again, (2.2) is only understood heuristically. The auto-correlation function of a CARMA process is for example given in Proposition 4.1, Paper G, and it offers a significant increase in the flexibility of the auto-correlation function compared to the Ornstein-Uhlenbeck process as shown in Figure 2, Paper G (the Ornstein-Uhlenbeck and CIR process have the same auto-correlation function).

2.3 Stochastic delay differential equations

A stochastic delay differential equation (SDDE) is an equation where the increments of a solution depend on the past of the solution (and in some case also the future) and some noise. In particular, we say $(X(t))_{t\in\mathbb{R}}$ solves an SDDE if

$$X_t - X_s = \int_s^t \int_{[0,\infty)} X_{u-v}\, \eta(\mathrm{d}v)\, \mathrm{d}u + Z_t - Z_s, \quad s < t, \tag{2.3}$$

where the noise process $(Z(t))_{t\in\mathbb{R}}$ is an integrable stationary increment process and the delay measure η is a finite signed measure. This type of equation is the primary object of study in Paper C-F. In [12, 15] SDDEs are also studied but we consider the general setup where η may have unbounded support which allows us to connect SDDEs and CARMA processes. This connection has not been explored before.

Paper C

This paper suggest two continuous-time models which exhibits an auto-regressive structure. The first model is the SDDE given in (2.3). The second model, the so-called level model, is given by

$$X(t) = \int_0^\infty X(t-u)\,\phi(\mathrm{d}u) + \int_{-\infty}^t \theta(t-u)\,\mathrm{d}L(u), \quad t \in \mathbb{R}$$

for a suitably integrable function $\theta : \mathbb{R} \to \mathbb{R}$. Most effort is dedicated to SDDEs where existence and uniqueness results are developed. Furthermore, the connection to CARMA processes is indicated and the possibility of introducing long memory through the noise is explored. For the level model, existence and uniqueness of a solution are considered and it is shown that the level model can be chosen to be a discrete-time ARMA model when observed discretely.

Paper D

With the results developed in Paper C, a very natural next step is to consider the connection between SDDEs and CARMA processes. This led to Paper D where SDDEs are treated in a multivariate setting and where the connection between SDDEs and CARMA processes is explored in detail. The multivariate version of SDDEs (MSDDEs) is considered both to make the model more general and as a key ingredient in the connection to CARMA processes. A CARMA process can be associated with a higher order SDDE (that is, where differentials of X also appear in (2.3)) and this setup falls naturally into the multivariate setting. Multivariate CARMA processes (MCARMA) are also connected to MSDDEs. The SDDE setup allows for a very general noise process, and we therefore also obtain connections between SDDEs and fractional integrated CARMA (FICARMA) processes and between MSDDEs and multivariate FICARMA processes. Besides the connection between SDDEs and CARMA processes, a prediction formula is also presented.

Paper E

A way to introduce long-memory (that is, a non integrable auto-correlation function) to the solution of an SDDE is through the noise process. In particular, if the process Z in (2.3) is a fractional Lévy process on the form $I^{\beta}L$ with $\beta > 0$ (see Section 1, Paper E, for notation) then the solution to the associated SDDE will have long-memory. While this can be a very convenient way to introduce long-memory, it also forces other, possibly undesirable, properties on to the model. For example, the behaviour of the auto-correlation function changes around zero and the solution process is not a semi-martingale. This motivates introducing long-memory through the delay measure η. By doing so, we obtain a model for stationary processes with long-memory, but without changes to the auto-correlation function around zero and where the solution is still a semi-martingale.

Paper F

As opposed to the previous papers where focus was on theoretical properties of SDDEs and their solutions, Paper F studies numerical aspects of SDDEs. In particular, the SDDE gives a straight-forward inversion formula. That is, from observation of the solution process, the increments of the noise process can be computed directly by exploiting the SDDE relation. This inversion is explored numerically where a simulated CARMA process is written as an SDDE and the driving noise is recovered. Inversion of CARMA processes is also studied in [7] using a different approach that is not based on SDDEs.

3 Wind power production

Following a demand for more renewable energy generation, wind power production has seen a rapid growth in the last decades. As of 2018, the world's wind power capacity was the world's second largest among the renewable energy sources, surpassed only by hydropower. The installed world capacity of wind energy in 2018 has increased 9.44% from the year before to 591 GW and thereby accounting for 24.85% of the worlds renewable energy capacity. Wind energy accounted for 28.18% of the world's increase of renewable energy capacity from 2017 to 2018. The other major source of increase in the world's capacity of renewable energy comes from solar photovoltaic which accounted for 55.25% of the increase in renewable energy capacity from 2017 to 2018.[1]

[1] see ren21.net

The increasing reliance on wind for energy generation puts emphasis on a better understanding of statistical properties of wind speed and wind energy generation. In this section focus will be on statistical models that are applicable in wind power production setting. We consider a model which we argue is likely to have an application to short-term wind forecasting (Paper G). Next, two multivariate continuous-time models for wind utilization in Germany are proposed where the application is aimed at hedging volumetric risk for wind power producers and on quantifying the risk premium (Paper H). Finally, we consider a surrogate model for extreme loads on wind turbines (Paper I). The models considered in Paper G and H are based on stationary processes with particular attention towards Ornstein-Uhlenbeck, CIR, and CARMA processes. As the aim of Paper I is to create a surrogate model, a different approach has been taken. In particular, the model is based on a non-parametric regression technique (k-nearest neighbours). This model is still an area of active research and currently, another non-parametric regression technique (random forest) is investigated with a promising performance.

The models considered in this section are not limited to applications to wind data. Most notably, the CIR-CARMA and CARMA process in Paper G are also considered in the context of pricing zero-coupon bonds.

Paper G

Modeling the short-term statistical behaviour of wind speed is of high relevance to several aspects of managing wind turbines. For example, a short term power production forecast and wind turbine control both benefit from a good short-term wind speed forecast. In [2] the CIR process is successfully applied to provide a short-term wind forecast. As we argue in Paper G, more flexibility in the auto-correlation function than the CIR process have a potential for further improving the forecast.

The fundamental idea of the CIR-CARMA process, introduced in Paper G, is to generalize the CIR process with inspiration from the way CARMA processes generalize Ornstein-Uhlenbeck processes. In this way, we obtain a process that closely resembles the CIR process but has a more flexible auto-covariance structure. It therefore seems that the CIR-CARMA process has a potential for improving the short-term wind forecast.

There are still open questions regarding the CIR-CARMA process, and more research into this class of processes would support many applications. In particular, a filtering technique for estimating the hidden Wishart process based on observation of the CIR-CARMA process is needed to calculate the zero-coupon bond price, and it would also be relevant in other applications (see the discussion after Theorem 3.5, Paper G).

We also consider a CARMA process driven by a compound Poisson process with exponential jumps, and show that this process also captures the auto-correlation structure induced by the wind speed data. The stationary distribution of certain CARMA processes driven by a compound Poisson process with exponential jumps is also investigated, where we show that it can be written as an infinite sum of independent gamma distributions.

Paper H

A wind power producer faces several risk factors in term of future revenue streams. Two major risks are the price of electricity (market risk) and utilization of the installed capacity which is uncertain due to the dependence on weather (volumetric risk). Market risk is addressed by so-called power purchase agreements and government subsidies. More recently, volumetric risk has been addressed in Germany with the introduction of exchange-traded wind power futures contracts. The futures contracts are written on an index (which we call the German wind index) that measure the overall utilization on a daily basis of the installed wind energy capacity in Germany. These contracts can be used to form a hedge against low utilization of the installed capacity at given wind site in Germany. Alternatively, the futures contracts can be used as the basis for wind site specific over-the-counter contract that can be sold to a wind power producer in order to completely alleviate the volumetric risk. The seller of the over-the-counter contract can hold several such instruments in a portfolio and use the exchange-traded wind power futures contracts to minimize the risk exposure.

The German wind index has been analysed in an one-dimensional model in [4] where focus is on pricing of derivatives and risk premiums. In Paper H two multivariate continuous-time models for three wind site utilization indexes and the German wind index are employed. The first model is based on a multivariate Ornstein-Uhlenbeck process driven by a multivariate compound Poisson process with exponential jumps. The second model is based on a multivariate Gaussian Ornstein-Uhlenbeck process. We show how both models capture several key statistical properties of the data, but also that there are some aspect of the data that the non-Gaussian model is better able to capture (see Table 3, Paper H). We argue that the models can be applied to determine a minimum variance hedge. Both an in-sample and out-of-sample context is considered, and a significant variance reduction is obtained thereby relieving a large part of the volumetric risk. We also analyse a portfolio consisting of all the three wind site indexes, and show how a further variance reduction can be obtained by combining indexes. This motivates the possibility of third party selling over-the-counter futures contracts for site specific wind utilization to wind power producers, and thereby obtaining a large portfolio of wind site indexes that can then be more effectively hedged against the German wind index.

Lastly, the models are applied to calculate risk premiums of futures contracts. Both the yearly and quarterly risk premiums are accessed, with the quarterly futures contract showing a seasonal pattern in the risk premium.

Paper I

A reliable estimate of the probabilistic properties of the extreme loads on a wind turbine is of importance to the design, control, and lifetime assessment of the turbine. The 10-minute maximum loads on a wind turbine in a certain environment can be found through a simulation tool but this is computationally heavy. The basic idea proposed in Paper I is to circumvent this computational burden by introducing a so-called surrogate model. The simulation tool generates the 10-minute maximum loads at different places on the wind turbine in a wide variety of environments. The simulation also generates associated covariates such as the 10-minute maximum, minimum, mean, and standard deviation of the generator speed, electrical power, and blade angle (see Paper I for a full list of covariates). The surrogate model learns the relation between the covariates and the distribution of the 10-minute maximum loads. The extreme loads at a given location can then quickly be evaluated based on observation of covariates on a site and the surrogate model.

In the paper, we consider two models that can be used as a surrogate model for the distribution of the 10-minute extreme loads. We evaluate the performance of the models on a physical wind turbine that, as a part of a measuring campaign, is equipped with extra loads measurement equipment. We argue that it is indeed possible for the surrogate model to capture the distribution of the extreme loads. With the distribution of the 10-minute maximum loads, the extreme loads can then be estimated by extrapolation (using extreme value distributions or a Pareto distribution with a peak-over-threshold method, see [16]).

References

[1] Barndorff-Nielsen, O.E. and N. Shephard (2001). Non-Gaussian Ornstein–Uhlenbeck-based models and some of their uses in financial economics. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 63(2), 167–241.

[2] Bensoussan, A. and A. Brouste (2016). Cox–Ingersoll–Ross model for wind speed modeling and forecasting. *Wind Energy* 19(7), 1355–1365.

[3] Benth, F.E., C. Klüppelberg, G. Müller and L. Vos (2014). Futures pricing in electricity markets based on stable CARMA spot models. *Energy Econ.* 44, 392–406.

[4] Benth, F.E. and A. Pircalabu (2018). A non-Gaussian Ornstein-Uhlenbeck model for pricing wind power futures. *Applied Mathematical Finance* 25(1), 36–65.

[5] Benth, F.E., J. Šaltytė-Benth and S. Koekebakker (2007). Putting a price on temperature. *Scand. J. Statist.* 34(4), 746–767. DOI: 10.1111/j.1467-9469.2007.00564.x.

[6] Brockwell, P.J. (2014). Recent results in the theory and applications of CARMA processes. *Ann. Inst. Statist. Math.* 66(4), 647–685. DOI: 10.1007/s10463-014-0468-7.

[7] Brockwell, P.J., R.A. Davis and Y. Yang (2011). Estimation for non-negative Lévy-driven CARMA processes. *J. Bus. Econom. Statist.* 29(2), 250–259. DOI: 10.1198/jbes.2010.08165.

[8] Carr, P., H. Geman, D.B. Madan and M. Yor (2002). The fine structure of asset returns: An empirical investigation. *The Journal of Business* 75(2), 305–332.

[9] Carr, P. and D. Madan (1999). Option valuation using the fast Fourier transform. *Journal of Computational Finance* 2(4), 61–73.

[10] Cox, J.C., J.E. Ingersoll Jr and S.A. Ross (2005). “A theory of the term structure of interest rates”. *Theory of Valuation*. World Scientific, 129–164.

[11] García, I., C. Klüppelberg and G. Müller (2011). Estimation of stable CARMA models with an application to electricity spot prices. *Stat. Model.* 11(5), 447–470.

[12] Gushchin, A.A. and U. Küchler (2000). On stationary solutions of delay differential equations driven by a Lévy process. *Stochastic Process. Appl.* 88(2), 195–211. DOI: 10.1016/S0304-4149(99)00126-X.

[13] Heston, S.L. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *The review of financial studies* 6(2), 327–343.

[14] Kyprianou, A.E. (2014). *Fluctuations of Lévy processes with applications: Introductory Lectures*. Springer Science & Business Media.

[15] Mohammed, S.E.A. and M.K.R. Scheutzow (1990). Lyapunov exponents and stationary solutions for affine stochastic delay equations. *Stochastics Stochastics Rep.* 29(2), 259–283.

[16] Ragan, P. and L. Manuel (2008). Statistical extrapolation methods for estimating wind turbine extreme loads. *Journal of Solar Energy Engineering* 130(3), 031011.

[17] Sato, K. (1999). *Lévy Processes and Infinitely Divisible Distributions*. Vol. 68. Cambridge Studies in Advanced Mathematics. Translated from the 1990 Japanese original, Revised by the author. Cambridge University Press.

[18] Todorov, V. (2009). Estimation of continuous-time stochastic volatility models with jumps using high-frequency data. *J. Econometrics* 148(2), 131–148.

Paper A

On Infinite Divisibility Of A Class Of Two-Dimensional Vectors In The Second Wiener Chaos

Andreas Basse-O'Connor, Jan Pedersen, and Victor Rohde

Abstract

Infinite divisibility of a class of two-dimensional vectors with components in the second Wiener chaos is studied. Necessary and sufficient conditions for infinite divisibility is presented as well as more easily verifiable sufficient conditions. The case where both components consist of a sum of two Gaussian squares is treated in more depth, and it is conjectured that such vectors are infinitely divisible.

MSC 2010: 60E07; 60G15; 62H05; 62H10

Keywords: sums of Gaussian squares; infinite divisibility; second Wiener chaos

1 Introduction

Paul Lévy [11] raised the question of infinite divisibility of Gaussian squares, that is, for a centered Gaussian vector $(X_1,\ldots,X_n)$ when can $(X_1^2,\ldots,X_n^2)$ be written as a sum of m independent identical distributed random vectors for any $m \in \mathbb{N}$? Several authors have studied this problem. We refer to [4, 5, 6, 7, 8, 13] and reference therein. These works include several novel approaches and gives a great understanding of when Gaussian squares are infinitely divisible. In this paper we will provide a characterization of infinite divisibility of sums of Gaussian squares which to the best of our knowledge has not been studied in the literature except in special cases. This problem is highly motivated by the fact that sums of Gaussian squares are the usual limits in many limit theorems in the presence of either long range dependence, see [2] or [16], or degenerate U-statistics, see [9]. In the following we will go in more details.

Let Y be random variable in the second (Gaussian) Wiener chaos, that is, the closed linear span in L^2 of $\{W(h)^2-1 : h \in H, \|h\| = 1\}$ for a real separable Hilbert space H and an isonormal Gaussian process W. For convenience, we assume H is infinite-dimensional. Then there exists a sequence of independent standard Gaussian variables (ξ_i) and a sequence of real numbers (α_i) such that

$$Y \stackrel{d}{=} \sum_{i=1}^{\infty} \alpha_i(\xi_i^2 - 1),$$

where the sum converges in L^2 (see for example [9, Theorem 6.1]). Since the ξ_i's are independent, $(\xi_1^2, \dots, \xi_d^2)$ is infinitely divisible for any $d \geq 1$ and therefore, Y is infinitely divisible. Such a sum of Gaussian squares appears as the limit of U-statistics in the degenerate case (see [9, Corollary 11.5]). In this case the α_i are certain binomial coefficients times the eigenvalues of operators associated to the U-statistics. We note that the sequence (ξ_i) depends heavily on Y, so one can not deduce joint infinite divisibility of random vectors with components in the second Wiener chaos. In particular, for a vector with dimension greater than or equal to three and components in the second Wiener chaos it is well known (cf. Theorem 1.1 below) that it need not be infinite divisibility. In between these two cases is the open question of infinite divisibility of a two-dimensional vector with components in the second Wiener chaos. Let $(X_1, \dots, X_{n_1+n_2})$ be a mean zero Gaussian vector for $n_1, n_2 \in \mathbb{N}$. That any two-dimensional vector in the second Wiener chaos is infinitely divisible is equivalent to

$$(d_1 X_1^2 + \cdots + d_{n_1} X_{n_1}^2, d_{n_1} X_{n_1+1}^2 + \cdots + d_{n_1+n_2} X_{n_1+n_2}^2) \tag{1.1}$$

being infinitely divisible for any $d_1, \dots, d_{n_1+n_2} = \pm 1$, any covariance structure of $(X_1, \dots, X_{n_1+n_2})$, and any $n_1, n_2 \in \mathbb{N}$ (something that follows by the definition of the second Wiener chaos).

The following theorem, which is due to Griffiths [8] and Bapat [1], is an important first result related to infinite divisibility in the second Wiener chaos. We refer to Marcus and Rosen [12, Theorem 13.2.1 and Lemma 14.9.4] for a proof.

Theorem 1.1 (Griffiths and Bapat). *Let $(X_1, \dots, X_n)$ be a mean zero Gaussian vector with positive definite covariance matrix Σ. Then $(X_1^2, \dots, X_n^2)$ is infinitely divisible if and only if there exists an $n \times n$ matrix U on the form* $\mathrm{diag}(\pm 1, \dots, \pm 1)$ *such that $U^t \Sigma^{-1} U$ has non-positive off-diagonal elements.*

This theorem resolved the question of infinite divisibility of Gaussian squares. For $n \geq 3$ there is an $n \times n$ positive definite matrix Σ where there does not exist an $n \times n$ matrix U on the form $\mathrm{diag}(\pm 1, \dots, \pm 1)$ such that $U^t \Sigma^{-1} U$ has non-positive off-diagonal elements. Consequently, there are mean zero Gaussian vectors $(X_1, \dots, X_n)$ such that $(X_1^2, \dots, X_n^2)$ is not infinite divisible whenever $n \geq 3$.

Eisenbaum [3] and Eisenbaum and Kapsi [5] found a connection between the condition of Griffiths and Bapat and the Green function of a Markov process. In particular, a Gaussian process has infinite divisible squares if and only if its covariance function (up to a constant function) can be associated with the Green function of a strongly symmetric transient Borel right Markov process.

When discussing the infinite divisibility of the Wishart distribution Shanbhag [15] showed that for any covariance structure of a mean zero Gaussian vector $(X_1, \dots, X_n)$,

$$(X_1^2, X_2^2 + \cdots + X_n^2)$$

is infinitely divisible. Furthermore, it was found that infinite divisibility of any bivariate marginals of a centered Wishart distribution can be reduced to infinite divisibility of $(X_1 X_2, X_3 X_4)$. By the polarization identity,

$$(X_1 X_2, X_3 X_4) = \tfrac{1}{4}((X_1 + X_2)^2 - (X_1 - X_2)^2, (X_3 + X_4)^2 - (X_3 - X_4)^2).$$

Consequently, infinite divisibility of any bivariate marginals of a centered Wishart distribution is again related to the question of infinite divisibility of a two-dimensional vector from the second Wiener chaos.

We will be interested in the infinite divisibility of

$$(X_1^2 + \cdots + X_{n_1}^2, X_{n_1+1}^2 + \cdots + X_{n_1+n_2}^2),$$

i.e., the case $d_1 = \cdots = d_{n_1+n_2} = 1$ in (1.1). The general case, where $d_i = -1$ for at least one i, seems to require new ideas going beyond the present paper. We will have a special interest in the case $n_1 = n_2 = 2$.

Despite the simplicity of the question, it has proven rather subtle, and a definite answer is not presented. Instead, we give easily verifiable conditions for infinite divisible in the case $n_1 = n_2 = 2$ as well as more complicated necessary and sufficient conditions in the general case that may or may not always hold. We will, in addition, investigate the infinite divisibility of $(X_1^2 + X_2^2, X_3^2 + X_4^2)$ numerically which, together with Theorem 2.4 (ii), leads us to conjecture that infinite divisibility of this vector always holds.

The main results without proofs are presented in Section 2. Section 3 contains two examples and a small numerical discussion. We end with Section 4 where the proofs of the results stated in Section 2 are given.

2 Main Results

We begin with a definition which is a natural extension to the present setup (see the proof of Corollary 2.7) of the terminology used by Bapat [1].

Definition 2.1. *Let $n_1, n_2 \in \mathbb{N}$. An $(n_1 + n_2) \times (n_1 + n_2)$ orthogonal matrix U is said to be an (n_1, n_2)-signature matrix if*

$$U = \begin{pmatrix} U_1 & 0 \\ 0 & U_2 \end{pmatrix}$$

where U_1 is an $n_1 \times n_1$ matrix and U_2 is an $n_2 \times n_2$ matrix, both orthogonal, and for 0's of suitable dimensions.

Let $n_1, n_2 \in \mathbb{N}$ and consider a mean zero Gaussian vector $(X_1, \dots, X_{n_1+n_2})$ with positive definite covariance matrix Σ. Now we present a necessary and sufficient condition for infinite divisibility of

$$(X_1^2 + \cdots + X_{n_1}^2, X_{n_1+1}^2 + \cdots + X_{n_1+n_2}^2). \tag{2.1}$$

For $a > 0$, let $Q = I - (I + a\Sigma)^{-1}$ and write

$$Q = \begin{pmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{pmatrix}$$

where Q_{11} is an $n_1 \times n_1$ matrix, Q_{22} is an $n_2 \times n_2$ matrix, and $Q_{12} = Q_{21}^t$ (where Q_{21}^t is the transpose of Q_{21}) is an $n_1 \times n_2$ matrix. Note that if λ is an eigenvalue of Σ, $\frac{a\lambda}{1+a\lambda}$ is an eigenvalue of Q. Since Q is symmetric and has positive eigenvalues, it is positive definite.

Theorem 2.2. *The vector in (2.1) is infinitely divisible if and only if for all $k, m \in \mathbb{N}_0$ and for all $a > 0$ sufficiently large,*

$$\begin{aligned} &\sum \operatorname{trace} Q_{11}^{k_1} Q_{12} Q_{22}^{m_1} Q_{21} Q_{11}^{k_2} \cdots Q_{11}^{k_d} Q_{12} Q_{22}^{m_d} Q_{21} Q_{11}^{k_{d+1}} \\ &\quad + \sum \operatorname{trace} Q_{22}^{m_1} Q_{21} Q_{11}^{k_1} Q_{12} Q_{22}^{m_2} \cdots Q_{22}^{m_{d-1}} Q_{21} Q_{11}^{k_d} Q_{12} Q_{22}^{m_{d+1}} \geq 0, \end{aligned} \tag{2.2}$$

where the first sum is over all $k_1, \ldots, k_{d+1}$ and $m_1, \ldots, m_d$ such that

$$k_1 + \cdots + k_{d+1} + d = k \quad \text{and} \quad m_1 + \cdots + m_d + d = m,$$

and the second sum is over all $m_1, \ldots, m_{d+1}$ and $k_1, \ldots, k_d$ such that

$$m_1 + \cdots + m_{d+1} + d = m \quad \text{and} \quad k_1 + \cdots + k_d + d = k.$$

Remark 2.3. By applying Theorem 2.2 we can give a new and simple proof of Shanbhag's [15] result that $(X_1^2, X_2^2 + \cdots + X_{1+n_2}^2)$ is infinite divisible. To see this, consider the case $n_1 = 1$ and $n_2 \in \mathbb{N}$. Then Q_{11} is a positive number and $Q_{12} Q_{22}^m Q_{21}$ is a non-negative number for any $m \in \mathbb{N}$. In particular, we have

$$\begin{aligned} &\operatorname{trace} Q_{11}^{k_1} Q_{12} Q_{22}^{m_1} Q_{21} \cdots Q_{12} Q_{22}^{m_d} Q_{21} Q_{11}^{k_{d+1}} \\ &\quad = Q_{11}^{k_1} \cdots Q_{11}^{k_{d+1}} Q_{12} Q_{22}^{m_1} Q_{21} \cdots Q_{12} Q_{22}^{m_d} Q_{21} \geq 0 \end{aligned}$$

for any $k_1, \ldots, k_{d+1}, m_1, \ldots, m_d \in \mathbb{N}_0$. Consequently, the first sum in (2.2) is a sum of non-negative numbers. A similar argument gives that the other sum is non-negative too. We conclude that $(X_1^2, X_2^2 + \cdots + X_{1+n_2}^2)$ is infinite divisible.

In order to get a concise formulation of the following results we will need some terminology and conventions. To this end, consider a 2×2 symmetric matrix A. Let v_1 and v_2 be the eigenvectors of A, and λ_1 and λ_2 be the corresponding eigenvalues. We say that v_i is associated with the largest eigenvalue if $\lambda_i \geq \lambda_j$ for $j = 1, 2$. Furthermore, whenever A is a multiple of the identity matrix, we fix $(1, 0)$ to be the eigenvector associated with the largest eigenvalue.

Now consider the special case $n_1 = n_2 = 2$, i.e., the vector

$$(X_1^2 + X_2^2, X_3^2 + X_4^2) \tag{2.3}$$

where (X_1, X_2, X_3, X_4) is a mean zero Gaussian vector with a 4×4 positive definite covariance matrix Σ. We still let $Q = I - (I + a\Sigma)^{-1}$ and write

$$Q = \begin{pmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{pmatrix}$$

where Q_{ij} is a 2×2 matrix for $i,j = 1,2$. Let W be a $(2,2)$-signature matrix such that

$$W^tQW = \begin{pmatrix} W_1^tQ_{11}W_1 & W_1^tQ_{12}W_2 \\ W_2^tQ_{21}W_1 & W_2^tQ_{22}W_2 \end{pmatrix} = \begin{pmatrix} q_{11} & 0 & q_{13} & q_{14} \\ 0 & q_{22} & q_{23} & q_{24} \\ q_{13} & q_{23} & q_{33} & 0 \\ q_{14} & q_{24} & 0 & q_{44} \end{pmatrix},$$

where $q_{11} \geq q_{22} > 0$ and $q_{33} \geq q_{44} > 0$ which exists by Lemma 4.1. Note that q_{ij} is not the (i,j)-th entry of Q but of W^tQW. Let $v_1 = (v_{11}, v_{21})$ be the eigenvector of $W_1^tQ_{12}Q_{21}W_1$ associated with the largest eigenvalue. If $q_{11} = q_{22}$ or $q_{33} = q_{44}$, any orthogonal W_1 or W_2 gives the desired form. In this case, we may always choose W_1 or W_2 such that $v_{11}q_{13}(v_{11}q_{13} + v_{21}q_{23}) \geq 0$ (see the proof of Lemma 4.2, (ii) $\Rightarrow$ (iii)), and it is such a choice we fix. The following theorem addresses the non-negativity of the sums in (2.2) when $n_1 = n_2 = 2$,.

Theorem 2.4. *Let $n_1 = n_2 = 2$. Then, in the notation above, we have the following.*

(i) *For all $d \in \mathbb{N}_0$ and $k_1, \dots, k_{d+1}, m_1, \dots, m_d \in \mathbb{N}_0$,*

$$\operatorname{trace} Q_{11}^{k_1}Q_{12}Q_{22}^{m_1}Q_{21}Q_{11}^{k_1}\cdots Q_{11}^{k_d}Q_{12}Q_{22}^{m_d}Q_{21}Q_{11}^{k_{d+1}} \geq 0$$

if and only if $v_{11}q_{13}(v_{11}q_{13} + v_{21}q_{23}) \geq 0$. In particular, (2.3) is infinitely divisible if the latter inequality is satisfied for all sufficiently large a.

(ii) *For any $k,m \in \mathbb{N}_0$ such that at least one of the following inequalities is satisfied: (i) $k \leq 2$, (ii) $m \leq 2$, or (iii) $k+m \leq 7$, the sum in (2.2) is non-negative.*

Remark 2.5. When $v_{11}q_{13}(v_{11}q_{13}+v_{21}q_{23}) < 0$, we know that there are $k,m \in \mathbb{N}_0$ such that (2.2) with $n_1 = n_2 = 2$ contains negative terms cf. Theorem 2.4 (i). If $k = 0$ or $m = 0$ then Theorem 2.4 (ii) gives that the sum in (2.2) is non-negative. If $k,m \geq 1$, the sum in (2.2) always contains terms on the form

$$\operatorname{trace} Q_{11}^{k_1}Q_{12}Q_{22}^{m_1}Q_{21}. \tag{2.4}$$

Since Q_{11} is positive definite and $\operatorname{trace} AB = \operatorname{trace} BA$ for any matrices A and B such that both sides make sense,

$$\operatorname{trace} Q_{11}^{k_1}Q_{12}Q_{22}^{m_1}Q_{21} = \operatorname{trace} Q_{11}^{k_1/2}Q_{12}Q_{22}^{m_1}Q_{21}Q_{11}^{k_1/2}.$$

Using $Q_{12} = Q_{21}^t$ we conclude that (2.4) is equal to the trace of a positive semi-definite matrix and therefore non-negative. Consequently, there are always non-negative terms in (2.2).

It is an open problem if there exists a positive definite matrix Q with eigenvalues less than 1 and $k,m \in \mathbb{N}_0$ such that (2.2) is negative, which would be an example of (2.3) not being infinite divisible, or if the non-negative terms always compensate for possible negative terms, which is equivalent to (2.3) always being infinitely divisible.

Continue to consider the case $n_1 = n_2 = 2$ and write

$$\Sigma^{-1} = \begin{pmatrix} \Sigma^{11} & \Sigma^{12} \\ \Sigma^{21} & \Sigma^{22} \end{pmatrix}$$

where Σ^{ij} is a 2×2 matrix for $i,j=1,2$. Let W be a $(2,2)$-signature matrix such that

$$W^t\Sigma^{-1}W=\begin{pmatrix} W_1^t\Sigma^{11}W_1 & W_1^t\Sigma^{12}W_2 \\ W_2^t\Sigma^{21}W_1 & W_2^t\Sigma^{22}W_2\end{pmatrix}=\begin{pmatrix}\sigma_{11} & 0 & \sigma_{13} & \sigma_{14}\\ 0 & \sigma_{22} & \sigma_{23} & \sigma_{24}\\ \sigma_{13} & \sigma_{23} & \sigma_{33} & 0\\ \sigma_{14} & \sigma_{24} & 0 & \sigma_{44}\end{pmatrix}$$

where $\sigma_{11}\geq\sigma_{22}>0$ and $\sigma_{33}\geq\sigma_{44}>0$ which exists by Lemma 4.1. Note that σ_{ij} is not the (i,j)-th entry of Σ^{-1} but of $W^t\Sigma^{-1}W$. Let $v_1=(v_{11},v_{21})$ be the eigenvector of $W_1^t\Sigma^{12}\Sigma^{21}W_1$ associated with the largest eigenvalue. If $\sigma_{11}=\sigma_{22}$ or $\sigma_{33}=\sigma_{44}$, any orthogonal W_1 or W_2 gives the desired form. In this case, we may chose W_1 or W_2 such that $v_{21}\sigma_{24}(v_{21}\sigma_{24}+v_{11}\sigma_{14})\geq 0$, and it is such a choice we fix. Then we have the following theorem.

Theorem 2.6. *The vector $(X_1^2+X_2^2,X_3^2+X_4^2)$ is infinitely divisible if one of the following equivalent conditions is satisfied.*

(i) *There exists a $(2,2)$-signature matrix U such that $U^t\Sigma^{-1}U$ has non-positive off-diagonal elements.*

(ii) *The inequality $v_{21}\sigma_{24}(v_{21}\sigma_{24}+v_{11}\sigma_{14})\geq 0$ holds.*

Example 3.2 builds intuition about condition (ii) above, in particular that the condition holds in cases where $(X_1^2,X_2^2,X_3^2,X_4^2)$ is not infinitely divisible, but also that it is not always satisfied.

Theorem 2.6 (i) holds for general $n_1,n_2\geq 1$ as the following result shows. We give the proof below since it is short and makes the need for signature matrices clear. The proof of the more applicable condition (ii) in Theorem 2.6 is postponed to Section 4 since it relies on results that will be establish in that section.

Corollary 2.7 (to Theorem 1.1). *Let $(X_1,\dots,X_{n_1+n_2})$ be a mean zero Gaussian vector with positive definite covariance matrix Σ. Then*

$$(X_1^2+\cdots+X_{n_1}^2,X_{n_1+1}^2+\cdots+X_{n_1+n_2}^2)\tag{2.5}$$

is infinitely divisible if there exists an (n_1,n_2)-signature matrix U such that $U^t\Sigma^{-1}U$ has non-positive off-diagonal elements.

Proof. Write $X=(X_1,\dots,X_{n_1})$ and $Y=(X_{n_1+1},\dots,X_{n_1+n_2})$, and note that

$$\begin{aligned}(X_1^2+\cdots+X_{n_1}^2,X_{n_1+1}^2+\cdots+X_{n_1+n_2}^2)&=(\|X\|^2,\|Y\|^2)\\ &=(\|U_1X\|^2,\|U_2Y\|^2)\end{aligned}\tag{2.6}$$

for any $n_1\times n_1$ orthogonal matrix U_1 and $n_2\times n_2$ orthogonal matrix U_2. Consequently, any property of the distribution of (2.5) is invariant under transformations of the form

$$\begin{pmatrix}U_1^t & 0\\ 0 & U_2^t\end{pmatrix}\Sigma\begin{pmatrix}U_1 & 0\\ 0 & U_2\end{pmatrix}$$

of the covariance matrix Σ. Therefore, when there exists an (n_1,n_2)-signature matrix U such that $U^t\Sigma^{-1}U$ has non-positive off-diagonal elements, Theorem 1.1 ensures infinite divisibility of (2.6). □

3 Examples and numerics

We begin this section by presenting two examples treating the inequalities in Theorem 2.2 (ii) and Theorem 2.6 (ii) in special cases. Then we calculate the sums in Theorem 2.2 numerically with $n_1 = n_2 = 2$ for a specific value of Q for k and m less than 60.

Example 3.1. Fix $a > 0$ and assume that Q is on the form

$$Q = \begin{pmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{pmatrix} = \begin{pmatrix} q_1 & 0 & \varepsilon & \varepsilon \\ 0 & q_2 & \varepsilon & -\delta \\ \varepsilon & \varepsilon & q_3 & 0 \\ \varepsilon & -\delta & 0 & q_4 \end{pmatrix}$$

where $\delta, \varepsilon > 0$, $q_1 > q_2 > 0$, and $q_3 > q_4 > 0$. Let $v_1 = (v_{11}, v_{21})$ be the eigenvector of

$$Q_{12}Q_{21} = \begin{pmatrix} 2\varepsilon^2 & \varepsilon(\varepsilon-\delta) \\ \varepsilon(\varepsilon-\delta) & \varepsilon^2+\delta^2 \end{pmatrix}$$

associated with the largest eigenvalue λ_1. We will argue that the inequality in Theorem 2.4 (i), which reads

$$v_{11}(v_{11} + v_{21}) \geq 0 \tag{3.1}$$

in this case, holds if and only if $\delta \leq \varepsilon$. Then the same theorem will imply that

$$\operatorname{trace} Q_{11}^{k_1} Q_{12} Q_{22}^{m_1} Q_{21} Q_{11}^{k_1} \cdots Q_{11}^{k_d} Q_{12} Q_{22}^{m_d} Q_{21} Q_{11}^{k_{d+1}} \geq 0$$

for all $d \in \mathbb{N}_0$ and $k_1, \ldots, k_{d+1} m_1, \ldots, m_d \in \mathbb{N}_0$ if and only if $\delta \leq \varepsilon$, and therefore also that the sum in (2.2) is non-negative whenever this is the case.

Since $-v_1$ also is an eigenvector of $Q_{12}Q_{21}$ associated with the largest eigenvalue, we assume $v_{11} \geq 0$ without loss of generality. Assume $\delta \leq \varepsilon$. If $\delta = \varepsilon$, $v_1 = (1, 0)$ and the inequality in (3.1) holds. Assume $\delta < \varepsilon$. Since λ_1 is the largest eigenvalue,

$$\lambda_1 = \sup_{|v|=1} v^t Q_{12} Q_{21} v \geq 2\varepsilon^2$$

which implies that

$$2\varepsilon^2 - \lambda_1 \leq 0 \leq \varepsilon(\varepsilon - \delta).$$

Since v_1 is an eigenvector, $(Q - \lambda_1)v_1 = 0$ and we therefore have that

$$0 = (2\varepsilon^2 - \lambda_1)v_{11} + \varepsilon(\varepsilon - \delta)v_{21} \leq \varepsilon(\varepsilon - \delta)(v_{11} + v_{21}).$$

We conclude that (3.1) holds.

On the other hand, assume $\delta > \varepsilon$ and $v_{11} \geq 0$. Since λ_1 is the largest eigenvalue, $\lambda_1 \geq \delta^2 + \varepsilon^2 > \delta\varepsilon + \varepsilon^2$ and therefore,

$$(\lambda_1 - 2\varepsilon^2) > \varepsilon(\delta - \varepsilon).$$

Note that v_{11} can not be zero since the off-diagonal element in $Q_{12}Q_{21}$ is non-zero. We conclude that

$$0 = (\lambda_1 - 2\varepsilon^2)v_{11} + \varepsilon(\delta - \varepsilon)v_{21} > \varepsilon(\delta - \varepsilon)(v_{11} + v_{21}).$$

This implies that (3.1) does not hold.

Example 3.2. Assume Σ^{-1} is on the form

$$\Sigma^{-1} = \begin{pmatrix} \Sigma^{11} & \Sigma^{12} \\ \Sigma^{21} & \Sigma^{22} \end{pmatrix} = \begin{pmatrix} \sigma_1 & 0 & -\delta & \varepsilon \\ 0 & \sigma_2 & \varepsilon & \varepsilon \\ -\delta & \varepsilon & \sigma_3 & 0 \\ \varepsilon & \varepsilon & 0 & \sigma_4 \end{pmatrix} \tag{3.2}$$

where $\sigma_1 > \sigma_2 > 0$, $\sigma_3 > \sigma_4 > 0$, and $\delta, \varepsilon > 0$. Let $v_1 = (v_{11}, v_{21})$ be the eigenvector of $\Sigma^{12}\Sigma^{21}$ associated with the largest eigenvalue. We will argue that the inequality in Theorem 2.6 (ii) holds if and only if $\delta \leq \varepsilon$. Then the same theorem implies that $(X_1^2 + X_2^2, X_3^2 + X_4^2)$ is infinitely divisible whenever $\delta \leq \varepsilon$. On the other hand, Theorem 1.1 implies that $(X_1^2, X_2^2, X_3^2, X_4^2)$ is never infinite divisible under (3.2) since there does not exists a matrix D on the form $\mathrm{diag}(\pm1, \pm1, \pm1, \pm1)$ such that $D\Sigma^{-1}D$ has non-positive off-diagonal elements. Indeed, for any two matrices D_1 and D_2 on the form $\mathrm{diag}(\pm1, \pm1)$, $D_1\Sigma^{12}D_2$ has either three negative and one positive or one negative and three positive entrances.

To see that $v_{21}(v_{11} + v_{21}) \geq 0$ if and only if $\delta \leq \varepsilon$, let

$$P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

and Q_{12} be given as in Example 3.1. Then $P\Sigma^{12}P = Q_{12}$, implying that (v_{21}, v_{11}) is the eigenvector associated with the largest eigenvalue of $Q_{12}Q_{21}$. We have argued in Example 3.1 that $v_{21}(v_{11} + v_{21}) \geq 0$ holds if and only if $\delta \leq \varepsilon$ which is the desired conclusion.

Now we investigate infinite divisibility of $(X_1^2 + X_2^2, X_3^2 + X_4^2)$ numerically. More specifically, we consider the sums in (2.2) with $n_1 = n_2 = 2$ for a specific choice of positive definite matrix and different values of k and m. We will scale Q to have its largest eigenvalue equal to one to avoid getting too close to zero. Due to Theorem 2.4 the case where $v_{11}q_{13}(v_{11}q_{13} + v_{21}q_{23}) < 0$ (in the notation from Theorem 2.4) is the only case where the infinite divisibility of $(X_1^2 + X_2^2, X_3^2 + X_4^2)$ is open.

Let

$$Q = \frac{1}{\lambda} \begin{pmatrix} 0.8 & 0 & 0.01 & 0.01 \\ 0 & 0.3 & 0.01 & -0.2 \\ 0.01 & 0.01 & 0.8 & 0 \\ 0.01 & -0.2 & 0 & 0.3 \end{pmatrix}$$

where $\lambda > 0$ is chosen such that Q has its largest eigenvalue equal to 1. Note that by Example 3.1, $v_{11}q_{13}(v_{11}q_{13} + v_{21}q_{23}) < 0$. In Figure 1 the logarithm of the sums in (2.2) for k and m between 0 and 60 is plotted. It is seen that the logarithm seems stable and therefore, that the sums in (2.2) remain positive in this case. A similar analysis have been done for other positive definite matrices, and we have not encountered any $k, m \in \mathbb{N}_0$ such that (2.2) is negative. This, together with Theorem 2.4 (ii), leads us to conjecture that $(X_1^2 + X_2^2, X_3^2 + X_4^2)$ is infinite divisible for any covariance structure of (X_1, X_2, X_3, X_4).

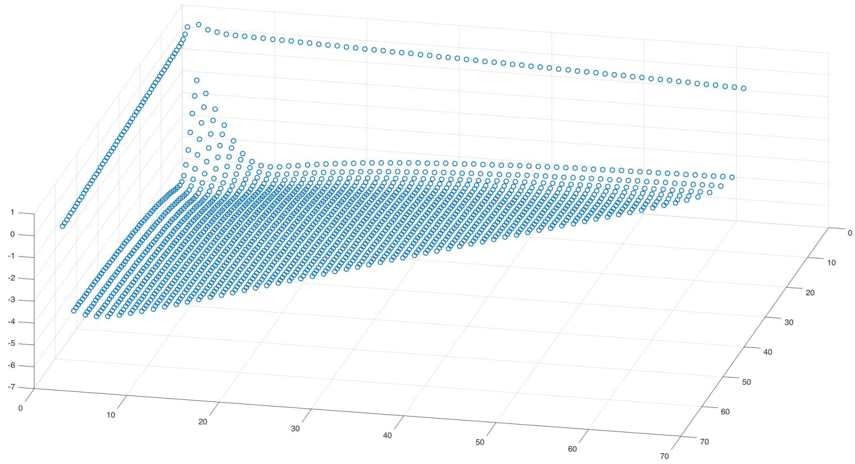


Figure 1: The logarithm of the sums in (2.2) for k and m between 0 and 60.

4 Proofs

We start this section with two lemmas on linear algebra. Lemma 4.2 will be very useful in the proofs that make up the rest of this section.

Lemma 4.1. *Let A be a $n \times n$ positive definite matrix. Let $n_1, n_2 \in \mathbb{N}$ be such that $n_1 + n_2 = n$ and write*

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$$

where A_{11} is an $n_1 \times n_1$ matrix, A_{22} is an $n_2 \times n_2$ matrix, and $A_{12} = A_{21}^t$ is an $n_1 \times n_2$ matrix. Then there exists an (n_1, n_2)-signature matrix W such that $W^t A W$ has the form

$$\begin{pmatrix} \tilde{A}_{11} & \tilde{A}_{12} \\ \tilde{A}_{21} & \tilde{A}_{22} \end{pmatrix}$$

where $\tilde{A}_{11} = \mathrm{diag}(a_1, \ldots, a_{n_1})$ and $\tilde{A}_{22} = \mathrm{diag}(a_{n_1+1}, \ldots, a_{n_1+n_2})$ with $a_i > 0$ for $i = 1, \ldots, n_1 + n_2$, and where $\tilde{A}_{12} = \tilde{A}_{21}^t$. Furthermore, we may choose W such that $a_1 \geq a_2 \geq \cdots \geq a_{n_1}$ and $a_{n_1+1} \geq a_{n_1+2} \geq \cdots \geq a_{n_1+n_2}$.

Proof. Since A is positive definite, A_{11} and A_{22} are positive definite. Consequently, by the spectral theorem (see for example [10, Corollary 6.4.7]), there exists an $n_1 \times n_1$ matrix W_1 and an $n_2 \times n_2$ matrix W_2, both orthogonal, such that $W_1^t A_{11} W_1$ and $W_2^t A_{22} W_2$ are diagonal with positive diagonal entries. Since permutation matrices are orthogonal matrices, we may assume the diagonal is ordered by size in both $W_1^t A_{11} W_1$ and $W_2^t A_{22} W_2$. Consequently, letting

$$W = \begin{pmatrix} W_1 & 0 \\ 0 & W_2 \end{pmatrix},$$

implies that $W^t A W$ has the right form. □

For a fixed eigenvector v_i we call the system $Av_i = \lambda_i v_i$, the system of eigenequations. The k'th equation in this system will be called the k'th eigenequation associated with v_i.

Let A be a 4×4 positive definite matrix, and let W be a $(2,2)$-signature such that

$$W^tAW = \begin{pmatrix} W_1^tA_{11}W_1 & W_1^tA_{12}W_2 \\ W_2^tA_{21}W_1 & W_2^tA_{22}W_2 \end{pmatrix} = \begin{pmatrix} a_{11} & 0 & a_{13} & a_{14} \\ 0 & a_{22} & a_{23} & a_{24} \\ a_{13} & a_{23} & a_{33} & 0 \\ a_{14} & a_{24} & 0 & a_{44} \end{pmatrix},$$

where $a_{11} \geq a_{22} > 0$ and $a_{33} \geq a_{44} > 0$ which exists by Lemma 4.1. Note that a_{ij} is not the (i,j)-th entry of A but of W^tAW. Let $v_1 = (v_{11}, v_{21})$ be the eigenvector associated with the largest eigenvalue of $W_1^tA_{12}A_{21}W_1$. If $a_{11} = a_{22}$ or $a_{33} = a_{44}$, any orthogonal W_1 or W_2 give the desired form. In this case, we may chose W_1 or W_2 such that $v_{11}a_{13}(v_{11}a_{13} + v_{21}a_{23}) \geq 0$, and it is such a choice we fix. Then the lemma below will play a central role in the proofs of the previously stated results.

Lemma 4.2. *In the notation above, the following are equivalent.*

(i) *There exists a $(2,2)$-signature matrix U such that U^tAU has all entries non-negative.*

(ii) *For any $d \in \mathbb{N}$ and $k_1,\dots,k_{d+1}, m_1,\dots m_d \in \mathbb{N}_0$,*
$$\operatorname{trace} A_{11}^{k_1}A_{12}A_{22}^{m_1}A_{21}A_{11}^{k_2}\cdots A_{11}^{k_d}A_{12}A_{22}^{m_d}A_{21}A_{11}^{k_{d+1}} \geq 0.$$

(iii) *The inequality $v_{11}a_{13}(v_{11}a_{13} + v_{21}a_{23}) \geq 0$ holds.*

Proof. (i) $\Rightarrow$ (ii). Let

$$U = \begin{pmatrix} U_1 & 0 \\ 0 & U_2 \end{pmatrix}$$

be such that $B_{ij} = U_i^tA_{ij}U_j$ has non-negative entries for $i,j = 1,2$. Then

$$\begin{aligned} &\operatorname{trace} A_{11}^{k_0}A_{12}A_{22}^{m_1}A_{21}A_{11}^{k_1}\cdots A_{11}^{k_{d-1}}A_{12}A_{22}^{m_d}A_{21}A_{11}^{k_d} \\ &\quad = \operatorname{trace} B_{11}^{k_0}B_{12}B_{22}^{m_1}B_{21}B_{11}^{k_1}\cdots B_{11}^{k_{d-1}}B_{12}B_{22}^{m_d}B_{21}B_{11}^{k_d}. \end{aligned}$$

This trace is non-negative since all matrices in the product only contain non-negative entries.

(ii) $\Rightarrow$ (iii). By the spectral theorem, we may write $W_1^tA_{12}A_{21}W_1 = V\Lambda V^t$ where V is a 2×2 orthogonal matrix and $\Lambda = \operatorname{diag}(\lambda_1, \lambda_2)$ with $\lambda_1 \geq \lambda_2 \geq 0$. Note that v_1, the eigenvector associated with largest eigenvalue of $W_1^tA_{12}A_{21}W_1$, is the first column of V. If $\lambda_1 = \lambda_2$, $v_1 = (1,0)$ and the inequality holds. If $a_{11} = a_{22}$ or $a_{33} = a_{44}$, $W_1^tA_{11}W_1 = A_{11}$ or $W_2^tA_{22}W_2 = A_{22}$, and choosing W_1 or W_2 such that $a_{23} = 0$ then ensures the inequality in (iii) holds.

Assume now that $\lambda_1 > \lambda_2$, $a_{11} > a_{22}$, and $a_{33} > a_{44}$. It follows by assumption that

$$\begin{aligned} 0 &\leq \frac{1}{a_{11}^k}\frac{1}{a_{33}^k}\frac{1}{\lambda_1^k}\operatorname{trace} A_{11}^kA_{12}A_{22}^kA_{21}(A_{12}A_{21})^k \\ &= \operatorname{trace}\begin{pmatrix} 1 & 0 \\ 0 & (\frac{a_{22}}{a_{11}})^k \end{pmatrix} W_1^tA_{12}W_2 \begin{pmatrix} 1 & 0 \\ 0 & (\frac{a_{44}}{a_{33}})^k \end{pmatrix} W_2^tA_{21}W_1V \begin{pmatrix} 1 & 0 \\ 0 & (\frac{\lambda_1}{\lambda_2})^k \end{pmatrix} V^t \\ &\to \operatorname{trace}\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} W_1^tA_{12}W_2 \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} W_2^tA_{21}W_1V \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} V^t \end{aligned}$$

as $k \to \infty$. This gives the inequality in (iii) since

$$\begin{aligned}&\operatorname{trace}\begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix} W_1^t A_{12} W_2 \begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix} W_2^t A_{21} W_1 V \begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix} V^t\\ &\quad = v_{11}a_{13}(v_{11}a_{13} + v_{21}a_{23}).\end{aligned}$$

(iii) $\Rightarrow$ (i). To ease the notation and without loss of generality assume that $W = I$. We are then pursuing two 2×2 orthogonal matrices U_1 and U_2 such that $U_1^t A_{11} U_1$, $U_1^t A_{12} U_2$, and $U_2^t A_{22} U_2$ all have non-negative entrances. Initially consider D_1 and D_2 on the form diag(±1, ±1). Then clearly, $D_1 A_{11} D_1 = A_{11}$ and $D_2 A_{22} D_2 = A_{22}$ since A_{11} and A_{22} are diagonal matrices. Next, note that either it is possible to find D_1 and D_2 such that $D_1 A_{12} D_2$ has all entrances non-negative or such that

$$D_1 A_{12} D_2 = \begin{pmatrix} a_{13} & a_{14}\\ a_{23} & -a_{24}\end{pmatrix} \tag{4.1}$$

where $a_{13}, a_{23}, a_{14}, a_{24} > 0$. Consequently, we will assume A_{12} is on the form in (4.1) since otherwise choosing $U_1 = D_1$ and $U_2 = D_2$ would be sufficient.

As one of two cases, assume $a_{13}a_{23} - a_{14}a_{24} \geq 0$, and define

$$U_2 = \begin{pmatrix} \alpha\frac{a_{14}a_{24}}{a_{23}} & \beta a_{23}\\ \alpha a_{14} & -\beta a_{24}\end{pmatrix}$$

where $\alpha, \beta > 0$ are chosen such that each column in U_2 has norm one. Then U_2 is orthogonal,

$$A_{12}U_2 = \begin{pmatrix} \alpha\left(a_{14}^2 + \frac{a_{13}a_{14}a_{24}}{a_{23}}\right) & \beta(a_{13}a_{23} - a_{14}a_{24})\\ 0 & \beta(a_{23}^2 + a_{24}^2)\end{pmatrix},$$

and

$$U_2^t A_{22} U_2 = \begin{pmatrix} \alpha^2\left(a_{33}\left(\frac{a_{14}a_{24}}{a_{23}}\right)^2 + a_{44}a_{14}^2\right) & \alpha\beta a_{14}a_{24}(a_{33} - a_{44})\\ \alpha\beta a_{14}a_{24}(a_{33} - a_{44}) & \beta^2 a_{23}^2 + \beta^2 a_{24}^2\end{pmatrix}.$$

Since $a_{33} \geq a_{44}$, all entries in $A_{12}U_2$ and $U_2^t A_{22} U_2$ are non-negative. Choosing $U_1 = I$ then gives a pair of orthogonal matrices with the desired property.

Now assume $a_{13}a_{23} - a_{14}a_{24} < 0$. Note that A_{12} on the form (4.1) can not be singular and consequently, there exists $\lambda_1 \geq \lambda_2 > 0$ and an orthogonal matrix V such that $A_{12}A_{21} = V\Lambda V^t$, where $\Lambda = \operatorname{diag}(\lambda_1, \lambda_2)$. Furthermore, since V contains the eigenvectors of $A_{12}A_{21}$ we may assume v_{11} and v_{12} have the same sign where v_{ij} is the (i,j)-th component of V. Define

$$W = A_{21}V(\Lambda^{1/2})^{-1}, \tag{4.2}$$

and note that this is an orthogonal matrix which, together with V, decomposes A_{12} into its singular value decomposition, that is, $V^t A_{12} W = \Lambda^{1/2}$. Then

$$V^t A_{11} V = \begin{pmatrix} a_{11}v_{11}^2 + a_{22}v_{21}^2 & v_{11}v_{12}(a_{11} - a_{22})\\ v_{11}v_{12}(a_{11} - a_{22}) & a_{11}v_{12}^2 + a_{22}v_{22}^2\end{pmatrix}.$$

All entries in $V^t A_{11} V$ are non-negative since we chose v_{11} and v_{12} to have the same sign, and since $a_{11} \geq a_{22} > 0$.

To see that $W^t A_{22} W$ also have all entries non-negative, consider the first line in the eigenequations for $A_{12}A_{21}$ associated with the eigenvector (v_{12}, v_{22}), the eigenvector associated with the smallest eigenvalue λ_2,

$$(a_{13}^2 + a_{14}^2 - \lambda_2)v_{12} + (a_{13}a_{23} - a_{14}a_{24})v_{22} = 0. \tag{4.3}$$

Since λ_2 is the smallest eigenvalue of $A_{12}A_{21}$,

$$\lambda_2 = \inf_{|v|=1} v^t A_{12} A_{21} v,$$

and since the off-diagonal elements in $A_{12}A_{21}$ are non-zero, $(1,0)$ and $(0,1)$ cannot be eigenvectors. Consequently, λ_2 is strictly smaller than any diagonal element of $A_{12}A_{21}$, and in particular $a_{13}^2 + a_{14}^2 - \lambda_2 > 0$. Since we also have $a_{13}a_{23} - a_{14}a_{24} < 0$, (4.3) gives that v_{12} and v_{22} need to have the same sign for the sum to equal zero. Let w_{ij} be the (i,j)-th component of W and note that by (4.2),

$$w_{11}w_{12} = \frac{v_{11}a_{13} + v_{21}a_{23}}{\lambda_1^{1/2}} \frac{v_{12}a_{13} + v_{22}a_{23}}{\lambda_2^{1/2}}.$$

The assumption $v_{11}a_{13}(v_{11}a_{13} + v_{21}a_{23}) \geq 0$ implies that $v_{11}a_{13} + v_{21}a_{23}$ and v_{11} have the same sign. Since v_{11} and v_{12} were chosen to have the same sign, and v_{12} and v_{22} have the same sign, we conclude that $(v_{11}a_{13} + v_{21}a_{23})(v_{12}a_{13} + v_{22}a_{23})$ is non-negative and therefore, $w_{11}w_{12}$ is non-negative too. Then writing

$$W^t A_{22} W = \begin{pmatrix} a_{33}w_{11}^2 + a_{44}w_{21}^2 & w_{11}w_{12}(a_{33} - a_{44}) \\ w_{11}w_{12}(a_{33} - a_{44}) & a_{33}w_{12}^2 + a_{44}w_{22}^2 \end{pmatrix}$$

makes it clear that $W^t A_{22} W$ has non-negative elements. Thus, letting $U_1 = V$ and $U_2 = W$ completes the proof. □

Corollary 4.3. *Let A and v_1 be given as in Lemma 4.2. Then there exists a $(2,2)$-signature matrix U such that $U^t A U$ has non-positive off-diagonal elements if and only if*

$$v_{21}a_{24}(v_{21}a_{24} + v_{11}a_{14}) \geq 0. \tag{4.4}$$

Proof. Let W be defined as in Lemma 4.2. Define

$$P_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \text{and} \quad P = \begin{pmatrix} P_1 & 0 \\ 0 & P_1 \end{pmatrix}.$$

Then $P_1 v_1 = (v_{21}, v_{11})$ is the eigenvector of $P_1 W_1^t A_{12} A_{21} W_1 P_1$ associated with the largest eigenvalue. Let

$$\tilde{A} = \begin{pmatrix} W_1^t A_{11} W_1 & P_1 W_1^t A_{12} W_2 P_1 \\ P_1 W_2^t A_{21} W_1 P_1 & W_2^t A_{22} W_2 \end{pmatrix} = \begin{pmatrix} a_{11} & 0 & a_{24} & a_{23} \\ 0 & a_{22} & a_{14} & a_{13} \\ a_{24} & a_{14} & a_{33} & 0 \\ a_{23} & a_{13} & 0 & a_{44} \end{pmatrix}.$$

By Lemma 4.2, there exists a $(2,2)$-signature matrix

$$\tilde{U} = \begin{pmatrix} \tilde{U}_1 & 0 \\ 0 & \tilde{U}_2 \end{pmatrix}$$

such that $\tilde{U}^t \tilde{A} \tilde{U}$ has non-negative entries if and only if $v_{21}a_{24}(v_{21}a_{24} + v_{11}a_{14}) \geq 0$. Define now the $(2,2)$-signature matrix U as

$$U = \begin{pmatrix} U_1 & 0 \\ 0 & U_2 \end{pmatrix} = \begin{pmatrix} -W_1 P_1 \tilde{U}_1 & 0 \\ 0 & W_2 P_1 \tilde{U}_2 \end{pmatrix}.$$

Let $\tilde{u}_{ij}$ be the (i,j)-th component of $\tilde{U}_1$. Since $\tilde{U}_1$ is orthogonal, $\tilde{u}_{12}\tilde{u}_{22} = -\tilde{u}_{11}\tilde{u}_{21}$ implying that

$$U_1^t A_{11} U_1 = \begin{pmatrix} \tilde{u}_{11}^2 a_{22} + \tilde{u}_{21}^2 a_{11} & \tilde{u}_{11}\tilde{u}_{12}(a_{22} - a_{11}) \\ \tilde{u}_{11}\tilde{u}_{12}(a_{22} - a_{11}) & \tilde{u}_{12}^2 a_{22} + \tilde{u}_{22}^2 a_{11} \end{pmatrix}$$

and

$$\tilde{U}_1^t W_1^t A_{11} W_1 \tilde{U}_1 = \begin{pmatrix} \tilde{u}_{11}^2 a_{11} + \tilde{u}_{21}^2 a_{22} & \tilde{u}_{11}\tilde{u}_{12}(a_{11} - a_{22}) \\ \tilde{u}_{11}\tilde{u}_{12}(a_{11} - a_{22}) & \tilde{u}_{12}^2 a_{11} + \tilde{u}_{22}^2 a_{22} \end{pmatrix}.$$

Consequently $\tilde{U}_1^t W_1^t A_{11} W_1 \tilde{U}_1$ has non-negative elements if and only if $U_1^t A_{11} U_1$ has non-positive off-diagonal elements. Similarly, $\tilde{U}_2^t W_2^t A_{22} W_2 \tilde{U}_2$ has non-negative elements if and only if $U_2^t A_{22} U_2$ has non-positive off-diagonal elements by a similar argument. Finally we note that

$$U_1^t A_{12} U_2 = -\tilde{U}_1^t P_1 W_1^t A_{12} W_2 P_1 \tilde{U}_2,$$

and it follows that $U^t A U$ has non-positive off-diagonal elements if and only if

$$\tilde{U}_1^t P_1 W_1^t A_{12} W_2 P_1 \tilde{U}_2, \quad \tilde{U}_1^t W_1^t A_{11} W_1 \tilde{U}_1 \quad \text{and,} \quad \tilde{U}_2^t W_2 A_{22} W_2 \tilde{U}_2$$

have all entries non-negative. We conclude that we can find a $(2,2)$-signature matrix U such that $U^t A U$ has non-positive off-diagonal element if and only if (4.4) holds. □

The following lemma will be useful in the proof of Theorem 2.2. A proof can be found in [12, Lemma 13.2.2].

Lemma 4.4. *Let $\psi : \mathbb{R}_+^n \to (0,\infty)$ be a continuous function. Suppose that, for all $a > 0$ sufficiently large, $\log \psi(a(1-s_1),\ldots,a(1-s_n))$ has a power series expansion for $s = (s_1,\ldots,s_n) \in [0,1]^n$ around $s = 0$ with all its coefficients non-negative, except for the constant term. Then ψ is the Laplace transform of an infinitely divisible random variable in $\mathbb{R}_+^n$.*

We now give the proof of Theorem 2.2, where all the main steps follow similar as in [12, Proof of Theorem 13.2.1], but with several modifications to adjust to a different setting. E.g. there is a difference in the S matrix appearing in the proof.

Proof (Proof of Theorem 2.2). By [12, Lemma 5.2.1],

$$\begin{aligned} P(s_1,s_2) &= \mathbb{E}\exp\{-\tfrac{1}{2}a((1-s_1)(X_1^2 + \cdots + X_{n_1}^2) + (1-s_2)(X_{n_1+1}^2 + \cdots + X_{n_2}^2))\} \\ &= \frac{1}{|I + \Sigma a(I-S)|^{1/2}}, \end{aligned}$$

where S is the $(n_1+n_2)\times(n_1+n_2)$ diagonal matrix with s_1 on the first n_1 diagonal entries and s_2 on the remaining n_2 diagonal entries. Recall that $Q = I-(I+a\Sigma)^{-1}$. Then

$$\begin{aligned} P(s_1,s_2)^2 &= |I+a\Sigma-a\Sigma S|^{-1} \\ &= |(I-Q)^{-1}-((I-Q)^{-1}-I)S|^{-1} \\ &= |I-Q||I-QS|^{-1}, \end{aligned}$$

from which it follows that

$$\begin{aligned} 2\log P(s_1,s_2) &= \log|I-Q| - \log|I-QS| \\ &= \log|I-Q| + \sum_{n=1}^{\infty} \frac{\mathrm{trace}\{(QS)^n\}}{n}, \end{aligned} \tag{4.5}$$

where the last equality follows from [12, p. 562]. Now assume that the vector $(X_1^2+\cdots+X_{n_1}^2, X_{n_1+1}^2+\cdots+X_{n_1+n_2}^2)$ is infinitely divisible, and write

$$(X_1^2+\cdots+X_{n_1}^2, X_{n_1+1}^2+\cdots+X_{n_1+n_2}^2) \stackrel{d}{=} Y_1^n+\cdots+Y_n^n$$

where $Y_1^n,\ldots Y_n^n$ are 2-dimensional independent identically distributed stochastic vectors. Let Y_{ij}^n be the j-th component of Y_i^n and note that $Y_{ij}^n \geq 0$ a.s. for all i,j,n. Then

$$P(s_1,s_2)^{1/n} = \mathbb{E}\exp\{-\tfrac{1}{2}a((1-s_1)Y_{11}^n+(1-s_2)Y_{12}^n)\}.$$

That $P^{1/n}(s_1,s_2)$ has a power series expansion with all coefficient non-negative follows from writing

$$\exp\{-\tfrac{1}{2}a((1-s_j)Y_{1j}^n)\} = \exp\{-\tfrac{1}{2}aY_{1j}^n)\})\sum_{k=0}^{\infty}\frac{(s_j a Y_{1j}^n)^k}{2^k k!}.$$

We have that

$$\log P(s_1,s_2) = \lim_{n\to\infty}(n(P^{1/n}(s_1,s_2)-1)). \tag{4.6}$$

Note that $(s_1,s_2)\mapsto n(P^{1/n}(s_1,s_2)-1)$ and all its derivatives converge uniformly on $[0,1)\times[0,1)$ by a Weierstrass M-test (see for example [14, Theorem 7.10]). Consequently, we may use [14, Theorem 7.17] to conclude that

$$\frac{\partial^{\alpha+\beta}}{\partial s_1^\alpha \partial s_2^\beta}\lim_{n\to\infty}(n(P^{1/n}(s_1,s_2)-1)) = \lim_{n\to\infty}\frac{\partial^{\alpha+\beta}}{\partial s_1^\alpha \partial s_2^\beta}(n(P^{1/n}(s_1,s_2)-1))$$

for any $\alpha,\beta\in\mathbb{N}_0$. Thus, that all the terms in the power series expansion of $P^{1/n}(s_1,s_2)$ are non-negative implies that all the terms in the power series representation of $\log P(s_1,s_2)$ except the constant term are non-negative by (4.6). By (4.5) we conclude that any coefficient in front of $s_1^k s_2^m$ in $\mathrm{trace}\{(QS)^{k+m}\}$ has to be non-negative for all $k,m\in\mathbb{N}$ and $a>0$. Expanding out the trace then gives that this is equivalent to non-negativity of the sum in (2.2) for all $k,m\in\mathbb{N}_0$.

On the other hand, if the sum in (2.2) is non-negative for all $k, m \in \mathbb{N}_0$ and $a > 0$ sufficiently large, (4.5) and Lemma 4.4 imply that

$$(X_1^2 + \cdots + X_{n_1}^2, X_{n_1+1}^2 + \cdots + X_{n_1+n_2}^2)$$

is infinitely divisible. □

Proof (Proof of Theorem 2.4). Lemma 4.2 implies the equivalence in (i). Now we set out to show (ii), i.e., to show that the sum in Theorem 2.2 is non-negative for $k, m \in \mathbb{N}_0$ such that $k \leq 2$, $m \leq 2$, or $k + m \leq 7$ in the special case $n_1 = n_2 = 2$. To this end, consider a 4×4 positive definite matrix Q and write

$$Q = \begin{pmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{pmatrix}$$

where Q_{ij} is a 2×2 matrix for $i, j = 1, 2$. Let W_1 and W_2 be two 2×2 orthogonal matrices and define $P_{ij} = W_i Q_{ij} W_j$. Then

$$\begin{aligned} &\operatorname{trace} Q_{11}^{k_1} Q_{12} Q_{22}^{m_1} Q_{21} \cdots Q_{12} Q_{22}^{m_d} Q_{21} Q_{11}^{k_{d+1}} \\ &= \operatorname{trace} P_{11}^{k_1} P_{12} P_{22}^{m_1} P_{21} \cdots P_{12} P_{22}^{m_d} P_{21} P_{11}^{k_{d+1}}. \end{aligned} \tag{4.7}$$

Consequently (see Lemma 4.1), we may assume, without loss of generality, that Q_{11} and Q_{22} are diagonal with the first diagonal element greater than or equal the other and all entries non-negative.

Either there exists D_1 and D_2 on the form $\operatorname{diag}(\pm 1, \pm 1)$ such that $D_1 Q_{12} D_2$ has all entries non-negative or such that

$$D_1 Q_{12} D_2 = \begin{pmatrix} q_{13} & q_{23} \\ q_{14} & -q_{24} \end{pmatrix}$$

where $q_{13}, q_{23}, q_{14}, q_{24} > 0$. If $D_1 Q_{12} D_2$ has all entries non-negative, writing as in (4.7) with W_i replaced by D_i implies non-negativity of each individual trace. We conclude that we may assume

$$Q = \begin{pmatrix} \lambda_1 & 0 & q_{13} & q_{14} \\ 0 & \lambda_2 & q_{23} & -q_{24} \\ q_{13} & q_{23} & \lambda_3 & 0 \\ q_{14} & -q_{24} & 0 & \lambda_4 \end{pmatrix},$$

where $\lambda_1 \geq \lambda_2 \geq 0$ and $\lambda_3 \geq \lambda_4 \geq 0$ and $q_{13}, q_{23}, q_{14}, q_{24} > 0$, without loss of generality.

We now write out the traces in (2.2) for specific values of k and m and show non-negativity in each case.

$k = 0$ **or** $m = 0$

Assume $k = 0$ and fix some $m \in \mathbb{N}$. Then the terms in the sum in Theorem 2.2 reduce to $\operatorname{trace} Q_{22}^m$. Since Q_{22} is positive definite, Q_{22}^m is positive definite. Consequently, $\operatorname{trace} Q_{22}^m > 0$. Similarly, when $m = 0$ and $k \in \mathbb{N}$, the terms in the sum in Theorem 2.2 reduce to $\operatorname{trace} Q_{11}^k$, which again is positive since Q_{11} is positive definite.

$k = 1$ **or** $m = 1$

Assume $k = 1$ and fix some $m \in \mathbb{N}$. Then (2.2) reduces to

$$\operatorname{trace} Q_{12} Q_{22}^{m} Q_{21} + \sum_{m_1=0}^{m-1} \operatorname{trace} Q_{22}^{m_1} Q_{21} Q_{12} Q_{22}^{m-1-m_1},$$

which equals

$$(m+1) \operatorname{trace} Q_{12} Q_{22}^{m} Q_{21}.$$

Since $Q_{12} = Q_{21}^t$ and Q_{22} is positive definite, $Q_{12} Q_{22}^m Q_{21}$ is positive semi-definite. We conclude that $\operatorname{trace} Q_{12} Q_{22}^m Q_{21} \geq 0$.

Assume $m = 1$ and fix some $k \in \mathbb{N}$. Similar to above, (2.2) reduces to

$$\operatorname{trace} Q_{21} Q_{11}^{k} Q_{12} + \sum_{k_1=0}^{k-1} \operatorname{trace} Q_{11}^{k_1} Q_{12} Q_{21} Q_{11}^{k-1-k_1}.$$

That this trace is non-negative follows by arguments similar to those above.

$k = 2$ **or** $m = 2$

Assume that $k = 2$ and let $m \in \mathbb{N}$. The case $m = 1$ is discussed above. Assume $m \geq 2$. Then (2.2) reduces to

$$\begin{aligned}
&\operatorname{trace} Q_{11} Q_{12} Q_{22}^{m-1} Q_{21} \\
&\quad + \sum_{m_1+m_2+1=m} \operatorname{trace} Q_{22}^{m_1} Q_{21} Q_{11} Q_{12} Q_{22}^{m_2} \\
&\quad + \sum_{m_1+m_2+2=m} \operatorname{trace} Q_{12} Q_{22}^{m_1} Q_{21} Q_{12} Q_{22}^{m_2} Q_{21} \\
&\quad + \sum_{m_1+m_2+m_3+2=m} \operatorname{trace} Q_{22}^{m_1} Q_{21} Q_{12} Q_{22}^{m_2} Q_{21} Q_{12} Q_{22}^{m_3}.
\end{aligned}$$

All the traces above are non-negative. To see this, consider for example

$$\operatorname{trace} Q_{22}^{m_1} Q_{21} Q_{12} Q_{22}^{m_2} Q_{21} Q_{12} Q_{22}^{m_3}$$

for some $m_1, m_2, m_3 \in \mathbb{N}_0$. Since Q_{22} is positive definite it has a unique positive definite square root $Q_{22}^{1/2}$. We conclude that

$$\begin{aligned}
&\operatorname{trace} Q_{22}^{m_1} Q_{21} Q_{12} Q_{22}^{m_2} Q_{21} Q_{12} Q_{22}^{m_3} \\
&\quad = \operatorname{trace} Q_{22}^{(m_1+m_3)/2} Q_{21} Q_{12} Q_{22}^{m_2} Q_{21} Q_{12} Q_{22}^{(m_1+m_3)/2}.
\end{aligned} \tag{4.8}$$

Note that

$$Q_{22}^{(m_1+m_3)/2} Q_{21} Q_{12} = (Q_{21} Q_{12} Q_{22}^{(m_1+m_3)/2})^t,$$

which implies that (4.8) is the trace of positive semi-definite matrix and therefore non-negative.

Non-negativity of the traces when $m = 2$ and $k \in \mathbb{N}$ follows by symmetry.

$k = 3$ **and** $m = 3$

In the following we will need to expand traces, and we therefore note that

$$\operatorname{trace} Q_{11}^k Q_{12} Q_{22}^m Q_{21} = \lambda_1^k \lambda_3^m q_{13}^2 + \lambda_1^k \lambda_4^m q_{14}^2 + \lambda_2^k \lambda_3^m q_{23}^2 + \lambda_2^k \lambda_4^m q_{24}^2 \tag{4.9}$$

for any $k, m \in \mathbb{N}$, and

$$\begin{aligned}
\operatorname{trace} & Q_{11}^{k_1} Q_{12} Q_{22}^{m_1} Q_{21} Q_{11}^{k_2} Q_{12} Q_{22}^{m_2} Q_{21} \\
&= \lambda_1^{k_1+k_2} \lambda_3^{m_1+m_2} q_{13}^4 + \lambda_1^{k_1+k_2} \lambda_4^{m_1+m_2} q_{14}^4 \\
&+ \lambda_2^{k_1+k_2} \lambda_3^{m_1+m_2} q_{23}^4 + \lambda_2^{k_1+k_2} \lambda_4^{m_1+m_2} q_{24}^4 \\
&+ \lambda_1^{k_1+k_2} (\lambda_3^{m_1} \lambda_4^{m_2} + \lambda_4^{m_1} \lambda_3^{m_2}) q_{13}^2 q_{14}^2 \\
&+ \lambda_2^{k_1+k_2} (\lambda_3^{m_1} \lambda_4^{m_2} + \lambda_3^{m_2} \lambda_4^{m_1}) q_{23}^2 q_{24}^2 \\
&+ \lambda_3^{m_1+m_2} (\lambda_1^{k_1} \lambda_2^{k_2} + \lambda_1^{k_2} \lambda_2^{k_1}) q_{13}^2 q_{23}^2 \\
&+ \lambda_4^{m_1+m_2} (\lambda_1^{k_1} \lambda_2^{k_2} + \lambda_1^{k_2} \lambda_2^{k_1}) q_{14}^2 q_{24}^2 \\
&- (\lambda_1^{k_1} \lambda_2^{k_2} + \lambda_1^{k_2} \lambda_2^{k_1})(\lambda_3^{m_1} \lambda_4^{m_2} + \lambda_3^{m_2} \lambda_4^{m_1}) q_{13} q_{23} q_{14} q_{24}
\end{aligned} \tag{4.10}$$

for any $k_1, k_2, m_1, m_2 \in \mathbb{N}$.

Assume now $k = 3$ and $m = 3$ and consider the sum in Theorem 2.2. The sum contains all terms on the form

$$\operatorname{trace} Q_{11}^{k_1} Q_{12} Q_{22}^2 Q_{21} Q_{11}^{k_2}$$

where $k_1 + k_2 = 2$ and

$$\operatorname{trace} Q_{22}^{m_1} Q_{21} Q_{11}^2 Q_{12} Q_{22}^{m_2}$$

where $m_1 + m_2 = 2$. All these traces equal

$$\operatorname{trace} Q_{11}^2 Q_{12} Q_{22}^2 Q_{21},$$

and there are all together 6 of these terms. Next, the sum in Theorem 2.2 also contains all terms on the form

$$\operatorname{trace} Q_{11}^{k_1} Q_{12} Q_{22}^{m_1} Q_{21} Q_{11}^{k_2} Q_{12} Q_{22}^{m_2} Q_{21} Q_{11}^{k_3}$$

where $k_1 + k_2 + k_3 = 1$ and $m_1 + m_2 = 1$, and

$$\operatorname{trace} Q_{22}^{m_1} Q_{21} Q_{11}^{k_1} Q_{12} Q_{22}^{m_2} Q_{21} Q_{11}^{k_2} Q_{12} Q_{22}^{m_3}$$

where $m_1 + m_2 + m_3 = 1$ and $k_1 + k_2 = 1$. Using both that $\operatorname{trace} AB = \operatorname{trace} BA$ and $\operatorname{trace} A^t = \operatorname{trace} A$ for any two square matrices A and B of the same dimensions we get that all these traces share the common trace

$$\operatorname{trace} Q_{11} Q_{12} Q_{21} Q_{12} Q_{22} Q_{21}.$$

All together there are 12 of these terms. Finally, the sum in Theorem 2.2 contains the two terms

$$\operatorname{trace}(Q_{12} Q_{21})^3 \quad \text{and} \quad \operatorname{trace}(Q_{21} Q_{12})^3,$$

which share a common trace. We conclude that the sum in Theorem 2.2 reads

$$\operatorname{trace}\{6Q_{11}^2Q_{12}Q_{22}^2Q_{21} + 12Q_{11}Q_{12}Q_{21}Q_{12}Q_{22}Q_{21} + 2(Q_{12}Q_{21})^3\}. \tag{4.11}$$

Since $Q_{12} = Q_{21}^t$, $Q_{12}Q_{21}$ is positive semi-definite and consequently, $\operatorname{trace}(Q_{12}Q_{21})^3 \geq 0$. Furthermore, we have

$$\operatorname{trace}\ Q_{11}^2Q_{12}Q_{22}^2Q_{21} = \operatorname{trace}\ Q_{11}Q_{12}Q_{22}^2Q_{21}Q_{11} \geq 0.$$

Contrarily, there exists a positive definite matrix Q such that

$$\operatorname{trace} Q_{11}Q_{12}Q_{21}Q_{12}Q_{22}Q_{21} < 0.$$

(To see this, consider Q on the form in Example 3.1 with ε small and δ large relative to ε.) We will now argue that despite this,(4.11) remains non-negative. Initially we note that

$$Q_{11}^{k_i}Q_{12}Q_{22}^{m_i}Q_{21} = \begin{pmatrix} \lambda_1^{k_i}(\lambda_3^{m_i}q_{13}^2 + \lambda_4^{m_i}q_{14}^2) & \lambda_1^{k_i}(\lambda_3^{m_i}q_{13}q_{23} - \lambda_4^{m_i}q_{14}q_{24}) \\ \lambda_2^{k_i}(\lambda_3^{m_i}q_{13}q_{23} - \lambda_4^{m_i}q_{14}q_{24}) & \lambda_2^{k_i}(\lambda_3^{m_i}q_{23}^2 + \lambda_4^{m_i}q_{24}^2) \end{pmatrix}$$

and

$$Q_{22}^{m_i}Q_{21}Q_{11}^{k_i}Q_{12} = \begin{pmatrix} \lambda_3^{m_i}(\lambda_1^{k_i}q_{13}^2 + \lambda_2^{k_i}q_{23}^2) & \lambda_3^{m_i}(\lambda_1^{k_i}q_{13}q_{14} - \lambda_2^{k_i}q_{23}q_{24}) \\ \lambda_4^{m_i}(\lambda_1^{k_i}q_{13}q_{14} - \lambda_2^{k_i}q_{23}q_{24}) & \lambda_4^{m_i}(\lambda_1^{k_i}q_{14}^2 + \lambda_2^{k_i}q_{24}^2) \end{pmatrix}.$$

Since $\lambda_1 \geq \lambda_2$ and $\lambda_3 \geq \lambda_4$, we see that if $q_{13}q_{14} \geq q_{23}q_{24}$ or $q_{13}q_{23} \geq q_{14}q_{24}$, then one of two matrices above have only non-negative entrances for any $k_i, m_i \in \mathbb{N}_0$. Consequently,

$$\operatorname{trace} Q_{11}^{k_1}Q_{12}Q_{22}^{m_1}Q_{21}Q_{11}^{k_2}Q_{12}Q_{22}^{m_2}Q_{21} = \operatorname{trace} Q_{22}^{m_1}Q_{21}Q_{11}^{k_1}Q_{12}Q_{22}^{m_2}Q_{21}Q_{11}^{k_2}Q_{12}$$

would be non-negative if this was the case. Especially, we would have

$$\operatorname{trace} Q_{11}Q_{12}Q_{21}Q_{12}Q_{22}Q_{21} \geq 0.$$

Assume now that $q_{13}q_{14} \leq q_{23}q_{24}$ and $q_{13}q_{23} \leq q_{14}q_{24}$. By (4.9) and (4.10),

$$\begin{aligned} &\operatorname{trace}\{\tfrac{1}{2}Q_{11}^2Q_{12}Q_{22}^2Q_{21} + Q_{11}Q_{12}Q_{22}Q_{21}Q_{12}Q_{21}\} \\ &\quad = \tfrac{1}{2}\lambda_1^2\lambda_3^2q_{13}^2 + \tfrac{1}{2}\lambda_1^2\lambda_4^2q_{14}^2 + \tfrac{1}{2}\lambda_2^2\lambda_3^2q_{23}^2 + \tfrac{1}{2}\lambda_2^2\lambda_4^2q_{24}^2 \\ &\quad + \lambda_1\lambda_3q_{13}^4 + \lambda_1\lambda_4q_{14}^4 + \lambda_2\lambda_3q_{23}^4 + \lambda_2\lambda_4q_{24}^4 \\ &\quad + \lambda_1(\lambda_3+\lambda_4)q_{13}^2q_{14}^2 + \lambda_2(\lambda_3+\lambda_4)q_{23}^2q_{24}^2 \\ &\quad + \lambda_3(\lambda_1+\lambda_2)q_{13}^2q_{23}^2 + \lambda_4(\lambda_1+\lambda_2)q_{14}^2q_{24}^2 \\ &\quad - (\lambda_1+\lambda_2)(\lambda_3+\lambda_4)q_{13}q_{23}q_{14}q_{24}. \end{aligned} \tag{4.12}$$

We are going to bound the term $(\lambda_1+\lambda_2)(\lambda_3+\lambda_4)q_{13}q_{23}q_{14}q_{24}$ by the positive terms to show non-negative of this trace. We recall that $\lambda_1 \geq \lambda_2 > 0$ and $\lambda_3 \geq \lambda_4 > 0$. Initially, note that

$$\begin{aligned} \lambda_2\lambda_3q_{13}q_{23}q_{14}q_{24} &\leq \lambda_2\lambda_3q_{13}^2q_{23}^2 \\ \lambda_2\lambda_4q_{13}q_{23}q_{14}q_{24} &\leq \lambda_1\lambda_4q_{13}^2q_{14}^2 \\ \lambda_1\lambda_4q_{13}q_{23}q_{14}q_{24} &\leq \lambda_1\lambda_3q_{13}^2q_{14}^2. \end{aligned}$$

This leaves only $\lambda_1\lambda_3 q_{13}q_{23}q_{14}q_{24}$ to be bounded. If $\lambda_1\lambda_3 q_{13}q_{23}q_{14}q_{24} \leq \frac{1}{2}\lambda_1^2\lambda_3^2 q_{13}^2$, we have a bounding term in (4.12). Therefore, assume $2q_{23}q_{14}q_{24} \geq \lambda_1\lambda_3 q_{13}$. Since Q was assumed positive definite, $\lambda_2\lambda_4 \geq q_{24}^2$. Consequently,

$$\begin{aligned}
\lambda_1\lambda_3 q_{13}q_{23}q_{14}q_{24} &\leq 2q_{23}^2 q_{14}^2 q_{24}^2 \\
&\leq 2\lambda_2\lambda_4 q_{23}^2 q_{13}^2 \\
&\leq \lambda_2\lambda_4(q_{23}^4 + q_{13}^4) \\
&\leq \lambda_2\lambda_3 q_{23}^4 + \lambda_1\lambda_3 q_{13}^4.
\end{aligned}$$

We conclude that (4.12) and hence (4.11) is non-negative.

$k + m = 7$

Now consider $k, m \in \mathbb{N}$ such that $k + m = 7$. Whenever $k, m = 1, 2$, we already know that the sum in Theorem 2.2 is non-negative. Let $k = 3$ and $m = 4$. Then the sum in Theorem 2.2 reads

$$\begin{aligned}
\operatorname{trace}\{&14Q_{11}Q_{12}Q_{21}Q_{12}Q_{22}^2Q_{21} + 7Q_{11}^2Q_{12}Q_{22}^3Q_{21} \\
&7Q_{11}(Q_{12}Q_{22}Q_{21})^2 + 7Q_{12}Q_{22}Q_{21}(Q_{12}Q_{21})^2\}.
\end{aligned} \tag{4.13}$$

Initially we note that

$$\operatorname{trace} Q_{11}(Q_{12}Q_{22}Q_{21})^2 \geq 0 \quad \text{and} \quad \operatorname{trace} Q_{12}Q_{22}Q_{21}(Q_{12}Q_{21})^2 \geq 0$$

since they both can be written as the trace of positive semi-definite matrices (see above for more details). Next, by (4.9) and (4.10),

$$\begin{aligned}
\operatorname{trace}&\{\tfrac{1}{2}Q_{11}^2Q_{12}Q_{22}^3Q_{21} + Q_{11}Q_{12}Q_{22}^2Q_{21}Q_{12}Q_{21}\} \\
&= \tfrac{1}{2}\lambda_1^2\lambda_3^3q_{13}^2 + \tfrac{1}{2}\lambda_1^2\lambda_4^3q_{14}^2 + \tfrac{1}{2}\lambda_2^2\lambda_3^3q_{23}^2 + \tfrac{1}{2}\lambda_2^2\lambda_4^3q_{24}^2 \\
&+ \lambda_1(\lambda_3^2 + \lambda_4^2)q_{13}^2q_{14}^2 + \lambda_2(\lambda_3^2 + \lambda_4^2)q_{23}^2q_{24}^2 \\
&+ \lambda_3^2(\lambda_1 + \lambda_2)q_{13}^2q_{23}^2 + \lambda_4^2(\lambda_1 + \lambda_2)q_{14}^2q_{24}^2 \\
&+ \lambda_1\lambda_3^2q_{13}^4 + \lambda_1\lambda_4^2q_{14}^4 + \lambda_2\lambda_3^2q_{23}^4 + \lambda_2\lambda_4^2q_{24}^4 \\
&- (\lambda_1 + \lambda_2)(\lambda_3^2 + \lambda_4^2)q_{13}q_{23}q_{14}q_{24}.
\end{aligned} \tag{4.14}$$

Again we bound the negative term by positive terms. Recall that $\lambda_1 \geq \lambda_2$ and $\lambda_3 \geq \lambda_4$, and that we may assume $q_{23}q_{24} \geq q_{13}q_{14}$ and $q_{14}q_{24} \geq q_{13}q_{23}$ without loss of generality. Consequently,

$$\begin{aligned}
\lambda_1\lambda_4^2q_{13}q_{23}q_{14}q_{24} &\leq \lambda_1\lambda_4^2q_{14}^2q_{24}^2 \\
\lambda_2\lambda_3^2q_{13}q_{23}q_{14}q_{24} &\leq \lambda_2\lambda_3^2q_{23}^2q_{24}^2 \\
\lambda_2\lambda_4^2q_{13}q_{23}q_{14}q_{24} &\leq \lambda_2\lambda_4^2q_{14}^2q_{24}^2,
\end{aligned}$$

leaving $\lambda_1\lambda_3^2q_{13}q_{23}q_{14}q_{24}$ to be bounded. First note that

$$\tfrac{1}{2}\lambda_1^2\lambda_3^3q_{13}^2 - \lambda_1\lambda_3^2q_{13}q_{23}q_{14}q_{24} = \lambda_1\lambda_3^2q_{13}(\tfrac{1}{2}\lambda_1\lambda_3q_{13} - q_{23}q_{14}q_{24}),$$

so that non-negativity holds if $\frac{1}{2}\lambda_1\lambda_3 q_{13} \geq q_{23}q_{14}q_{24}$. Assume $\lambda_1\lambda_3 q_{13} \leq 2q_{23}q_{14}q_{24}$ and recall that $\lambda_2\lambda_4 \geq q_{24}^2$ since Q is positive definite. Then

$$\begin{aligned}
\lambda_1\lambda_3^2 q_{13}q_{23}q_{14}q_{24} &\leq 2\lambda_3 q_{23}^2 q_{14}^2 q_{24}^2 \\
&\leq 2\lambda_2\lambda_3\lambda_4 q_{23}^2 q_{14}^2 \\
&\leq \lambda_2\lambda_3^2 q_{23}^4 + \lambda_2\lambda_4^2 q_{14}^4 \\
&\leq \lambda_2\lambda_3^2 q_{23}^4 + \lambda_1\lambda_4^2 q_{14}^4
\end{aligned}$$

so we have found bounding terms for the last expression. We conclude that (4.14) is non-negative and therefore, (4.13) is non-negative too. The case $k = 4$ and $m = 3$ follows by symmetry. It follows that the sum in Theorem 2.2 is non-negative for $k + m = 7$. □

Proof (Proof of Theorem 2.6). Corollary 4.3 gives that (i) and (ii) are equivalent and Corollary 2.7 gives that (i) implies infinite divisibility. □

Acknowledgments. This research was supported by the Danish Council for Independent Research (Grant DFF - 4002-00003).

References

[1] Bapat, R.B. (1989). Infinite divisibility of multivariate gamma distributions and M-matrices. *Sankhyā Ser. A* 51(1), 73–78.

[2] Dobrushin, R.L. and P. Major (1979). Non-central limit theorems for nonlinear functionals of Gaussian fields. *Z. Wahrsch. Verw. Gebiete* 50(1), 27–52.

[3] Eisenbaum, N. (2003). On the infinite divisibility of squared Gaussian processes. *Probab. Theory Related Fields* 125(3), 381–392.

[4] Eisenbaum, N. (2014). Characterization of positively correlated squared Gaussian processes. *Ann. Probab.* 42(2), 559–575.

[5] Eisenbaum, N. and H. Kaspi (2006). A characterization of the infinitely divisible squared Gaussian processes. *Ann. Probab.* 34(2), 728–742.

[6] Evans, S.N. (1991). Association and infinite divisibility for the Wishart distribution and its diagonal marginals. *J. Multivariate Anal.* 36(2), 199–203.

[7] Griffiths, R.C. (1970). Infinitely divisible multivariate gamma distributions. *Sankhyā Ser. A* 32, 393–404.

[8] Griffiths, R.C. (1984). Characterization of infinitely divisible multivariate gamma distributions. *J. Multivariate Anal.* 15(1), 13–20.

[9] Janson, S. (1997). *Gaussian Hilbert spaces*. Vol. 129. Cambridge Tracts in Mathematics. Cambridge University Press, Cambridge, x+340.

[10] Leon, S.J. (1980). *Linear algebra with applications*. Macmillan, Inc., New York; Collier-Macmillan Publishers, London, xii+338.

[11] Lévy, P. (1948). The arithmetic character of the Wishart distribution. *Proc. Cambridge Philos. Soc.* 44, 295–297.

[12] Marcus, M.B. and J. Rosen (2006). *Markov processes, Gaussian processes, and local times*. Vol. 100. Cambridge Studies in Advanced Mathematics. Cambridge University Press, Cambridge, x+620.

[13] Moran, P.A. and D. Vere-Jones (1969). The infinite divisibility of of multi-gamma distributions. *Sankhyā Ser. A* 31, 191–194.

[14] Rudin, W. (1976). *Principles of mathematical analysis*. Third. International Series in Pure and Applied Mathematics. McGraw-Hill Book Co., New York-Auckland-Düsseldorf, x+342.

[15] Shanbhag, D.N. (1976). On the structure of the Wishart distribution. *J. Multivariate Anal.* 6(3), 347–355.

[16] Taqqu, M.S. (1979). Convergence of integrated processes of arbitrary Hermite rank. *Z. Wahrsch. Verw. Gebiete* 50(1), 53–83.

PAPER B

The Polylogarithmic Distribution

Victor Rohde

Abstract

This study introduces a family of infinitely divisible distributions: the polylogarithmic distribution. The polylogarithmic distribution is closely connected to the stationary and conditional distribution of Ornstein-Uhlenbeck processes, and we explore this connection. We show that the characteristic function and moment-generating function of an Ornstein-Uhlenbeck process driven by a polylogarithmic Lévy process can be calculated analytically. Furthermore, the stationary and conditional distribution of a CARMA process driven by a polylogarithmic Lévy process is examined, and, under certain conditions, algorithms for fast calculation of the characteristic and moment-generating function are proposed. Algorithms for fast computation of the polylogarithm of order 2 and 3 are suggested.

Keywords: polylogarithm; Ornstein-Uhlenbeck processes; CARMA processes; infinitely divisible distributions

1 Introduction

Continuous-time Lévy driven moving averages are a valuable tool for modeling many types of phenomena. They introduce a wide range of auto-correlation structures and, for certain classes of kernels, they offer a lot tractability. If the kernel of the moving average is non-negative and the Lévy process has non-negative increments, then the moving average will too be non-negative. This is an attractive property when, for example, modeling wind power, stochastic volatility of a financial asset, temperature, or the price of electricity.

When the Lévy process is not a Brownian motion, the stationary distribution of the moving average can be difficult to characterize. This can be a hurdle for employing a moving average model. For example, if we consider the moving average,

$$X(t) = \int_{\mathbb{R}} g(t-u)\,\mathrm{d}L(u)$$

then the characteristic function of X is given by

$$\phi_X(s) = \mathbb{E}[\exp\{isX(0)\}] = \exp\left\{\int_{\mathbb{R}} \psi_L(isg(u))\,\mathrm{d}u\right\} \tag{1.1}$$

where $\psi_L(z) = \log \mathbb{E}[\exp\{zL(1)\}]$ is the cumulant-generating function of $L(1)$, see [11]. (Here, we are ignoring technical assumption about g and L.) To do, for example, Fourier pricing (see [8]) using this model, it is crucial to calculate the Fourier transform fast, and the integral in (1.1) may prohibit this. It would therefore be of large benefit to have an analytical expression or a fast approximation scheme. We will consider a class of Lévy processes with non-negative increments where the integral in (1.1) can be calculated analytically for an Ornstein-Uhlenbeck (OU) kernel and where a fast approximation scheme can be formulated for a continuous-time autoregressive moving average (CARMA) kernel under some conditions. This class of Lévy processes include the compound Poisson process with exponential jumps and the gamma Lévy process. An OU or CARMA porcess driven by a compound Poisson process with exponential jumps or a gamma Lévy process have, for example, been used to model stochastic volatility of financial assets and wind power utilization, see [1, 3, 4, 7].

2 Motivation

Let $(L(t))_{t\in\mathbb{R}}$ be a Lévy process with log moments (that is, $\mathbb{E}[\log(1 \vee |L(1)|)]) < \infty$). Then we define a Lévy driven OU process $(X(t))_{t\in\mathbb{R}}$ by

$$X(t) = \int_{-\infty}^{t} e^{-\lambda(t-u)}\,\mathrm{d}L(u) \quad t \in \mathbb{R}, \tag{2.1}$$

for $\lambda > 0$. The class of possible distributions for (2.1) is exactly the self-decomposable distributions, see [12]. For fixed $s \in \mathbb{R}$, we also consider the process $(X_s(t))_{t\geq s}$ defined by

$$X_s(t) = \int_{s}^{t} e^{-\lambda(t-u)}\,\mathrm{d}L(u), \quad t \geq s. \tag{2.2}$$

This process is important to consider when considering the distribution of $X(t)$ conditional on $\sigma\{X(u) : u \leq s\}$ since

$$X(t) = \int_{-\infty}^{t} e^{-\lambda(t-u)}\,\mathrm{d}L(u) = X_s(t) + e^{-\lambda(t-s)}X(s).$$

We will also consider a subclass of CARMA processes. We will not go into a detailed introduction of CARMA processes but refer to [6] for a thorough treatment. In [5, Remark 4.9] it is argued that the subclass considered covers a large class of the CARMA processes employed in the literature. We will consider a CARMA$(p, p-1)$ process, $p \in \mathbb{N}$, $(Y(t))_{t\in\mathbb{R}}$ given by

$$Y(t) = \int_{-\infty}^{t} \sum_{i=1}^{p} a_i e^{-b_i(t-u)}\,\mathrm{d}L(u), \quad t \in \mathbb{R}, \tag{2.3}$$

and, again for fixed $s \in \mathbb{R}$, the process $(Y_s(t))_{t \geq s}$ defined by

$$Y_s(t) = \int_s^t \sum_{i=1}^p a_i e^{-b_i(t-u)} \, \mathrm{d}L(u), \quad t \geq s, \tag{2.4}$$

where $a_i, b_i > 0$, $i = 1, \dots, p$ and $\sum_{i=1}^p a_i = 1$. We have

$$Y(t) = \int_{-\infty}^t \sum_{i=1}^p a_i e^{-b_i(t-u)} \, \mathrm{d}L(u) = Y_s(t) + \int_{-\infty}^s \sum_{i=1}^p a_i e^{-b_i(t-u)} \, \mathrm{d}L(u).$$

If we impose the additional assumption that $(Y(t))_{t \in \mathbb{R}}$ is invertible (see [2, Remark 4.9]), then $L(v) - L(u)$, $s \geq v > u$, is measurable with respect to $\sigma\{Y(u) : u \leq s\}$, and therefore so is $\int_{-\infty}^s \sum_{i=1}^p a_i e^{-b_i(t-u)} \, \mathrm{d}L(u)$. We therefore see that $Y_s(t)$ is the critical part for assessing the conditional distribution of $Y(t)$.

3 The Polylogarithmic distribution

In this section we introduce a family of infinitely divisible distributions. This family of distributions, as we will explore later, is closely connected to the Ornstein-Uhlenbeck process. Two special cases include the distribution of a compound Poisson process with exponential jumps and the gamma distribution.

3.1 The polylogarithm

The polylogarithm (see [10] for a review on the polylogarithm) of order $s \in \mathbb{R}$, Li_s, is defined as the power series

$$\mathrm{Li}_s(z) = \sum_{k=1}^\infty \frac{z^k}{k^s} \tag{3.1}$$

for $z \in \mathbb{C}$ with $|z| < 1$. It is possible to analytically extend the polylogarithm to $|z| \geq 1$ except the line $[1, \infty)$. The polylogarithm can be shown to satisfy the integral relation

$$\mathrm{Li}_{s+1}(z) = \int_0^z \frac{\mathrm{Li}_s(t)}{t} \, \mathrm{d}t, \quad z \in \mathbb{C} \setminus [1, \infty) \tag{3.2}$$

or equivalently,

$$\frac{\mathrm{d}}{\mathrm{d}z} \mathrm{Li}_{s+1}(z) = \frac{\mathrm{Li}_s(z)}{z}, \quad z \in \mathbb{C} \setminus (\{0\} \cup [1, \infty)). \tag{3.3}$$

Special cases include

$$\mathrm{Li}_1(z) = -\log(1-z), \quad \mathrm{Li}_0(z) = \frac{z}{1-z}, \quad \text{and} \quad \mathrm{Li}_{-1}(z) = \frac{z}{(1-z)^2}. \tag{3.4}$$

Using (3.3) and (3.4), closed expression for Li_s when $s = -2, -3, \dots$ is readily attainable. On the other hand, there are no closed form expressions for Li_s when $s > 1$ or $s \notin \mathbb{Z}$.

The polylogarithm has a long history in the field of mathematics, first appearing in one of the letters from Leibniz to Johann Bernoulli in 1696. It has later been studied by some of the great mathematicians of the past such as Euler, Abel, and Ramanujan (cf. [10]). The polylogarithm is also related to the Riemann zeta function ζ. In particular, $\zeta(s) = \mathrm{Li}_s(1)$.

3.2 Connection to Ornstein-Uhlenbeck processes

For a random variable U, the cumulant-generating function, ψ_U is defined by

$$\psi_U(z) = \log \mathbb{E}[\exp\{zU\}],$$

for $z \in \mathbb{C}$ such that the expectation is well-defined and finite. We will say that U has a polylogarithmic distribution of order $n \in \mathbb{N}_0$ and with parameters $\alpha > 0$ and $\beta > 0$ if

$$\psi_U(z) = \alpha \operatorname{Li}_n(z/\beta), \quad z \in \mathbb{C} \setminus ([1,\infty) \times i\mathbb{R}) \tag{3.5}$$

and, with a slight abuse of notation, write $U \sim \operatorname{Li}_n(\alpha, \beta)$. In Theorem 3.3 below we will argue that there exists random variables with a polylogarithmic distribution of order n for any $n \in \mathbb{N}_0$. Assuming it exists, we note that it is an infinitely divisible distribution. To see this, consider $U_1 \sim \operatorname{Li}_n(\alpha_1, \beta)$ and $U_2 \sim \operatorname{Li}_n(\alpha_2, \beta)$ independent. Then $U_1 + U_2 \sim \operatorname{Li}_1(\alpha_1 + \alpha_2, \beta)$ by the definition of the cumulant-generating function. Since the polylogarithmic distribution is infinitely divisible, we can associate a Lévy process to it (for more on Lévy processes see for example [13]).

Definition 3.1. *A Lévy process $(L(t))_{t\in\mathbb{R}}$ is said to be a polylogarithmic Lévy process of order $n \in \mathbb{N}_0$ and with parameters $\alpha > 0$ and $\beta > 0$ if $L(1) \sim \operatorname{Li}_n(\alpha, \beta)$.*

Remark 3.2. Let $(L(t))_{t\in\mathbb{R}}$ be a compound Poisson process with intensity $\alpha > 0$ and exponential distributed jumps with parameter $\beta > 0$. Then it is not too difficult to show that the cumulant-generating function of $L(1)$, $\psi_{L(1)}$, is

$$\psi_{L(1)}(z) = \log \mathbb{E}[e^{zL(1)}] = \alpha \frac{z}{\beta - z} = \alpha \operatorname{Li}_0(z/\beta), \quad z \in \mathbb{C} \setminus ([\beta,\infty) \times i\mathbb{R})$$

by (3.4). We therefore see that $(L(t))_{t\in\mathbb{R}}$ is a polylogarithmic Lévy process of order 0 with parameters α and β.

Additionally, we recognize $\alpha \operatorname{Li}_1(z/\beta) = -\alpha \log(1 - z/\beta)$ as the cumulant-generating function for the gamma distribution with shape α and rate β, and therefore conclude that the polylogarithmic distribution of order 1 is the gamma distribution.

The following result is a central reason for studying polylogarithmic distributions. One of the implications is that an OU process driven by a polylogarithmic Lévy process of order n has a polylogarithmic distribution of order $n+1$.

Theorem 3.3. *Let $(X(t))_{t\in\mathbb{R}}$ be the OU process in* (2.1) *and $(X_s(t))_{t\geq s}$ be given by* (2.2) *where $(L(t))_{t\in\mathbb{R}}$ is a polylogarithmic Lévy process of order $n \in \mathbb{N}_0$ and with parameters $\alpha > 0$ and $\beta > 0$. Then*

$$\psi_{X(0)}(z) = \log \mathbb{E}[\exp\{zX(0)\}] = \frac{\alpha}{\lambda} \operatorname{Li}_{n+1}(z/\beta) \tag{3.6}$$

and

$$\psi_{X_s(t)}(z) = \log \mathbb{E}[\exp\{zX_s(t)\}] = \frac{\alpha}{\lambda} \left(\operatorname{Li}_{n+1}(z/\beta) - \operatorname{Li}_{n+1}\left(ze^{-\lambda(t-s)}/\beta\right)\right) \tag{3.7}$$

for $z \in \mathbb{C} \setminus ([\beta,\infty) \times i\mathbb{R})$.

Proof. Initially, using (3.3), we find

$$\frac{\mathrm{d}}{\mathrm{d}u}\frac{\mathrm{Li}_{n+1}(ze^{-\lambda u}/\beta)}{-\lambda} = \mathrm{Li}_n(ze^{-\lambda u}/\beta).$$

Now we have

$$\begin{aligned}
\log \mathbb{E}[\exp\{zX(t)\}] &= \log \mathbb{E}\left[\exp\left\{z\int_{-\infty}^{t} e^{-\lambda(t-u)}\,\mathrm{d}L(u)\right\}\right] \\
&= \alpha\int_0^\infty \mathrm{Li}_n(ze^{-\lambda u}/\beta)\,\mathrm{d}u \\
&= \left[\frac{\mathrm{Li}_{n+1}(ze^{-\lambda u}/\beta)}{-\lambda}\right]_0^\infty \\
&= \frac{\mathrm{Li}_{n+1}(z/\beta)}{\lambda}
\end{aligned}$$

and

$$\begin{aligned}
\log \mathbb{E}[\exp\{zX_s(t)\}] &= \log \mathbb{E}\left[\exp\left\{z\int_{s}^{t} e^{-\lambda(t-u)}\,\mathrm{d}L(u)\right\}\right] \\
&= \alpha\int_0^{t-s} \mathrm{Li}_n(ze^{-\lambda u}/\beta)\,\mathrm{d}u \\
&= \alpha\left[\frac{\mathrm{Li}_{n+1}(ze^{-\lambda u}/\beta)}{-\lambda}\right]_0^{t-s} \\
&= \frac{\alpha}{\lambda}\left(\mathrm{Li}_{n+1}(z/\beta) - \mathrm{Li}_{n+1}\left(ze^{-\lambda(t-s)}/\beta\right)\right).
\end{aligned}$$

□

Remark 3.4. From Theorem 3.3 we conclude that the polylogarithmic distribution has support on $[0,\infty)$ by an induction argument: the polylogarithm of order 0 has support on $[0,\infty)$ because it is the distribution of a compound Poisson process with exponential jumps. An OU process driven by a non-negative Lévy process is again non-negative, and it therefore follows from Theorem 3.3 that the polylogarithmic distribution of order 1 has support on $[0,\infty)$ (this also follows since it is a gamma distribution). This argument can now be continued.

Remark 3.5. Since the polylogarithmic distribution of order $n \in \mathbb{N}$ can be realized as the stationary distribution of an OU process, it follows that it is a self-decomposable distribution.

Example 3.6. Let $(L(t))_{t\in\mathbb{R}}$ be a gamma Lévy process with shape $\alpha > 0$ and rate $\beta > 0$, that is,

$$\psi_{L(1)}(z) = -\alpha\log(1 - z/\beta) = \alpha\,\mathrm{Li}_1(z/\beta), \quad z \in \mathbb{C}\setminus([\beta,\infty)\times i\mathbb{R}).$$

Then the cumulant-generating function of the OU process with mean reversion $\lambda > 0$ and driven by $(L(t))_{t\in\mathbb{R}}$ is

$$z \mapsto \frac{\alpha}{\lambda}\mathrm{Li}_2(z/\beta), \quad z \in \mathbb{C}\setminus([\beta,\infty)\times i\mathbb{R}).$$

The function Li_2 is known as the Dilogarithm or Spence's function. It does not have a closed form but the integral representation

$$\mathrm{Li}_2(z) = -\int_0^z \frac{\log(1-y)}{y}\,dy, \quad , \quad z \in \mathbb{C} \setminus ([1,\infty) \times i\mathbb{R}).$$

An efficient algorithm for computing $\mathrm{Li}_2(z)$ is presented in Section 5.

3.3 Properties of the polylogarithmic distribution

We now turn to proving some useful properties of the polylogarithmic distribution. We start by giving the first moment, and the second and third centralized moments. In principle, any centralized moment could be calculated using the approach of the proof, but the calculations quickly become rather cumbersome and we therefore restrain ourselves.

Proposition 3.7. *Let U be a random variable with a polylogarithmic distribution of order $n \in \mathbb{N}_0$ and with parameters $\alpha > 0$ and $\beta > 0$. Then*

$$\begin{aligned} \mathbb{E}[U] &= \frac{\alpha}{\beta} \\ \mathbb{E}[(U - \mathbb{E}[U])^2] &= \frac{\alpha}{2^{n-1}\beta^2} \\ \mathbb{E}[(U - \mathbb{E}[U])^3] &= \frac{2\alpha}{3^{n-1}\beta^3}. \end{aligned}$$

Proof. Since U has a polylogarithmic distribution of order n and with parameters α and β,

$$\psi_U(ix) = \exp(\alpha\,\mathrm{Li}_n(ix/\beta)).$$

We find that

$$\begin{aligned} \frac{\frac{\mathrm{d}}{\mathrm{d}x}\psi_U(ix)}{\psi_U(ix)} &= \frac{\alpha\,\mathrm{Li}_{n-1}(ix/\beta)}{ix} \\ \frac{\frac{\mathrm{d}^2}{\mathrm{d}x^2}\psi_U(ix)}{\psi_U(ix)} &= \alpha\left(\frac{\mathrm{Li}_{n-2}(ix/\beta) - \mathrm{Li}_{n-1}(ix/\beta)}{(ix)^2}\right) + \left(\frac{\alpha\,\mathrm{Li}_{n-1}(ix/\beta)}{ix}\right)^2 \\ \frac{\frac{\mathrm{d}^3}{\mathrm{d}x^3}\psi_U(ix)}{\psi_U(ix)} &= \alpha\left(\frac{\mathrm{Li}_{n-3}(ix/\beta) - 3\,\mathrm{Li}_{n-2}(ix/\beta) + 2\,\mathrm{Li}_{n-1}(ix/\beta)}{(ix)^3}\right) \\ &\quad + 3\alpha^2\frac{\mathrm{Li}_{n-1}(ix/\beta)}{ix}\left(\frac{\mathrm{Li}_{n-2}(ix/\beta) - \mathrm{Li}_{n-1}(ix/\beta)}{(ix)^2}\right) \\ &\quad + \left(\frac{\alpha\,\mathrm{Li}_{n-1}(ix/\beta)}{ix}\right)^3. \end{aligned} \tag{3.8}$$

It follows from (3.1) that

$$\begin{aligned} \mathrm{Li}_{n-1}(x) &= x + O(x^2) \\ \mathrm{Li}_{n-2}(x) - \mathrm{Li}_{n-1}(x) &= \frac{x^2}{2^{n-1}} + O(x^3) \\ \mathrm{Li}_{n-3}(x) - 3\,\mathrm{Li}_{n-2}(x) + 2\,\mathrm{Li}_{n-1}(x) &= \frac{2x^3}{3^{n-1}} + O(x^4) \end{aligned}$$

□

for $x \to 0$. The result now follow from letting $x \to 0$ in (3.8).

The next result is a direct consequence of the definition of the polylogarithmic distribution.

Proposition 3.8. *For $\alpha_1, \alpha_2, \beta > 0$ and $n \in \mathbb{N}_0$, let $U_1 \sim \mathrm{Li}_n(\alpha_1, \beta)$ and $U_2 \sim \mathrm{Li}_n(\alpha_2, \beta)$ be independent. Consider $c > 0$. Then*

$$cU_1 \sim \mathrm{Li}_n(\alpha_1, \beta/c) \quad \text{and} \quad U_1 + U_2 \sim \mathrm{Li}_n(\alpha_1 + \alpha_2, \beta).$$

Furthermore, U_1 has exponential moment of any order below β. That is,

$$\mathbb{E}[\exp\{xU_1\}] < \infty$$

for any $x < \beta$.

Proof. This is a consequence of (3.5). □

Finally, using Fourier inversion, we plot some densities of the polylogarithmic distribution in Figure 1. The polylogarithmic distribution of order 0 is not included since it is not absolutely continuous with respect to the Lebesgue measure. All densities correspond to a polylogarithmic distribution with mean and variance equal to 1.

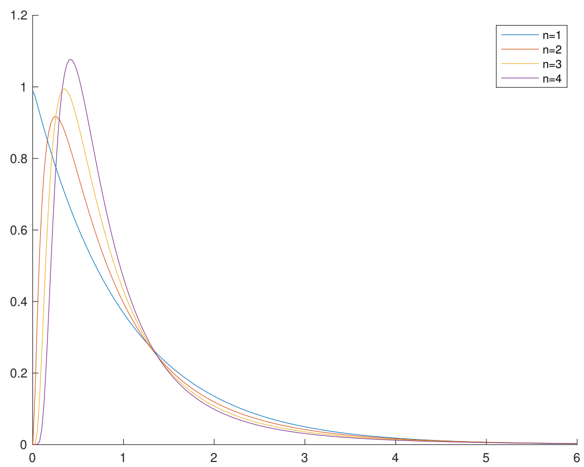


Figure 1: The densities of a polylogarithmic distribution of order $n = 1, 2, 3, 4$ with mean and variance equal to 1.

4 CARMA processes

This section develops a theory for fast approximation of the cumulant-generating function of a CARMA process driven by a polylogarithmic Lévy process under some conditions. This can then, in turn, be used to do fast approximation of the corresponding characteristic function and moment-generating function. In [5] the question of the stationary distribution of a CARMA process driven by a compound Poisson process with exponential jumps is also treated, and the result we present here can be seen as an extension to [5, Section 4]. Furthermore, the relation in (4.1) plays a role in the calculation of covariances in [3] where it helps speed up the computation time.

4.1 Distribution of CARMA processes driven by polylogarithmic Lévy processes

The following theorem gives an expression for the cumulant-generating function of a subclass of CARMA processes driven by a polylogarithmic Lévy process.

Theorem 4.1. *Let $(Y(t))_{t\in\mathbb{R}}$ be the CARMA process in* (2.3) *and $(Y_s(t))_{t\geq s}$ be given by* (2.4) *where $(L(t))_{t\in\mathbb{R}}$ is a polylogarithmic Lévy process of order $n\in\mathbb{N}_0$ and with parameters $\alpha>0$ and $\beta>0$. Let $g(t)=\sum_{i=1}^p a_i e^{-b_i t}$ be the kernel function in* (2.3) *and* (2.4)*. Then*

$$\begin{aligned}
\psi_{Y(0)}(z) &= \log\mathbb{E}[\exp\{zY(0)\}]\\
&= -\alpha\,\mathrm{Li}_{n+1}\left(\frac{z}{\beta}\right)\frac{1}{g'(0)} - \alpha\int_0^\infty \mathrm{Li}_{n+1}\left(\frac{zg(s)}{\beta}\right)\left(\frac{g(s)}{g'(s)}\right)'\mathrm{d}s\\
&= -\alpha\,\mathrm{Li}_{n+1}\left(\frac{z}{\beta}\right)\frac{1}{g'(0)} + \alpha\,\mathrm{Li}_{n+2}\left(\frac{zg(0)}{\beta}\right)\frac{1}{g'(0)}\left(\frac{1}{g'(0)}\right)'\\
&\quad+\alpha\int_0^\infty \mathrm{Li}_{n+2}\left(\frac{zg(s)}{\beta}\right)\left(\frac{g(s)}{g'(s)}\left(\frac{g(s)}{g'(s)}\right)'\right)'\mathrm{d}s
\end{aligned}\tag{4.1}$$

and

$$\begin{aligned}
\psi_{Y_s(t)}(z) &= \log\mathbb{E}[\exp\{zY_s(t)\}]\\
&= \alpha\,\mathrm{Li}_{n+1}\left(\frac{zg(t-s)}{\beta}\right)\frac{g(t-s)}{g'(t-s)} - \alpha\,\mathrm{Li}_{n+1}\left(\frac{z}{\beta}\right)\frac{1}{g'(0)}\\
&\quad-\alpha\int_0^{t-s}\mathrm{Li}_{n+1}\left(\frac{zg(u)}{\beta}\right)\left(\frac{g(u)}{g'(u)}\right)'\mathrm{d}u\\
&= \alpha\,\mathrm{Li}_{n+1}\left(\frac{zg(t-s)}{\beta}\right)\frac{g(t-s)}{g'(t-s)} - \alpha\,\mathrm{Li}_{n+1}\left(\frac{z}{\beta}\right)\frac{1}{g'(0)}\\
&\quad-\alpha\,\mathrm{Li}_{n+2}\left(\frac{zg(t-s)}{\beta}\right)\frac{g(t-s)}{g'(t-s)}\left(\frac{g(t-s)}{g'(t-s)}\right)' + \alpha\,\mathrm{Li}_{n+2}\left(\frac{zg(0)}{\beta}\right)\frac{1}{g'(0)}\left(\frac{1}{g'(0)}\right)'\\
&\quad+\alpha\int_0^{t-s}\mathrm{Li}_{n+2}\left(\frac{zg(u)}{\beta}\right)\left(\frac{g(u)}{g'(u)}\left(\frac{g(u)}{g'(u)}\right)'\right)'\mathrm{d}u,
\end{aligned}\tag{4.2}$$

for $z\in\mathbb{C}\setminus([\beta,\infty)\times i\mathbb{R})$, where

$$\left|\left(\frac{g(u)}{g'(u)}\right)'\right| \leq \frac{\max_{i,j}(b_i-b_j)^2}{2\min_i b_i^2}\tag{4.3}$$

and

$$\left|\left(\frac{g(u)}{g'(u)}\left(\frac{g(u)}{g'(u)}\right)'\right)'\right| \leq \frac{\max_{i,j}(b_i-b_j)^4}{4\min_i b_i^4} + \frac{\max_{j,i>j,k}\left|(b_i-b_j)^2\left(b_k-\frac{b_ib_j}{b_k}\right)\right|}{2\min_i |b_i|^3}\tag{4.4}$$

for all $u\in\mathbb{R}$.

Proof. We start by proving (4.2). First, note that

$$\frac{\mathrm{d}}{\mathrm{d}u}\operatorname{Li}_{n+1}\left(\frac{zg(u)}{\beta}\right) = \operatorname{Li}_n\left(\frac{zg(u)}{\beta}\right)\frac{g'(u)}{g(u)}$$

for any $n \in \mathbb{N}_0$, where we have used (3.3). Note $g(0) = 1$ since $\sum_{i=1}^d a_i = 1$. Then, using integration by parts twice,

$$\begin{aligned}
\log \mathbb{E}[\exp\{zY_s(t)\}] &= \alpha \int_0^{t-s} \operatorname{Li}_n\left(\frac{zg(u)}{\beta}\right)\mathrm{d}u \\
&= \alpha \int_0^{t-s} \frac{\mathrm{d}}{\mathrm{d}u}\operatorname{Li}_{n+1}\left(\frac{zg(u)}{\beta}\right)\frac{g(u)}{g'(u)}\,\mathrm{d}u \\
&= \alpha \operatorname{Li}_{n+1}\left(\frac{zg(t-s)}{\beta}\right)\frac{g(t-s)}{g'(t-s)} - \alpha \operatorname{Li}_{n+1}\left(\frac{z}{\beta}\right)\frac{1}{g'(0)} \\
&\quad - \alpha \int_0^{t-s} \operatorname{Li}_{n+1}\left(\frac{zg(u)}{\beta}\right)\left(\frac{g(u)}{g'(u)}\right)'\mathrm{d}u \\
&= \alpha \operatorname{Li}_{n+1}\left(\frac{zg(t-s)}{\beta}\right)\frac{g(t-s)}{g'(t-s)} - \alpha \operatorname{Li}_{n+1}\left(\frac{z}{\beta}\right)\frac{1}{g'(0)} \\
&\quad - \alpha \operatorname{Li}_{n+2}\left(\frac{zg(t-s)}{\beta}\right)\frac{g(t-s)}{g'(t-s)}\left(\frac{g(t-s)}{g'(t-s)}\right)' + \alpha \operatorname{Li}_{n+2}\left(\frac{z}{\beta}\right)\frac{1}{g'(0)}\left(\frac{1}{g'(0)}\right)' \\
&\quad + \alpha \int_0^{t-s} \operatorname{Li}_{n+2}\left(\frac{zg(u)}{\beta}\right)\left(\frac{g(u)}{g'(u)}\left(\frac{g(u)}{g'(u)}\right)'\right)'\mathrm{d}u
\end{aligned} \tag{4.5}$$

We now prove (4.3). Initially, we have

$$\left(\frac{g(u)}{g'(u)}\right)' = \frac{(g'(u))^2 - g(u)g''(u)}{(g'(u))^2}. \tag{4.6}$$

We find

$$\begin{aligned}
(g'(u))^2 - g(u)g''(u) &= \left(\sum_{i=1}^d a_i b_i e^{-b_i u}\right)^2 - \left(\sum_{i=1}^d a_i e^{-b_i u}\right)\left(\sum_{i=1}^d a_i b_i^2 e^{-b_i u}\right) \\
&= -\sum_{j=1}^{d-1}\sum_{i=j+1}^{d} a_i a_j (b_i - b_j)^2 e^{-(b_i+b_j)u}
\end{aligned} \tag{4.7}$$

and therefore,

$$\begin{aligned}
\left|\left(\frac{g(u)}{g'(u)}\right)'\right| &\le \frac{\max_{i,j}(b_i-b_j)^2}{\min_i b_i^2}\frac{\sum_{j=1}^{d-1}\sum_{i=j+1}^{d} a_i a_j e^{-(b_i+b_j)t}}{\left(\sum_{i=1}^d a_i e^{-b_i t}\right)^2} \\
&= \frac{\max_{i,j}(b_i-b_j)^2}{\min_i b_i^2}\frac{\sum_{j=1}^{d-1}\sum_{i=j+1}^{d} a_i a_j e^{-(b_i+b_j)t}}{\sum_{i=1}^d a_i^2 e^{-b_i t} + 2\sum_{j=1}^{d-1}\sum_{i=j+1}^{d} a_i a_j e^{-(b_i+b_j)t}} \\
&\le \frac{\max_{i,j}(b_i-b_j)^2}{2\min_i b_i^2}.
\end{aligned}$$

Next we consider (4.4). Let $f(u) = g(u)/g'(u)$. Then

$$\begin{aligned}\left|\Big(f(u)(f(u))'\Big)'\right| &\leq |(f'(u))^2| + |f(u)f''(u)| \\ &\leq \frac{\max_{i,j}(b_i - b_j)^4}{4\min_i b_i^4} + \frac{1}{\min_i |b_i|}|f''(u)|\end{aligned}$$

where we have used (4.3). Next, note

$$f''(u) = \frac{g''(u)((g'(u))^2 - g(u)g''(u)) - g(u)((g''(u))^2 - g'(u)g'''(u))}{(g'(u))^3}.$$

From (4.7) it follows that

$$\begin{aligned}&|g''(u)((g'(u))^2 - g(u)g''(u)) - g(u)((g''(u))^2 - g'(u)g'''(u))| \\ &\quad= \left|\sum_{j=1}^{d-1}\sum_{i=j+1}^{d} a_i a_j e^{-(b_i+b_j)u} \sum_{k=1}^{d} a_k e^{-b_k u}(b_i - b_j)^2(b_k^2 - b_i b_j)\right| \\ &\quad\leq \max_{j,i>j,k}\left|(b_i - b_j)^2\left(b_k - \frac{b_i b_j}{b_k}\right)\right|\left|\sum_{j=1}^{d-1}\sum_{i=j+1}^{d} a_i a_j e^{-(b_i+b_j)u}\right|\left|\sum_{k=1}^{d} a_k b_k e^{-b_k u}\right|\end{aligned}$$

The bound now follows since

$$\begin{aligned}|g'(u)^3| &= \left|\sum_{i=1}^{d} a_i^2 b_i^2 e^{-b_i u} + 2\sum_{j=1}^{d-1}\sum_{i=j+1}^{d} a_i a_j b_i b_j e^{-(b_i+b_j)u}\right|\left|\sum_{k=1}^{d} a_k b_k e^{-b_k u}\right| \\ &\geq 2\min_i b_i^2\left|\sum_{j=1}^{d-1}\sum_{i=j+1}^{d} a_i a_j e^{-(b_i+b_j)u}\right|\left|\sum_{k=1}^{d} a_k b_k e^{-b_k u}\right|.\end{aligned}$$

Finally, to prove (4.1) we initially note that

$$\log \mathbb{E}[\exp\{zY(0)\}] = \alpha\int_0^\infty \mathrm{Li}_n\left(\frac{zg(u)}{\beta}\right)\mathrm{d}u = \lim_{s\to-\infty}\alpha\int_0^{t-s}\mathrm{Li}_n\left(\frac{zg(u)}{\beta}\right)\mathrm{d}u \tag{4.8}$$

Now (4.3) and (4.4), and that $\mathrm{Li}_n(z) = O(z)$ as $|z| \to 0$, which follows from (3.1), justifies letting $s \to -\infty$ in (4.5) to conclude (4.1). □

4.2 Comments on approximation

Considering Theorem 4.1 and the proof there are three approaches to calculate $\psi_{Y(0)}$ or $\psi_{Y_s(t)}$. We focus on $\psi_{Y(0)}$. We have

$$\begin{aligned}\psi_{Y(0)}(z) &= \alpha\int_0^\infty \mathrm{Li}_n\left(\frac{zg(u)}{\beta}\right)\mathrm{d}u && (4.9)\\ &= -\alpha\,\mathrm{Li}_{n+1}\left(\frac{z}{\beta}\right)\frac{1}{g'(0)} - \alpha\int_0^\infty \mathrm{Li}_{n+1}\left(\frac{zg(u)}{\beta}\right)\left(\frac{g(u)}{g'(u)}\right)'\mathrm{d}u && (4.10)\\ &= -\alpha\,\mathrm{Li}_{n+1}\left(\frac{z}{\beta}\right)\frac{1}{g'(0)} + \alpha\,\mathrm{Li}_{n+2}\left(\frac{z}{\beta}\right)\frac{1}{g'(0)}\left(\frac{1}{g'(0)}\right)' \\ &\quad+ \alpha\int_0^\infty \mathrm{Li}_{n+2}\left(\frac{zg(u)}{\beta}\right)\left(\frac{g(u)}{g'(u)}\left(\frac{g(u)}{g'(u)}\right)'\right)'\mathrm{d}u && (4.11)\end{aligned}$$

From (4.3) and (4.4) we conclude the integrals in (4.10) and (4.11) are small when the b_i's are close to each other. When the integrals are small, a more coarse approximation of them is possible for the same level of accuracy. Thus, the integrals in (4.10) and (4.11) can be approximated fast when the b_i's are close. On the other hand, the polylogarithm in the kernels of (4.9), (4.10), and (4.11) have different order. This makes a direct comparison more difficult but we have $\mathrm{Li}_n(z) = z + O(|z|^2)$ as $|z| \to 0$ for any $n \in \mathbb{N}_0$ (by (3.1)), so the polylogarithms are close, at least for small arguments.

We can also argue that (4.9) is larger than the integral in (4.10) for any real and positive z, that is,

$$\left|\int_0^\infty \mathrm{Li}_{n+1}\left(\frac{xg(u)}{\beta}\right)\left(\frac{g(u)}{g'(u)}\right)' \mathrm{d}u\right| < \left|\int_0^\infty \mathrm{Li}_n\left(\frac{xg(u)}{\beta}\right)\mathrm{d}u\right|$$

for $x \in (0,\beta)$. To see this, first note that $g'(0) < 0$. By (4.6) and (4.7), we also have $\left(\frac{g(u)}{g'(u)}\right)' < 0$. Finally, from (3.1), it is apparent that the polylogarithm is positive for positive input, and since $g(u) \geq 0$ we therefore have that both terms in (4.10) are positive from which the claim follows.

With the parameters estimated in [3] (where b_i should be considered as the i'th mean reversion λ_i) the bound in (4.3) is 0.0695 and the bound in (4.4) is 0.0307 which makes the integrals in (4.1) close to negligible. Theorem 4.1 is therefore suitable for decreasing the compute time in the situation considered in [3].

5 Algorithms for fast calculation of the polylogarithm

To apply the results of this paper it is crucial to be able to calculate the polylogarithm fast. We will only consider the polylogarithm of order 2 and 3 here. Polylogarithms of order 1 or a smaller integer can be found analytically as argued in Section 3.1. Polylogarithms of order larger than 3 are increasingly difficult to study, and the literature becomes more sparse on this topic.

The polylogarithm of order 2 satisfies the relations (see [10]):

$$\mathrm{Li}_2(z) = -\mathrm{Li}_2\left(\frac{z}{z-1}\right) - \tfrac{1}{2}\log^2(1-z) \tag{5.1}$$

$$\mathrm{Li}_2(z) = \mathrm{Li}_2\left(\frac{1}{z-1}\right) - \tfrac{1}{6}\pi^2 - \log(-z)\log(1-z) + \tfrac{1}{2}\log^2(1-z). \tag{5.2}$$

for $z \in \mathbb{C}$ with $\mathrm{Re}(z) \leq 0$ and $z \neq 0$. These two equations together with (3.1) makes it possible to make fast numerical calculations of the function $(-\infty, 0] \times i\mathbb{R} \ni z \mapsto \mathrm{Li}_2(z)$ which is useful for calculating the characteristic function and the moment-generating function for non-positive argument of the polylogarithmic distribution of second order. In particular, note that the convergence of (3.1) is fastest for $|z|$ small. Thus, when $|z| < 1$ we use (5.1) to calculate $\mathrm{Li}_2(z)$ and when $|z| \geq 1$ we use (5.2). Furthermore, $\mathrm{Li}_2(z)$ may also be calculated for $z \in [0,1)$ using (3.1) directly however, as z approaches 1, an increasing number of terms in the sum in (3.1) is required to reach a reasonable level of accuracy. This can hurt the computational speed.

As an example we consider 10,000 evaluations of $x \mapsto \mathrm{Li}_2(ix)$ $(x = 1, \ldots, 1000)$, which takes approximately 0.03 seconds using (5.1) and (5.2) where the sum in (3.1) is cut at $k = 30$ giving a maximal error of $3 \cdot 10^{-8}$.[1]

[1] implemented in MATLAB R2018b on a standard laptop.

The polylogarithm of order 3 satisfies (see [9, Section 5.2])

$$\mathrm{Li}_3(z) = \mathrm{Li}_3(1/z) - \frac{\pi^2}{6}\log(-z) - \frac{1}{6}\log^3(-z)$$

for $z \in \mathbb{C}$ with $\mathrm{Re}(z) \leq 0$ and $z \neq 0$. Again, the same 10,000 evaluations as above have a computational time of approximately 0.03 seconds.

References

[1] Barndorff-Nielsen, O.E. and N. Shephard (2001). Non-Gaussian Ornstein–Uhlenbeck-based models and some of their uses in financial economics. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 63(2), 167–241.

[2] Basse-O'Connor, A., M.S. Nielsen, J. Pedersen and V. Rohde (2018). Multivariate stochastic delay differential equations and CAR representations of CARMA processes. *Stochastic Process. Appl.* Forthcoming. DOI: 10.1016/j.spa.2018.11.011.

[3] Benth, F.E., T.S. Christensen and V. Rohde (2019). Multivariate continuous-time modeling of wind indexes and hedging of wind risk.

[4] Benth, F.E. and A. Pircalabu (2018). A non-Gaussian Ornstein-Uhlenbeck model for pricing wind power futures. *Applied Mathematical Finance* 25(1), 36–65.

[5] Benth, F.E. and V. Rohde (2019). On non-negative modeling with CARMA processes. *Journal of Mathematical Analysis and Applications* 476(1), 196–214.

[6] Brockwell, P.J. (2014). Recent results in the theory and applications of CARMA processes. *Ann. Inst. Statist. Math.* 66(4), 647–685. DOI: 10.1007/s10463-014-0468-7.

[7] Brockwell, P.J., R.A. Davis and Y. Yang (2011). Estimation for non-negative Lévy-driven CARMA processes. *J. Bus. Econom. Statist.* 29(2), 250–259. DOI: 10.1198/jbes.2010.08165.

[8] Carr, P. and D. Madan (1999). Option valuation using the fast Fourier transform. *Journal of Computational Finance* 2(4), 61–73.

[9] Kölbig, K.S. (1986). Nielsen's generalized polylogarithms. *SIAM Journal on Mathematical Analysis* 17(5), 1232–1258.

[10] Maximon, L.C. (2003). The dilogarithm function for complex argument. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences* 459(2039), 2807–2819.

[11] Rajput, B.S. and J. Rosiński (1989). Spectral representations of infinitely divisible processes. *Probab. Theory Related Fields* 82(3), 451–487.

[12] Rocha-Arteaga, A. and K.-i. Sato (2001). Topics in infinitely divisible distributions and Lévy processes. *Communicaciónes del CIMAT, Guanajuato*.

[13] Sato, K. (1999). *Lévy Processes and Infinitely Divisible Distributions*. Vol. 68. Cambridge Studies in Advanced Mathematics. Translated from the 1990 Japanese original, Revised by the author. Cambridge University Press.

Paper C

Stochastic Delay Differential Equations And Related Autoregressive Models

Andreas Basse-O'Connor, Mikkel Slot Nielsen, Jan Pedersen, and Victor Rohde

Abstract

In this paper we suggest two continuous-time models which exhibit an autoregressive structure. We obtain existence and uniqueness results and study the structure of the solution processes. One of the models, which corresponds to general stochastic delay differential equations, will be given particular attention. We use the obtained results to link the introduced processes to both discrete-time and continuous-time ARMA processes.

MSC 2010: 60G10, 60G22, 60H10, 60H20

Keywords: autoregressive structures, stochastic delay differential equations, processes of Ornstein-Uhlenbeck type, long-range dependence, CARMA processes, moving averages.

1 Introduction

Let $(L_t)_{t\in\mathbb{R}}$ be a two-sided Lévy process and $\psi:\mathbb{R}\to\mathbb{R}$ some measurable function which is integrable with respect to $(L_t)_{t\in\mathbb{R}}$ (in the sense of [23]). Processes of the form

$$X_t = \int_{\mathbb{R}} \psi(t-u)\,dL_u, \quad t\in\mathbb{R}, \tag{1.1}$$

are known as (stationary) continuous-time moving averages and have been studied extensively. Their popularity may be explained by the Wold-Karhunen decomposition: up to a drift term, essentially any stationary and square integrable process admits a representation of the form (1.1) with $(L_t)_{t\in\mathbb{R}}$ replaced by a process with second order stationary and orthogonal increments. For details on this type of representations, see

[28, Section 26.2] and [2, Theorem 4.1]. Note that the model (1.1) nests the discrete-time moving average with filter $(\psi_j)_{j\in\mathbb{Z}}$ (at least when it is driven by an infinitely divisible noise), since one can choose $\psi(t) = \sum_{j\in\mathbb{Z}} \psi_j \mathbb{1}_{(j-1,j]}(t)$. Another example of (1.1) is the Ornstein-Uhlenbeck process corresponding to $\psi(t) = e^{-\lambda t}\mathbb{1}_{[0,\infty)}(t)$ for $\lambda > 0$. Ornstein-Uhlenbeck processes often serve as building blocks in stochastic modeling, e.g. in stochastic volatility models for option pricing as illustrated in [4] or in models for (log) spot price of many different commodities, e.g., as in [26]. A generalization of the Ornstein-Uhlenbeck process, which is also of the form (1.1), is the CARMA process. To be concrete, for two real polynomials P and Q, of degree p and q ($p > q$) respectively, with no zeroes on $\{z \in \mathbb{C} : \mathrm{Re}(z) = 0\}$, choosing $\psi : \mathbb{R} \to \mathbb{R}$ to be the function characterized by

$$\int_{\mathbb{R}} e^{-ity}\psi(t)\,dt = \frac{Q(iy)}{P(iy)}, \quad y \in \mathbb{R},$$

results in a CARMA process. CARMA processes have found many applications, and extensions to account for long memory and to a multivariate setting have been made. For more on CARMA processes and their extensions, see [9, 10, 14, 19, 27]. Many general properties of continuous-time moving averages are well understood. This includes when they have long memory and have sample paths of finite variation (or, more generally, are semimartingales). For an extensive treatment of these processes and further examples we refer to [5, 6] and [3], respectively.

Instead of specifying the kernel ψ in (1.1) directly it is often preferred to view $(X_t)_{t\in\mathbb{R}}$ as a solution to a certain equation. For instance, as an alternative to (1.1), the Ornstein-Uhlenbeck process with parameter $\lambda > 0$, respectively the discrete-time moving average with filter $\psi_j = \alpha^j \mathbb{1}_{j\geq 1}$ for some $\alpha \in \mathbb{R}$ with $|\alpha| < 1$, may be characterized as the unique stationary process that satisfies

$$dX_t = -\lambda X_t\,dt + dL_t, \quad t \in \mathbb{R}, \tag{1.2}$$

respectively,

$$X_t = \alpha X_{t-1} + L_t - L_{t-1}, \quad t \in \mathbb{R}. \tag{1.3}$$

The representations (1.2)-(1.3) are useful in many aspects, e.g., in the understanding of the evolution of the process over time, to study properties of $(L_t)_{t\in\mathbb{R}}$ through observations of $(X_t)_{t\in\mathbb{R}}$ or to compute prediction formulas (which, enventually, may be used to estimate the models). Therefore, we aim at generalizing equations (1.2)-(1.3) in a suitable way and studying the corresponding solutions. Through this study we will argue that these generalizations lead to a wide class of stationary processes, which enjoy many of the same properties as the solutions to (1.2)-(1.3).

The two models of interest: Let η and ϕ be finite signed measures concentrated on $[0,\infty)$ and $(0,\infty)$, respectively, and let $\theta : \mathbb{R} \to \mathbb{R}$ be some measurable function (typically chosen to have a particularly simple structure) which is integrable with respect to $(L_t)_{t\in\mathbb{R}}$. Moreover, suppose that $(Z_t)_{t\in\mathbb{R}}$ is a measurable and integrable process with stationary increments. The equations of interest are

$$dX_t = \Big(\int_{[0,\infty)} X_{t-u}\,\eta(du)\Big)dt + dZ_t, \quad t \in \mathbb{R}, \tag{1.4}$$

and

$$X_t = \int_0^\infty X_{t-u}\,\phi(du) + \int_{-\infty}^t \theta(t-u)\,dL_u, \quad t \in \mathbb{R}. \tag{1.5}$$

We see that (1.2) is a special case of (1.4) with $\eta = -\lambda\delta_0$ and $Z_t = L_t$, and (1.3) is a special case of (1.5) with $\phi = \alpha\delta_1$ and $\theta = \mathbb{1}_{(0,1]}$. Here δ_c refers to the Dirac measure at $c \in \mathbb{R}$. Equation (1.4) is known in the literature as a stochastic delay differential equation (SDDE), and existence and (distributional) uniqueness results have been obtained when η is compactly supported and $(Z_t)_{t\in\mathbb{R}}$ is a Lévy process (see [13, 16]). As indicated above, models of the type (1.4) are useful for recovering the increments of $(Z_t)_{t\in\mathbb{R}}$ as well as prediction and estimation. We refer to [7, 17, 21] for details.

Another generalization of the noise term is given in [24]. Other parametrizations of ϕ in (1.5) that we will study in Examples 3.4 and 3.6 are

$$\phi(du) = \alpha e^{-\beta u}\mathbb{1}_{[0,\infty)}(u)\,du$$

for $\alpha \in \mathbb{R}$ and $\beta > 0$ and

$$\phi = \sum_{j=1}^{p} \phi_j \delta_j$$

for $\phi_j \in \mathbb{R}$. As far as we know, equations of the type (1.5) have not been studied before. We will refer to (1.5) as a level model, since it specifies X_t directly (rather than its increments, $X_t - X_s$). Although the level model may seem odd at first glance as the noise term is forced to be stationary, one of its strengths is that it can be used as a model for the increments of a stationary increment process. We present this idea in Example 3.5 where a stationary increment solution to (1.4) is found when no stationary solution exists.

Our main results: In Section 2 we prove existence and uniqueness in the model (1.4) under the assumptions that

$$\int_{[0,\infty)} u^2\,|\eta|(du) < \infty \quad \text{and} \quad iy - \int_{[0,\infty)} e^{-iuy}\,\eta(du) \neq 0$$

for all $y \in \mathbb{R}$ ($|\eta|$ being the variation of η). In relation to this result we provide several examples of choices of η and $(Z_t)_{t\in\mathbb{R}}$. Among other things, we show that long memory in the sense of a hyperbolically decaying autocovariance function can be incorporated through the noise process $(Z_t)_{t\in\mathbb{R}}$, and we indicate how invertible CARMA processes can be viewed as solutions to SDDEs. Moreover, in Corollary 2.6 it is observed that as long as $(Z_t)_{t\in\mathbb{R}}$ is of the form

$$Z_t = \int_{\mathbb{R}} \big[\theta(t-u) - \theta_0(-u)\big]\,dL_u, \quad t \in \mathbb{R},$$

for suitable kernels $\theta, \theta_0 : \mathbb{R} \to \mathbb{R}$, the solution to (1.4) is a moving average of the type (1.1). On the other hand, Example 2.14 provides an example of $(Z_t)_{t\in\mathbb{R}}$ where the solution is not of the form (1.1). Next, in Section 3, we briefly discuss existence and uniqueness of solutions to (1.5) and provide a few examples. Section 4 contains some technical results together with proofs of all the presented results.

Our proofs rely heavily on the theory of Fourier (and, more generally, bilateral Laplace) transforms, in particular it concerns functions belonging to certain Hardy spaces (or to slight modifications of such). Specific types of Musielak-Orlicz spaces will also play an important role in order to show our results.

Definitions and conventions: For $p \in (0,\infty]$ and a (non-negative) measure μ on the Borel σ-field $\mathscr{B}(\mathbb{R})$ on $\mathbb{R}$ we denote by $L^p(\mu)$ the usual L^p space relative to μ. If μ is the Lebesgue measure, we will suppress the dependence on the measure and write $f \in L^p$. By a finite signed measure we refer to a set function $\mu : \mathscr{B}(\mathbb{R}) \to \mathbb{R}$ of the form $\mu = \mu^+ - \mu^-$, where μ^+ and μ^- are two finite measures which are mutually singular. Integration of a function $f : \mathbb{R} \to \mathbb{R}$ is defined in an obvious way whenever $f \in L^1(|\mu|)$, where $|\mu| := \mu^+ + \mu^-$. For any given finite signed measure μ set and $z \in \mathbb{C}$ such that

$$\int_{\mathbb{R}} e^{\mathrm{Re}(z)u}\, |\mu|(du) < \infty,$$

we define the bilateral Laplace transform $\mathcal{L}[\mu](z)$ of μ at z by

$$\mathcal{L}[\mu](z) = \int_{\mathbb{R}} e^{zu}\, \mu(du).$$

In particular, the Fourier transform $\mathcal{F}[\mu](y) := \mathcal{L}[\mu](iy)$ is well-defined for all $y \in \mathbb{R}$. (Note that the Laplace and Fourier transforms are often defined with a minus in the exponent; we have chosen this alternative definition so that $\mathcal{F}[\mu]$ coincides with the traditional definition of the characteristic function.) If $f \in L^1$ we define $\mathcal{L}[f] := \mathcal{L}[f(u)\,du]$. We note that $\mathcal{F}[f] \in L^2$ when $f \in L^1 \cap L^2$ and that $\mathcal{F}$ can be extended to an isometric isomorphism from L^2 onto L^2 by Plancherel's Theorem.

For two finite signed measures μ and ν we define the convolution $\mu * \nu$ as

$$\mu * \nu(B) = \int_{\mathbb{R}} \int_{\mathbb{R}} \mathbb{1}_B(u+v)\, \mu(du)\, \nu(dv)$$

for any Borel set B. Moreover, if $f : \mathbb{R} :\to \mathbb{R}$ is a measurable function such that $f(t - \cdot) \in L^1(|\mu|)$ we define the convolution $f * \mu(t)$ at $t \in \mathbb{R}$ by

$$f * \mu(t) = \int_{\mathbb{R}} f(t-u)\, \mu(du).$$

Recall also that a process $(L_t)_{t\in\mathbb{R}}$, $L_0 = 0$, is called a (two-sided) Lévy process if it has stationary and independent increments and càdlàg sample paths (for details, see [25]). Let $(L_t)_{t\in\mathbb{R}}$ be a centered Lévy process with Gaussian component σ^2 and Lévy measure ν. Then, for any measurable function $f : \mathbb{R} \to \mathbb{R}$ satisfying

$$\int_{\mathbb{R}} \Big(f(u)^2 \sigma^2 + \int_{\mathbb{R}} \big(|xf(u)|^2 \wedge |xf(u)| \big)\, \nu(dx) \Big)\, du < \infty, \tag{1.6}$$

the integral of f with respect to $(L_t)_{t\in\mathbb{R}}$ is well-defined and belongs to $L^1(\mathbb{P})$ (see [23, Theorem 3.3]).

2 The SDDE setup

Recall that, for a given finite signed measure η on $[0,\infty)$ and a measurable process $(Z_t)_{t\in\mathbb{R}}$ with stationary increments and $\mathbb{E}[|Z_t|] < \infty$ for all t, we are interested in the existence and uniqueness of a measurable and stationary process $(X_t)_{t\in\mathbb{R}}$ with $\mathbb{E}[|X_0|] < \infty$ which satisfies

$$X_t - X_s = \int_s^t \int_{[0,\infty)} X_{u-v}\,\eta(dv)\,du + Z_t - Z_s \tag{2.1}$$

almost surely for each $s < t$.

Remark 2.1. In the literature, (2.1) is often solved on $[0,\infty)$ given an initial condition $(X_u)_{u\leq 0}$. However, since we will be interested in (possibly) non-causal solutions, it turns out to be convenient to solve (2.1) on $\mathbb{R}$ with no initial condition (see [12, p. 46 and Section 3.2] for details).

In line with [13], we will construct a solution as a convolution of $(Z_t)_{t\in\mathbb{R}}$ and a deterministic kernel $x_0 : \mathbb{R} \to \mathbb{R}$ characterized through η. This kernel is known as the differential resolvent (of η) in the literature. Although many (if not all) of the statements of Lemma 2.2 concerning x_0 should be well-known, we have not been able to find a precise reference, and hence we have chosen to include a proof here. The core of Lemma 2.2 as well as further properties of differential resolvents can be found in [12, Section 3.3].

In the formulation we will say that η has n-th moment, $n \in \mathbb{N}$, if $v \mapsto v^n \in L^1(|\eta|)$ and that η has an exponential moment of order $\delta \geq 0$ if $v \mapsto e^{\delta v} \in L^1(|\eta|)$. Finally, we will make use of the function

$$h(z) := -z - \mathcal{L}[\eta](z), \tag{2.2}$$

which is always well-defined for $z \in \mathbb{C}$ with $\mathrm{Re}(z) \leq \delta$ if η admits an exponential moment of order $\delta \geq 0$.

Lemma 2.2. *Suppose that $h(iy) \neq 0$ for all $y \in \mathbb{R}$. Then there exists a unique function $x_0 : \mathbb{R} \to \mathbb{R}$, which meets $u \mapsto x_0(u)e^{cu} \in L^2$ for all $c \in [a,0]$ and a suitably chosen $a < 0$, and satisfies*

$$x_0(t) = \mathbb{1}_{[0,\infty)}(t) + \int_{-\infty}^t \int_{[0,\infty)} x_0(u-v)\,\eta(dv)\,du \tag{2.3}$$

for all $t \in \mathbb{R}$. Furthermore, x_0 is characterized by $\mathcal{L}[x_0](z) = 1/h(z)$ for $z \in \mathbb{C}$ with $Re(z) \in (a,0)$, and the following statements hold:

(i) *If η has n-th moment for some $n \in \mathbb{N}$, then $(u \mapsto x_0(u)u^n) \in L^2$. In particular, $x_0 \in L^q$ for all $q \in [1/n,\infty]$.*

(ii) *If η has an exponential moment of order $\delta > 0$, then there exists $\varepsilon \in (0,\delta]$ such that $u \mapsto x_0(u)e^{cu} \in L^2$ for all $c \in [a,\varepsilon]$ and, in particular, $x_0 \in L^q$ for all $q \in (0,\infty]$.*

(iii) *If $h(z) \neq 0$ for all $z \in \mathbb{C}$ with $Re(z) \leq 0$, then $x_0(t) = 0$ for all $t < 0$.*

By (2.3) it follows that x_0 induces a Lebesgue-Stieltjes measure μ_{x_0}. From Lemma 2.2 we deduce immediately the following properties of μ_{x_0}:

Corollary 2.3. *Suppose that $h(iy) \neq 0$ for all $y \in \mathbb{R}$. Then x_0 defines a Lebesgue-Stieltjes measure, and it is given by*

$$\mu_{x_0}(du) = \delta_0(du) + \Big(\int_{[0,\infty)} x_0(u-v)\,\eta(dv)\Big)du.$$

A function $\theta : \mathbb{R} \to \mathbb{R}$ is integrable with respect to μ_{x_0} if and only if

$$\int_{[0,\infty)} \int_{\mathbb{R}} |\theta(u+v)x_0(u)|\, du\, |\eta|(dv) < \infty. \tag{2.4}$$

Example 2.4. Let the setup be as in Corollary 2.3. We will here discuss a few implications of this result.

(i) Suppose that η has n-th moment for some $n \in \mathbb{N}$. By using the inequality $|u+v|^{n-1} \leq 2^{n-1}(|u|^{n-1} + |v|^{n-1})$ we establish that

$$\begin{aligned} &\frac{1}{2^{n-1}} \int_{[0,\infty)} \int_{\mathbb{R}} |(u+v)^{n-1} x_0(u)|\, du\, |\eta|(dv) \\ &\quad \leq |\eta|([0,\infty)) \int_{\mathbb{R}} |x_0(u) u^{n-1}|\, du + \int_{[0,\infty)} |v|^{n-1}\, |\eta|(dv) \int_{\mathbb{R}} |x_0(u)|\, du. \end{aligned} \tag{2.5}$$

The last term on the right-hand side of (2.5) is finite, since $x_0 \in L^1$ by Lemma 2.2((i)). The Cauchy-Schwarz inequality and the same lemma once again imply

$$\left(\int_{|u|>1} |x_0(u) u^{n-1}|\, du \right)^2 \leq \int_{|u|>1} \left(x_0(u) u^n \right)^2 du \int_{|u|>1} u^{-2}\, du < \infty.$$

Consequently, since $u \mapsto x_0(u)u^{n-1}$ is locally bounded, we deduce that $(u \mapsto x_0(u)u^{n-1}) \in L^1$ and that the first term on the right-hand side of (2.5) is also finite. It follows that (2.4) is satisfied for $\theta(u) = |u|^{n-1}$, so μ_{x_0} has moments up to order $n-1$.

(ii) Suppose that η has an exponential moment of order $\delta > 0$. Let γ be any number in $(0, \varepsilon)$, where $\varepsilon \in (0, \delta)$ is chosen as in Lemma 2.2((ii)). With this choice it is straightforward to check that $(u \mapsto x_0(u)e^{\gamma u}) \in L^1$, and hence

$$\int_{[0,\infty)} \int_{\mathbb{R}} e^{\gamma(u+v)} |x_0(u)|\, du\, |\eta|(dv) = \int_{[0,\infty)} e^{\gamma u}\, |\eta|(dv) \int_{\mathbb{R}} |x_0(u)| e^{\gamma u}\, du < \infty.$$

This shows that (2.4) holds with $\theta(u) = e^{\gamma u}$, so μ_{x_0} has as an exponential moment of order $\gamma > 0$.

(iii) Whenever η has first moment, x_0 is bounded (cf. Lemma 2.2((i))). Thus, under this assumption, a sufficient condition for (2.4) to hold is that $\theta \in L^1$.

With the differential resolvent in hand we present our main result of this section:

Theorem 2.5. *Let $(Z_t)_{t\in\mathbb{R}}$ be a measurable process which has stationary increments and satisfies $\mathbb{E}[|Z_t|] < \infty$ for all t. Suppose that η is a finite signed measure with second moment and $h(iy) \neq 0$ for all $y \in \mathbb{R}$. Then the process*

$$X_t = Z_t + \int_{\mathbb{R}} Z_{t-u} \int_{[0,\infty)} x_0(u-v)\, \eta(dv)\, du, \quad t \in \mathbb{R}, \tag{2.6}$$

is well-defined and the unique integrable stationary solution (up to modification) of equation (2.1). If $h(z) \neq 0$ for all $z \in \mathbb{C}$ with $Re(z) \leq 0$, $(X_t)_{t\in\mathbb{R}}$ admits the following causal representation:

$$X_t = \int_0^\infty [Z_{t-u} - Z_t] \int_{[0,\infty)} x_0(u-v)\, \eta(dv)\, du, \quad t \in \mathbb{R}. \tag{2.7}$$

Often, $(Z_t)_{t\in\mathbb{R}}$ is given by

$$Z_t = \int_{\mathbb{R}} \Big[\theta(t-u) - \theta(-u)\Big]\, dL_u, \quad t \in \mathbb{R}, \tag{2.8}$$

for some integrable Lévy process $(L_t)_{t\in\mathbb{R}}$ with $\mathbb{E}[L_1] = 0$ and measurable function $\theta : \mathbb{R} \to \mathbb{R}$ such that $u \mapsto \theta(t+u) - \theta(u)$ satisfies (1.6) for $t > 0$. The next result shows that the (unique) solution to (2.1) is a Lévy-driven moving average in this particular setup.

Corollary 2.6. *Let the setup be as in Theorem 2.5 and suppose that $(Z_t)_{t\in\mathbb{R}}$ is of the form* (2.8). *Then the unique integrable and stationary solution to* (2.1) *is given by*

$$X_t = \int_{\mathbb{R}} \theta * \mu_{x_0}(t-u)\, dL_u, \quad t \in \mathbb{R}. \tag{2.9}$$

In particular if $Z_t = L_t$ for $t \in \mathbb{R}$, we have that

$$X_t = \int_{\mathbb{R}} x_0(t-u)\, dL_u, \quad t \in \mathbb{R}.$$

Remark 2.7. Let the situation be as in Corollary 2.6 with $h(z) \neq 0$ whenever $\mathrm{Re}(z) \leq 0$. In this case we know from Theorem 2.5 that $(X_t)_{t\in\mathbb{R}}$ has the causal representation (2.7) with respect to $(Z_t)_{t\in\mathbb{R}}$. Now, if $(Z_t)_{t\in\mathbb{R}}$ is causal with respect to $(L_t)_{t\in\mathbb{R}}$ in the sense that $\theta(t) = 0$ for $t < 0$, $(X_t)_{t\in\mathbb{R}}$ admits the following causal representation with respect to $(L_t)_{t\in\mathbb{R}}$:

$$X_t = \int_{-\infty}^{t} \theta * \mu_{x_0}(t-u)\, dL_u, \quad t \in \mathbb{R}.$$

This follows from (2.9) and the fact that $\theta * \mu_{x_0}(t) = 0$ for $t < 0$ (using Lemma 2.2((iii))).

Remark 2.8. The assumption $h(0) = -\eta([0,\infty)) \neq 0$ is rather crucial in order to find stationary solutions. It may be seen as the analogue of assuming that the AR coefficients in a discrete-time ARMA setting do not sum to zero. For instance, the setup where $\eta \equiv 0$ will satisfy $h(iy) \neq 0$ for all $y \in \mathbb{R} \setminus \{0\}$, but if $(Z_t)_{t\in\mathbb{R}}$ is a Lévy process, the SDDE (2.1) cannot have stationary solutions. In Example 3.5, we show how one can find solutions with stationary increments for a reasonably large class of delay measures η with $\eta([0,\infty)) = 0$.

Remark 2.9. It should be stressed that for more restrictive choices of η, and in case $(Z_t)_{t\in\mathbb{R}}$ is a Lévy process, solutions sometimes exist even when $\mathbb{E}[|Z_1|] = \infty$. Indeed, if η is compactly supported and $\mathrm{Re}(z) \leq 0$ implies $h(z) \neq 0$, one only needs that $\mathbb{E}[\log^+ |Z_1|] < \infty$ to ensure that a stationary solution exists. We refer to [13, 24] for more details.

We now present some concrete examples of SDDEs. The first three examples concern the specification of the delay measure and the last two concern the specification of the noise.

Example 2.10. Let $\lambda \neq 0$ and consider the equation

$$X_t - X_s = -\lambda \int_s^t X_u\, du + Z_t - Z_s, \quad s < t. \tag{2.10}$$

In the setup of (2.1) this corresponds to $\eta = -\lambda\delta_0$. With h given by (2.2), we have $h(z) = \lambda - z \neq 0$ for every $z \in \mathbb{C}$ with $\mathrm{Re}(z) \neq \lambda$, and hence Theorem 2.5 implies that there exists a stationary process $(X_t)_{t\in\mathbb{R}}$ with $\mathbb{E}[|X_0|] < \infty$ satisfying (2.10). According to Lemma 2.2 the differential resolvent x_0 can be determined through its Laplace transform on $\{z \in \mathbb{C} : a < \mathrm{Re}(z) < 0\}$ for a suitable $a < 0$ as

$$\mathcal{L}[x_0](z) = \frac{1}{\lambda - z} = \begin{cases} \mathcal{L}\Big[\mathbb{1}_{[0,\infty)}e^{-\lambda\cdot}\Big](z) & \text{if} \quad \lambda > 0 \\ \mathcal{L}\Big[-\mathbb{1}_{(-\infty,0)}e^{-\lambda\cdot}\Big](z) & \text{if} \quad \lambda < 0. \end{cases}$$

Consequently, by Theorem 2.5,

$$X_t = \begin{cases} Z_t - \lambda e^{-\lambda t}\int_{-\infty}^t Z_u e^{\lambda u}\,du & \text{if} \quad \lambda > 0 \\ Z_t + \lambda e^{-\lambda t}\int_t^{\infty} Z_u e^{\lambda u}\,du & \text{if} \quad \lambda < 0. \end{cases} \tag{2.11}$$

Ornstein-Uhlenbeck processes satisfying (2.10) have already been studied in the literature, and representations of the stationary solution have been given, see e.g. [2, Theorem 2.1, Proposition 4.2].

Example 2.11. Let $(L_t)_{t\in\mathbb{R}}$ be a Lévy process with $\mathbb{E}[|L_1|] < \infty$. Recall that $(X_t)_{t\in\mathbb{R}}$ is said to be a CARMA$(2,1)$ process if

$$X_t = \int_{-\infty}^t g(t-u)\,dL_u, \quad t \in \mathbb{R},$$

where the kernel g is characterized by

$$\mathcal{F}[g](y) = \frac{b_0 - iy}{-y^2 - a_1 iy - a_2}, \quad y \in \mathbb{R},$$

for suitable $b_0, a_1, a_2 \in \mathbb{R}$, such that $z \mapsto z^2 + a_1 z + a_2$ has no roots on $\{z \in \mathbb{C} : \mathrm{Re}(z) \geq 0\}$. To relate the CARMA$(2,1)$ process to a solution to an SDDE we will suppose that the invertibility assumption $b_0 > 0$ is satisfied. In particular, $b_0 - iy \neq 0$ for all $y \in \mathbb{R}$ and, thus, we may write

$$\mathcal{F}[g](y) = \frac{1}{-iy + a_1 - b_0 + \frac{a_2 - b_0(a_1 - b_0)}{b_0 - iy}}, \quad y \in \mathbb{R}.$$

By choosing $\eta(dv) = (b_0 - a_1)\delta_0(dv) - (a_2 - b_0(a_1 - b_0))e^{-b_0 v}\mathbb{1}_{[0,\infty]}(v)\,dv$ (a finite signed measure with exponential moment of any order $\delta < b_0$) it is seen that the function h given in (2.2) satisfies $1/h(iy) = \mathcal{F}[g](y)$ for $y \in \mathbb{R}$. Consequently, we conclude from Theorem 2.5 that the CARMA$(2,1)$ process with parameter vector (b_0, a_1, a_2) is the unique solution to the SDDE (2.1) with delay measure η. In fact, any CARMA(p,q) process ($p, q \in \mathbb{N}_0$ and $p > q$) satisfying a suitable invertibility condition can be represented as the solution to an equation of the SDDE type. See [7, Theorem 4.8] for a precise statement.

Example 2.12. In this example we consider a delay measure η where the corresponding solution to the SDDE in (2.1) may be regarded as a CARMA process with fractional polynomials. Specifically, consider

$$\eta(dv) = \alpha_1\delta_0(dv) + \frac{\alpha_2}{\Gamma(\beta)}\mathbb{1}_{[0,\infty)}(v)v^{\beta-1}e^{-\gamma v}\,dv,$$

where $\beta,\gamma > 0$ and Γ is the gamma function. In this case, $h(z) = -z-\alpha_1-\alpha_2(\gamma-z)^{-\beta}$, and hence h is of the form $P_1(\gamma-\cdot)/P_2(\gamma-\cdot)$, where $P_i(z) = z^{a_i}+b_i z^{c_i}+d_i$ for suitable $a_i, c_i > 0$ and $b_i, d_i \in \mathbb{R}$. In this way, one may think of h as a ratio of fractional polynomials (recall from Example 2.11 that the solution to (2.1) will sometimes be a regular CARMA process when $\beta \in \mathbb{N}$). By Lemma 2.2 and Theorem 2.5 the associated SDDE has a unique solution with differential resolvent x_0 satisfying $x_0(t) = 0$ for $t < 0$, if

$$\mathrm{Re}(z) \leq 0 \quad \Longrightarrow \quad -z-\alpha_1-\alpha_2(\gamma-z)^{-\beta} \neq 0. \tag{2.12}$$

Each of the following two cases is sufficient for (2.12) to be satisfied:

(i) $\alpha_1 + |\alpha_2|\gamma^{-\beta} < 0$: In this case we have in particular that $\alpha_1 < 0$, so

$$|-z-\alpha_1-\alpha_2(\gamma-z)^{-\beta}| \geq -\alpha_1 - |\alpha_2||(\gamma-z)^{-\beta}| \geq -\alpha_1 - |\alpha_2|\gamma^{-\beta} > 0$$

whenever $\mathrm{Re}(z) \leq 0$.

(ii) $\alpha_1, \alpha_2 < 0$ and $\beta < 1$: In this case $\mathrm{Re}((\gamma-z)^{-\beta}) > 0$ and, thus, $\mathrm{Re}(-z-\alpha_1-\alpha_2(\gamma-z)^{-\beta}) > 0$ as long as $\mathrm{Re}(z) \leq 0$

Example 2.13. Let η be any finite signed measure with second moment, which satisfies $h(iy) \neq 0$ for all $y \in \mathbb{R}$. Consider the case where $(Z_t)_{t\in\mathbb{R}}$ is a fractional Lévy process, that is,

$$Z_t = \frac{1}{\Gamma(1+d)}\int_{\mathbb{R}} \Big[(t-u)_+^d - (-u)_+^d\Big]\, dL_u, \quad t \in \mathbb{R},$$

where $d \in (0,1/2)$ and $(L_t)_{t\in\mathbb{R}}$ is a centered and square integrable Lévy process. Let

$$\theta(t) = \frac{1}{\Gamma(1+d)} t_+^d, \quad t \in \mathbb{R}.$$

Then it follows by Corollary 2.6 that the solution to (2.1) takes the form

$$X_t = \int_{\mathbb{R}} \theta * \mu_{x_0}(t-u)\, dL_u, \quad t \in \mathbb{R}.$$

It is not too difficult to show that $\theta * \mu_{x_0}$ coincides with the left-sided Riemann-Liouville fractional integral of x_0, and hence $X_t = \int_{\mathbb{R}} x_0(t-u)\, dZ_u$, where the integral with respect to $(Z_t)_{t\in\mathbb{R}}$ is defined as in [18]. Consequently, we can use the proof of [18, Theorem 6.3] to deduce that $(X_t)_{t\in\mathbb{R}}$ has long memory in the sense that its autocovariance function is hyperbolically decaying at ∞:

$$\gamma_X(t) := \mathbb{E}[X_t X_0] \sim \frac{\Gamma(1-2d)}{\Gamma(d)\Gamma(1-d)} \frac{\mathbb{E}[L_1^2]}{h(0)^2} t^{2d-1}, \quad t \to \infty. \tag{2.13}$$

In particular, (2.13) shows that $\gamma_X \notin L^1$.

Our last example, presented below, deals with a situation where Theorem 2.5 is applicable, but $(Z_t)_{t\in\mathbb{R}}$ is not of the form (2.8). It is closely related to [2, Corollary 2.3].

Example 2.14. Consider a filtration $\mathscr{F} = (\mathcal{F}_t)_{t\in\mathbb{R}}$ and let $(B_t)_{t\in\mathbb{R}}$ be an $\mathscr{F}$-Brownian motion. Moreover, let $(\sigma_t)_{t\in\mathbb{R}}$ be a predictable process with $\sigma_0 \in L^2(\mathbb{P})$. Finally, assume that $(\sigma_t, B_t)_{t\in\mathbb{R}}$ and $(\sigma_{t+u}, B_{t+u} - B_u)_{t\in\mathbb{R}}$ have the same finite-dimensional marginal distributions for all $u \in \mathbb{R}$. In this case

$$Z_t = \int_0^t \sigma_s dB_s, \quad t \in \mathbb{R},$$

is well-defined, continuous and square integrable, and it has stationary increments. Here we use the convention $\int_0^t := -\int_t^0$ when $t < 0$. Under the assumptions that $h(z) \neq 0$ for all $z \in \mathbb{C}$ with $\mathrm{Re}(z) \leq 0$ and η has second moment, Theorem 2.5 implies that there exists a unique stationary solution $(X_t)_{t\in\mathbb{R}}$ to (2.1) and, since $x_0(t) = 0$ for $t < 0$, it is given by

$$\begin{aligned}
X_t &= Z_t + \int_0^\infty Z_{t-s} \int_{[0,\infty)} x_0(s-v)\,\eta(dv)\,ds \\
&= -\int_0^\infty \int_{t-s}^t \sigma_u\,dB_u \int_{[0,\infty)} x_0(s-v)\,\eta(dv)\,ds \\
&= -\int_{-\infty}^t \sigma_u \int_{t-u}^\infty \int_{[0,\infty)} x_0(s-v)\,\eta(dv)\,ds\,dB_u \\
&= \int_{-\infty}^t x_0(t-u)\sigma_u\,dB_u
\end{aligned}$$

for $t \in \mathbb{R}$, where we have used Corollary 2.3, (4.10) and an extension of the stochastic Fubini given in [22, Chapter IV, Theorem 65] to integrals over unbounded intervals.

3 The level model

In this section we consider the equation

$$X_t = \int_0^\infty X_{t-u}\,\phi(du) + \int_{-\infty}^t \theta(t-u)\,dL_u, \tag{3.1}$$

where ϕ is a finite signed measure on $(0,\infty)$, $(L_t)_{t\in\mathbb{R}}$ is an integrable Lévy process with $\mathbb{E}[L_1] = 0$ and $\theta : \mathbb{R} \to \mathbb{R}$ is a measurable function, which vanishes on $(-\infty, 0)$ and satisfies (1.6).

Remark 3.1. Due to the extreme flexibility of the model (3.1), one should require that ϕ and θ take a particular simple form. To elaborate, under the assumptions of Theorem 3.2 or Remark 3.3, a solution to (3.1) associated to the pair (ϕ, θ) is a causal moving average with kernel ψ. On the other hand, this solution could also have been obtained using the pair $(0, \psi)$. However, it might be that ϕ and θ have a simple form while ψ has not, and hence (3.1) should be used to obtain parsimonious representations of a wide range of processes. This idea is similar to that of the discrete-time ARMA processes, which could as well have been represented as an MA(∞) process or (under an invertibility assumption) an AR(∞) process.

Equation (3.1) can be solved using the backward recursion method under the contraction assumption $|\phi|((0,\infty)) < 1$, and this is how we obtain Theorem 3.2. For the noise term we will put the additional assumption that $\mathbb{E}[L_1^2] < \infty$, and hence (in view of (1.6)) that $\theta \in L^2$. In the formulation we will denote by ϕ^{*n} the n-fold convolution of ϕ, that is, $\phi^{*n} := \phi * \cdots * \phi$ for $n \in \mathbb{N}$ and $\phi^{*0} = \delta_0$.

Theorem 3.2. *Let $(L_t)_{t\in\mathbb{R}}$ be a Lévy process with $\mathbb{E}[L_1] = 0$ and $\mathbb{E}[L_1^2] < \infty$, and suppose that $\theta \in L^2$, and $|\phi|((0,\infty)) < 1$. Then there exists a unique square integrable solution to* (3.1). *It is given by*

$$X_t = \int_{-\infty}^{t} \psi(t-u)\,dL_u, \quad t \in \mathbb{R},$$

where $\psi := \sum_{n=0}^{\infty} \theta * \phi^{*n}$ exists as a limit in L^2 and vanishes on $(-\infty, 0)$.

Remark 3.3. One can ask for the existence of solutions to (3.1) under weaker conditions on ϕ than $|\phi|((0,\infty)) < 1$ (as imposed in Theorem 3.2). In particular, suppose still that $\mathbb{E}[L_1] = 0$, $\mathbb{E}[L_1^2] < \infty$ and $\theta \in L^2$, but instead of $|\phi|((0,\infty)) < 1$ suppose for some $a < 0$ that $\mathcal{L}[\phi](z) \neq 1$ whenever $\mathrm{Re}(z) \in (a,0)$ and

$$\sup_{a<x<0} \int_{\mathbb{R}} \left| \frac{\mathcal{L}[\theta](x+iy)}{1-\mathcal{L}[\phi](x+iy)} \right|^2 dy < \infty. \tag{3.2}$$

Under these assumptions one can find a function $\psi \in L^2$, such that $(u \mapsto e^{cu}\psi(u)) \in L^1$ for all $c \in (a,0)$ and

$$\mathcal{L}[\psi] = \frac{\mathcal{L}[\theta]}{1-\mathcal{L}[\phi]} \quad \text{on} \quad \{z \in \mathbb{C} : a < \mathrm{Re}(z) < 0\}. \tag{3.3}$$

This is shown in Lemma 4.1. With this choice of ψ it follows that $\mathcal{L}[\psi](z) = \mathcal{L}[\psi](z)\mathcal{L}[\phi](z) + \mathcal{L}[\theta](z)$, and hence

$$\begin{aligned}\mathcal{L}[\psi(t-\cdot)](-z) &= e^{-zt}\Big(\mathcal{L}[\psi](z)\mathcal{L}[\phi](z) + \mathcal{L}[\theta](z)\Big) \\ &= \mathcal{L}\left[\int_0^\infty \psi(t-u-\cdot)\,\phi(du) + \theta(t-\cdot)\right](-z)\end{aligned}$$

for each fixed $t \in \mathbb{R}$ and all $z \in \mathbb{C}$ with $\mathrm{Re}(z) \in (a,0)$. By uniqueness of Laplace transforms, this establishes that

$$\psi(t-r) = \int_0^\infty \psi(t-u-r)\,\phi(du) + \theta(t-r) \tag{3.4}$$

for Lebesgue almost all $r \in \mathbb{R}$ and each fixed $t \in \mathbb{R}$. By integrating both sides of (3.4) with respect to dL_r and using a stochastic Fubini result (e.g., [2, Theorem 3.1]) it follows that the moving average $X_t = \int_{\mathbb{R}} \psi(t-r)\,dL_r$, $t \in \mathbb{R}$, is a solution to (3.1).

To see that the conditions on ϕ imposed here are weaker than $|\phi|((0,\infty)) < 1$ as imposed in Theorem 3.2, observe that $\mathcal{L}[\phi](z) \neq 1$ whenever $\mathrm{Re}(z) \in (a,0)$ by the inequality $|\mathcal{L}[\phi](z)| \leq |\phi|((0,\infty))$, and

$$\sup_{a<x<0} \int_{\mathbb{R}} \left| \frac{\mathcal{L}[\theta](x+iy)}{1-\mathcal{L}[\phi](x+iy)} \right|^2 dy \leq \frac{2\pi}{(1-|\phi|((0,\infty)))^2} \int_0^\infty \theta(u)^2\,du. \tag{3.5}$$

In (3.5) we have made use of Plancherel's Theorem. Suppose that $|\phi|((0,\infty)) < 1$ so that Theorem 3.2 is applicable, let ψ be defined through (3.3) and set $\tilde{\psi} = \sum_{n=0}^{\infty} \theta * \phi^{*n}$. Then it follows by uniqueness of solutions to (3.1) and the isometry property of the integral map that

$$0 = \mathbb{E}\left[\left(\int_{\mathbb{R}} \psi(t-u)\,dL_u - \int_{\mathbb{R}} \tilde{\psi}(t-u)\,dL_u\right)^2\right] = \mathbb{E}[L_1^2] \int_{\mathbb{R}} (\psi(u) - \tilde{\psi}(u))^2\,du.$$

This shows that $\psi = \tilde{\psi}$ almost everywhere and that $\sum_{n=0}^{\infty} \theta * \phi^{*n}$ is an alternative characterization of ψ when $|\phi|((0,\infty)) < 1$. Another argument, which would not rely on the uniqueness of solutions to (3.1), would be to show that ψ and $\tilde{\psi}$ have the same Fourier transform.

Example 3.4. Suppose that $(L_t)_{t\in\mathbb{R}}$ is a Lévy process with $\mathbb{E}[L_1] = 0$ and $\mathbb{E}[L_1^2] < \infty$, and let $\theta \in L^2$. For $\alpha \in \mathbb{R}$ and $\beta > 0$, consider

$$\phi(du) = \alpha e^{-\beta u} \mathbb{1}_{[0,\infty)}(u)\, du$$

and define the measure

$$\xi(du) = e^{\alpha u}\phi(du) = \alpha e^{-(\beta-\alpha)u} \mathbb{1}_{[0,\infty)}(u)\, du.$$

We will argue that a solution to (3.1) exists as long as $\alpha/\beta < 1$ by considering the two cases (i) $-1 < \alpha/\beta < 1$ and (ii) $\alpha/\beta \leq -1$ separately.

(i) $-1 < \alpha/\beta < 1$: In this case $|\phi|((0,\infty)) = |\alpha|/\beta < 1$, and the existence of a solution is ensured by Theorem 3.2. To determine the solution kernel ψ, note that $\phi^{*n}(du) = \frac{\alpha^n}{(n-1)!} u^{n-1} e^{-\beta u} \mathbb{1}_{[0,\infty)}(u)\, du$ for $n \in \mathbb{N}$ and, thus,

$$\sum_{n=0}^{N} \theta * \phi^{*n}(t) = \theta(t) + \alpha \int_0^t \theta(t-u) e^{-\beta u} \sum_{n=0}^{N-1} \frac{(\alpha u)^n}{n!}\, du \to \theta(t) + \theta * \xi(t)$$

as $N \to \infty$ by Lebesgue's Theorem on Dominated Convergence. This shows that $\psi = \theta + \theta * \xi$.

(ii) $\alpha/\beta \leq -1$: In this case $|\phi|((0,\infty)) \geq 1$, so Theorem 3.2 does not apply. However, observe that $\mathcal{L}[\phi](z) = \alpha/(\beta - z) \neq 1$ and

$$\frac{\mathcal{L}[\theta](z)}{1 - \mathcal{L}[\phi](z)} = \frac{\mathcal{L}[\theta](z)}{1 - \alpha \frac{1}{\beta - z}} = \mathcal{L}[\theta](z) + \mathcal{L}[\theta](z) \frac{\alpha}{\beta - \alpha - z} = \mathcal{L}[\theta + \theta * \xi](z)$$

when $\mathrm{Re}(z) < 0$. The latter observation shows that

$$\sup_{x<0} \int_{\mathbb{R}} \left| \frac{\mathcal{L}[\theta](x+iy)}{1 - \mathcal{L}[\phi](x+iy)} \right|^2 dy \leq 2\pi \int_{\mathbb{R}} (\theta(u) + \theta * \xi(u))^2\, du < \infty$$

by Plancherel's Theorem. Now Remark 3.3 implies that a solution to (3.1) also exists in this case and $\psi = \theta + \theta * \xi$ is the solution kernel.

The next example relates (3.1) to (2.1) in a certain setup.

Example 3.5. We will give an example of an SDDE where Theorem 2.5 does not provide a solution, but where a solution can be found by considering an associated level model. Consider the SDDE model (2.1) in the case where η is absolutely continuous and its cumulative distribution function $F_\eta(t) := \eta([0,t])$, $t \geq 0$, satisfies

$$\int_0^\infty |F_\eta(t)|\, dt < 1. \tag{3.6}$$

This means in particular that $\eta([0,\infty)) = \lim_{t\to\infty} F_\eta(t) = 0$, and hence h defined in (2.2) satisfies $h(0) = 0$ and Theorem 2.5 does not apply (cf. Remark 2.8). In fact, using a stochastic Fubini theorem (such as [2, Theorem 3.1]) and integration by parts on the delay term, the equation may be written as

$$X_t - X_s = \int_0^\infty \Big[X_{t-u} - X_{s-u}\Big] F_\eta(u)\,du + Z_t - Z_s, \quad s < t. \tag{3.7}$$

This shows that uniqueness does not hold, since if $(X_t)_{t\in\mathbb{R}}$ is a solution then so is $(X_t + \xi)_{t\in\mathbb{R}}$ for any $\xi \in L^1(\mathbb{P})$. Moreover, as noted in Remark 2.8, we cannot expect to find stationary solutions in this setup. In the following let us restrict the attention to the case where

$$Z_t = \int_{\mathbb{R}} [f(t-u) - f_0(-u)]\,dL_u, \quad t \in \mathbb{R},$$

for a given Lévy process $(L_t)_{t\in\mathbb{R}}$ with $\mathbb{E}[L_1] = 0$ and $\mathbb{E}[L_1^2] < \infty$, and for some functions $f, f_0 : \mathbb{R} \to \mathbb{R}$, vanishing on $(-\infty, 0)$, such that $u \mapsto f(t+u) - f_0(u)$ belongs to L^2. Using Theorem 3.2 we will now argue that there always exists a centered and square integrable solution with stationary increments in this setup and that the increments of any two such solutions are identical.

To show the uniqueness part, suppose that $(X_t)_{t\in\mathbb{R}}$ is a centered and square integrable stationary increment process which satisfies (3.7). Then, for any given $s > 0$, we have that the increment process $X(s)_t = X_t - X_{t-s}$, $t \in \mathbb{R}$, is a stationary, centered and square integrable solution to the level equation (3.1) with $\phi(du) = F_\eta(u)\,du$ and $\theta = f - f(\cdot - s)$. By the uniqueness part of Theorem 3.2 and (3.6) it follows that

$$X(s)_t = \int_{\mathbb{R}} \psi_s(t-u)\,dL_u, \quad t \in \mathbb{R},$$

where $\psi_s(t) = \sum_{n=0}^\infty \int_0^\infty [f(t-u) - f(t-s-u)]\,\phi^{*n}(du)$ (the sum being convergent in L^2). Consequently, by a stochastic Fubini result, $(X_t)_{t\in\mathbb{R}}$ must take the form

$$X_t = \xi + \sum_{n=0}^\infty \int_0^\infty [Z_{t-u} - Z_{-u}]\,\phi^{*n}(du), \quad t \in \mathbb{R}, \tag{3.8}$$

for a suitable $\xi \in L^2(\mathbb{P})$ with $\mathbb{E}[\xi] = 0$. Conversely, if one defines $(X_t)_{t\in\mathbb{R}}$ by (3.8) we can use the same reasoning as above to conclude that $(X_t)_{t\in\mathbb{R}}$ is a stationary increment solution to (2.1). It should be stressed that one can find other representations of the solution than (3.8) (e.g., in a similar manner as in Example 3.4). For more on non-stationary solutions to (2.1), see [20].

A nice property of the model (3.1) is that it can recover the discrete-time ARMA(p,q) process. Example 3.6 gives (well-known) results for ARMA processes by using Remark 3.3. For an extensive treatment of ARMA processes, see e.g. [8].

Example 3.6. Let $p, q \in \mathbb{N}_0$ and define the polynomials $\Phi, \Theta : \mathbb{C} \to \mathbb{C}$ by

$$\Phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p \quad \text{and} \quad \Theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q,$$

where the coefficients are assumed to be real. Let $(L_t)_{t\in\mathbb{R}}$ be a Lévy process with $\mathbb{E}[L_1] = 0$ and $\mathbb{E}[L_1^2] < \infty$, and consider choosing $\phi(du) = \sum_{j=1}^p \phi_j \delta_j(du)$ and $\theta(u) =$

$\mathbb{1}_{[0,1)}(u) + \sum_{j=1}^{q} \theta_j \mathbb{1}_{[j,j+1)}(u)$. In this case (3.1) reads

$$X_t = \sum_{i=1}^{p} \phi_i X_{t-i} + Z_t + \sum_{i=1}^{q} \theta_i Z_{t-i}, \quad t \in \mathbb{R}, \tag{3.9}$$

with $Z_t = L_t - L_{t-1}$. In particular, if $(X_t)_{t\in\mathbb{R}}$ is a solution to (3.9), $(X_t)_{t\in\mathbb{Z}}$ is a usual ARMA process. Suppose that $\Phi(z) \neq 0$ for all $z \in \mathbb{C}$ with $|z| = 1$. Then, by continuity of Φ, there exists $a < 0$ such that $1 - \mathcal{L}[\phi](z) = \Phi(e^z)$ is strictly separated from 0 for $z \in \mathbb{C}$ with $\mathrm{Re}(z) \in (a, 0)$. Thus, since $\theta \in L^2$, Remark 3.3 implies that there exists a stationary solution to (3.1), and it is given by $X_t = \int_{\mathbb{R}} \psi(t-u)\, dL_u$, $t \in \mathbb{R}$, where ψ is characterized by (3.3). Choose a small $\varepsilon > 0$ and $(\psi_j)_{j\in\mathbb{Z}}$ so that the relation

$$\frac{\Theta(z)}{\Phi(z)} = \sum_{j=-\infty}^{\infty} \psi_j z^j$$

holds true for all $z \in \mathbb{C}$ with $1 - \varepsilon < |z| < 1 + \varepsilon$. Then

$$\mathcal{L}[\psi](z) = \mathcal{L}[\mathbb{1}_{[0,1)}](z)\frac{\Theta(e^z)}{\Phi(e^z)} = \sum_{j=-\infty}^{\infty} \psi_j \mathcal{L}[\mathbb{1}_{[j,j+1)}](z) = \mathcal{L}\Bigg[\sum_{j=-\infty}^{\infty} \psi_j \mathbb{1}_{[j,j+1)}\Bigg](z)$$

for all $z \in \mathbb{C}$ with a negative real part sufficiently close to zero. Thus, we have the well-known representation $X_t = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j}$ for $t \in \mathbb{R}$.

4 Proofs and technical results

The first result is closely related to the characterization of the so-called Hardy spaces and some of the Paley-Wiener theorems. For more on these topics, see e.g. [11, Section 2.3] and [15, Chapter VI (Section 7)]. We will use the notation $\mathcal{S}_{a,b} = \{z \in \mathbb{C} : a < \mathrm{Re}(z) < b\}$ throughout this section.

Lemma 4.1. *Let $-\infty \le a < b \le \infty$. Suppose that $F : \mathbb{C} \to \mathbb{C}$ is a function which is analytic on the strip $\mathcal{S}_{a,b}$ and satisfies*

$$\sup_{a<x<b} \int_{\mathbb{R}} |F(x+iy)|^2 dy < \infty. \tag{4.1}$$

Then there exists a function $f : \mathbb{R} \to \mathbb{C}$ such that $(u \mapsto f(u)e^{cu}) \in L^1$ for $c \in (a,b)$, $(u \mapsto f(u)e^{cu}) \in L^2$ for $c \in [a,b]$, and $\mathcal{L}[f](z) = F(z)$ for $z \in \mathcal{S}_{a,b}$.

Remark 4.2. If $a = -\infty$, the property $u \mapsto f(u)e^{au} \in L^2$ is understood as $f(u) = 0$ for almost all $u < 0$ and similarly, $f(u) = 0$ for almost all $u > 0$ if $u \mapsto f(u)e^{bu} \in L^2$ for $b = \infty$.

Proof (Proof of Lemma 4.1). Fix $c_1, c_2 \in (a,b)$ with $c_1 < c_2$. For any $y > 0$ and $u \in \mathbb{R}$, consider (anti-clockwise) integration of $z \mapsto e^{-zu}F(z)$ along a rectangular contour R_y

with vertices $c_1 - iy$, $c_2 - iy$, $c_2 + iy$, and $c_1 + iy$:

$$\begin{aligned} 0 &= \oint_{R_y} e^{-zu} F(z)\,dz \\ &= \int_{c_1}^{c_2} e^{-(x-iy)u} F(x-iy)\,dx + ie^{-c_2 u} \int_{-y}^{y} e^{-ixu} F(c_2+ix)\,dx \\ &\quad - \int_{c_1}^{c_2} e^{-(x+iy)u} F(x+iy)\,dx - ie^{-c_1 u} \int_{-y}^{y} e^{-ixu} F(c_1+ix)\,dx. \end{aligned} \tag{4.2}$$

Since

$$\begin{aligned} &\int_{\mathbb{R}} \left| \int_{c_1}^{c_2} e^{-(x+iy)u} F(x+iy) dx \right|^2 dy \\ &\qquad \leq e^{-2(c_1 u \wedge c_2 u)} (c_2 - c_1)^2 \sup_{a<x<b} \int_{\mathbb{R}} |F(x+iy)|^2 dy < \infty, \end{aligned}$$

we deduce the existence of a sequence $(y_n)_{n\in\mathbb{N}} \subseteq (0,\infty)$, such that $y_n \to \infty$ and

$$\int_{c_1}^{c_2} e^{-(x \pm iy_n)u} F(x \pm iy_n) dx \to 0.$$

Furthermore, for $k = 1,2$ it holds that

$$\left(u \mapsto \int_{-y}^{y} e^{-ixu} F(c_k+ix)\,dx \right) \to \left(u \mapsto 2\pi \mathcal{F}^{-1}[F(c_k + i\cdot)](u) \right)$$

in L^2 as $y \to \infty$ by Plancherel's Theorem. In particular, this convergence holds along the sequence $(y_n)_{n\in\mathbb{N}}$ and, eventually by only considering a subsequence of $(y_n)_{n\in\mathbb{N}}$, we may assume that

$$\lim_{n\to\infty} \int_{-y_n}^{y_n} e^{-ixu} F(c_k + ix)\,dx = 2\pi \mathcal{F}^{-1}[F(c_k + i\cdot)](u), \quad k = 1,2,$$

for almost all $u \in \mathbb{R}$. Combining this with (4.2) yields $e^{-c_1 u} \mathcal{F}^{-1}[F(c_1 + i\cdot)](u) = e^{-c_2 u} \mathcal{F}^{-1}[F(c_2 + i\cdot)](u)$ for almost all $u \in \mathbb{R}$. Consequently, there exists a function $f : \mathbb{R} \to \mathbb{C}$ with the property that $f(u) = e^{-cu} \mathcal{F}^{-1}[F(c + i\cdot)](u)$ for almost all $u \in \mathbb{R}$ for any given $c \in (a,b)$. For such c we compute

$$\int_{\mathbb{R}} |e^{cu} f(u)|^2 du = \int_{\mathbb{R}} |\mathcal{F}^{-1}[F(c+i\cdot)](u)|^2 du \leq \sup_{x\in(a,b)} \int_{\mathbb{R}} |F(x+iy)|^2 dy < \infty.$$

Consequently, $(u \mapsto f(u)e^{cu}) \in L^2$ for any $c \in (a,b)$ and by Fatou's Lemma, this holds as well for $c = a$ and $c = b$. Furthermore, if $c \in (a,b)$, we can choose $\varepsilon > 0$ such that $c \pm \varepsilon \in (a,b)$ as well, from which we get that

$$\begin{aligned} &\left(\int_{\mathbb{R}} |f(u)| e^{cu} du \right)^2 \\ &\quad \leq \left(\int_0^{\infty} |f(u) e^{(c+\varepsilon)u}|^2 du + \int_{-\infty}^{0} |f(u) e^{(c-\varepsilon)u}|^2 du \right) \int_0^{\infty} e^{-2\varepsilon u} du < \infty \end{aligned}$$

by Hölder's inequality. This shows that $(u \mapsto f(u)e^{cu}) \in L^1$. Finally, we find for $z = x + iy \in \mathcal{S}_{a,b}$ (by definition of f) that

$$\mathcal{L}[f](z) = \int_{\mathbb{R}} e^{iyu} e^{xu} f(u)\,du = \mathcal{F}[\mathcal{F}^{-1}[F(x+i\cdot)]](y) = F(z),$$

and this completes the proof. □

Proof (Proof of Lemma 2.2). Observe that, generally, $h(z) \neq 0$ if $\mathrm{Re}(z) \leq 0$ and $|z| > |\eta|([0,\infty))$, and thus, under the assumption that $h(iy) \neq 0$ for all $y \in \mathbb{R}$ and by continuity of h there must be an $a < 0$ such that $h(z) \neq 0$ for all $z \in \mathcal{S}_{a,0}$. The fact that $|h(z)| \sim |z|$ as $|z| \to \infty$ when $\mathrm{Re}(z) \leq 0$ and, once again, the continuity of h imply that (4.1) is satisfied for $1/h$, and thus we get the existence of a function $\tilde{x}_0 : \mathbb{R} \to \mathbb{R}$ such that $\mathcal{L}[\tilde{x}_0] = 1/h$ on $\mathcal{S}_{a,0}$ and $t \mapsto e^{ct}\tilde{x}_0(t) \in L^2$ for all $c \in [a,0]$. Observe that this gives in particular that $\tilde{x}_0 \mathbb{1}_{(-\infty,0]} \in L^1$ and thus, since $\tilde{x}_0 \in L^2$, we also get that $\tilde{x}_0 \mathbb{1}_{(-\infty,t]} \in L^1$ for all $t \in \mathbb{R}$. This ensures that $x_0 : \mathbb{R} \to \mathbb{R}$ given by

$$x_0(t) = \mathbb{1}_{[0,\infty)}(t) + \int_{-\infty}^{t} \int_{[0,\infty)} \tilde{x}_0(u-v)\,\eta(dv)\,du, \quad t \in \mathbb{R},$$

is a well-defined function. To establish the first part of the statement (in particular (2.3)) it suffices to argue that $\mathcal{L}[x_0] = 1/h$ on $\mathcal{S}_{a,0}$. However, this follows from the following calculation, which holds for an arbitrary $z \in \mathcal{S}_{a,0}$:

$$\begin{aligned}
&\mathcal{L}\Big[\mathbb{1}_{[0,\infty)} + \int_{-\infty}^{\cdot} \int_{[0,\infty)} \tilde{x}_0(u-v)\,\eta(dv)\,du\Big](z) \\
&\qquad = -z^{-1}\Big[1 + \mathcal{L}[\tilde{x}_0](z)\mathcal{L}[\eta](z)\Big] = -z^{-1}\frac{z}{z + \mathcal{L}[\eta](z)} = \frac{1}{h(z)}.
\end{aligned}$$

Suppose now that η has n-th moment for some $n \in \mathbb{N}$ and note that

$$|D^k h(iy)| \leq 1 + \int_{[0,\infty)} v^k |\eta|(dv) < \infty,$$

for $k \in \{1,\dots,n\}$ (D^k denoting the k-th order derivative with respect to y). Since $D^k[1/h(iy)]$ will be a sum of terms of the form $D^l h(iy)/h(iy)^m$, $l,m = 1,\dots,k$, and $(y \mapsto 1/h(iy)) \in L^2$, this means in turn that $D^k[1/h(i\cdot)] \in L^2$ for $k = 1,\dots,n$. Since $\mathcal{F}^{-1}$ maps L^2 functions to L^2 functions, $\mathcal{F}^{-1}[D^n[1/h(i\cdot)]] \in L^2$. Moreover, it is well-known that if $f, Df \in L^1$, we have the formula $\mathcal{F}^{-1}[Df](u) = iu\mathcal{F}^{-1}[f](u)$ for $u \in \mathbb{R}$, and by an approximation argument it holds when $f, Df \in L^2$ as well (although only for almost all u), cf. [1, Corollary 3.23]. Hence, by induction we establish that

$$\mathcal{F}^{-1}\Big[D^n \frac{1}{h(i\cdot)}\Big](u) = (iu)^n \mathcal{F}^{-1}\Big[\frac{1}{h(i\cdot)}\Big](u) = (iu)^n x_0(u).$$

This shows the first part of ((i)). For any given $q \in [1/n, 2)$ it follows by Hölder's inequality that

$$\begin{aligned}
&\int_{\mathbb{R}} |x_0(u)|^q\,du \\
&\qquad \leq \Big(\int_{\mathbb{R}} \big(x_0(u)(1+|u|^n)\big)^2\,du\Big)^{q/2} \Big(\int_{\mathbb{R}} (1+|u|^n)^{-2q/(2-q)}\,du\Big)^{1-q/2} < \infty,
\end{aligned}$$

which shows $x_0 \in L^q$. By using the relation (2.3), which was verified just above, we obtain

$$|x_0(t)| \leq 1 + \int_{-\infty}^{t} \int_{[0,\infty)} |x_0(u-v)|\,|\eta|(dv)\,du \leq |\eta|([0,\infty)) \int_{\mathbb{R}} |x_0(u)|\,du.$$

Since $x_0 \in L^1$, the inequalities above imply $x_0 \in L^\infty$, and thus we get $x_0 \in L^q$ for $q \in [1/n, \infty]$, which shows the second part of ((i)). If η has an exponential moment of order δ then we can find $a < 0 < b \leq \delta$ such that $1/h$ satisfies (4.1) and therefore, we have that $u \mapsto x_0(u)e^{cu} \in L^2$ for $c \in [a,b]$. If $h(z) \neq 0$ for all $z \in \mathbb{C}$ with $\mathrm{Re}(z) \leq 0$ we can argue that (4.1) holds with $a = -\infty$ (and $b = 0$) in the same way as above and, thus, Lemma 4.1 implies $x_0(u) = 0$ for $u < 0$. □

The following lemma is used to ensure uniqueness of solutions to (2.1):

Lemma 4.3. *Fix $s \in \mathbb{R}$. Suppose that $h(iy) \neq 0$ for all $y \in \mathbb{R}$ and that, given $(Y_t)_{t \leq s}$, a process $(X_t)_{t\in\mathbb{R}}$ satisfies*

$$X_t = \begin{cases} X_s + \int_s^t \int_{[0,\infty)} X_{u-v}\,\eta(dv)\,du, & t \geq s \\ Y_t, & t < s \end{cases} \tag{4.3}$$

almost surely for each $t \in \mathbb{R}$ (the $\mathbb{P}$-null sets are allowed to depend on t) and $\sup_{t\in\mathbb{R}} \mathbb{E}[|X_t|] < \infty$. *Then*

$$X_t = X_s x_0(t-s) + \int_s^\infty \int_{(u-s,\infty)} Y_{u-v}\,\eta(dv)\,x_0(t-u)\,du \tag{4.4}$$

for Lebesgue almost all $t \geq s$ outside a $\mathbb{P}$-null set.

Proof. Observe that, by Fubini's Theorem, we can remove a $\mathbb{P}$-null set and have that (4.3) is satisfied for Lebesgue almost all $t \in \mathbb{R}$. Let $a < 0$ be such that $h(z) \neq 0$ for all $z \in \mathcal{S}_{a,0}$ (this is possible due to the assumption $h(iy) \neq 0$ for all $y \in \mathbb{R}$). Note that

$$\mathbb{E}\left[\int_s^\infty e^{-n^{-1}t}|X_t|\,dt\right] \leq n \sup_{t\in\mathbb{R}} \mathbb{E}[|X_t|] < \infty$$

for any given $n \in \mathbb{N}$ by Tonelli's Theorem. This means that $\int_s^\infty e^{-n^{-1}t}|X_t|\,dt < \infty$ for all n almost surely and, hence, $\mathcal{L}[X\mathbb{1}_{[s,\infty)}]$ is well-defined on $\mathcal{S}_{a,0}$ outside a $\mathbb{P}$-null set. For $z \in \mathcal{S}_{a,0}$ we compute

$$\begin{aligned}
\mathcal{L}[X\mathbb{1}_{[s,\infty)}](z) =& \mathcal{L}\left[\mathbb{1}_{[s,\infty)}\left\{X_s + \int_s^{\cdot} \int_{[0,\infty)} X_{u-v}\,\eta(dv)\,du\right\}\right](z) \\
=& -\frac{X_s e^{zs}}{z} + \int_s^\infty e^{zt} \int_s^t \int_{[0,\infty)} X_{u-v}\,\eta(dv)\,du\,dt \\
=& -\frac{X_s e^{zs}}{z} + \int_{[0,\infty)} \int_s^\infty X_{u-v} \int_u^\infty e^{zt}\,dt\,du\,\eta(dv) \\
=& -\frac{1}{z}\left(X_s e^{zs} + \int_{[0,\infty)} \int_{s-v}^\infty X_u e^{z(u+v)}\,du\,\eta(dv)\right) \\
=& -\frac{1}{z}\Bigg(X_s e^{zs} + \mathcal{L}[\eta](z)\mathcal{L}[X\mathbb{1}_{[s,\infty)}](z) \\
&+ \mathcal{L}\left[\mathbb{1}_{[s,\infty)} \int_{(\cdot-s,\infty)} Y_{\cdot-v}\,\eta(dv)\right](z)\Bigg).
\end{aligned}$$

In the calculations above we have used Fubini's Theorem several times; specifically, in the third and fifth equality. These calculations are valid (at least after removing yet another $\mathbb{P}$-null set) by the same type of argument as used to establish that $\mathcal{L}[X\mathbb{1}_{[s,\infty)}]$ is well-defined on $\mathcal{S}_{a,0}$ almost surely. For instance, Fubini's Theorem is applicable in the third equality above for any $z\in\mathcal{S}_{a,0}$ almost surely, since

$$\begin{aligned}\mathbb{E}\Big[\int_s^\infty\int_s^t\int_{[0,\infty)} e^{-n^{-1}t}|X_{u-v}||\eta|(dv)\,du\,dt\Big]\\ = |\eta|([0,\infty))\int_s^\infty (t-s)e^{-n^{-1}t}\,dt\,\sup_{t\in\mathbb{R}}\mathbb{E}[|X_t|]<\infty\end{aligned}$$

for an arbitrary $n\in\mathbb{N}$. Returning to the computations, we find by rearranging terms that

$$\mathcal{L}[X\mathbb{1}_{[s,\infty)}](z) = \frac{X_s e^{zs}}{h(z)} + \frac{\mathcal{L}\Big[\mathbb{1}_{[s,\infty)}\int_{(\cdot-s,\infty)} Y_{\cdot-v}\,\eta(dv)\Big](z)}{h(z)}. \tag{4.5}$$

By applying the expectation operator, we note that

$$\int_s^\infty\int_{(u-s,\infty)} |Y_{u-v}||\eta|(dv)|x_0(t-u)|\,du<\infty \tag{4.6}$$

almost surely for each $t\in\mathbb{R}$ if $\int_s^\infty |\eta|((u-s,\infty))|x_0(t-u)|\,du<\infty$. Since $|\eta|([0,\infty))<\infty$, it is sufficient that $x_0\mathbb{1}_{(-\infty,t-s]}\in L^1$, but this is indeed the case (see the beginning of the proof of Lemma 2.2). Consequently, Tonelli's Theorem implies that (4.6) holds for Lebesgue almost all $t\in\mathbb{R}$ outside a $\mathbb{P}$-null set. Furthermore, again by Lemma 2.2, there exists $\varepsilon>0$ such that

$$\int_{\mathbb{R}} e^{-\varepsilon t}\int_s^\infty |x_0(t-u)|\,du\,dt = \int_{\mathbb{R}} e^{-\varepsilon t}|x_0(t)|\,dt\int_s^\infty e^{-\varepsilon u}\,du<\infty.$$

From this it follows that, almost surely, $\int_s^\infty\int_{(u-s,\infty)} Y_{u-v}\,\eta(dv)\,x_0(t-u)\,du$ is well-defined and that its Laplace transform exists on $\mathcal{S}_{-\varepsilon,0}$. We conclude that

$$\mathcal{L}\Big[\int_s^\infty\int_{(u-s,\infty)} Y_{u-v}\,\eta(dv)\,x_0(\cdot-u)\,du\Big](z) = \frac{\mathcal{L}\Big[\mathbb{1}_{[s,\infty)}\int_{(\cdot-s,\infty)} Y_{\cdot-v}\,\eta(dv)\Big](z)}{h(z)}$$

for $z\in\mathcal{S}_{-\varepsilon,0}$, and the result follows since we also have $\mathcal{L}[x_0(\cdot-s)](z)=e^{zs}/h(z)$ for $z\in\mathcal{S}_{-\varepsilon,0}$. □

When proving Theorem 2.5, [2, Corollary A.3] will play a crucial role, and for reference we have chosen to include (a suitable version of) it here:

Corollary 4.4 ([2, Corollary A.3]). *Let $p\geq 1$ and $(X_t)_{t\in\mathbb{R}}$ be a measurable process with stationary increments and $\mathbb{E}[|X_t|^p]<\infty$ for all $t\in\mathbb{R}$. Then $(X_t)_{t\in\mathbb{R}}$ is continuous in $L^p(\mathbb{P})$, and there exist $\alpha,\beta>0$ such that $\mathbb{E}[|X_t|^p]^{1/p}\leq\alpha+\beta|t|$ for all $t\in\mathbb{R}$.*

Proof (Proof of Theorem 2.5). We start by noting that if $(X_t)_{t\in\mathbb{R}}$ and $(Y_t)_{t\in\mathbb{R}}$ are two measurable, stationary and integrable ($\mathbb{E}[|X_0|],\mathbb{E}[|Y_0|]<\infty$) solutions to (2.1) then, for fixed $s\in\mathbb{R}$,

$$U_t = U_s + \int_s^t\int_{[0,\infty)} U_{u-v}\,\eta(dv)\,du \tag{4.7}$$

almost surely for each $t \in \mathbb{R}$, when we set $U_t := X_t - Y_t$. In particular, for a given $t \in \mathbb{R}$, we get by Lemma 4.3,

$$U_r = U_s x_0(r-s) + \int_s^\infty \int_{(u-s,\infty)} U_{u-v}\,\eta(dv)\,x_0(r-u)\,du \tag{4.8}$$

for Lebesgue almost all $r > t-1$ and all $s \in \mathbb{Q}$ with $s \le t-1$. For any such r we observe that the right-hand side of (4.8) tends to zero in $L^1(\mathbb{P})$ as $\mathbb{Q} \ni s \to -\infty$, from which we deduce $U_r = 0$ or, equivalently, $X_r = Y_r$ almost surely. By Corollary 4.4 it follows that $(U_r)_{r\in\mathbb{R}}$ is continuous in $L^1(\mathbb{P})$, and hence we get that $X_t = Y_t$ almost surely as well. This shows that a solution to (2.1) is unique up to modification.

We have $\mathbb{E}[|Z_u|] \le a + b|u|$ for any $u, v \in \mathbb{R}$ with suitably chosen $a, b > 0$ (see Corollary 4.4), and this implies that

$$\begin{aligned}
&\mathbb{E}\left[\int_{\mathbb{R}} |Z_u| \int_{[0,\infty)} |x_0(t-u-v)|\,|\eta|(dv)\,du\right] \\
&\le a|\eta|([0,\infty)) \int_{\mathbb{R}} |x_0(u)|\,du + b\int_{\mathbb{R}} |u| \int_{[0,\infty)} |x_0(t-u-v)|\,|\eta|(dv)\,du \\
&\le \left(a|\eta|([0,\infty)) + b\int_{[0,\infty)} v|\eta|(dv)\right)\int_{\mathbb{R}} |x_0(u)|\,du \\
&\quad + b|\eta|([0,\infty))\int_{\mathbb{R}} (|t|+|u|)|x_0(u)|\,du.
\end{aligned}$$

This is finite by Lemma 2.2 and Example 2.4, and $\int_{\mathbb{R}} Z_u \int_{[0,\infty)} x_0(t-u-v)\,\eta(dv)\,du$ is therefore almost surely well-defined.

To argue that $X_t = Z_t + \int_{\mathbb{R}} Z_u \int_{[0,\infty)} x_0(t-u-v)\,\eta(dv)\,du$, $t \in \mathbb{R}$, satisfies (2.1), let $s < t$ and note that by Lemma 2.2 we have

$$\begin{aligned}
&\int_s^t \int_{[0,\infty)} X_{u-v}\,\eta(dv)\,du - \int_s^t \int_{[0,\infty)} Z_{u-v}\,\eta(dv)\,du \\
&= \int_s^t \int_{[0,\infty)} \int_{\mathbb{R}} Z_r \int_{[0,\infty)} x_0(u-v-r-w)\,\eta(dw)\,dr\,\eta(dv)\,du \\
&= \int_{\mathbb{R}} Z_r \int_{[0,\infty)} \int_{s-r-w}^{t-r-w} \int_{[0,\infty)} x_0(u-v)\,\eta(dv)\,du\,\eta(dw)\,dr \\
&= \int_{\mathbb{R}} Z_r \int_{[0,\infty)} [x_0(t-r-w) - x_0(s-r-w)]\,\eta(dw)\,dr \\
&- \int_{\mathbb{R}} \int_{[0,\infty)} Z_r[\mathbb{1}_{[0,\infty)}(t-r-w) - \mathbb{1}_{[0,\infty)}(s-r-w)]\,\eta(dw)\,dr \\
&= \int_{\mathbb{R}} Z_r \int_{[0,\infty)} [x_0(t-r-w) - x_0(s-r-w)]\,\eta(dw)\,dr \\
&- \int_s^t \int_{[0,\infty)} Z_{r-w}\,\eta(dw)\,dr.
\end{aligned}$$

Next, we write

$$X_t = \int_{\mathbb{R}} (Z_t - Z_{t-u}) \int_{[0,\infty)} x_0(u-v)\,\eta(dv)\,du, \quad t \in \mathbb{R}, \tag{4.9}$$

using that

$$\int_{\mathbb{R}}\int_{[0,\infty)} x_0(u-v)\,\eta(dv)\,du = \int_{\mathbb{R}} x_0(u)\,du\,\eta([0,\infty)) = h(0)\eta([0,\infty)) = -1 \tag{4.10}$$

Since $(Z_t)_{t\in\mathbb{R}}$ is continuous in $L^1(\mathbb{P})$, one shows that the process

$$X_t^n := \int_{-n}^{n} (Z_t - Z_{t-u}) \int_{[0,\infty)} x_0(u-v)\,\eta(dv)\,du, \quad t\in\mathbb{R},$$

is stationary by approximating it by Riemann sums in $L^1(\mathbb{P})$. Subsequently, due to the fact that $X_t^n \to X_t$ almost surely as $n\to\infty$ for any $t\in\mathbb{R}$, we conclude that $(X_t)_{t\in\mathbb{R}}$ is stationary. This type of approximation arguments are carried out in detail in [7, p. 20]. In case $h(z)\neq 0$ for all $z\in\mathbb{C}$ with $\mathrm{Re}(z)\leq 0$, the causal representation (2.7) of $(X_t)_{t\in\mathbb{R}}$ follows from (4.9) and the fact that $x_0(t)=0$ for $t<0$ by Lemma 2.2((iii)). This completes the proof. □

Proof (Proof of Corollary 2.6). It follows from (4.10) and Corollary 2.3 that

$$\begin{aligned}
&Z_t + \int_{\mathbb{R}} Z_{t-u} \int_{[0,\infty)} x_0(u-v)\,\eta(dv)\,du \\
&\quad= \int_{\mathbb{R}} [Z_{t-u} - Z_t] \int_{[0,\infty)} x_0(u-v)\,\eta(dv)\,du \\
&\quad= \int_{\mathbb{R}}\int_{\mathbb{R}} [\theta(t-u-r) - \theta(t-r)][\mu_{x_0}(du) - \delta_0(du)]dL_r \\
&\quad= \int_{\mathbb{R}}\int_{\mathbb{R}} \theta(t-u-r)\,\mu_{x_0}(du)\,dL_r \\
&\quad= \int_{\mathbb{R}} \theta * x_0(t-r)\,dL_r,
\end{aligned}$$

where we have used that $\mu_{x_0}(\mathbb{R}) = 0$ since $x_0(t)\to 0$ for $t\to\pm\infty$ by (2.3). □

Proof (Proof of Theorem 3.2). First, observe that there exists $\psi:\mathbb{R}\to\mathbb{R}$ such that $\sum_{k=0}^n \theta*\phi^{*k}\to\psi$ in L^2 as $n\to\infty$. To see this, set $\psi_n = \sum_{k=0}^n \theta*\phi^{*k}$, let $m<n$ and note that

$$\int_{\mathbb{R}} (\psi_n(t)-\psi_m(t))^2\,dt = \frac{1}{2\pi}\int_{\mathbb{R}} \left|\mathcal{F}\left[\sum_{k=m+1}^{n} \theta*\phi^{*k}\right](y)\right|^2 dy \tag{4.11}$$

for $m<n$ by Plancherel's Theorem. For any $y\in\mathbb{R}$ we have that

$$\left|\mathcal{F}\left[\sum_{k=m+1}^{n}\theta*\phi^{*k}\right](y)\right| \leq |\mathcal{F}[\theta](y)| \sum_{k=m+1}^{n} |\phi|((0,\infty))^k \leq \frac{|\mathcal{F}[\theta](y)|}{1-|\phi|((0,\infty))}. \tag{4.12}$$

The first inequality in (4.12) shows that $|\mathcal{F}[\sum_{k=m+1}^n \theta*\phi^{*k}](y)|\to 0$ as $n,m\to\infty$, and hence we can use the second inequality of (4.12) and dominated convergence together with the relation (4.11) to deduce that $(\psi_n)_{n\in\mathbb{N}}$ is a Cauchy sequence in L^2. This establishes the existence of ψ. Due to the fact that ψ_n is real-valued and vanishes on $(-\infty,0)$ for all $n\in\mathbb{N}$, the same holds for ψ almost everywhere.

Suppose now that we have a square integrable stationary solution $(X_t)_{t\in\mathbb{R}}$. Then, using a stochastic Fubini (e.g., [2, Theorem 3.1]), it follows that for each $t \in \mathbb{R}$ and almost surely,

$$\begin{aligned} X_t &= X * \phi^{*n}(t) + \sum_{k=0}^{n-1}\left(\int_{\mathbb{R}} \theta(\cdot - u)\, dL_u\right) * \phi^{*k}(t) \\ &= X * \phi^{*n}(t) + \int_{\mathbb{R}} \psi_{n-1}(t-u)\, dL_u \end{aligned} \tag{4.13}$$

for an arbitrary $n \in \mathbb{N}$. By Jensen's inequality and stationarity of $(X_u)_{u\in\mathbb{R}}$,

$$\mathbb{E}[X * \phi^{*n}(t)^2] \le \mathbb{E}\left[\left(\int_0^\infty |X_{t-u}||\phi^{*n}|(du)\right)^2\right] \le |\phi^{*n}|((0,\infty))\mathbb{E}[X_0^2].$$

Since $\mathbb{E}[X_0^2] < \infty$ and $|\phi^{*n}|((0,\infty)) = |\phi|((0,\infty))^n \to 0$ as $n \to \infty$, it follows that $X * \phi^{*n}(t) \to 0$ in $L^2(\mathbb{P})$ as $n \to \infty$. Consequently, (4.13) shows that $\int_{\mathbb{R}} \psi_n(t-u)\, dL_u \to X_t$ in $L^2(\mathbb{P})$ as $n \to \infty$. On the other hand, by the isometry property of the stochastic integral we also have that

$$\mathbb{E}\left[\left(\int_{\mathbb{R}} \psi(t-u)\, dL_u - \int_{\mathbb{R}} \psi_n(t-u)\, dL_u\right)^2\right] = \mathbb{E}[L_1^2]\int_{\mathbb{R}} (\psi(u) - \psi_n(u))^2\, du \to 0$$

as $n \to \infty$, and hence $X_t = \int_{\mathbb{R}} \psi(t-u)\, dL_u$ almost surely by uniqueness of limits in $L^2(\mathbb{P})$.

Conversely, define a square integrable stationary process $(X_t)_{t\in\mathbb{R}}$ by $X_t = \int_{\mathbb{R}} \psi(t-u)\, dL_u$ for $t \in \mathbb{R}$. After noting that $\psi_n * \phi = \sum_{k=1}^{n+1} \theta * \phi^{*k} = \psi_{n+1} - \theta$ for all n, we find

$$\begin{aligned} 0 &= \limsup_{n\to\infty} \int_0^\infty \left[\psi_{n+1}(u) - \theta(u) - \psi_n * \phi(u)\right]^2 du \\ &= \int_0^\infty \left[\psi(u) - \theta(u) - \psi * \phi(u)\right]^2 du \\ &= \mathbb{E}\left[\left(X_t - X * \phi(t) - \int_{\mathbb{R}} \theta(t-u)\, dL_u\right)^2\right]\mathbb{E}[L_1^2]^{-1}. \end{aligned}$$

Thus, $(X_t)_{t\in\mathbb{R}}$ satisfies (3.1). □

Acknowledgments

We thank the referees for constructive and detailed reports. Their comments and suggestions have helped us to improve the paper significantly. This work was supported by the Danish Council for Independent Research (Grant DFF - 4002 - 00003).

References

[1] Adams, R.A. and J.J.F. Fournier (2003). *Sobolev spaces*. Second. Vol. 140. Pure and Applied Mathematics (Amsterdam). Elsevier/Academic Press, Amsterdam, xiv+305.

[2] Barndorff-Nielsen, O.E. and A. Basse-O'Connor (2011). Quasi Ornstein–Uhlenbeck processes. *Bernoulli* 17(3), 916–941. DOI: 10.3150/10-BEJ311.

[3] Barndorff-Nielsen, O.E., F.E. Benth and A.E.D. Veraart (2018). *Ambit stochastics*. Vol. 88. Probability Theory and Stochastic Modelling. Springer, Cham. DOI: 10.1007/978-3-319-94129-5.

[4] Barndorff-Nielsen, O.E. and N. Shephard (2001). Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial economics. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 63(2), 167–241.

[5] Basse-O'Connor, A. and J. Rosiński (2013). Characterization of the finite variation property for a class of stationary increment infinitely divisible processes. *Stochastic Process. Appl.* 123(6), 1871–1890. DOI: 10.1016/j.spa.2013.01.014.

[6] Basse-O'Connor, A. and J. Rosiński (2016). On infinitely divisible semimartingales. *Probab. Theory Related Fields* 164(1–2), 133–163. DOI: 10.1007/s00440-014-0609-1.

[7] Basse-O'Connor, A., M.S. Nielsen, J. Pedersen and V. Rohde (2018). Multivariate stochastic delay differential equations and CAR representations of CARMA processes. *Stochastic Process. Appl.* Forthcoming. DOI: 10.1016/j.spa.2018.11.011.

[8] Brockwell, P.J. and R.A. Davis (2006). *Time series: theory and methods*. Springer Series in Statistics. Reprint of the second (1991) edition. Springer, New York.

[9] Brockwell, P.J. and A. Lindner (2009). Existence and uniqueness of stationary Lévy-driven CARMA processes. *Stochastic Process. Appl.* 119(8), 2660–2681. DOI: 10.1016/j.spa.2009.01.006.

[10] Brockwell, P.J. and T. Marquardt (2005). Lévy-driven and fractionally integrated ARMA processes with continuous time parameter. *Statist. Sinica* 15(2), 477–494.

[11] Dym, H. and H.P McKean (1976). *Gaussian processes, function theory, and the inverse spectral problem*. Probability and Mathematical Statistics, Vol. 31. New York: Academic Press [Harcourt Brace Jovanovich Publishers].

[12] Gripenberg, G., S.-O. Londen and O. Staffans (1990). *Volterra integral and functional equations*. Vol. 34. Encyclopedia of Mathematics and its Applications. Cambridge University Press. DOI: 10.1017/CBO9780511662805.

[13] Gushchin, A.A. and U. Küchler (2000). On stationary solutions of delay differential equations driven by a Lévy process. *Stochastic Process. Appl.* 88(2), 195–211. DOI: 10.1016/S0304-4149(99)00126-X.

[14] Jones, R.H. and L.M Ackerson (1990). Serial correlation in unequally spaced longitudinal data. *Biometrika* 77(4), 721–731. DOI: 10.1093/biomet/77.4.721.

[15] Katznelson, Y. (2004). *An introduction to harmonic analysis*. Third. Cambridge Mathematical Library. Cambridge University Press. DOI: 10.1017/CBO9781139165372.

[16] Küchler, U. and B. Mensch (1992). Langevin's stochastic differential equation extended by a time-delayed term. *Stochastics Stochastics Rep.* 40(1-2), 23–42. DOI: 10.1080/17442509208833780.

[17] Küchler, U. and M. Sørensen (2013). Statistical inference for discrete-time samples from affine stochastic delay differential equations. *Bernoulli* 19(2), 409–425. DOI: 10.3150/11-BEJ411.

[18] Marquardt, T. (2006). Fractional Lévy processes with an application to long memory moving average processes. *Bernoulli* 12(6), 1099–1126.

[19] Marquardt, T. and R. Stelzer (2007). Multivariate CARMA processes. *Stochastic Process. Appl.* 117(1), 96–120. DOI: 10.1016/j.spa.2006.05.014.

[20] Nielsen, M.S. (2019). On non-stationary solutions to MSDDEs: representations and the cointegration space. arXiv: 1903.02066.

[21] Nielsen, M.S. and V.U. Rohde (2017). Recovering the background noise of a Lévy-driven CARMA process using an SDDE approach. *Proceedings ITISE 2017* 2, 707–718.

[22] Protter, P.E. (2004). *Stochastic Integration and Differential Equations*. Second. Vol. 21. Applications of Mathematics (New York). Stochastic Modelling and Applied Probability. Berlin: Springer-Verlag.

[23] Rajput, B.S. and J. Rosiński (1989). Spectral representations of infinitely divisible processes. *Probab. Theory Related Fields* 82(3), 451–487.

[24] Reiß, M., M. Riedle and O. van Gaans (2006). Delay differential equations driven by Lévy processes: stationarity and Feller properties. *Stochastic Process. Appl.* 116(10), 1409–1432. DOI: 10.1016/j.spa.2006.03.002.

[25] Sato, K. (1999). *Lévy Processes and Infinitely Divisible Distributions*. Vol. 68. Cambridge Studies in Advanced Mathematics. Translated from the 1990 Japanese original, Revised by the author. Cambridge University Press.

[26] Schwartz, E.S. (1997). The stochastic behavior of commodity prices: Implications for valuation and hedging. *J. Finance* 52(3), 923–973.

[27] Todorov, V. and G. Tauchen (2006). Simulation methods for Lévy-driven continuous-time autoregressive moving average (CARMA) stochastic volatility models. *J. Bus. Econom. Statist.* 24(4), 455–469. DOI: 10.1198/073500106000000260.

[28] Yaglom, A.M (1987). *Correlation theory of stationary and related random functions. Vol. I*. Springer Series in Statistics. Basic results. New York: Springer-Verlag.

PAPER D

Multivariate Stochastic Delay Differential Equations And Their Relation To CARMA Processes

Andreas Basse-O'Connor, Mikkel Slot Nielsen, Jan Pedersen, and Victor Rohde

Abstract

In this study we show how to represent a continuous time autoregressive moving average (CARMA) as a higher order stochastic delay differential equation, which can be thought of as an AR(∞) process. Furthermore, we show how the AR(∞) representation gives rise to a prediction formula for CARMA processes. To be used in the above mentioned results we develop a general theory for multivariate stochastic delay differential equations, which will be of independent interest, and where we focus on existence, uniqueness and representations.

MSC 2010: 60G10, 60G22, 60H10, 60H20

Keywords: multivariate stochastic delay differential equations; multivariate Ornstein-Uhlenbeck processes; CARMA processes; FICARMA processes; MCARMA processes; noise recovery; prediction; long memory; continuous time

1 Introduction and main ideas

The class of autoregressive moving averages (ARMA) is one of the most popular classes of stochastic processes for modeling time series in discrete time. This class goes back to the thesis of Whittle in 1951 and was popularized in [5]. The continuous time analogue of an ARMA process is called a CARMA process, and it is the formal solution $(X_t)_{t\in\mathbb{R}}$ to the equation

$$P(D)X_t = Q(D)DZ_t, \quad t \in \mathbb{R}, \tag{1.1}$$

where P and Q are polynomials of degree p and q, respectively. Furthermore, D denotes differentiation with respect to t, and $(Z_t)_{t\in\mathbb{R}}$ is a Lévy process, the continuous time analogue of a random walk. In the following we will assume that $p > q$ and $P(z), Q(z) \neq 0$ whenever $\mathrm{Re}(z) \geq 0$. In this case one can give precise meaning to $(X_t)_{t\in\mathbb{R}}$ as a causal stochastic process through a state-space representation as long as $(Z_t)_{t\in\mathbb{R}}$ has log moments. Lévy-driven CARMA processes have found many applications, for example, in modeling temperature, electricity, and stochastic volatility, cf. [4, 14, 26]. Moreover, there exists a vast amount of literature on theoretical results for CARMA processes (and variations of these), and a few references are [7, 10, 6, 8, 18, 19, 25].

It is well-known that any causal CARMA process has a moving average representation of MA(∞) type

$$X_t = \int_{-\infty}^{t} g(t-u)\,dZ_u, \qquad t \in \mathbb{R},$$

cf. Section 4.3. This representation may be very convenient for studying many of their properties. A main contribution of our work is that we obtain an autoregressive representation of CARMA processes of AR(∞) type

$$R(D)X_t = \int_0^\infty X_{t-u} f(u)\,du + DZ_t, \quad t \in \mathbb{R}, \tag{1.2}$$

where R is a polynomial of order $p-q$ and $f : \mathbb{R} \to \mathbb{R}$ is a deterministic function, both defined through P and Q. Since $(X_t)_{t\in\mathbb{R}}$ is $p-q-1$ times differentiable, see [19, Proposition 3.32], the relation (1.2) is well-defined if we integrate both sides. A heuristic argument why (1.2) is a reasonable continuous time equivalent of the discrete time AR(∞) representation is as follows. If $q = 0$, Q is constant and we get the (finite order) AR representation immediately:

$$P(D)X_t = DZ_t.$$

If $q \geq 1$, let $(\phi_j)_{j\in\mathbb{N}_0}$ be the coefficients in the power series expansion of P/Q on $\{z \in \mathbb{C} : \mathrm{Re}(z) > -\delta\}$ for a sufficiently small $\delta > 0$ and write (1.1) as

$$\sum_{j=0}^{\infty} \phi_j D^j X_t = DZ_t. \tag{1.3}$$

Except in trivial cases, $(X_t)_{t\in\mathbb{R}}$ is no more than $p-q-1$ times differentiable, and thus we still need to make sense of the left-hand side of (1.3). By polynomial long division we may choose a polynomial R of order $p-q$ such that

$$S(z) := Q(z)R(z) - P(z), \qquad z \in \mathbb{C},$$

is a polynomial of at most order $q-1$. Now observe that

$$\begin{aligned}
\mathcal{F}\Big[\sum_{j=0}^{\infty} \phi_j D^j X\Big](y) &= \Big(\sum_{j=0}^{\infty} \phi_j (-iy)^j\Big)\mathcal{F}[X](y) \\
&= \Big(R(-iy) - \frac{S(-iy)}{Q(-iy)}\Big)\mathcal{F}[X](y) \\
&= \mathcal{F}[R(D)X](y) - \mathcal{F}[f](y)\mathcal{F}[X](y),
\end{aligned}$$

where $f:\mathbb{R}\to\mathbb{R}$ is the L^2 function characterized by $\mathcal{F}[f](y) = S(-iy)/Q(-iy)$ for $y\in\mathbb{R}$. (In fact, we even know that f is vanishing on $(-\infty,0)$ and decays exponentially fast at ∞, cf. Remark 4.10.) Combining this identity with (1.3) results in the representation (1.2).

We show in Theorem 4.8 that (1.2) does indeed hold true for any invertible (Lévy-driven) CARMA process. Similar relations are shown to hold for invertible fractionally integrated CARMA (FICARMA) processes, where $(Z_t)_{t\in\mathbb{R}}$ is a fractional Lévy process, and also for their multi-dimensional counterparts, which we will refer to as MCARMA and MFICARMA processes, respectively. We use these representations to obtain a prediction formula for general CARMA type processes (see Corollary 4.11). A prediction formula for invertible one-dimensional Lévy-driven CARMA processes is given in [9, Theorem 2.7], but prediction of MCARMA processes has, to the best of our knowledge, not been studied in the literature.

AR representations such as (1.2) are useful for several reasons. To give a few examples, they separate the noise $(Z_t)_{t\in\mathbb{R}}$ from $(X_t)_{t\in\mathbb{R}}$ and hence provide a recipe for recovering increments of the noise from the observed process, they ease the task of prediction (and thus estimation), and they clarify the dynamic behavior of the process. These facts motivate the idea of defining a broad class of processes, including the CARMA type processes above, which all admit an AR representation, and it turns out that a well-suited class to study is the one formed by solutions to multi-dimensional stochastic delay differential equations (MSDDEs). To be precise, for an integrable n-dimensional (measurable) process $Z_t = (Z_t^1,\dots,Z_t^n)^T$, $t\in\mathbb{R}$, with stationary increments and a finite signed $n\times n$ matrix-valued measure η, concentrated on $[0,\infty)$, a stationary process $X_t = (X_t^1,\dots,X_t^n)^T$, $t\in\mathbb{R}$, is a solution to the associated MSDDE if it satisfies

$$dX_t = \eta * X(t)\,dt + dZ_t. \tag{1.4}$$

By equation (1.4) we mean that

$$X_t^j - X_s^j = \sum_{k=1}^n \int_s^t \int_{[0,\infty)} X_{u-v}^k\,\eta_{jk}(dv)\,du + Z_t^j - Z_s^j, \quad j=1,\dots,n, \tag{1.5}$$

almost surely for each $s<t$. This system of equations is an extension of the stochastic delay differential equation (SDDE) in [3, Section 3.3] to the multivariate case. The overall structure of (1.4) is also in line with earlier literature such as [16, 20] on univariate SDDEs, but here we allow for infinite delay (η is allowed to have unbounded support) which is a key property in order to include the CARMA type processes in the framework.

The structure of the paper is as follows: In Section 2 we introduce the notation used throughout this paper. Next, in Section 3, we develop the general theory for MSDDEs with particular focus on existence, uniqueness and prediction. The general results of Section 3 are then specialized in Section 4 to various settings. Specifically, in Section 4.1 we consider the case where the noise process gives rise to a reasonable integral, and in Section 4.2 we demonstrate how to derive results for higher order SDDEs by nesting them into MSDDEs. Finally, in Section 4.3 we use the above mentioned findings to represent CARMA processes and generalizations thereof as solutions to higher order SDDEs and to obtain the corresponding prediction formulas.

2 Notation

Let $f : \mathbb{R} \to \mathbb{C}^{m\times k}$ be a measurable function and μ a $k \times n$ (non-negative) matrix measure, that is,

$$\mu = \begin{bmatrix} \mu_{11} & \cdots & \mu_{1n} \\ \vdots & \ddots & \vdots \\ \mu_{k1} & \cdots & \mu_{kn} \end{bmatrix}$$

where each μ_{jl} is a measure on $\mathbb{R}$. Then, we will write $f \in L^p(\mu)$ if

$$\int_{\mathbb{R}} |f_{il}(u)|^p \mu_{lj}(du) < \infty$$

for $l = 1,\ldots,k$, $i = 1,\ldots,m$ and $j = 1,\ldots,n$. Provided that $f \in L^1(\mu)$, we set

$$\int_{\mathbb{R}} f(u)\mu(du) = \sum_{l=1}^{k} \begin{bmatrix} \int_{\mathbb{R}} f_{1l}(u)\mu_{l1}(du) & \cdots & \int_{\mathbb{R}} f_{1l}(u)\mu_{ln}(du) \\ \vdots & \ddots & \vdots \\ \int_{\mathbb{R}} f_{ml}(u)\mu_{l1}(du) & \cdots & \int_{\mathbb{R}} f_{ml}(u)\mu_{ln}(du) \end{bmatrix}. \tag{2.1}$$

If μ is the Lebesgue measure, we will suppress the dependence on the measure and write $f \in L^p$, and in case f is measurable and bounded Lebesgue almost everywhere, $f \in L^\infty$. For two (matrix) measures μ^+ and μ^- on $\mathbb{R}$, where at least one of them are finite, we call the set function $\mu(B) := \mu^+(B) - \mu^-(B)$, defined for any Borel set B, a signed measure (and, from this point, simply referred to as a measure). We may and do assume that the two measures μ^+ and μ^- are singular. To the measure μ we will associate its variation measure $|\mu| := \mu^+ + \mu^-$, and when $|\mu|(\mathbb{R}) < \infty$, we will say that μ is finite. Integrals with respect to μ are defined in a natural way from (2.1) whenever $f \in L^1(\mu) := L^1(|\mu|)$. If f is one-dimensional, respectively if μ is one-dimensional, we will write $f \in L^1(\mu)$ if $f \in L^1(|\mu_{ij}|)$ for all $i = 1,\ldots,k$ and $j = 1,\ldots,n$, respectively if $f_{ij} \in L^1(|\mu|)$ for all $i = 1,\ldots,m$ and $j = 1,\ldots,k$. The associated integral is defined in an obvious manner.

We define the convolution at a given point $t \in \mathbb{R}$ by

$$f * \mu(t) = \int_{\mathbb{R}} f(t-u)\mu(du)$$

provided that $f(t-\cdot) \in L^1(\mu)$. In case that μ is the Lebesgue-Stieltjes measure of a function $g : \mathbb{R} \to \mathbb{R}^{k\times n}$ we will also write $f * g(t)$ instead of $f * \mu(t)$ (not to be confused with the standard convolution between functions). For a given measure μ we set

$$D(\mu) = \left\{ z \in \mathbb{C} : \int_{\mathbb{R}} e^{\mathrm{Re}(z)u} |\mu_{ij}|(du) < \infty \quad \text{for } i = 1,\ldots,k \text{ and } j = 1,\ldots,n \right\}$$

and define its Laplace transform $\mathcal{L}[\mu]$ as

$$\mathcal{L}[\mu]_{ij}(z) = \int_{\mathbb{R}} e^{zu}\,\mu_{ij}(du), \quad \text{for} \quad i = 1,\ldots,k,\ j = 1,\ldots,n,$$

for every $z \in D(\mu)$. If μ is a finite measure, we will also refer to the Fourier transform $\mathcal{F}[\mu]$ of μ, which is given as $\mathcal{F}[\mu](y) = \mathcal{L}[\mu](iy)$ for $y \in \mathbb{R}$. If $\mu(du) = f(u)\,du$ for some measurable function f, we write $\mathcal{L}[f]$ and $\mathcal{F}[f]$ instead. We will also use that the Fourier transform $\mathcal{F}$ extends from L^1 to $L^1 \cup L^2$, and it maps L^2 onto L^2. We will say that μ has a moment of order $p \in \mathbb{N}_0$ if

$$\int_{\mathbb{R}} |u|^p \, |\mu_{jk}|(du) < \infty$$

for all $j,k = 1,\dots,n$. Finally, for two functions $f,g : \mathbb{R} \to \mathbb{R}$ and $a \in [-\infty,\infty]$, we write $f(t) = o(g(t))$, $f(t) \sim g(t)$ and $f(t) = O(g(t))$ as $t \to a$ if

$$\lim_{t\to a} \frac{f(t)}{g(t)} \to 0, \quad \lim_{t\to a} \frac{f(t)}{g(t)} = 1 \quad \text{and} \quad \limsup_{t\to a} \left|\frac{f(t)}{g(t)}\right| < \infty,$$

respectively.

3 Stochastic delay differential equations

Consider the general MSDDE in (1.4). The first main result provides sufficient conditions to ensure existence and uniqueness of a solution. To obtain such results we need to put assumptions on the delay measure η. In order to do so, we associate to η the function $h : D(\eta) \to \mathbb{C}^{n\times n}$ given by

$$h(z) = -zI_n - \mathcal{L}[\eta](z). \tag{3.1}$$

where I_n is the $n \times n$ identity matrix.

Theorem 3.1. *Let h be given in (3.1) and suppose that* $\det(h(iy)) \neq 0$ *for all $y \in \mathbb{R}$. Suppose further that η has second moment. Then there exists a function $g : \mathbb{R} \to \mathbb{R}^{n\times n}$ in L^2 characterized by*

$$\mathcal{F}[g](y) = h(iy)^{-1}, \tag{3.2}$$

the convolution

$$g * Z(t) := Z_t + \int_{\mathbb{R}} g * \eta(t-u) Z_u \, du \tag{3.3}$$

is well-defined for each $t \in \mathbb{R}$ almost surely, and $X_t = g * Z(t)$, $t \in \mathbb{R}$, is the unique (up to modification) stationary and integrable solution to (1.4). If, in addition to the above stated assumptions, $\det(h(z)) \neq 0$ *for all $z \in \mathbb{C}$ with* $\mathrm{Re}(z) \leq 0$ *then the solution in* (3.3) *is casual in the sense that $(X_t)_{t\in\mathbb{R}}$ is adapted to the filtration*

$$\{\sigma(Z_t - Z_s \,:\, s < t)\}_{t\in\mathbb{R}}.$$

The solution $(X_t)_{t\in\mathbb{R}}$ to (1.4) will very often take form as a $(Z_t)_{t\in\mathbb{R}}$-driven moving average, that is,

$$X_t = \int_{\mathbb{R}} g(t-u) \, dZ_u \tag{3.4}$$

for each $t \in \mathbb{R}$ (cf. Section 4.1). This fact justifies the notation $g * Z$ introduced in (3.3). In case $n = 1$, equation (1.4) reduces to the usual first order SDDE, and then the existence condition becomes $h(iy) = -iy - \mathcal{F}[\eta](y) \neq 0$ for all $y \in \mathbb{R}$, and the kernel driving the solution is characterized by $\mathcal{F}[g](y) = 1/h(iy)$. This is consistent with earlier literature (cf. [3, 16, 20]).

The second main result concerns prediction of MSDDEs. In particular, the content of the result is that we can compute a prediction of future values of the observed process if we are able to compute the same type of prediction of the noise.

Theorem 3.2. *Suppose that* $\det(h(z)) \neq 0$ *for all* $z \in \mathbb{C}$ *with* $Re(z) \leq 0$ *and that* η *has second moment. Furthermore, let* $(X_t)_{t\in\mathbb{R}}$ *be the stationary and integrable solution to* (1.4) *and let* g *be given by* (3.2)*. Fix* $s < t$*. Then, if we set*

$$\hat{Z}_u = \mathbb{E}[Z_u - Z_s \mid Z_s - Z_r,\, r < s], \quad u > s, \tag{3.5}$$

it holds that

$$\begin{aligned}&\mathbb{E}[X_t \mid X_u,\, u \leq s]\\ &\quad = g(t-s)X_s + \int_s^t g(t-u)\eta * \Big\{\mathbb{1}_{(-\infty,s]}X\Big\}(u)\,du + g * \Big\{\mathbb{1}_{(s,\infty)}\hat{Z}\Big\}(t),\end{aligned}$$

using the notation

$$\begin{aligned}\Big(\eta * \{\mathbb{1}_{(-\infty,s]}X\}(u)\Big)_j &:= \sum_{k=1}^n \int_{[u-s,\infty)} X_{u-v}^k\, \eta_{jk}(dv) \quad \textit{and}\\ \Big(g * \{\mathbb{1}_{(s,\infty)}\hat{Z}\}(u)\Big)_j &:= \sum_{k=1}^n \int_{[0,u-s)} \hat{Z}_{u-v}^k\, g_{jk}(dv)\end{aligned}$$

for $u > s$ *and* $j = 1,\dots,n$.

Remark 3.3. In case $(Z_t)_{t\in\mathbb{R}}$ is a Lévy process, the prediction formula in Theorem 3.2 simplifies, since $\hat{Z}_u = (u-s)\mathbb{E}[Z_1]$ and thus

$$\begin{aligned}&\mathbb{E}[X_t \mid X_u,\, u \leq s]\\ &\quad = g(t-s)X_s + \int_s^t g(t-u)\eta * \Big\{\mathbb{1}_{(-\infty,s]}X\Big\}(u)\,du + \int_s^t g(t-u)\,du\,\mathbb{E}[Z_1],\end{aligned}$$

using integration by parts. Obviously, the formula takes an even simpler form if $\mathbb{E}[Z_1] = 0$. If instead we are in a long memory setting and $(Z_t)_{t\in\mathbb{R}}$ is a fractional Brownian motion, we can rely on [15] to obtain $(\hat{Z}_u)_{s<u\leq t}$ and then use the formula given in Theorem 3.2 to compute the prediction $\mathbb{E}[X_t \mid X_u,\, u \leq s]$.

In Section 4.3 we use this prediction formula combined with the relation between MSDDEs and MCARMA processes to obtain a prediction formula for any invertible MCARMA process.

4 Examples and further results

In this section we will consider several examples of MSDDEs and give some additional results. We begin by defining what we mean by a regular integrator, since this makes it possible to have the compact form (3.4) of the solution to (1.4) in most cases. Next, we show how one can nest higher order MSDDEs in the (first order) MSDDE framework. Finally, we show that invertible MCARMA processes (and some generalizations) form a particular subclass of solutions to higher order MSDDEs.

4.1 Regular integrators and moving average representations

When considering the form of the solution in Theorem 3.1 it is natural to ask if this can be seen as a moving average of the kernel g with respect to the noise $(Z_t)_{t\in\mathbb{R}}$, that is, if

$$X_t^j = \left(\int_{\mathbb{R}} g(t-u)\,dZ_u\right)_j = \sum_{k=1}^n \int_{\mathbb{R}} g_{jk}(t-u)\,dZ_u^k, \quad t\in\mathbb{R}, \tag{4.1}$$

for $j = 1,\dots,n$. The next result shows that the answer is positive if $(Z_t^k)_{t\in\mathbb{R}}$ is a "reasonable" integrator for a suitable class of deterministic integrands for each $k = 1,\dots,n$.

Proposition 4.1. *Let h be the function given in (3.1) and suppose that, for all $y\in\mathbb{R}$, $\det(h(iy))\neq 0$. Suppose further that η has second moment and let $(X_t)_{t\in\mathbb{R}}$ be the solution to* (1.4) *given by* (3.3)*. Finally assume that, for each $k = 1,\dots,n$, there exists a linear map $I_k : L^1\cap L^2 \to L^1(\mathbb{P})$ which has the following properties:*

(i) *For all $s<t$, $I_k(\mathbb{1}_{(s,t]}) = Z_t^k - Z_s^k$.*

(ii) *If μ is a finite Borel measure on $\mathbb{R}$ having first moment then*

$$I_k\left(\int_{\mathbb{R}} f_r(t-\cdot)\,\mu(dr)\right) = \int_{\mathbb{R}} I_k(f_r(t-\cdot))\,\mu(dr) \tag{4.2}$$

almost surely for all $t\in\mathbb{R}$, where $f_r = \mathbb{1}_{[0,\infty)}(\cdot - r) - \mathbb{1}_{[0,\infty)}$ for $r\in\mathbb{R}$.

Then it holds that

$$X_t^j = \sum_{k=1}^n I_k(g_{jk}(t-\cdot)), \quad j = 1,\dots,n, \tag{4.3}$$

almost surely for each $t\in\mathbb{R}$. In this case, $(Z_t)_{t\in\mathbb{R}}$ will be called a regular integrator and we will write $\int \cdot\, dZ^k = I_k$.

The typical example of a regular integrator is a multi-dimensional Lévy process:

Example 4.2. Suppose that $(Z_t)_{t\in\mathbb{R}}$ is an n-dimensional integrable Lévy process. Then, in particular, each $(Z_t^j)_{t\in\mathbb{R}}$ is an integrable (one-dimensional) Lévy process, and in [3, Lemma 5.3] it is shown that the integral $\int_{\mathbb{R}} f(u)\,dZ_u^j$ is well-defined in the sense of [21] and belongs to $L^1(\mathbb{P})$ if $f\in L^1\cap L^2$. Moreover, the stochastic Fubini result given in [2, Theorem 3.1] implies in particular that condition (ii) of Proposition 4.1 is satisfied, which shows that $(Z_t)_{t\in\mathbb{R}}$ is a regular integrator and that (4.1) holds.

We will now show that a class of multi-dimensional fractional Lévy processes can serve as regular integrators as well (cf. Example 4.4 below). Fractional noise processes are often used as a tool to incorporate (some variant of) long memory in the corresponding solution process. As will appear, the integration theory for fractional Lévy processes we will use below relies on the ideas of [17], but is extended to allow for symmetric stable Lévy processes as well. For more on fractional stable Lévy processes, the so-called linear fractional stable motions, we refer to [22, p. 343]. First, however, we will need the following observation:

Proposition 4.3. *Let $f : \mathbb{R} \to \mathbb{R}$ be a function in $L^1 \cap L^\alpha$ for some $\alpha \in (1,2]$. Then the right-sided Riemann-Liouville fractional integral*

$$I_-^\beta f : t \mapsto \frac{1}{\Gamma(\beta)} \int_t^\infty f(u)(u-t)^{\beta-1}\,du \tag{4.4}$$

is well-defined and belongs to L^α for any $\beta \in (0, 1-1/\alpha)$.

Example 4.4. Let $\alpha = (\alpha_1,\dots,\alpha_n)$ with $\alpha_j \in (1,2]$ and $f = (f_{jk}) : \mathbb{R} \to \mathbb{R}^{n\times n}$ be a function such that $f_{jk} \in L^1 \cap L^{\alpha_k}$ for $j,k = 1,\dots,n$. Consider an n-dimensional Lévy process $(L_t)_{t\in\mathbb{R}}$ where its j-th coordinate is symmetric α_j-stable if $\alpha_j \in (1,2)$ and mean zero and square integrable if $\alpha_j = 2$. Then, for a given vector $\beta = (\beta_1,\dots,\beta_n)$ with $\beta_j \in (0, 1-1/\alpha_j)$ for $j = 1,\dots,n$ the corresponding fractional Lévy process $(Z_t)_{t\in\mathbb{R}}$ with parameter β is defined as

$$\begin{aligned} Z_t^j &= \int_{\mathbb{R}} \Big(I_-^{\beta_j}[\mathbb{1}_{(-\infty,t]} - \mathbb{1}_{(-\infty,0]}]\Big)(u)\,dL_u^j \\ &= \frac{1}{\Gamma(1+\beta_j)} \int_{\mathbb{R}} \Big[(t-u)_+^{\beta_j} - (-u)_+^{\beta_j}\Big]\,dL_u^j \end{aligned}$$

for $t \in \mathbb{R}$ and $j = 1,\dots,n$, and where $x_+ = \max\{x,0\}$. In light of Proposition 4.3, this definition makes it natural to define the integral of a function $f : \mathbb{R} \to \mathbb{R}$ in $L^1 \cap L^{\alpha_j}$ (particularly in $L^1 \cap L^2$) with respect to $(Z_t^j)_{t\in\mathbb{R}}$ as

$$\int_{\mathbb{R}} f(u)\,dZ_u^j = \int_{\mathbb{R}} \Big(I_-^{\beta_j} f\Big)(u)\,dL_u^j$$

for $j = 1,\dots,n$. Note that the integral belongs to $L^2(\mathbb{P})$ for $\alpha_j = 2$ and to $L^\gamma(\mathbb{P})$ for any $\gamma < \alpha_j$ if $\alpha_j \in (1,2)$. Using Proposition 4.3 and the stochastic Fubini result given in [2, Theorem 3.1] for $(L_t^j)_{t\in\mathbb{R}}$ it is straightforward to verify that assumption (ii) of Proposition 4.1 is satisfied as well, and thus $(Z_t)_{t\in\mathbb{R}}$ is a regular integrator and the solution $(X_t)_{t\in\mathbb{R}}$ to (1.4) takes the moving average form (4.1).

At this point it should be clear that the conditions for being a regular integrator are mild, hence they will, besides the examples mentioned above, also be satisfied for a wide class of semimartingales with stationary increments.

4.2 Higher order (multivariate) SDDEs

An advantage of introducing the multivariate setting (1.4) is that we can nest higher order MSDDEs in this framework. Effectively, as usual and as will be demonstrated below, it is done by increasing the dimension accordingly.

Let $\varpi_0, \varpi_1, \dots, \varpi_{m-1}$ be (entrywise) finite $n\times n$ measures concentrated on $[0,\infty)$ which all admit second moment, and let $(Z_t)_{t\in\mathbb{R}}$ be an n-dimensional integrable stochastic process with stationary increments. For convenience we will assume that $(Z_t)_{t\in\mathbb{R}}$ is a regular integrator in the sense of Proposition 4.1. We will say that an n-dimensional stationary, integrable and measurable process $(X_t)_{t\in\mathbb{R}}$ satisfies the corresponding m-th order MSDDE if it is $m-1$ times differentiable and

$$dX_t^{(m-1)} = \sum_{j=0}^{m-1} \varpi_j * X^{(j)}(t)\,dt + dZ_t \tag{4.5}$$

where $(X_t^{(j)})_{t\in\mathbb{R}}$ denotes the entrywise j-th derivative of $(X_t)_{t\in\mathbb{R}}$ with respect to t. By (4.5) we mean that

$$\left(X_t^{(m-1)}\right)^k - \left(X_s^{(m-1)}\right)^k = \sum_{j=0}^{m-1}\sum_{l=1}^{n}\int_s^t \int_{[0,\infty)} \left(X_{u-v}^{(j)}\right)^l (\varpi_j)_{kl}(dv)\,du + Z_t^k - Z_s^k$$

for $k = 1,\dots,n$ and each $s < t$ almost surely. Equation (4.5) corresponds to the mn-dimensional MSDDE in (1.4) with noise $(0,\dots,0,Z_t^T)^T \in \mathbb{R}^{mn}$ and

$$\eta = \begin{bmatrix} 0 & I_n\delta_0 & 0 & \cdots & 0 \\ 0 & 0 & I_n\delta_0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & I_n\delta_0 \\ \varpi_0 & \varpi_1 & \varpi_2 & \cdots & \varpi_{m-1} \end{bmatrix}. \tag{4.6}$$

(If $n = 1$ then $\eta = \varpi_0$.) With η given by (4.6) it follows that

$$D(\eta) = \bigcap_{j=0}^{m-1} D(\varpi_j)$$

and

$$h(z) = -\begin{bmatrix} I_n z & I_n & 0 & \cdots & 0 \\ 0 & I_n z & I_n & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & I_n z & I_n \\ \mathcal{L}[\varpi_0](z) & \mathcal{L}[\varpi_1](z) & \cdots & \mathcal{L}[\varpi_{m-2}](z) & I_n z + \mathcal{L}[\varpi_{m-1}](z) \end{bmatrix}$$

for $z \in D(\eta)$. In general, we know from Theorem 3.1 that a solution to (4.5) exists if $\det(h(iy)) \neq 0$ for all $y \in \mathbb{R}$, and in this case the unique solution is given by

$$X_t = \int_{\mathbb{R}} g_{1m}(t-u)\,dZ_u, \quad t \in \mathbb{R}, \tag{4.7}$$

where $\mathcal{F}[g_{1m}]$ is characterized as entrance $(1,m)$ in the $n\times n$ block representation of $h(i\cdot)^{-1}$. In other words, if e_j denotes the j-th canonical basisvector of $\mathbb{R}^m$ and $\otimes$ the Kronecker product,

$$\mathcal{F}[g_{1m}](y) = (e_1 \otimes I_n)^T h(iy)^{-1}(e_m \otimes I_n)$$

for $y \in \mathbb{R}$. However, due to the particular structure of η in (4.6) we can simplify these expressions:

Theorem 4.5. *Let the setup be as above. Then it holds that*

$$\det(h(z)) = \det\left(I_n(-z)^m - \sum_{j=0}^{m-1} \mathcal{L}[\varpi_j](z)(-z)^j\right) \tag{4.8}$$

for all $z \in D(\eta)$*, and if* $\det(h(iy)) \neq 0$ *for all* $y \in \mathbb{R}$*, there exists a unique solution to (4.5) and it is given as (4.7) where* $g : \mathbb{R} \to \mathbb{R}^{n\times n}$ *is characterized by*

$$\mathcal{F}[g_{1m}](y) = \left(I_n(-iy)^m - \sum_{j=0}^{m-1} \mathcal{F}[\varpi_j](y)(-iy)^j \right)^{-1} \tag{4.9}$$

for $y \in \mathbb{R}$*. The solution is causal if* $\det(h(z)) \neq 0$ *whenever* $Re(z) \leq 0$*.*

Observe that, as should be the case, we are back to the first order MSDDE when $m = 1$ and (4.8)-(4.9) agree with Theorem 3.1. As we will see in Section 4.3 below, one motivation for introducing higher order MSDDEs of the form (4.5) and to study the structure of the associated solutions, is their relation to MCARMA processes. However, we start with the multivariate CAR(p) process, where no delay term will be present, as an example:

Example 4.6. Let $P(z) = I_n z^p + A_1 z^{p-1} + \cdots + A_p$, $z \in \mathbb{C}$, for suitable $A_1, \ldots, A_p \in \mathbb{R}^{n\times n}$. The associated CAR($p$) process $(X_t)_{t\in\mathbb{R}}$ with noise $(Z_t)_{t\in\mathbb{R}}$ can be thought of as formally satisfying $P(D)X_t = DZ_t$, $t \in \mathbb{R}$, where D denotes differentiation with respect to t. Integrating both sides and rearranging terms gives

$$dX_t^{(p-1)} = -\sum_{j=0}^{p-1} A_{p-j} X_t^{(j)}\, dt + dZ_t, \quad t \in \mathbb{R}, \tag{4.10}$$

which is of the form (4.5) with $m = p$ and $\varpi_j = -A_{p-j}\delta_0$ for $j = 0, 1, \ldots, p-1$. Proposition 4.5 shows that a unique solution exists if

$$\det\left(I_n(iy)^p + \sum_{j=0}^{p-1} A_{p-j}(iy)^j \right) = \det(P(iy)) \neq 0$$

for all $y \in \mathbb{R}$, and in this case $\mathcal{F}[g_{1m}](y) = P(-iy)^{-1}$ for $y \in \mathbb{R}$. This agrees with the rigorous definition of the CAR(p) process, see e.g. [19]. In case $p = 1$, (4.10) collapses to the multivariate Ornstein-Uhlenbeck equation

$$dX_t = -A_1 X_t\, dt + dZ_t, \quad t \in \mathbb{R},$$

and if the eigenvalues of A_1 are all positive, it is easy to check that $g_{1m}(t) = e^{-A_1 t}\mathbb{1}_{[0,\infty)}(t)$ so that the unique solution $(X_t)_{t\in\mathbb{R}}$ is causal and takes the well-known form

$$X_t = \int_{-\infty}^{t} e^{-A_1(t-u)}\, dZ_u \tag{4.11}$$

for $t \in \mathbb{R}$. Lévy-driven multivariate Ornstein-Uhlenbeck processes have been studied extensively in the literature, and the moving average structure (4.11) of the solution is well-known when $(Z_t)_{t\in\mathbb{R}}$ is a Lévy process. We refer to [1, 23, 24] for further details. The one-dimensional case where $(Z_t)_{t\in\mathbb{R}}$ is allowed to be a general stationary increment process has been studied in [2].

4.3 Relations to MCARMA processes

Let $p \in \mathbb{N}$ and define the polynomials $P, Q : \mathbb{C} \to \mathbb{C}^{n\times n}$ by

$$\begin{aligned} P(z) &= I_n z^p + A_1 z^{p-1} + \cdots + A_p \quad \text{and} \\ Q(z) &= B_0 + B_1 z + \cdots + B_{p-1} z^{p-1} \end{aligned} \tag{4.12}$$

for $z \in \mathbb{C}$ and suitable $A_1, \dots, A_p, B_0, \dots, B_{p-1} \in \mathbb{R}^{n\times n}$. We will also fix $q \in \mathbb{N}_0$, $q < p$, and set $B_q = I_n$ and $B_j = 0$ for all $q < j < p$. It will always be assumed that $\det(P(iy)) \neq 0$ for all $y \in \mathbb{R}$. Under this assumption there exists a function $\tilde{g} : \mathbb{R} \to \mathbb{R}^{n\times n}$ which is in $L^1 \cap L^2$ and

$$\mathcal{F}[\tilde{g}](y) = P(-iy)^{-1} Q(-iy) \tag{4.13}$$

for every $y \in \mathbb{R}$. Consequently, for any regular integrator $(Z_t)_{t\in\mathbb{R}}$ in the sense of Proposition 4.1, the n-dimensional stationary and integrable process $(X_t)_{t\in\mathbb{R}}$ given by

$$X_t = \int_{\mathbb{R}} \tilde{g}(t-u)\, dZ_u, \quad t \in \mathbb{R}, \tag{4.14}$$

is well-defined. If it is additionally assumed that $\det(P(z)) \neq 0$ for $z \in \mathbb{C}$ with $\mathrm{Re}(z) \geq 0$ then it is argued in [19] that

$$\tilde{g}(t) = \mathbb{1}_{[0,\infty)}(t)(e_1^p \otimes I_n)^T e^{At} E \tag{4.15}$$

where

$$A = \begin{bmatrix} 0 & I_n & 0 & \cdots & 0 \\ 0 & 0 & I_n & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & I_n \\ -A_p & -A_{p-1} & \cdots & -A_2 & -A_1 \end{bmatrix} \quad \text{and} \quad E = \begin{bmatrix} E_1 \\ \vdots \\ E_p \end{bmatrix},$$

with $E(z) = E_1 z^{p-1} + \cdots + E_p$ chosen such that

$$z \mapsto P(z)E(z) - Q(z)z^p$$

is at most of degree $p-1$. (Above, and henceforth, we use the notation e_j^k for the j-th canonical basis vector of $\mathbb{R}^k$.) We will refer to the process $(X_t)_{t\in\mathbb{R}}$ as a $(Z_t)_{t\in\mathbb{R}}$-driven MCARMA(p,q) process. For instance, when $(Z_t)_{t\in\mathbb{R}}$ is an n-dimensional Lévy process, $(X_t)_{t\in\mathbb{R}}$ is a (Lévy-driven) MCARMA(p,q) process as introduced in [19]. If $(L_t)_{t\in\mathbb{R}}$ is an n-dimensional second order Lévy process with mean zero, and

$$Z_t^j = \frac{1}{\Gamma(1+\beta_j)} \int_{\mathbb{R}} \left[(t-u)_+^{\beta_j} - (-u)_+^{\beta_j} \right] dL_u^j, \quad t \in \mathbb{R},$$

for $\beta_j \in (0, 1/2)$ and $j = 1, \dots, n$, then $(X_t)_{t\in\mathbb{R}}$ is an MFICARMA(p,β,q) process, $\beta = (\beta_1, \dots, \beta_n)$, as studied in [18]. For the univariate case ($n = 1$), the processes above correspond to the CARMA(p,q) and FICARMA(p,β_1,q) process, respectively. The class of CARMA processes has been studied extensively, and we refer to the references in the introduction for details.

Remark 4.7. Observe that, generally, Lévy-driven MCARMA (hence CARMA) processes are defined even when $(Z_t)_{t\in\mathbb{R}}$ has no more than log moments. However, it relies heavily on the fact that $\tilde{g}$ and $(Z_t)_{t\in\mathbb{R}}$ are well-behaved enough to ensure that the process in (4.14) remains well-defined. At this point, a setup where the noise does not admit a first moment has not been integrated in a framework as general as that of (1.4).

In the following our aim is to show that, under a suitable invertibility assumption, the $(Z_t)_{t\in\mathbb{R}}$-driven MCARMA(p,q) process given in (4.14) is the unique solution to a certain (possibly higher order) MSDDE of the form (4.5). Before formulating the main result of this section we introduce some notation. To P and Q defined in (4.12) we will associate the unique polynomial $R(z) = I_n z^{p-q} + C_{p-q-1} z^{p-q-1} + \cdots + C_0$, $z \in \mathbb{C}$ and $C_0, C_1, \ldots, C_{p-q-1} \in \mathbb{R}^{n\times n}$, having the property that

$$z \mapsto Q(z)R(z) - P(z) \tag{4.16}$$

is a polynomial of at most order $q-1$ (see the introduction for an intuition about why this property is desirable).

Theorem 4.8. *Let P and Q be given as in (4.12), and let $(X_t)_{t\in\mathbb{R}}$ be the associated $(Z_t)_{t\in\mathbb{R}}$-driven MCARMA(p,q) process. Suppose that* $\det(Q(z)) \neq 0$ *for all $z \in \mathbb{C}$ with* $\mathrm{Re}(z) \geq 0$. *Then $(X_t)_{t\in\mathbb{R}}$ is the unique solution to (4.5) with*

$$m = p - q, \quad \varpi_0(du) = -C_0\delta_0(du) + f(u)\,du, \quad \textit{and} \quad \varpi_j = -C_j\delta_0,$$

for $1 \leq j \leq m-1$ or, written out,

$$dX_t^{(m-1)} = -\sum_{j=0}^{m-1} C_j X_t^{(j)}\,dt + \left(\int_0^\infty X_{t-u}^T f(u)^T\,du\right)^T dt + dZ_t, \tag{4.17}$$

where $C_0, \ldots, C_{m-1} \in \mathbb{R}^{n\times n}$ are defined as in (4.16) above, $(X_t^{(j)})_{t\in\mathbb{R}}$ is the j-th derivative of $(X_t)_{t\in\mathbb{R}}$, and where $f : \mathbb{R} \to \mathbb{R}^{n\times n}$ is characterized by

$$\mathcal{F}[f](y) = R(-iy) - Q(-iy)^{-1}P(-iy). \tag{4.18}$$

It follows from Theorem 4.8 that $p-q$ is the order of the (possibly multivariate) SDDE we can associate with a (possibly multivariate) CARMA process. Thus, this seems as a natural extension of [3], where the univariate first order SDDE is studied and related to the univariate CARMA$(p,p-1)$ process.

Remark 4.9. An immediate consequence of Theorem 4.8 is that we obtain an inversion formula for $(Z_t)_{t\in\mathbb{R}}$-driven MCARMA processes. In other words, it shows how to recover the increments of $(Z_t)_{t\in\mathbb{R}}$ from observing $(X_t)_{t\in\mathbb{R}}$. For this reason it seems natural to impose the invertibility assumption $\det(Q(z)) \neq 0$ for all $z \in \mathbb{C}$ with $\mathrm{Re}(z) \geq 0$, which is the direct analogue of the one for discrete time ARMA processes (or, more generally, moving averages). It is usually referred to as the minimum phase property in signal processing. The inversion problem for (Lévy-driven) CARMA processes has been studied in [3, 7, 8, 9] and for (Lévy-driven) MCARMA processes in [11]. In both cases a different approach, which does not rely on MSDDEs, is used.

Remark 4.10. Since the Fourier transform $\mathcal{F}[f]$ of the function f defined in Theorem 4.8 is rational, one can determine f explicitly (e.g., by using the partial fraction expansion of $\mathcal{F}[f]$). Indeed, since the Fourier transform of f is of the same form as the Fourier transform of the solution kernel $\tilde{g}$ of the MCARMA process we can deduce that

$$f(t) = (e_1^q \otimes I_n)^T e^{Bt} F, \quad t \geq 0, \tag{4.19}$$

with

$$B = \begin{bmatrix} 0 & I_n & 0 & \cdots & 0 \\ 0 & 0 & I_n & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & I_n \\ -B_0 & -B_1 & \cdots & -B_{q-2} & -B_{q-1} \end{bmatrix} \quad \text{and} \quad F = \begin{bmatrix} F_1 \\ \vdots \\ F_q \end{bmatrix},$$

where $F(z) = F_1 z^{q-1} + \cdots + F_q$ is chosen such that

$$z \mapsto Q(z)F(z) - [Q(z)R(z) - P(z)]z^q$$

is at most of degree $q-1$ (see (4.13) and (4.15)).

In Corollary 4.11 we formulate the prediction formula in Theorem 3.2 in the special case where $(X_t)_{t\in\mathbb{R}}$ is a $(Z_t)_{t\in\mathbb{R}}$-driven MCARMA process. In the formulation we use the definition

$$\hat{Z}_u = \mathbb{E}[Z_u - Z_s \mid Z_s - Z_r,\, r < s], \quad u > s,$$

in line with (3.5).

Corollary 4.11. *Let $(X_t)_{t\in\mathbb{R}}$ be a $(Z_t)_{t\in\mathbb{R}}$-driven MCARMA process and set*

$$\tilde{g}_j(t) = (e_1^p \otimes I_n)^T e^{At} \sum_{k=j}^{p-q} A^{k-j} E C_k, \quad t \geq 0,$$

for $j = 1, \dots, p-q$, where $C_0, \dots, C_{p-q-1}$ are given in (4.16) *and $C_{p-q} = I_n$. Suppose that* $\det(P(z)) \neq 0$ *and* $\det(Q(z)) \neq 0$ *for all $z \in \mathbb{C}$ with* $\mathrm{Re}(z) \geq 0$. *Fix $s < t$. Then the following prediction formula holds*

$$\begin{aligned} \mathbb{E}[X_t \mid X_u,\, u \leq s] = {} & \sum_{j=1}^{p-q} \tilde{g}_j(t-s) X_s^{(j-1)} \\ & + \int_{-\infty}^{s} \int_{s}^{t} \tilde{g}(t-u) f(u-v)\, du\, X_v\, dv + \tilde{g} * \{\hat{Z} \mathbb{1}_{(s,\infty)}\}(t), \end{aligned}$$

where $\tilde{g}$ and f are given in (4.15) *and* (4.19), *respectively, and*

$$\tilde{g} * \{\hat{Z} \mathbb{1}_{(s,\infty)}\}(t) = \mathbb{1}_{\{p=q+1\}} \hat{Z}_u + (e_1^p \otimes I_n)^T A e^{At} \int_s^t e^{-Av} E \hat{Z}_v\, dv.$$

Example 4.12. To illustrate the results above we will consider an n-dimensional $(Z_t)_{t\in\mathbb{R}}$-driven MCARMA(3,1) process $(X_t)_{t\in\mathbb{R}}$ with P and Q polynomials given by

$$\begin{aligned}P(z) &= I_n z^3 + A_1 z^2 + A_2 z + A_3,\\ Q(z) &= B_0 + I_n z\end{aligned}$$

for matrices $B_0, A_1, A_2, A_3 \in \mathbb{R}^{n\times n}$ such that $\det(P(z)) \neq 0$ and $\det(Q(z)) \neq 0$ for all $z \in \mathbb{C}$ with $\mathrm{Re}(z) \geq 0$. According to (4.15), $(X_t)_{t\in\mathbb{R}}$ may be written as

$$X_t = \int_{-\infty}^t (e_1^3 \otimes I_n)^T e^{A(t-u)} E \, dZ_u$$

where $E_1 = 0$, $E_2 = I_n$, and $E_3 = B_0 - A_1$. With

$$C_1 = A_1 - B_0, \quad C_0 = A_2 + B_0(B_0 - A_1)$$

and

$$F = B_0(A_2 - B_0(A_1 - B_0)) - A_3,$$

Theorem 4.8 and Remark 4.10 imply that

$$dX_t^{(1)} = -C_1 X_t^{(1)} \, dt - C_0 X_t \, dt + \left(\int_0^\infty (F X_{t-u})^T e^{-B_0^T u} \, du \right)^T dt + dZ_t.$$

Moreover, by Corollary 4.11, we have the prediction formula

$$\begin{aligned}\mathbb{E}[X_t \mid X_u,\, u \leq s] = &(e_1^3 \otimes I_n)^T e^{At} \Bigg[(EC_1 + AE) X_s + E X_s^{(1)} \\ &+ \int_s^t e^{-Au} E \left(e^{B_0 u} \int_{-\infty}^s e^{-B_0 v} F X_v \, dv + \hat{Z}_u \right) du \Bigg].\end{aligned}$$

5 Proofs and auxiliary results

We will start this section by discussing some technical results. These results will then be used in the proofs of all the results stated above.

Recall the function $h : D(\eta) \to \mathbb{C}^{n\times n}$ defined in (3.1). Note that we always have $\{z \in \mathbb{C} : \mathrm{Re}(z) \leq 0\} \subseteq D(\eta)$ and $h(iy) = -iy I_n - \mathcal{F}[\eta](y)$ for $y \in \mathbb{R}$. Provided that η is sufficiently nice, Proposition 5.1 below ensures the existence of a kernel $g : \mathbb{R} \to \mathbb{R}^{n\times n}$ which will drive the solution to (1.4).

Proposition 5.1. *Let h be given as in (3.1) and suppose that* $\det(h(iy)) \neq 0$ *for all $y \in \mathbb{R}$. Then there exists a function $g = (g_{jk}) : \mathbb{R} \to \mathbb{R}^{n\times n}$ in L^2 characterized by*

$$\mathcal{F}[g](y) = h(iy)^{-1} \tag{5.1}$$

for $y \in \mathbb{R}$. Moreover, the following statements hold:

(i) *The function g satisfies*

$$g(t-r) - g(s-r) = \mathbb{1}_{(s,t]}(r) I_n + \int_s^t g * \eta(u-r) \, du$$

for almost all $r \in \mathbb{R}$ and each fixed $s < t$.

(ii) *If η has moment of order $p \in \mathbb{N}$, then $g \in L^q$ for all $q \in [1/p, \infty]$, and*

$$g(t) = \mathbb{1}_{[0,\infty)}(t) I_n + \int_{-\infty}^{t} g * \eta(u)\, du \tag{5.2}$$

for almost all $t \in \mathbb{R}$. In particular,

$$\int_{\mathbb{R}} g * \eta(u)\, du = -I_n. \tag{5.3}$$

(iii) *If $\int_{[0,\infty)} e^{\delta u} |\eta_{jk}|(du) < \infty$ for all $j,k = 1,\dots,n$ and some $\delta > 0$, then there exists $\varepsilon > 0$ such that*

$$\sup_{t\in\mathbb{R}} \max_{j,k=1,\dots,n} |g_{jk}(t)| e^{\varepsilon |t|} \leq C$$

for a suitable constant $C > 0$.

(iv) *If $\det(h(z)) \neq 0$ for all $z \in \mathbb{C}$ with $Re(z) \leq 0$ then g is vanishing on $(-\infty, 0)$ almost everywhere.*

Proof. In order to show the existence of g it suffices to argue that

$$y \mapsto \left(h(iy)^{-1}\right)_{jk} \text{ is in } L^2 \text{ for } j,k = 1,\dots,n, \tag{5.4}$$

since the Fourier transform $\mathcal{F}$ maps L^2 onto L^2. (Here $(h(iy)^{-1})_{jk}$ refers to the (j,k)-th entry in the matrix $h(iy)^{-1}$.) Indeed, in this case we just set $g_{jk} = \mathcal{F}^{-1}[(h(i\cdot)^{-1})_{jk}]$.

Let $\widehat{h(iy)}$ denote the $n \times n$ matrix which has the same rows as $h(iy)$, but where the j-th column is replaced by the k-th canonical basis vector (that is, the vector with all entries equal to zero except of the k-th entry which equals one). Then it follows by Cramer's rule that

$$\left(h(iy)^{-1}\right)_{jk} = \frac{\det(\widehat{h(iy)})}{\det(h(iy))}.$$

Recalling that $h(iy) = -iyI - \mathcal{F}[\eta](y)$ and that $\mathcal{F}[\eta](y)$ is bounded in y we get by the Leibniz formula that $|\det(h(iy))| \sim |y|^n$ and $|\det(\widehat{h(iy)})| = O(|y|^{n-1})$ as $|y| \to \infty$. This shows in particular that

$$\left|\left(h(iy)^{-1}\right)_{jk}\right| = O\left(|y|^{-1}\right) \tag{5.5}$$

as $|y| \to \infty$. Since j and k were arbitrarily chosen we get by continuity of (all the entries of) $y \mapsto h(iy)^{-1}$ that (5.4) holds, which ensures the existence part. The fact that $\overline{\mathcal{F}[g](-y)} = \mathcal{F}[g](y)$, $y \in \mathbb{R}$, implies that g takes values in $\mathbb{R}^{n\times n}$.

To show ((i)), we fix $s < t$ and apply the Fourier transform to obtain

$$\begin{aligned}
&\mathcal{F}\left[g(t-\cdot) - g(s-\cdot) - \int_s^t g * \eta(u-\cdot)\, du\right](y) \\
&\quad = (e^{ity} - e^{isy})\mathcal{F}[g](-y) - \mathcal{F}[\mathbb{1}_{(s,t]}](y)\mathcal{F}[g](-y)\mathcal{F}[\eta](-y) \\
&\quad = \mathcal{F}[\mathbb{1}_{(s,t]}](y) h(-iy)^{-1}(iyI - \mathcal{F}[\eta](-y)) \\
&\quad = \mathcal{F}[\mathbb{1}_{(s,t]}](y) I_n,
\end{aligned}$$

which verifies the result.

We will now show ((ii)) and for this we suppose that η has a moment of order $p \in \mathbb{N}$. Then it follows that $\tilde{h}: y \mapsto h(iy)$ is (entry-wise) p times differentiable with the m-th derivative given by

$$-\bigg(i\delta_0(\{m-1\}) + i^m \int_{[0,\infty)} e^{iuy} u^m \eta_{jk}(du)\bigg), \quad m = 1,\ldots,p,$$

and in particular all the the entries of $(D^m\tilde{h})(y)$ are bounded in y. Observe that, clearly, if a function $A: \mathbb{R} \to \mathbb{C}^{n\times n}$ takes the form

$$A(t) = B(t)C(t)D(t), \quad t \in \mathbb{R}, \tag{5.6}$$

where all the entries of $B, D: \mathbb{R} \to \mathbb{C}^{n\times n}$ decay at least as $|y|^{-1}$ as $|y| \to \infty$ and all the entries of $C: \mathbb{R} \to \mathbb{C}^{n\times n}$ are bounded, then all the entries of A decay at least as $|y|^{-1}$ as $|y| \to \infty$. Using the product rule for differentiation and the fact that

$$\big(D\tilde{h}^{-1}\big)(y) = -\tilde{h}(y)^{-1}(D\tilde{h})(y)\tilde{h}(y)^{-1}, \quad y \in \mathbb{R},$$

it follows recursively that $D^m\tilde{h}^{-1}$ is a sum of functions of the form (5.6), thus all its entries decay at least as $|y|^{-1}$ as $|y| \to \infty$, for $m = 1,\ldots,p$. Since the entries of $D^m\tilde{h}^{-1}$ are continuous as well, they belong to L^2, and we can use the inverse Fourier transform $\mathcal{F}^{-1}$ to conclude that

$$\mathcal{F}^{-1}[D^p\tilde{h}](t) = (it)^p\mathcal{F}^{-1}[\tilde{h}](t) = (it)^p g(t), \quad t \in \mathbb{R},$$

is an L^2 function. This implies in turn that $t \mapsto g_{jk}(t)(1+|t|)^p \in L^2$ and, thus,

$$\int_{\mathbb{R}} |g_{jk}(t)|^q\,dt \le \bigg(\int_{\mathbb{R}} \Big(g_{jk}(t)(1+|t|)^p\Big)^2 dt\bigg)^{\frac{q}{2}} \bigg(\int_{\mathbb{R}} (1+|t|)^{-\frac{2pq}{2-q}}\,dt\bigg)^{1-\frac{q}{2}} < \infty$$

for any $q \in [1/p, 2)$ and $j,k = 1,\ldots,n$. By using the particular observation that $g \in L^1$ and ((i)) we obtain that

$$g(t) = \mathbb{1}_{[0,\infty)}(t)I + \int_{-\infty}^t g * \eta(u)\,du \tag{5.7}$$

for (almost) all $t \in \mathbb{R}$. This shows that

$$|g_{jk}(t)| \le 1 + \int_{\mathbb{R}} |(g*\eta(u))_{jk}|\,du \le 1 + \sum_{l=1}^n \int_{\mathbb{R}} |g_{jl}(u)|\,du\,|\eta_{lk}|([0,\infty))$$

for all $t \in \mathbb{R}$ and for every $j,k = 1,\ldots,n$ which implies $g \in L^\infty$ and, thus, $g \in L^q$ for all $q \in [1/p, \infty]$. Since $g(t) \to 0$ entrywise as $t \to \infty$, we get by (5.7) that

$$\int_{\mathbb{R}} g*\eta(u)\,du = -I_n,$$

which concludes the proof of ((ii)).

Now suppose that $\int_{[0,\infty)} e^{\delta u}\,|\eta_{jk}|(du) < \infty$ for all $j,k = 1,\ldots,n$ and some $\delta > 0$. In this case, $\mathcal{S}_\delta := \{z \in \mathbb{C} : \mathrm{Re}(z) \in [-\delta,\delta]\} \subseteq D(\eta)$ and

$$z \mapsto \det(h(z)) = \det\bigg(-zI - \int_{[0,\infty)} e^{zu}\,\eta(du)\bigg)$$

is strictly separated from 0 when $|z|$, $z \in \mathcal{S}_\delta$, is sufficiently large. Indeed, the dominating term in $\det(h(z))$ is $(-1)^n z^n$ when $|z|$ is large, since

$$\left|\left(\int_{[0,\infty)} e^{zu}\,\eta(du)\right)_{jk}\right| \leq \max_{l,m=1,\dots,n} \int_{[0,\infty)} e^{\delta u}\,|\eta_{lm}|(du)$$

for $j,k = 1,\dots,n$. Using this together with the continuity of $z \mapsto \det(h(z))$ implies that there exists $\tilde{\delta} \in (0,\delta]$ so that $z \mapsto \det(h(z))$ is strictly separated from 0 on $\mathcal{S}_{\tilde{\delta}} := \{z \in \mathbb{C} : \operatorname{Re}(z) \in [-\tilde{\delta},\tilde{\delta}]\}$. In particular, $z \mapsto (h(z)^{-1})_{jk}$ is bounded on any compact set of $\mathcal{S}_{\tilde{\delta}}$, and by using Cramer's rule and the Leibniz formula as in (5.5) we get that $|(h(z)^{-1})_{jk}| = O(|z|^{-1})$ as $|z| \to \infty$ provided that $z \in \mathcal{S}_{\tilde{\delta}}$. Consequently,

$$\sup_{x\in[-\tilde{\delta},\tilde{\delta}]} \int_{\mathbb{R}} \left|\left(h(x+iy)^{-1}\right)_{jk}\right|^2 dy < \infty,$$

and this implies by [3, Lemma 5.1] that $t \mapsto g_{jk}(t)e^{\varepsilon t} \in L^1$ for all $\varepsilon \in (-\tilde{\delta},\tilde{\delta})$. Fix any $\varepsilon \in (0,\tilde{\delta})$ and $j,k \in \{1,\dots,n\}$, and observe from (5.7) that g_{jk} is absolutely continuous on both $[0,\infty)$ and $(-\infty,0)$ with density $(g*\eta)_{jk}$. Consequently, for fixed $t > 0$, integration by parts yields

$$|g_{jk}(t)|e^{\varepsilon t} \leq |g_{jk}(0)| + \int_{\mathbb{R}} |(g*\eta(u))_{jk}|e^{\varepsilon u}\,du + \varepsilon \int_{\mathbb{R}} |g_{jk}(u)|e^{\varepsilon u}\,du. \tag{5.8}$$

Since

$$\int_{\mathbb{R}} |(g*\eta(u))_{jk}|e^{\varepsilon u}\,du \leq \sum_{l=1}^n \int_{\mathbb{R}} |g_{jl}(u)|e^{\varepsilon u}\,du \int_{[0,\infty)} e^{\varepsilon u}\,|\eta_{lk}|(du)$$

it follows from (5.8) that

$$\max_{j,k=1,\dots,n} |g_{jk}(t)| \leq Ce^{-\varepsilon t}$$

for all $t > 0$ with

$$\begin{aligned} C := 1& \\ &+ \max_{j,k=1,\dots,n}\left(\sum_{l=1}^n \int_{\mathbb{R}} |g_{jl}(u)|e^{\varepsilon|u|}\,du \int_{[0,\infty)} e^{\varepsilon u}\,|\eta_{lk}|(du) + \varepsilon\int_{\mathbb{R}} |g_{jk}(u)|e^{\varepsilon|u|}\,du\right). \end{aligned}$$

By considering $-\varepsilon$ rather than ε in the above calculations one reaches the conclusion that

$$\max_{j,k=1,\dots,n} |g_{jk}(t)| \leq Ce^{\varepsilon t}, \quad t < 0,$$

and this verifies ((iii)).

Finally, suppose that $\det(h(z)) \neq 0$ for all $z \in \mathbb{C}$ with $\operatorname{Re}(z) \leq 0$. Then it holds that h, and thus $z \mapsto h(z)^{-1}$, is continuous on $\{z \in \mathbb{C} : \operatorname{Re}(z) \leq 0\}$ and analytic on $\{z \in \mathbb{C} : \operatorname{Re}(z) < 0\}$. Moreover, arguments similar to those in (5.5) show that $|(h(z)^{-1})_{jk}| = O(|z|^{-1})$ as $|z| \to \infty$, and thus we may deduce that

$$\sup_{x<0} \int_{\mathbb{R}} |(h(x+iy)^{-1})_{jk}|\,dy < \infty.$$

From the theory on Hardy spaces, see [3, Lemma 5.1], [12, Section 2.3] or [13], this implies that g is vanishing on $(-\infty,0)$ almost everywhere, which verifies ((iv)) and ends the proof. □

From Proposition 5.1 it becomes evident that we may and, thus, do choose the kernel g to satisfy (5.2) pointwise, so that the function induces a finite Lebesgue-Stieltjes measure $g(du)$. We summarize a few properties of this measure in the corollary below.

Corollary 5.2. *Let h be the function introduced in (3.1) and suppose that* $\det(h(iy)) \neq 0$ *for all $y \in \mathbb{R}$. Suppose further that η has first moment. Then the kernel $g : \mathbb{R} \to \mathbb{R}^{n\times n}$ characterized in (5.1) induces an $n\times n$ finite Lebesgue-Stieltjes measure, which is given by*

$$g(du) = I_n\delta_0(du) + g * \eta(u)\,du. \tag{5.9}$$

A function $f = (f_{jk}) : \mathbb{R} \to \mathbb{C}^{m\times n}$ is in $L^1(g(du))$ if

$$\int_{\mathbb{R}} |f_{jl}(u)(g*\eta)_{lk}(u)|\,du < \infty, \quad l = 1,\dots,n,$$

for $j = 1,\dots,m$ and $k = 1,\dots,n$. Moreover, the measure $g(du)$ has $(p-1)$-th moment whenever η has p-th moment for any $p \in \mathbb{N}$.

Proof. The fact that g induces a Lebesgue-Stieltjes measure of the form (5.9) is an immediate consequence of (5.2). For a measurable function $f = (f_{jk}) : \mathbb{R} \to \mathbb{C}^{m\times n}$ to be integrable with respect to $g(du) = (g_{jk}(du))$ we require that $f_{jl} \in L^1(|g_{lk}(du)|)$, $l = 1,\dots,n$, for each choice of $j = 1,\dots,m$ and $k = 1,\dots,n$. Since the variation measure $|g_{lk}|(du)$ of $g_{lk}(du)$ is given by

$$|g_{lk}|(du) = \delta_0(\{l-k\})\delta_0(du) + |(g*\eta(u))_{lk}|\,du,$$

we see that this condition is equivalent to the statement in the result. Finally, suppose that η has p-th moment for some $p \in \mathbb{N}$. Then, for any $j,k \in \{1,\dots,n\}$, we get that

$$\begin{aligned}\int_{\mathbb{R}} |u|^{p-1}\,|g_{jk}|(du) \le \sum_{l=1}^{n} \Bigg(&|\eta_{lk}|([0,\infty)) \int_{\mathbb{R}} |u^{p-1} g_{jl}(u)|\,du \\ &+ \int_{[0,\infty)} |v|^{p-1}\,|\eta_{lk}|(dv) \int_{\mathbb{R}} |g_{jl}(u)|\,du \Bigg).\end{aligned}$$

From the assumptions on η and Proposition 5.1((ii)) we get immediately that $|\eta_{lk}|([0,\infty))$, $\int_{[0,\infty)} |v|^{p-1}\,|\eta_{lk}|(dv)$ and $\int_{\mathbb{R}} |g_{jl}(u)|\,du$ are finite for all $l = 1,\dots,n$. Moreover, for any such l we compute that

$$\begin{aligned}&\int_{\mathbb{R}} |u^{p-1} g_{jl}(u)|\,du \\ &\quad\le \int_{\{|u|\le 1\}} |u^{p-1} g_{jl}(u)|\,du + \left(\int_{\{|u|>1\}} u^{-2}\,du\right)^{\frac{1}{2}} \left(\int_{\{|u|>1\}} (u^p g_{jl}(u))^2\,du\right)^{\frac{1}{2}}\end{aligned}$$

which is finite since $u \mapsto u^p g_{jl}(u) \in L^2$, according to the proof of Proposition 5.1((ii)), and hence we have shown the last part of the result. □

We now give a result that both will be used to prove the uniqueness part of Theorem 3.1 and Theorem 3.2.

Lemma 5.3. *Suppose that* $\det(h(iy)) \neq 0$ *for all* $y \in \mathbb{R}$ *and that* η *is a finite measure with second moment, and let* g *be given by* (3.2). *Furthermore, let* $(X_t)_{t\in\mathbb{R}}$ *be a measurable process, which is bounded in* $L^1(\mathbb{P})$ *and satisfies* (1.5) *almost surely for all* $s < t$. *Then, for each* $s \in \mathbb{R}$ *and almost surely,*

$$\begin{aligned} X_t =& g(t-s)X_s + \int_s^\infty g(t-u)\,\eta * \Big\{\mathbb{1}_{(-\infty,s]}X\Big\}(u)\,du \\ &+ g * \Big\{\mathbb{1}_{(s,\infty)}(Z-Z_s)\Big\}(t) \end{aligned} \tag{5.10}$$

for Lebesgue almost all $t > s$, *using the notation*

$$(\eta * \{\mathbb{1}_A X\})_j(t) := \sum_{k=1}^n \int_{[0,\infty)} \mathbb{1}_A(t-u) X^k_{t-u}\,\eta_{jk}(du) \quad \textit{and}$$

$$(g * \{\mathbb{1}_{(s,\infty)}(Z-Z_s)\})_j(t) := \sum_{k=1}^n \int_{\mathbb{R}} \mathbb{1}_{(s,\infty)}(t-u)\Big(Z^k_{t-u} - Z^k_s\Big) g_{jk}(du)$$

for $j = 1,\dots,n$ *and* $t \in \mathbb{R}$.

Proof. By arguments similar to those in the proof of Proposition 5.1((iii)) we get that the assumption $\det(h(iy)) \neq 0$ for all $y \in \mathbb{R}$ implies that we can choose $\delta \in (0,\varepsilon)$, such that $\det(h(z)) \neq 0$ for all $z \in \mathbb{C}$ with $-\delta < \operatorname{Re}(z) \leq 0$ and

$$\sup_{x\in(-\delta,0)} \int_{\mathbb{R}} \Big|\Big(h(x+iy)^{-1}\Big)_{jk}\Big|^2\,dy < \infty.$$

for all $j,k = 1,\dots,n$. Thus, [3, Lemma 5.1] ensures that $\mathcal{L}[g](z) = h(z)^{-1}$ when $\operatorname{Re}(z) \in (-\delta, 0)$. From this point we will fix such z and let $s \in \mathbb{R}$ be given. Since $(X_t)_{t\in\mathbb{R}}$ satisfies (1.4),

$$\mathbb{1}_{(s,\infty)}(t)X_t = \mathbb{1}_{(s,\infty)}(t)X_s + \int_{-\infty}^t \mathbb{1}_{(s,\infty)}(u)\,\eta * X(u)\,du + \mathbb{1}_{(s,\infty)}(t)(Z_t - Z_s)$$

for Lebesgue almost all $t \in \mathbb{R}$ outside a $\mathbb{P}$-null set (which is a consequence of Tonelli's theorem). In particular, this shows that

$$\begin{aligned} &-z\mathcal{L}[\mathbb{1}_{(s,\infty)}X](z) \\ &\quad= -z\Bigg\{X_s\mathcal{L}[\mathbb{1}_{(s,\infty)}](z) + \mathcal{L}\Bigg[\int_{-\infty}^{\cdot} \mathbb{1}_{(s,\infty)}(u)\,\eta * X(u)\,du\Bigg](z) \\ &\qquad + \mathcal{L}[\mathbb{1}_{(s,\infty)}(Z-Z_s)](z)\Bigg\} \\ &\quad= \mathcal{L}[X_s\delta_0(\cdot - s)](z) + \mathcal{L}[\mathbb{1}_{(s,\infty)}\,\eta * X](z) - z\mathcal{L}[\mathbb{1}_{(s,\infty)}(Z-Z_s)](z). \end{aligned}$$

By noticing that

$$\begin{aligned} \mathcal{L}[\mathbb{1}_{(s,\infty)}\,\eta * X](z) &= \mathcal{L}\Big[\mathbb{1}_{(s,\infty)}\,\eta * \Big\{\mathbb{1}_{(-\infty,s]}X\Big\}\Big](z) + \mathcal{L}\Big[\eta * \Big\{\mathbb{1}_{(s,\infty)}X\Big\}\Big](z) \\ &= \mathcal{L}\Big[\mathbb{1}_{(s,\infty)}\,\eta * \Big\{\mathbb{1}_{(-\infty,s]}X\Big\}\Big](z) + \mathcal{L}[\eta](z)\mathcal{L}\Big[\Big\{\mathbb{1}_{(s,\infty)}X\Big\}\Big](z) \end{aligned}$$

it thus follows that

$$
\begin{aligned}
&h(z)\mathcal{L}[\mathbb{1}_{(s,\infty)}X](z) \\
&\quad=\mathcal{L}\Big[X_s\delta_0(\cdot - s) + \mathbb{1}_{(s,\infty)}\,\eta * \Big\{\mathbb{1}_{(-\infty,s]}X\Big\}\Big](z) - z\mathcal{L}[\mathbb{1}_{(s,\infty)}(Z-Z_s)](z).
\end{aligned}
$$

(The reader should observe that since both $(X_t)_{t\in\mathbb{R}}$ and $(Z_t)_{t\in\mathbb{R}}$ are bounded in $L^1(\mathbb{P})$, the Laplace transforms above are all well-defined almost surely. We refer to the beginning of the proof of Theorem 3.1 where details for a similar argument are given.) Now, using that $\mathcal{L}[g](z) = h(z)^{-1}$, we notice

$$
\begin{aligned}
-zh(z)^{-1}\mathcal{L}[\mathbb{1}_{(s,\infty)}(Z-Z_s)](z) &= \mathcal{L}[g(du)](z)\mathcal{L}[\mathbb{1}_{(s,\infty)}(Z-Z_s)](z) \\
&= \mathcal{L}\Big[g * \Big\{\mathbb{1}_{(s,\infty)}(Z-Z_s)\Big\}\Big](z),
\end{aligned}
$$

and thus

$$
X_t = g(t-s)X_s + \int_s^\infty g(t-u)\,\eta * \Big\{\mathbb{1}_{(-\infty,s]}X\Big\}(u)\,du + g * \Big\{\mathbb{1}_{(s,\infty)}(Z-Z_s)\Big\} \qquad \square
$$

for Lebesgue almost all $t > s$ with probability one.

With Lemma 5.3 in hand we are now ready to prove the general result, Theorem 3.1, for existence and uniqueness of solutions to the MSDDE (1.4).

Proof (Proof of Theorem 3.1). Fix $t \in \mathbb{R}$. The convolution in (3.3) is well-defined if $u \mapsto Z_{t-u}^T$ is g^T-integrable (by Corollary 5.2) which means that $u \mapsto Z_{t-u}^k$ belongs to $L^1(|g_{jk}|(du))$ for all $j,k = 1,\dots,n$. Observe that, since $(Z_u^k)_{u\in\mathbb{R}}$ is integrable and has stationary increments, [2, Corollary A.3] implies that there exists $\alpha,\beta > 0$ such that $\mathbb{E}[|Z_u^k|] \le \alpha + \beta|u|$ for all $u \in \mathbb{R}$. Consequently,

$$
\mathbb{E}\Bigg[\int_{\mathbb{R}} |Z_{t-u}^k|\,\mu(du)\Bigg] \le (\alpha + \beta|t|)\mu(\mathbb{R}) + \beta\int_{\mathbb{R}} |u|\,\mu(du) < \infty
$$

for any (non-negative) measure μ which has first moment. This shows that $u \mapsto Z_{t-u}^k$ will be integrable with respect to such measure almost surely, in particular with respect to $|g_{jk}|(du)$, $j = 1,\dots,n$, according to Corollary 5.2 as η has second moment.

We will now argue that $(X_t)_{t\in\mathbb{R}}$ defined by (3.3) does indeed satisfy (1.4), and thus we fix $s < t$. Due to the fact that

$$
\int_s^t X^T * \eta^T(u)\,du = \int_s^t Z^T * \eta^T(u)\,du + \int_s^t \bigg(\int_{\mathbb{R}} g * \eta(r) Z_{\cdot - r}\,du\bigg)^T * \eta^T(u)\,du
$$

it is clear by the definition of $(X_t)_{t\in\mathbb{R}}$ that it suffices to argue that

$$
\begin{aligned}
&\int_s^t \bigg(\int_{\mathbb{R}} g * \eta(r) Z_{\cdot - r}\,du\bigg)^T * \eta^T(u)\,du \\
&\qquad = \int_{\mathbb{R}} Z_r^T [g * \eta(t-r) - g * \eta(s-r)]^T\,dr - \int_s^t Z^T * \eta^T(r)\,dr.
\end{aligned}
$$

We do this componentwise, so we fix $i \in \{1,\dots,n\}$ and compute that

$$
\begin{aligned}
&\left(\int_s^t \left(\int_{\mathbb{R}} g * \eta(r) Z_{\cdot - r}\, dr\right)^T * \eta^T(u)\, du\right)_i \\
&= \sum_{j=1}^n \sum_{k=1}^n \sum_{l=1}^n \int_s^t \left(\int_{\mathbb{R}} g_{jl} * \eta_{lk}(v) Z^k_{\cdot - r}\, dr\right) * \eta_{ij}(u)\, du \\
&= \sum_{j=1}^n \sum_{k=1}^n \sum_{l=1}^n \int_{\mathbb{R}} Z^k_r \int_{[0,\infty)} \int_s^t \int_{[0,\infty)} g_{jl}(u - v - r - w)\, \eta_{ij}(dv)\, du\, \eta_{lk}(dw)\, dr \\
&= \sum_{k=1}^n \sum_{l=1}^n \int_{\mathbb{R}} Z^k_r \int_{[0,\infty)} \int_s^t (g * \eta)_{il}(u - r - w)\, du\, \eta_{lk}(dw)\, dr \\
&= \sum_{k=1}^n \sum_{l=1}^n \Bigg(\int_{\mathbb{R}} Z^k_r \int_{[0,\infty)} [g_{il}(t - r - w) - g_{il}(s - r - w)]\, \eta_{lk}(dw)\, dr \\
&\quad - \int_{\mathbb{R}} Z^k_r \int_{[0,\infty)} \delta_0(\{i - l\}) \mathbb{1}_{(s,t]}(r + w)\, \eta_{lk}(dw)\, dr \Bigg) \\
&= \sum_{k=1}^n \left(\int_{\mathbb{R}} Z^k_r [(g * \eta)_{ik}(t - r) - (g * \eta)_{ik}(s - r)]\, dr - \int_s^t Z^k * \eta_{ik}(r)\, dr \right) \\
&= \left(\int_{\mathbb{R}} Z^T_r [g * \eta(t - r) - g * \eta(s - r)]^T\, dr - \int_s^t Z^T * \eta^T(r)\, dr \right)_i
\end{aligned}
$$

where we have used ((i)) in Proposition 5.1 and the fact that g and η commute in a convolution sense, $g * \eta = (g^T * \eta^T)^T$ (compare the associated Fourier transforms).

Next, we need to argue that $(X_t)_{t\in\mathbb{R}}$ is stationary. Here we will use (5.3) to write the solution as

$$X_t = \int_{\mathbb{R}} g * \eta(u)\, [Z_{t-u} - Z_t]\, du$$

for each $t \in \mathbb{R}$. Fix $m \in \mathbb{R}$. Let $-m = t^k_0 < t^k_1 < \cdots < t^k_k = m$ be a partition of $[-m, m]$ with $\max_{j=1,\dots,k}(t^k_j - t^k_{j-1}) \to 0$, $k \to \infty$, and define the Riemann sum

$$X^{m,k}_t = \sum_{j=1}^k g * \eta(t^k_{j-1})\, [Z_{t - t^k_{j-1}} - Z_t]\, (t^k_j - t^k_{j-1}).$$

Observe that $(X^{m,k}_t)_{t\in\mathbb{R}}$ is stationary. Moreover, the i-th component of $X^{m,k}_t$ converges to the i-th component of

$$X^m_t = \int_{-m}^m g * \eta(u)\, [Z_{t-u} - Z_t]\, du$$

in $L^1(\mathbb{P})$ as $k \to \infty$. To see this, we start by noting that

$$
\begin{aligned}
\mathbb{E}\Big[\Big|\big(X^m_t\big)_i - \big(X^{m,k}_t\big)_i\Big|\Big] \le \sum_{j=1}^n \int_{\mathbb{R}} \sum_{l=1}^k \mathbb{1}_{(t^k_{l-1}, t^k_l]}(u)\, \mathbb{E}\Big[\Big|(g * \eta)_{ij}(u)\Big[Z^j_{t-u} - Z^j_t\Big] \\
- (g * \eta)_{ij}\big(t^k_{l-1}\big)\Big[Z^j_{t - t^k_{l-1}} - Z^j_t\Big]\Big|\Big]\, du.
\end{aligned}
$$

Then, for each $j \in \{1,\dots,n\}$,

$$\begin{aligned}\max_{l=1,\dots,k} \mathbb{1}_{(t^k_{l-1},t^k_l]}(u)\mathbb{E}\Big[\big|(g*\eta)_{ij}(u)\big[Z^j_{t-u}-Z^j_t\big]-(g*\eta)_{ij}\big(t^k_{l-1}\big)\big[Z^j_{t-t^k_{l-1}}-Z^j_t\big]\big|\Big]\\ \leq \max_{l=1,\dots,k} \mathbb{1}_{(t^k_{l-1},t^k_l]}(u)\Big(\big|(g*\eta)_{ij}(u)\big|\mathbb{E}\Big[\big|Z^j_{t-u}-Z^j_{t-t^k_{l-1}}\big|\Big]\\ +\mathbb{E}\Big[\big|Z^j_{t-t^k_{l-1}}-Z^j_t\big|\Big]\big|(g*\eta)_{ij}(u)-(g*\eta)_{ij}\big(t^k_{l-1}\big)\big|\Big)\to 0\end{aligned}$$

as $k\to\infty$ for almost all $u\in\mathbb{R}$ using that $(Z^j_t)_{t\in\mathbb{R}}$ is continuous in $L^1(\mathbb{P})$ (cf. [2, Corollary A.3]) and that $(g*\eta)_{ij}$ is càdlàg. Consequently, Lebesgue's theorem on dominated convergence implies that $X^{m,k}_t\to X^m_t$ entrywise in $L^1(\mathbb{P})$ as $k\to\infty$, thus $(X^m_t)_{t\in\mathbb{R}}$ inherits the stationarity property from $(X^{m,k}_t)_{t\in\mathbb{R}}$. Finally, since $X^m_t\to X_t$ (entrywise) almost surely as $m\to\infty$, we obtain that $(X_t)_{t\in\mathbb{R}}$ is stationary as well.

To show the uniqueness part, we let $(U_t)_{t\in\mathbb{R}}$ and $(V_t)_{t\in\mathbb{R}}$ be two stationary, integrable and measurable solutions to (1.4). Then $X_t := U_t - V_t$, $t\in\mathbb{R}$, is bounded in $L^1(\mathbb{P})$ and satisfies an MSDDE without noise. Consequently, Lemma 5.3 implies that

$$X_t = g(t-s)X_s + \int_s^\infty g(t-u)\eta*\{\mathbb{1}_{(-\infty,s]}X\}(u)\,du$$

holds for each $s\in\mathbb{R}$ and Lebesgue almost all $t>s$. For a given j we thus find that

$$\mathbb{E}\big[\big|X^j_t\big|\big]\leq C\sum_{k=1}^n\Big(|g_{jk}(t-s)|+\sum_{l=1}^n\int_s^\infty |g_{jk}(t-u)||\eta_{kl}|([u-s,\infty))\,du\Big)$$

where $C := \max_k \mathbb{E}[|U^k_0|+|V^k_0|]$. It follows by Proposition 5.1(ii) that $g(t)$ converges as $t\to\infty$, and since $g\in L^1$ it must be towards zero. Using this fact together with Lebesgue's theorem on dominated convergence it follows that the right-hand side of the expression above converges to zero as s tends to $-\infty$, from which we conclude that $U_t = V_t$ almost surely for Lebesgue almost all t. By continuity of both processes in $L^1(\mathbb{P})$ (cf. [2, Corollary A.3]), we get the same conclusion for all t.

Finally, under the assumption that $\det(h(z))\neq 0$ for $z\in\mathbb{C}$ with $\mathrm{Re}(z)\leq 0$ it follows from Proposition 5.1((iv)) that $g*\eta$ is vanishing on $(-\infty,0)$, and hence we get that the solution $(X_t)_{t\in\mathbb{R}}$ defined by (3.3) is causal since

$$X_t = Z_t + \int_0^\infty g*\eta(u)Z_{t-u}\,du = -\int_0^\infty g*\eta(u)[Z_t - Z_{t-u}]\,du$$

for $t\in\mathbb{R}$ by (5.3). □

Proof (Proof of Theorem 3.2). Since $(X_t)_{t\in\mathbb{R}}$ is a solution to an MSDDE,

$$\sigma(X_u : u\leq s) = \sigma(Z_s - Z_u : u\leq s)$$

and the theorem therefore follows by Lemma 5.3. □

Proof (Proof of Proposition 4.1). We start by arguing why (4.2) is well-defined. To see that this is the case, note initially that $I_k(f_r(t-\cdot)) = Z^k_t - Z^k_{t-r}$ and thus, since

$(Z_t^k)_{t\in\mathbb{R}}$ is integrable and has stationary increments, there exists $\alpha,\beta>0$ such that $\mathbb{E}[|I_k(f_r(t-\cdot))|]\leq \alpha+\beta|r|$ for all $r\in\mathbb{R}$ (see, e.g., [2, Corollary A.3]). In particular

$$\mathbb{E}\Big[\int_{\mathbb{R}} |I_k(f_r(t-\cdot))|\,|\mu|(dr)\Big] \leq \alpha|\mu|(\mathbb{R})+\beta\int_{\mathbb{R}}|r|\,|\mu|(dr)<\infty,$$

which shows that $I_k(f_r(t-\cdot))$ is integrable with respect to μ, thus the right-hand side of (4.2) is well-defined, almost surely for each $t\in\mathbb{R}$. To show that the left-hand side is well-defined, it suffices to note that $u\mapsto \int_{\mathbb{R}} f_r(u)\,\mu(dr)$ belongs to $L^1\cap L^2$ by an application of Jensen's inequality and Tonelli's theorem.

To show (4.3) we start by fixing $t\in\mathbb{R}$ and $j,k\in\{1,\dots,n\}$, and by noting that $\mu(dr)=(g*\eta)_{jk}(r)\,dr$ is a finite measure with having first moment according to Corollary 5.2. Consequently, we can use assumptions (i)-(ii) on I_k to get

$$\begin{aligned}\int_{\mathbb{R}}(g*\eta)_{jk}(r)\big[Z_{t-r}^k-Z_t^k\big]\,dr &= \int_{\mathbb{R}} I_k(\mathbb{1}_{(t,t-r]})(g*\eta)_{jk}(r)\,dr\\ &= I_k\Big(\int_{\mathbb{R}}\mathbb{1}_{(t,t-r]}(g*\eta)_{jk}(r)\,dr\Big)\\ &= I_k\Big(\delta_0(\{j-k\})\mathbb{1}_{[0,\infty)}(t-\cdot)+\int_{-\infty}^{t-\cdot}(g*\eta)_{jk}(u)\,du\Big)\\ &= I_k(g_{jk}(t-\cdot))\end{aligned}$$

using (5.2) and the convention that $\mathbb{1}_{(a,b]}=-\mathbb{1}_{(b,a]}$ when $a>b$. By combining this relation with (5.3) and (3.3) we obtain

$$X_t^j=\sum_{k=1}^n\int_{\mathbb{R}}(g*\eta)_{jk}(r)[Z_{t-r}^k-Z_t^k]\,dr=\sum_{k=1}^n I_k(g_{jk}(t-\cdot)). \qquad \square$$

Proof (Proof of Proposition 4.3). Let $\alpha\in(1,2]$ and $\beta\in(0,1-1/\alpha)$, and consider a function $f:\mathbb{R}\to\mathbb{R}$ in $L^1\cap L^\alpha$. We start by noticing that

$$\int_t^\infty |f(u)|(u-t)^{\beta-1}\,du=\int_0^1|f(t+u)|u^{\beta-1}\,du+\int_1^\infty|f(t+u)|u^{\beta-1}\,du.$$

For the left term we find that

$$\begin{aligned}&\int_{\mathbb{R}}\Big(\int_0^1|f(t+u)|u^{\beta-1}\,du\Big)^\alpha dt\\ &\quad\leq\Big(\int_0^1 u^{\beta-1}\,du\Big)^{\alpha-1}\int_{\mathbb{R}}\int_0^1|f(t+u)|^\alpha u^{\beta-1}\,du\,dt\\ &\quad=\Big(\int_0^1 u^{\beta-1}\,du\Big)^{\alpha}\int_{\mathbb{R}}|f(t)|^\alpha\,dt<\infty.\end{aligned}$$

For the right term we find

$$\begin{aligned}&\int_{\mathbb{R}}\Big(\int_1^\infty|f(t+u)|u^{\beta-1}\,du\Big)^\alpha dt\\ &\quad\leq\Big(\int_{\mathbb{R}} f(u)du\Big)^{\alpha-1}\int_{\mathbb{R}}\int_1^\infty|f(t+u)|u^{\alpha(\beta-1)}\,du\,dt\\ &\quad=\Big(\int_{\mathbb{R}} f(u)du\Big)^{\alpha}\int_1^\infty u^{\alpha(\beta-1)}\,du<\infty.\end{aligned}$$

We conclude that $\left(I_-^\beta f\right)(u) \in L^\alpha$. □

Proof (Proof of Theorem 4.5). The identity (4.8) is just a matter of applying standard computation rules for determinants. For instance, one may prove the result when $z \neq 0$ by induction using the block representation

$$-h(z) = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \tag{5.11}$$

with $A = I_n z$, $B = (e_1 \otimes I_n)^T \in \mathbb{R}^{n\times(m-1)n}$, $C = e_{m-1} \otimes \mathcal{L}[\varpi_0](z) \in \mathbb{R}^{(m-1)n\times n}$, and

$$D = \begin{bmatrix} I_n z & I_n & 0 & \cdots & 0 \\ 0 & I_n z & I_n & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & I_n z & I_n \\ \mathcal{L}[\varpi_1](z) & \mathcal{L}[\varpi_2](z) & \cdots & \mathcal{L}[\varpi_{m-2}](z) & I_n z + \mathcal{L}[\varpi_{m-1}](z) \end{bmatrix}.$$

Here e_1 and e_{m-1} refer to the firs and last canonical basis vector of $\mathbb{R}^{m-1}$, respectively. The case where $z = 0$ follows directly from the Leibniz formula. In case $\det(h(iy)) \neq 0$ for all $y \in \mathbb{R}$, we may write $h(iy)^{-1}$ as an $m\times m$ matrix, where each element $(h^{-1}(iy))_{jk}$ is an $n \times n$ matrix. We then know from Theorem 3.1 that the unique solution to (4.5) is a $(Z_t)_{t\in\mathbb{R}}$-driven moving average of the form (4.7) with $\mathcal{F}[g_{1m}](y) = (h^{-1}(iy))_{1m}$. Similar to the computation of $\det(h(z))$, when $h(z)$ is invertible, block $(1,m)$ of $h(z)^{-1}$ can inductively be shown to coincide with

$$\left(I_n(-z)^m - \sum_{j=0}^{m-1} \mathcal{L}[\varpi_j](z)(-z)^j\right)^{-1}$$

using the representation (5.11) and standard rules for inverting block matrices. This means in particular that (4.9) is true. □

Proof (Proof of Theorem 4.8). We start by arguing that that there exists a function f with the Fourier transform in (4.18). Note that, since $z \mapsto \det(Q(z))$ is just a polynomial (of order nq), the assumption that $\det(Q(z)) \neq 0$ whenever $\mathrm{Re}(z) \geq 0$ implies in fact that

$$H(z) := R(-z) - Q(-z)^{-1}P(-z) = Q(-z)^{-1}[Q(-z)R(-z) - P(-z)]$$

is well-defined for all $z \in \mathcal{S}_\delta := \{x + iy : x \leq \delta, y \in \mathbb{R}\}$ and a suitably chosen $\delta > 0$. According to [3, Lemma 5.1] it suffices to argue that there exists $\varepsilon \in (0,\delta]$ such that

$$\sup_{x<\varepsilon} \int_{\mathbb{R}} |H(x+iy)_{jk}|^2 \, dy < \infty \tag{5.12}$$

for all $j,k = 1,\dots,n$. Let $\|\cdot\|$ denote any sub-multiplicative norm on $\mathbb{C}^{n\times n}$ and note that $|H(z)_{jk}| \leq \|Q(-z)^{-1}\|\|Q(-z)R(-z) - P(-z)\|$. Thus, since $\|Q(z)R(z) - P(z)\| \sim c_1|z|^{q-1}$ and $\|Q(z)^{-1}\| \sim c_2|z|^{-q}$ as $|z| \to \infty$ for some $c_1, c_2 \geq 1$ (the former by the choice of R and the latter by Cramer's rule), $|H(z)_{jk}| = O(|z|^{-1})$. Consequently, the continuity of H ensures that (5.12) is satisfied for a suitable $\varepsilon \in (0,\delta]$, and we have established the existence of f with the desired Fourier transform. This also establishes that the $n \times n$ measures $\varpi_0, \varpi_1, \dots, \varpi_{p-q-1}$ defined as in the statement of the theorem are finite and have moments of any order. Associate to these measures the $n(p-q) \times n(p-q)$ measure η given in (4.6). Then it follows from (4.8) that

$$\det(h(iy)) = \det\left(I_n(-iy)^{p-q} + \sum_{j=0}^{p-q-1} R_j(-iy)^j - \mathcal{F}[f](y)\right) = \frac{\det(P(-iy))}{\det(Q(-iy))},$$

and hence is non-zero for all $y \in \mathbb{R}$. In light of Proposition 4.5, in particular (4.9), we may therefore conclude that the unique solution to (4.5) is a $(Z_t)_{t\in\mathbb{R}}$-driven moving average, where the driving kernel has Fourier transform

$$\left(I_n(-iy) + \sum_{j=0}^{p-q-1} R_j(-iy)^j - \mathcal{F}[f](y)\right)^{-1} = P(-iy)^{-1}Q(-iy)$$

for $y \in \mathbb{R}$. In other words, the unique solution is the $(Z_t)_{t\in\mathbb{R}}$-driven MCARMA(p,q) process associated to the polynomials P and Q. □

Before giving the proof of Corollary 4.11 we will need the following lemma:

Lemma 5.4. *Let $C_0, \dots, C_{p-q-1}$ be given in* (4.16) *and $C_{p-q} = I_n$. Define*

$$R_j(z) = \sum_{k=j}^{p-q} C_k z^{k-j}, \quad j = 1, \dots, p-q-1.$$

Then $\tilde{g}$ is $p-q-2$ times differentiable and $D^{p-q-2}\tilde{g}$ has a density with respect to the Lebesgue measure which we denote $D^{p-q-1}\tilde{g}$. Furthermore, we have that

$$(e_1^{p-q} \otimes I_n)^T g = (\tilde{g}R_1(D), \dots, \tilde{g}R_{p-q-1}(D), \tilde{g}) \tag{5.13}$$

where

$$\begin{aligned} \tilde{g}R_j(D)(t) &= \sum_{k=j}^{p-q} D^{k-j}\tilde{g}(t)C_k \\ &= \mathbb{1}_{[0,\infty)}(t)(e_1^p \otimes I_n)^T e^{At} \sum_{k=j}^{p-q} A^{k-j}EC_k \end{aligned} \tag{5.14}$$

for $j = 1, \dots, p-q-1$ and $g : \mathbb{R} \to \mathbb{R}^{n\times n}$ is characterized by $\mathcal{F}[g](y) = h(iy)^{-1}$ with $h : \mathbb{C} \to \mathbb{C}^{n(p-q)\times n(p-q)}$ given by

$$h(-z) = \begin{bmatrix} I_n z & -I_n & 0 & \cdots & 0 \\ 0 & I_n z & -I_n & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & I_n z & -I_n \\ Q^{-1}(z)P(z) - zR_1(z) & C_1 & \cdots & C_{p-q-2} & I_n z + C_{p-q-1} \end{bmatrix}.$$

Proof. That $\tilde{g}$ is $p-q-2$ times differentiable and $D^{p-q-2}\tilde{g}$ has a density with respect to the Lebesgue measure follows form the relation in (5.2). Furthermore, by Theorem 4.8 we know that $\mathcal{F}[\tilde{g}](y) = P^{-1}(-iy)Q(-iy)$. Consequently, (5.13) follows since

$$(P^{-1}(-iy)Q(-iy)R_1(-iy), \\ \dots, P^{-1}(-iy)Q(-iy)R_{p-q-1}(-iy), P^{-1}(-iy)Q(-iy))h(z) = (e_1^{p-q} \otimes I_n)^T.$$

The relation in (5.14) follows by the representation of $\tilde{g}$ given in (4.15). □

Proof (Proof of Corollary 4.11). The prediction formula is a consequence of Lemma 5.4 combined with Theorem 3.2 and Theorem 4.8. Furthermore, to get the expression for $\tilde{g} * \{\hat{Z}\mathbb{1}_{(s,\infty)}\}$, note that

$$\tilde{g}(dv) = \mathbb{1}_{\{p=q+1\}}\delta_0(dv) + (e_1^p \otimes I_n)^T e^{Av} AE\, dv,$$

which follows from the representation of $\tilde{g}$ in (4.15) □

References

[1] Barndorff-Nielsen, O.E., J.L. Jensen and M. Sørensen (1998). Some stationary processes in discrete and continuous time. *Adv. in Appl. Probab.* 30(4), 989–1007. DOI: 10.1239/aap/1035228204.

[2] Barndorff-Nielsen, O.E. and A. Basse-O'Connor (2011). Quasi Ornstein–Uhlenbeck processes. *Bernoulli* 17(3), 916–941. DOI: 10.3150/10-BEJ311.

[3] Basse-O'Connor, A., M.S. Nielsen, J. Pedersen and V. Rohde (2019). Stochastic delay differential equations and related autoregressive models. *Stochastics.* Forthcoming. DOI: 10.1080/17442508.2019.1635601.

[4] Benth, F.E., J. Šaltytė-Benth and S. Koekebakker (2007). Putting a price on temperature. *Scand. J. Statist.* 34(4), 746–767. DOI: 10.1111/j.1467-9469.2007.00564.x.

[5] Box, G.E.P and G.M Jenkins (1970). *Times series analysis. Forecasting and control.* Holden-Day, San Francisco, Calif.-London-Amsterdam.

[6] Brockwell, P.J. (2001). Lévy-driven CARMA processes. *Ann. Inst. Statist. Math.* 53(1). Nonlinear non-Gaussian models and related filtering methods (Tokyo, 2000), 113–124. DOI: 10.1023/A:1017972605872.

[7] Brockwell, P.J. (2014). Recent results in the theory and applications of CARMA processes. *Ann. Inst. Statist. Math.* 66(4), 647–685. DOI: 10.1007/s10463-014-0468-7.

[8] Brockwell, P.J., R.A. Davis and Y. Yang (2011). Estimation for non-negative Lévy-driven CARMA processes. *J. Bus. Econom. Statist.* 29(2), 250–259. DOI: 10.1198/jbes.2010.08165.

[9] Brockwell, P.J. and A. Lindner (2015). Prediction of Lévy-driven CARMA processes. *J. Econometrics* 189(2), 263–271.

[10] Brockwell, P.J. and T. Marquardt (2005). Lévy-driven and fractionally integrated ARMA processes with continuous time parameter. *Statist. Sinica* 15(2), 477–494.

[11] Brockwell, P.J. and E. Schlemm (2013). Parametric estimation of the driving Lévy process of multivariate CARMA processes from discrete observations. *J. Multivariate Anal.* 115, 217–251. doi: 10.1016/j.jmva.2012.09.004.

[12] Doetsch, G. (1937). Bedingungen für die Darstellbarkeit einer Funktion als Laplace-integral und eine Umkehrformel für die Laplace-Transformation. *Math. Z.* 42(1), 263–286. doi: 10.1007/BF01160078.

[13] Dym, H. and H.P McKean (1976). *Gaussian processes, function theory, and the inverse spectral problem*. Probability and Mathematical Statistics, Vol. 31. New York: Academic Press [Harcourt Brace Jovanovich Publishers].

[14] García, I., C. Klüppelberg and G. Müller (2011). Estimation of stable CARMA models with an application to electricity spot prices. *Stat. Model.* 11(5), 447–470. doi: 10.1177/1471082X1001100504.

[15] Gripenberg, G. and I. Norros (1996). On the prediction of fractional Brownian motion. *J. Appl. Probab.* 33(2), 400–410.

[16] Gushchin, A.A. and U. Küchler (2000). On stationary solutions of delay differential equations driven by a Lévy process. *Stochastic Process. Appl.* 88(2), 195–211. doi: 10.1016/S0304-4149(99)00126-X.

[17] Marquardt, T. (2006). Fractional Lévy processes with an application to long memory moving average processes. *Bernoulli* 12(6), 1099–1126.

[18] Marquardt, T. (2007). Multivariate fractionally integrated CARMA processes. *Journal of Mult. Anal.* 98(9), 1705–1725.

[19] Marquardt, T. and R. Stelzer (2007). Multivariate CARMA processes. *Stochastic Process. Appl.* 117(1), 96–120. doi: 10.1016/j.spa.2006.05.014.

[20] Mohammed, S.E.A. and M.K.R. Scheutzow (1990). Lyapunov exponents and stationary solutions for affine stochastic delay equations. *Stochastics Stochastics Rep.* 29(2), 259–283.

[21] Rajput, B.S. and J. Rosiński (1989). Spectral representations of infinitely divisible processes. *Probab. Theory Related Fields* 82(3), 451–487.

[22] Samorodnitsky, G. and M.S. Taqqu (1994). *Stable Non-Gaussian Random Processes*. Stochastic Modeling. Stochastic models with infinite variance. New York: Chapman & Hall.

[23] Sato, K., T. Watanabe and M. Yamazato (1994). Recurrence conditions for multidimensional processes of Ornstein–Uhlenbeck type. *J. Math. Soc. Japan* 46(2), 245–265.

[24] Sato, K. and M. Yamazato (1983). "Stationary processes of Ornstein–Uhlenbeck type". *Probability theory and mathematical statistics (Tbilisi, 1982)*. Vol. 1021. Lecture Notes in Math. Springer, Berlin, 541–551. doi: 10.1007/BFb0072949.

[25] Stelzer, R. (2011). CARMA Processes driven by Non-Gaussian Noise. arXiv: 1201.0155.

[26] Todorov, V. (2009). Estimation of continuous-time stochastic volatility models with jumps using high-frequency data. *J. Econometrics* 148(2), 131–148.

PAPER E

Stochastic Differential Equations With A Fractionally Filtered Delay: A Semimartingale Model For Long-Range Dependent Processes

Richard A. Davis, Mikkel Slot Nielsen, and Victor Rohde

Abstract

In this paper we introduce a model, the stochastic fractional delay differential equation (SFDDE), which is based on the linear stochastic delay differential equation and produces stationary processes with hyperbolically decaying autocovariance functions. The model departs from the usual way of incorporating this type of long-range dependence into a short-memory model as it is obtained by applying a fractional filter to the drift term rather than to the noise term. The advantages of this approach are that the corresponding long-range dependent solutions are semimartingales and the local behavior of the sample paths is unaffected by the degree of long memory. We prove existence and uniqueness of solutions to the SFDDEs and study their spectral densities and autocovariance functions. Moreover, we define a suitable subclass of SFDDEs which we study in detail and relate to the well-known fractionally integrated CARMA processes. Finally, we consider the task of simulating from the defining SFDDEs.

Keywords: long-range dependence; stochastic delay differential equations; moving average processes; semimartingales

1 Introduction

Models for time series producing slowly decaying autocorrelation functions (ACFs) have been of interest for more than 50 years. Such models were motivated by the empirical findings of Hurst in the 1950s that were related to the levels of the Nile

River. Later, in the 1960s, Benoit Mandelbrot referred to a slowly decaying ACF as the Joseph effect or long-range dependence. Since then, a vast amount of literature on theoretical results and applications have been developed. We refer to [6, 12, 24, 27, 28] and references therein for further background.

A very popular discrete-time model for long-range dependence is the *autoregressive fractionally integrated moving average* (ARFIMA) process, introduced by [14] and [18], which extends the ARMA process to allow for a hyperbolically decaying ACF. Let B be the backward shift operator and for $\gamma > -1$, define $(1-B)^\gamma$ by means of the binomial expansion,

$$(1-B)^\gamma = \sum_{j=0}^{\infty} \pi_j B^j$$

where $\pi_j = \prod_{0<k\leq j} \frac{k-1-\gamma}{k}$. An ARFIMA process $(X_t)_{t\in\mathbb{Z}}$ is characterized as the unique purely non-deterministic process (as defined in [8, p. 189]) satisfying

$$P(B)(1-B)^\beta X_t = Q(B)\varepsilon_t, \quad t \in \mathbb{Z}, \tag{1.1}$$

where P and Q are real polynomials with no zeroes on $\{z \in \mathbb{C} : |z| \leq 1\}$, $(\varepsilon_t)_{t\in\mathbb{Z}}$ is an i.i.d. sequence with $\mathbb{E}[\varepsilon_0] = 0$, $\mathbb{E}[\varepsilon_0^2] < \infty$, and $\beta \in (0, 1/2)$. The ARFIMA equation (1.1) is sometimes represented as an ARMA equation with a fractionally integrated noise, that is,

$$P(B)X_t = Q(B)(1-B)^{-\beta}\varepsilon_t, \quad t \in \mathbb{Z}. \tag{1.2}$$

In (1.1) one applies a fractional filter to $(X_t)_{t\in\mathbb{Z}}$, while in (1.2) one applies a fractional filter to $(\varepsilon_t)_{t\in\mathbb{Z}}$. One main feature of the solution to (1.1), equivalently (1.2), is that the autocovariance function $\gamma_X(t) := \mathbb{E}[X_0 X_t]$ satisfies

$$\gamma_X(t) \sim c t^{2\beta-1}, \quad t \to \infty, \tag{1.3}$$

for some constant $c > 0$.

A simple example of a continuous-time stationary process which exhibits long-memory in the sense of (1.3) is an Ornstein-Uhlenbeck process $(X_t)_{t\in\mathbb{R}}$ driven by a fractional Lévy process, that is, $(X_t)_{t\in\mathbb{R}}$ is the unique stationary solution to

$$dX_t = -\kappa X_t\, dt + dI^\beta L_t, \quad t \in \mathbb{R}, \tag{1.4}$$

where $\kappa > 0$ and

$$I^\beta L_t := \frac{1}{\Gamma(1+\beta)} \int_{-\infty}^{t} \left[(t-u)^\beta - (-u)_+^\beta\right] dL_u, \quad t \in \mathbb{R}, \tag{1.5}$$

with $(L_t)_{t\in\mathbb{R}}$ being a Lévy process which satisfies $\mathbb{E}[L_1] = 0$ and $\mathbb{E}[L_1^2] < \infty$. In (1.5), Γ denotes the gamma function and we have used the notation $x_+ = \max\{x, 0\}$ for $x \in \mathbb{R}$. The way to obtain long memory in (1.4) is by applying a fractional filter to the noise, which is in line with (1.2). To demonstrate the idea of this paper, consider the equation obtained from (1.4) but by applying a fractional filter to the drift term instead, i.e.,

$$X_t - X_s = -\frac{\kappa}{\Gamma(1-\beta)} \int_{-\infty}^{t} \left[(t-u)^{-\beta} - (s-u)_+^{-\beta}\right] X_u\, du + L_t - L_s, \quad s < t. \tag{1.6}$$

One can write (1.6) compactly as

$$dX_t = -\kappa D^\beta X_t\, dt + dL_t, \quad t \in \mathbb{R}, \tag{1.7}$$

with $(D^\beta X_t)_{t\in\mathbb{R}}$ being a suitable fractional derivative process of $(X_t)_{t\in\mathbb{R}}$ defined in Proposition 3.6. The equations (1.6)-(1.7) may be seen as akin to (1.1). It turns out that a unique purely non-deterministic process (as defined in (3.10)) satisfying (1.7) exists and has the following properties:

(i) The memory is long and controlled by β in the sense that $\gamma_X(t) \sim ct^{2\beta-1}$ as $t \to \infty$ for some $c > 0$.

(ii) The $L^2(\mathbb{P})$-Hölder continuity of the sample paths is not affected by β in the sense that $\gamma_X(0) - \gamma_X(t) \sim ct$ as $t \downarrow 0$ for some $c > 0$ (the notion of Hölder continuity in $L^2(\mathbb{P})$ is indeed closely related to the behavior of the ACF at zero; see Remark 3.9 for a precise relation).

(iii) $(X_t)_{t\in\mathbb{R}}$ is a semimartingale.

While both processes in (1.4) and (1.7) exhibit long memory in the sense of ((i)), one should keep in mind that models for long-memory processes obtained by applying a fractional filter to the noise will generally not meet ((ii))-((iii)), since they inherit various properties from the fractional Lévy process $(I^\beta L_t)_{t\in\mathbb{R}}$ rather than from the underlying Lévy process $(L_t)_{t\in\mathbb{R}}$. In particular, this observation applies to the fractional Ornstein-Uhlenbeck process (1.4) which is known not to possess the semimartingale property for many choices of $(L_t)_{t\in\mathbb{R}}$, and for which it holds that $\gamma_X(0) - \gamma_X(t) \sim ct^{2\beta+1}$ as $t \downarrow 0$ for some $c > 0$ (see [20, Theorem 4.7] and [1, Proposition 2.5]). The latter property, the behavior of γ_X near 0, implies an increased $L^2(\mathbb{P})$-Hölder continuity relative to (1.7). See Example 4.4 for details about the models (1.4) and (1.7).

The properties ((ii))-((iii)) may be desirable to retain in many modeling scenarios. For instance, if a stochastic process $(X_t)_{t\in\mathbb{R}}$ is used to model a financial asset, the semimartingale property is necessary to accommodate the No Free Lunch with Vanishing Risk condition according to the (First) Fundamental Theorem of Asset Pricing, see [10, Theorem 7.2]. Moreover, if $(X_t)_{t\in\mathbb{R}}$ is supposed to serve as a "good" integrator, it follows by the Bichteler-Dellacherie Theorem ([7, Theorem 7.6]) that $(X_t)_{t\in\mathbb{R}}$ must be a semimartingale. Also, the papers [4, 5] find evidence that the sample paths of electricity spot prices and intraday volatility of the E-mini S&P500 futures contract are rough, suggesting less smooth sample paths than what is induced by models such as the fractional Ornstein-Uhlenbeck process (1.4). In particular, the local smoothness of the sample paths should not be connected to the strength of long memory.

Several extensions to the fractional Ornstein-Uhlenbeck process (1.4) exist. For example, it is worth mentioning that the class of *fractionally integrated continuous-time autoregressive moving average* (FICARMA) processes were introduced in [9], where it is assumed that P and Q are real polynomials with $\deg(P) > \deg(Q)$ which have no zeroes on $\{z \in \mathbb{C} : \mathrm{Re}(z) \geq 0\}$. The FICARMA process associated to P and Q is then defined as the moving average process

$$X_t = \int_{-\infty}^t g(t-u)\, dI^\beta L_u, \quad t \in \mathbb{R}, \tag{1.8}$$

with $g : \mathbb{R} \to \mathbb{R}$ being the L^2 function characterized by

$$\mathcal{F}[g](y) := \int_{\mathbb{R}} e^{iyu} g(u)\, du = \frac{Q(-iy)}{P(-iy)}, \quad y \in \mathbb{R}.$$

In line with (1.2) for the ARFIMA process, a common way of viewing a FICARMA process is that it is obtained by applying a CARMA filter to fractional noise, that is, $(X_t)_{t\in\mathbb{R}}$ given by (1.8) is the solution to the formal equation

$$P(D)X_t = Q(D)DI^\beta L_t, \quad t \in \mathbb{R}.$$

(See, e.g., [20].) Another class, related to the FICARMA process, consists of solutions $(X_t)_{t\in\mathbb{R}}$ to fractional *stochastic delay differential equations* (SDDEs), that is, $(X_t)_{t\in\mathbb{R}}$ is the unique stationary solution to

$$dX_t = \int_{[0,\infty)} X_{t-u}\, \eta(du)\, dt + dI^\beta L_t, \quad t \in \mathbb{R}, \tag{1.9}$$

for a suitable finite signed measure η. See [3, 21] for details about fractional SDDEs. Note that the fractional Ornstein-Uhlenbeck process (1.4) is a FICARMA process with polynomials $P(z) = z + \kappa$ and $Q(z) = 1$ and a fractional SDDE with $\eta = -\kappa\delta_0$, δ_0 being the Dirac measure at zero.

The model we present includes (1.6) and extends this process in the same way as the fractional SDDE (1.9) extends the fractional Ornstein-Uhlenbeck (1.4). Specifically, we will be interested in a stationary process $(X_t)_{t\in\mathbb{R}}$ satisfying

$$X_t - X_s = \frac{1}{\Gamma(1-\beta)} \int_{-\infty}^{t} \left[(t-u)^{-\beta} - (s-u)_+^{-\beta}\right] \int_{[0,\infty)} X_{u-v}\, \eta(dv)\, du + L_t - L_s \tag{1.10}$$

almost surely for each $s < t$, where η is a given finite signed measure. We will refer to (1.10) as a *stochastic fractional delay differential equation* (SFDDE). Equation (1.10) can be compactly written as

$$dX_t = \int_{[0,\infty)} D^\beta X_{t-u}\, \eta(du)\, dt + dL_t, \quad t \in \mathbb{R}, \tag{1.11}$$

with $(D^\beta X_t)_{t\in\mathbb{R}}$ defined in Proposition 3.6. Representation (1.11) is, for instance, convenient in order to argue that solutions are semimartingales.

In Section 3 we show that, for a wide range of measures η, there exists a unique purely non-deterministic process $(X_t)_{t\in\mathbb{R}}$ satisfying the SFDDE (1.10). In addition, we study the behavior of the autocovariance function and the spectral density of $(X_t)_{t\in\mathbb{R}}$ and verify that ((i))-((ii)) hold. We end Section 3 by providing an explicit (prediction) formula for computing $\mathbb{E}[X_t \mid X_u, u \le s]$. In Section 4 we focus on delay measures η of exponential type, that is,

$$\eta(du) = -\kappa\delta_0(du) + f(u)\, du, \tag{1.12}$$

where $f(t) = \mathbb{1}_{[0,\infty)}(t) b^T e^{At} e_1$ with $e_1 = (1, 0, \ldots, 0)^T \in \mathbb{R}^n$, $b \in \mathbb{R}^n$, and A an $n \times n$ matrix with a spectrum contained in $\{z \in \mathbb{C} : \mathrm{Re}(z) < 0\}$. Besides relating this subclass to the FICARMA processes, we study two special cases of (1.12) in detail, namely the Ornstein-Uhlenbeck type presented in (1.7) and

$$dX_t = \int_0^\infty D^\beta X_{t-u} f(u)\, du\, dt + dL_t, \quad t \in \mathbb{R}. \tag{1.13}$$

Equation (1.13) is interesting to study as it collapses to an ordinary SDDE (cf. Propostion 4.2), and hence constitutes an example of a long-range dependent solution to equation (1.9) with $I^\beta L_t - I^\beta L_s$ replaced by $L_t - L_s$. While (1.13) falls into the overall setup of [2], the results obtained in that paper do, however, not apply. Finally, based on the two examples (1.6) and (1.13), we investigate some numerical aspects in Section 5, including the task of simulating $(X_t)_{t\in\mathbb{R}}$ from the defining equation. The proofs of all the results presented in Section 3 and 4 are given in Section 6.2. We start with a preliminary section which recalls a few definitions and results that will be used repeatedly.

2 Preliminaries

For a measure μ on the Borel σ-field $\mathscr{B}(\mathbb{R})$ on $\mathbb{R}$, let $L^p(\mu)$ denote the L^p space relative to μ. If μ is the Lebesgue measure we suppress the dependence on μ and write L^p instead of $L^p(\mu)$. By a finite signed measure we refer to a set function $\mu : \mathscr{B}(\mathbb{R}) \to \mathbb{R}$ of the form $\mu = \mu^+ - \mu^-$, where μ^+ and μ^- are two finite singular measures. Integration of a function f with respect to μ is defined (in an obvious way) whenever $f \in L^1(|\mu|)$ where $|\mu| := \mu^+ + \mu^-$. The convolution of two measurable functions $f, g : \mathbb{R} \to \mathbb{C}$ is defined as

$$f * g(t) = \int_{\mathbb{R}} f(t-u)g(u)\,du$$

whenever $f(t-\cdot)g \in L^1$. Similarly, if μ is a finite signed measure, we set

$$f * \mu(t) = \int_{\mathbb{R}} f(t-u)\,\mu(du)$$

if $f(t-\cdot) \in L^1(|\mu|)$. For such μ set

$$D(\mu) = \left\{ z \in \mathbb{C} \,:\, \int_{\mathbb{R}} e^{\mathrm{Re}(z)u}\,|\mu|(du) < \infty \right\}.$$

Then we define the bilateral Laplace transform $\mathcal{L}[\mu] : D(\mu) \to \mathbb{C}$ of μ by

$$\mathcal{L}[\mu](z) = \int_{\mathbb{R}} e^{zu}\,\mu(du), \quad z \in D(\mu),$$

and the Fourier transform by $\mathcal{F}[\mu](y) = \mathcal{L}[f](iy)$ for $y \in \mathbb{R}$. If $f \in L^1$ we will write $\mathcal{L}[f] = \mathcal{L}[f(u)\,du]$ and $\mathcal{F}[f] = \mathcal{F}[f(u)\,du]$. We also note that $\mathcal{F}[f] \in L^2$ when $f \in L^1 \cap L^2$ and that $\mathcal{F}$ can be extended to an isometric isomorphism from L^2 onto L^2 by Plancherel's theorem.

Recall that a Lévy process is the continuous-time analogue to the (discrete time) random walk. More precisely, a one-sided Lévy process $(L_t)_{t\geq 0}$, $L_0 = 0$, is a stochastic process having stationary independent increments and cádlág sample paths. From these properties it follows that the distribution of L_1 is infinitely divisible, and the distribution of $(L_t)_{t\geq 0}$ is determined from L_1 via the relation $\mathbb{E}[e^{iyL_t}] = \exp\{t \log \mathbb{E}[e^{iyL_1}]\}$ for $y \in \mathbb{R}$ and $t \geq 0$. The definition is extended to a two-sided Lévy process $(L_t)_{t\in\mathbb{R}}$ by taking a one-sided Lévy process $(L_t^1)_{t\geq 0}$ together with an independent copy $(L_t^2)_{t\geq 0}$ and setting $L_t = L_t^1$ if $t \geq 0$ and $L_t = -L^2_{(-t)-}$ if $t < 0$. If $\mathbb{E}[L_1^2] < \infty$, $\mathbb{E}[L_1] = 0$ and $f \in L^2$, the integral $\int_{\mathbb{R}} f(u)\,dL_u$ is well-defined as an L^2 limit of integrals of step functions, and the following isometry property holds:

$$\mathbb{E}\Big[\Big(\int_{\mathbb{R}} f(u)\,dL_u\Big)^2\Big] = \mathbb{E}[L_1^2]\int_{\mathbb{R}} f(u)^2\,du.$$

For more on Lévy processes and integrals with respect to these, see [25, 30]. Finally, for two functions $f,g:\mathbb{R}\to\mathbb{R}$ and $a\in[-\infty,\infty]$ we write $f(t)=o(g(t))$, $f(t)=O(g(t))$ and $f(t)\sim g(t)$ as $t\to a$ if

$$\lim_{t\to a}\frac{f(t)}{g(t)}=0,\quad \limsup_{t\to a}\Big|\frac{f(t)}{g(t)}\Big|<\infty \quad\text{and}\quad \lim_{t\to a}\frac{f(t)}{g(t)}=1,$$

respectively.

3 The stochastic fractional delay differential equation

Let $(L_t)_{t\in\mathbb{R}}$ be a Lévy process with $\mathbb{E}[L_1^2]<\infty$ and $\mathbb{E}[L_1]=0$, and let $\beta\in(0,1/2)$. Without loss of generality we will assume that $\mathbb{E}[L_1^2]=1$. Denote by η a finite (possibly signed) measure on $[0,\infty)$ with

$$\int_{[0,\infty)} u\,|\eta|(du)<\infty. \tag{3.1}$$

Then we will say that a process $(X_t)_{t\in\mathbb{R}}$ with $\mathbb{E}[|X_0|]<\infty$ is a solution to the corresponding SFDDE if it is stationary and satisfies

$$X_t-X_s=\frac{1}{\Gamma(1-\beta)}\int_{-\infty}^{t}\Big[(t-u)^{-\beta}-(s-u)_+^{-\beta}\Big]\int_{[0,\infty)} X_{u-v}\,\eta(dv)\,du+L_t-L_s \tag{3.2}$$

almost surely for each $s<t$. Note that equation (3.2) is indeed well-defined, since η is finite, $(X_t)_{t\in\mathbb{R}}$ is bounded in $L^1(\mathbb{P})$ and

$$\big(D_-^{\beta}\mathbb{1}_{(s,t]}\big)(u):=\frac{1}{\Gamma(1-\beta)}\Big[(t-u)_+^{-\beta}-(s-u)_+^{-\beta}\Big],\quad u\in\mathbb{R}, \tag{3.3}$$

belongs to L^1. In line with [12] we write $D_-^{\beta}\mathbb{1}_{(s,t]}$ rather than $D^{\beta}\mathbb{1}_{(s,t]}$ in (3.3) to emphasize that it is the right-sided version of the (Riemann-Liouville) fractional derivative of $\mathbb{1}_{(s,t]}$. As noted in the introduction, we will often write (3.2) shortly as

$$dX_t=\int_{[0,\infty)} D^{\beta}X_{t-u}\,\eta(du)\,dt+dL_t,\quad t\in\mathbb{R}, \tag{3.4}$$

where $(D^{\beta}X_t)_{t\in\mathbb{R}}$ is a suitable fractional derivative of $(X_t)_{t\in\mathbb{R}}$ (defined in Proposition 3.6).

In order to study which choices of η that lead to a stationary solution to (3.2) we introduce the function $h=h_{\beta,\eta}:\{z\in\mathbb{C}:\mathrm{Re}(z)\le 0\}\to\mathbb{C}$ given by

$$h(z)=(-z)^{1-\beta}-\int_{[0,\infty)} e^{zu}\,\eta(du). \tag{3.5}$$

Here, and in the following, we define $z^{\gamma}=r^{\gamma}e^{i\gamma\theta}$ using the polar representation $z=re^{i\theta}$ for $r>0$ and $\theta\in(-\pi,\pi]$. This definition corresponds to $z^{\gamma}=e^{\gamma\log z}$, using the principal branch of the complex logarithm, and hence $z\mapsto z^{\gamma}$ is analytic on $\mathbb{C}\setminus\{z\in\mathbb{R}:z\le 0\}$. In particular, this means that h is analytic on $\{z\in\mathbb{C}:\mathrm{Re}(z)<0\}$.

Proposition 3.1. *Suppose that $h(z)$ defined in (3.5) is non-zero for every $z \in \mathbb{C}$ with $Re(z) \le 0$. Then there exists a unique $g : \mathbb{R} \to \mathbb{R}$, which belongs to L^γ for $(1-\beta)^{-1} < \gamma \le 2$ and is vanishing on $(-\infty, 0)$, such that*

$$\mathcal{F}[g](y) = \frac{(-iy)^{-\beta}}{h(iy)} \tag{3.6}$$

for $y \in \mathbb{R}$. Moreover, the following statements hold:

(i) *For $t > 0$ the Marchaud fractional derivative $D^\beta g(t)$ at t of g given by*

$$D^\beta g(t) = \frac{\beta}{\Gamma(1-\beta)} \lim_{\delta \downarrow 0} \int_\delta^\infty \frac{g(t) - g(t-u)}{u^{1+\beta}}\, du \tag{3.7}$$

exists, $D^\beta g \in L^1 \cap L^2$ and $\mathcal{F}[D^\beta g](y) = 1/h(iy)$ for $y \in \mathbb{R}$.

(ii) *The function g is the Riemann-Liouville fractional integral of $D^\beta g$, that is,*

$$g(t) = \frac{1}{\Gamma(\beta)} \int_0^t D^\beta g(u)(t-u)^{\beta-1}\, du$$

for $t > 0$.

(iii) *The function g satisfies*

$$g(t) = 1 + \int_0^t \left(D^\beta g\right) * \eta(u)\, du, \quad t \ge 0, \tag{3.8}$$

and, for $v \in \mathbb{R}$ and with $D_-^\beta \mathbb{1}_{(s,t]}$ given in (3.3),

$$g(t-v) - g(s-v) = \int_{-\infty}^t \left(D_-^\beta \mathbb{1}_{(s,t]}\right)(u)\, g * \eta(u-v)\, du + \mathbb{1}_{(s,t]}(v). \tag{3.9}$$

Before formulating our main result, Theorem 3.2, recall that a stationary process $(X_t)_{t\in\mathbb{R}}$ with $\mathbb{E}[X_0^2] < \infty$ and $\mathbb{E}[X_0] = 0$ is said to be purely non-deterministic if

$$\bigcap_{t\in\mathbb{R}} \overline{\mathrm{sp}}\{X_s : s \le t\} = \{0\}, \tag{3.10}$$

see [1, Section 4]. Here $\overline{\mathrm{sp}}$ denotes the $L^2(\mathbb{P})$-closure of the linear span.

Theorem 3.2. *Suppose that $h(z)$ defined in (3.5) is non-zero for every $z \in \mathbb{C}$ with $Re(z) \le 0$ and let g be the function introduced in Proposition 3.1. Then the process*

$$X_t = \int_{-\infty}^t g(t-u)\, dL_u, \quad t \in \mathbb{R}, \tag{3.11}$$

is well-defined, centered and square integrable, and it is the unique purely non-deterministic solution to the SFDDE (3.2).

Remark 3.3. Note that we cannot hope to get a uniqueness result without imposing a condition such as (3.10). For instance, the fact that

$$\int_{-\infty}^t \left[(t-u)^{-\beta} - (s-u)_+^{-\beta}\right] du = 0,$$

shows together with (3.2) that $(X_t + U)_{t\in\mathbb{R}}$ is a solution for any $U \in L^1(\mathbb{P})$ as long as $(X_t)_{t\in\mathbb{R}}$ is a solution. Moreover, uniqueness relative to condition (3.10) is similar to that of discrete-time ARFIMA processes, see [8, Theorem 13.2.1].

Remark 3.4. It is possible to generalize (3.2) and Theorem 3.2 to allow for a heavy-tailed distribution of the noise. Specifically, suppose that $(L_t)_{t\in\mathbb{R}}$ is a symmetric α-stable Lévy process for some $\alpha \in (1,2)$, that is, $(L_t)_{t\in\mathbb{R}}$ is a Lévy process and

$$\mathbb{E}\Big[e^{iyL_1}\Big] = e^{-\sigma^\alpha |y|^\alpha}, \quad y \in \mathbb{R},$$

for some $\sigma > 0$. To define the process $(X_t)_{t\in\mathbb{R}}$ in (3.11) it is necessary and sufficient that $g \in L^\alpha$, which is indeed the case if $\beta \in (1, 1-1/\alpha)$ by Proposition 3.1. From this point, using (3.9), we only need a stochastic Fubini result (which can be found in [1, Theorem 3.1]) to verify that (3.2) is satisfied. One will need another notion (and proof) of uniqueness, however, as our approach relies on L^2 theory. For more on stable distributions and corresponding definitions and results, we refer to [29].

Remark 3.5. The process (3.11) and other well-known long-memory processes do naturally share parts of their construction. For instance, they are typically viewed as "borderline" stationary solutions to certain equations. To be more concrete, the ARFIMA process can be viewed as an ARMA process, but where the autoregressive polynomial P is replaced by $\tilde{P} : z \mapsto P(z)(1-z)^\beta$. Although an ordinary ARMA process exists if and only if P is non-zero on the unit circle (and, in the positive case, will be a short memory process), the autoregressive function $\tilde{P}$ of the ARFIMA model will always have a root at $z = 1$. The analogue to the autoregressive polynomial in the non-fractional SDDE model (that is, (3.2) with $D_-^\beta \mathbb{1}_{(s,t]}$ replaced by $\mathbb{1}_{(s,t]}$) is

$$z \mapsto -z - \mathcal{L}[\eta](z), \tag{3.12}$$

where the critical region is on the imaginary axis $\{iy : y \in \mathbb{R}\}$ rather than on the unit circle $\{z \in \mathbb{C} : |z| = 1\}$ (see [3]). The SFDDE corresponds to replacing (3.12) by $z \mapsto -z - (-z)^\beta \mathcal{L}[\eta](z)$, which will always have a root at $z = 0$. However, to ensure existence both in the ARFIMA model and in the SFDDE model, assumptions are made such that these roots will be the only ones in the critical region and their order will be β. For a treatment of ARFIMA processes, we refer to [8, Section 13.2].

The solution $(X_t)_{t\in\mathbb{R}}$ of Theorem 3.2 is causal in the sense that X_t only depends on past increments of the noise $L_t - L_s$, $s \le t$. An inspection of the proof of Theorem 3.2 reveals that one only needs to require that $h(iy) \neq 0$ for all $y \in \mathbb{R}$ for a (possibly non-causal) stationary solution to exist. The difference between the condition that $h(z)$ is non-zero when $\mathrm{Re}(z) = 0$ rather than when $\mathrm{Re}(z) \le 0$ in terms of causality is similar to that of non-fractional SDDEs (see, e.g., [3]).

The next result shows why one may view (3.2) as (3.4). In particular, it reveals that the corresponding solution $(X_t)_{t\in\mathbb{R}}$ is a semimartingale with respect to (the completion of) its own filtration or equivalently, in light of (3.2) and (3.11), the one generated from the increments of $(L_t)_{t\in\mathbb{R}}$.

Proposition 3.6. *Suppose that $h(z)$ is non-zero for every $z \in \mathbb{C}$ with $Re(z) \le 0$ and let $(X_t)_{t\in\mathbb{R}}$ be the solution to (3.2) given in Theorem 3.2. Then, for $t \in \mathbb{R}$, the limit*

$$D^\beta X_t := \frac{\beta}{\Gamma(1-\beta)} \lim_{\delta \downarrow 0} \int_\delta^\infty \frac{X_t - X_{t-u}}{u^{1+\beta}}\, du \tag{3.13}$$

exists in $L^2(\mathbb{P})$*,* $D^\beta X_t = \int_{-\infty}^t D^\beta g(t-u)\,dL_u$*, and it holds that*

$$\begin{aligned}\frac{1}{\Gamma(1-\beta)}\int_{-\infty}^t \Big[(t-u)^{-\beta} - (s-u)_+^{-\beta}\Big]\int_{[0,\infty)} X_{u-v}\,\eta(dv)\,du \\ = \int_s^t \int_{[0,\infty)} D^\beta X_{u-v}\,\eta(dv)\,du\end{aligned} \tag{3.14}$$

almost surely for each $s < t$*.*

We will now provide some properties of the solution $(X_t)_{t\in\mathbb{R}}$ to (3.2) given in (3.11). Since the autocovariance function γ_X takes the form

$$\gamma_X(t) = \int_{\mathbb{R}} g(t+u)g(u)\,du, \quad t \in \mathbb{R}, \tag{3.15}$$

it follows by Plancherel's theorem that $(X_t)_{t\in\mathbb{R}}$ admits a spectral density f_X which is given by

$$f_X(y) = |\mathcal{F}[g](y)|^2 = \frac{1}{|h(iy)|^2}|y|^{-2\beta}, \quad y \in \mathbb{R}. \tag{3.16}$$

(See the appendix for a brief recap of the spectral theory.) The following result concerning γ_X and f_X shows that solutions to (3.2) exhibit a long-memory behavior and that the degree of memory can be controlled by β.

Proposition 3.7. *Suppose that* $h(z)$ *is non-zero for every* $z \in \mathbb{C}$ *with* $Re(z) \le 0$ *and let* γ_X *and* f_X *be the functions introduced in (3.15)-(3.16). Then it holds that*

$$\gamma_X(t) \sim \frac{\Gamma(1-2\beta)}{\Gamma(\beta)\Gamma(1-\beta)\eta([0,\infty))^2}t^{2\beta-1} \quad \textit{and} \quad f_X(y) \sim \frac{1}{\eta([0,\infty))^2}|y|^{-2\beta}$$

as $t \to \infty$ *and* $y \to 0$*, respectively. In particular,* $\int_{\mathbb{R}} |\gamma_X(t)|\,dt = \infty$*.*

While the behavior of $\gamma_X(t)$ as $t \to \infty$ is controlled by β, the content of Proposition 3.8 is that the behavior of $\gamma_X(t)$ as $t \to 0$, and thus the $L^2(\mathbb{P})$-Hölder continuity of the sample paths of $(X_t)_{t\in\mathbb{R}}$ (cf. Remark 3.9), is unaffected by β.

Proposition 3.8. *Suppose that* $h(z)$ *is non-zero for every* $z \in \mathbb{C}$ *with* $Re(z) \le 0$*, let* $(X_t)_{t\in\mathbb{R}}$ *be the solution to (3.2) and denote by* ρ_X *its ACF. Then it holds that* $1 - \rho_X(t) \sim t$ *as* $t \downarrow 0$*.*

Remark 3.9. Recall that for a given $\gamma > 0$, a centered and square integrable process $(X_t)_{t\in\mathbb{R}}$ with stationary increments is said to be locally γ-Hölder continuous in $L^2(\mathbb{P})$ if there exists a constant $C > 0$ such that

$$\frac{\mathbb{E}\Big[(X_t - X_0)^2\Big]}{t^{2\gamma}} \le C$$

for all sufficiently small $t > 0$. By defining the semi-variogram

$$\gamma_V(t) := \tfrac{1}{2}\mathbb{E}[(X_t - X_0)^2], \quad t \in \mathbb{R},$$

we see that $(X_t)_{t\in\mathbb{R}}$ is locally γ-Hölder continuous if and only if $\gamma_V(t) = O(t^{2\gamma})$ as $t \to 0$. When $(X_t)_{t\in\mathbb{R}}$ is stationary we have the relation $\gamma_V = \gamma_X(0)(1-\rho_X)$, from which it follows that the $L^2(\mathbb{P})$ notion of Hölder continuity can be characterized in terms of the behavior of the ACF at zero. In particular, Proposition 3.8 shows that the solution $(X_t)_{t\in\mathbb{R}}$ to (3.2) is locally γ-Hölder continuous if and only if $\gamma \leq 1/2$. The behavior of the ACF at zero has been used as a measure of roughness of the sample paths in for example [4, 5].

Remark 3.10. As a final comment on the path properties of the solution $(X_t)_{t\in\mathbb{R}}$ to (3.2), observe that

$$X_t - X_s = \int_s^t \int_{[0,\infty)} D^\beta X_{u-v}\, \eta(dv)\, du + L_t - L_s$$

for each $s < t$ almost surely by Proposition 3.6. This shows that $(X_t)_{t\in\mathbb{R}}$ can be chosen so that it has jumps at the same time (and of the same size) as $(L_t)_{t\in\mathbb{R}}$. This is in contrast to models driven by a fractional Lévy process, such as (1.9), since $(I^\beta L_t)_{t\in\mathbb{R}}$ is continuous in t (see [20, Theorem 3.4]).

We end this section by providing a formula for computing $\mathbb{E}[X_t \mid X_u,\, u \leq s]$ for any $s < t$. One should compare its form to those obtained for other fractional models (such as the one in [2, Theorem 3.2] where, as opposed to Proposition 3.11, the prediction is expressed not only in terms of its own past, but also the past noise).

Proposition 3.11. *Suppose that $h(z)$ is non-zero for every $z \in \mathbb{C}$ with $Re(z) \leq 0$ and let $(X_t)_{t\in\mathbb{R}}$ denote the solution to (3.2). Then, for any $s < t$, it holds that*

$$\begin{aligned}\mathbb{E}[X_t \mid X_u,\, u \leq s] &\\ &= g(t-s)X_s + \int_{[0,t-s)} \int_{-\infty}^s X_w \int_{[0,\infty)} \Big(D_-^\beta \mathbb{1}_{(s,t-u]}\Big)(v+w)\, \eta(dv)\, dw\, g(du)\end{aligned}$$

where $g(du) = \delta_0(du) + (D^\beta g) * \eta(u)\, du$ is the Lebesgue-Stieltjes measure induced by g.

4 Delays of exponential type

Let A be an $n \times n$ matrix where all its eigenvalues belong to $\{z \in \mathbb{C} : \mathrm{Re}(z) < 0\}$, and let $b \in \mathbb{R}^n$ and $\kappa \in \mathbb{R}$. In this section we restrict our attention to measures η of the form

$$\eta(du) = -\kappa\delta_0(du) + f(u)\, du, \quad \text{with} \quad f(u) = b^T e^{Au} e_1, \tag{4.1}$$

where $e_1 := (1, 0, \ldots, 0)^T \in \mathbb{R}^n$. Note that e_1 is used as a normalization; the effect of replacing e_1 by any $c \in \mathbb{R}^n$ can be incorporated in the choice of A and b. It is well-known that the assumption on the eigenvalues of A imply that all the entries of e^{Au} decay exponentially fast as $u \to \infty$, so that η is a finite measure on $[0,\infty)$ with moments of any order. Since the Fourier transform $\mathcal{F}[f]$ of f is given by

$$\mathcal{F}[f](y) = -b^T (A + iyI_n)^{-1} e_1, \quad y \in \mathbb{R},$$

it admits a fraction decomposition; that is, there exist real polynomials $Q, R : \mathbb{C} \to \mathbb{C}$, Q being monic with the eigenvalues of A as its roots and being of larger degree than R, such that

$$\mathcal{F}[f](y) = -\frac{R(-iy)}{Q(-iy)} \tag{4.2}$$

for $y \in \mathbb{R}$. (This is a direct consequence of the inversion formula $B^{-1} = \operatorname{adj}(B)/\det(B)$.) By assuming that Q and R have no common roots, the pair (Q,R) is unique. The following existence and uniqueness result is simply an application of Theorem 3.2 to the particular setup in question:

Corollary 4.1. *Let Q and R be given as in (4.2). Suppose that $\kappa + b^T A^{-1} e_1 \neq 0$ and*

$$Q(z)\Big[z + \kappa z^\beta\Big] + R(z) z^\beta \neq 0 \tag{4.3}$$

for all $z \in \mathbb{C} \setminus \{0\}$ with $Re(z) \geq 0$. Then there exists a unique purely non-deterministic solution $(X_t)_{t\in\mathbb{R}}$ to (3.2) with η given by (4.1) and it is given by (3.11) with $g : \mathbb{R} \to \mathbb{R}$ characterized through the relation

$$\mathcal{F}[g](y) = \frac{Q(-iy)}{Q(-iy)\Big[-iy + \kappa(-iy)^\beta\Big] + R(-iy)(-iy)^\beta}, \quad y \in \mathbb{R}. \tag{4.4}$$

Before giving examples we state Proposition 4.2, which shows that the general SFDDE (3.2) can be written as

$$dX_t = -\kappa D^\beta X_t\, dt + \int_0^\infty X_{t-u} D^\beta f(u)\, du\, dt + dL_t, \quad t \in \mathbb{R}, \tag{4.5}$$

when η is of the form (4.1). In case $\kappa = 0$, (4.5) is a (non-fractional) SDDE. However, the usual existence results obtained in this setting (for instance, those in [3] and [17]) are not applicable, since the delay measure $D^\beta f(u)\, du$ has unbounded support and zero total mass $\int_0^\infty D^\beta f(u)\, du = 0$.

Proposition 4.2. *Let f be of the form (4.1). Then $D^\beta f : \mathbb{R} \to \mathbb{R}$ defined by $D^\beta f(t) = 0$ for $t \leq 0$ and*

$$D^\beta f(t) = \frac{1}{\Gamma(1-\beta)} b^T \Big(A e^{At} \int_0^t e^{-Au} u^{-\beta}\, du + t^{-\beta} I_n \Big) e_1$$

for $t > 0$ belongs to $L^1 \cap L^2$. If in addition, (4.3) holds, $\kappa + b^T A^{-1} e_1 \neq 0$, and $(X_t)_{t\in\mathbb{R}}$ is the solution given in Corollary 4.1, then

$$\int_0^\infty D^\beta X_{t-u} f(u)\, du = \int_0^\infty X_{t-u}\, D^\beta f(u)\, du$$

almost surely for any $t \in \mathbb{R}$.

Remark 4.3. Due to the structure of the function g in (4.4) one may, in line with the interpretation of CARMA processes, think of the corresponding solution $(X_t)_{t\in\mathbb{R}}$ as a stationary process that satisfies the formal equation

$$\Big(Q(D)\Big[D + \kappa D^\beta\Big] + R(D) D^\beta\Big) X_t = Q(D) D L_t, \quad t \in \mathbb{R}, \tag{4.6}$$

where D denotes differentiation with respect to t and D^β is a suitable fractional derivative. Indeed, by heuristically applying the Fourier transform $\mathcal{F}$ to (4.6) and using computation rules such as $\mathcal{F}[DX](y) = (-iy)\mathcal{F}[X](y)$ and $\mathcal{F}[D^\beta X](y) = (-iy)^\beta \mathcal{F}[X](y)$, one ends up concluding that $(X_t)_{t\in\mathbb{R}}$ is of the form (3.11) with g characterized by (4.4). For two monic polynomials P and Q with $q := \deg(Q) = \deg(P) - 1$ and all their roots contained in $\{z \in \mathbb{C} : \mathrm{Re}(z) < 0\}$, consider the FICARMA$(q+1,\beta,q)$ process $(X_t)_{t\in\mathbb{R}}$. Heuristically, by applying $\mathcal{F}$ as above, $(X_t)_{t\in\mathbb{R}}$ may be thought of as the solution to $P(D)D^\beta X_t = Q(D)DL_t$, $t \in \mathbb{R}$. By choosing the polynomial R and the constant κ such that $P(z) = Q(z)[z+\kappa] + R(z)$ we can think of $(X_t)_{t\in\mathbb{R}}$ as the solution to the formal equation

$$\Big(Q(D)\Big[D^{1+\beta} + \kappa D^\beta\Big] + R(D)D^\beta\Big)X_t = Q(D)DL_t, \quad t \in \mathbb{R}. \tag{4.7}$$

It follows that (4.6) and (4.7) are closely related, the only difference being that $D + \kappa D^\beta$ is replaced by $D^{1+\beta} + \kappa D^\beta$. In particular, one may view solutions to SFDDEs corresponding to measures of the form (4.1) as being of the same type as FICARMA processes. While the considerations above apply only to the case where $\deg(P) = q+1$, it should be possible to extend the SFDDE framework so that solutions are comparable to the FICARMA processes in the general case $\deg(P) > q$ by following the lines of [2], where similar theory is developed for the SDDE setting.

We will now give two examples of (4.5).

Example 4.4. Consider choosing $\eta = -\kappa\delta_0$ for some $\kappa > 0$ so that (3.2) becomes

$$X_t - X_s = -\frac{\kappa}{\Gamma(1-\beta)} \int_{-\infty}^t \Big[(t-u)^{-\beta} - (s-u)_+^{-\beta}\Big] X_u \, du + L_t - L_s \tag{4.8}$$

for $s < t$ or, in short,

$$dX_t = -\kappa D^\beta X_t \, dt + dL_t, \quad t \in \mathbb{R}. \tag{4.9}$$

To argue that a unique purely non-deterministic solution exists, we observe that $Q(z) = 1$ and $R(z) = 0$ for all $z \in \mathbb{C}$. Thus, in light of Corollary 4.1 and (4.3), it suffices to argue that $z + \kappa z^\beta \neq 0$ for all $z \in \mathbb{C}\setminus\{0\}$ with $\mathrm{Re}(z) \geq 0$. By writing such z as $z = re^{i\theta}$ for a suitable $r > 0$ and $\theta \in [-\pi/2, \pi/2]$, the condition may be written as

$$\Big(r\cos(\theta) + \kappa r^\beta \cos(\beta\theta)\Big) + i\Big(r\sin(\theta) + \kappa r^\beta \sin(\beta\theta)\Big) \neq 0. \tag{4.10}$$

If the imaginary part of the left-hand side of (4.10) is zero it must be the case that $\theta = 0$, since $\kappa > 0$ while $\sin(\theta)$ and $\sin(\beta\theta)$ are of the same sign. However, if $\theta = 0$, the real part of the left-hand side of (4.10) is $r + \kappa r^\beta > 0$. Consequently, Corollary 4.1 implies that a solution to (4.9) is characterized by (3.11) and $\mathcal{F}[g](y) = ((-iy)^\beta\kappa - iy)^{-1}$ for $y \in \mathbb{R}$. In particular, γ_X takes the form

$$\gamma_X(t) = \int_{\mathbb{R}} \frac{e^{ity}}{y^2 + 2\kappa\sin(\frac{\beta\pi}{2})|y|^{1+\beta} + \kappa^2|y|^{2\beta}} \, dy. \tag{4.11}$$

In Figure 1 we have plotted the ACF of $(X_t)_{t\in\mathbb{R}}$ using (4.11) with $\kappa = 1$ and $\beta \in \{0.1, 0.2, 0.3, 0.4\}$. We compare it to the ACF of the corresponding fractional Ornstein-Uhlenbeck process (equivalently, the FICARMA$(1,\beta,0)$ process) which was presented in (1.4). To do so, we use that its autocovariance function γ_β is given by

$$\gamma_\beta(t) = \int_{\mathbb{R}} \frac{e^{ity}}{|y|^{2(1+\beta)} + \kappa^2|y|^{2\beta}} \, dy. \tag{4.12}$$

From these plots it becomes evident that, although the ACFs share the same behavior at infinity, they behave differently near zero. In particular, we see that the ACF of $(X_t)_{t\in\mathbb{R}}$ decays more rapidly around zero, which is in line with Proposition 3.8 and the fact that the $L^2(\mathbb{P})$-Hölder continuity of the fractional Ornstein-Uhlenbeck process increases as β increases (cf. the introduction).

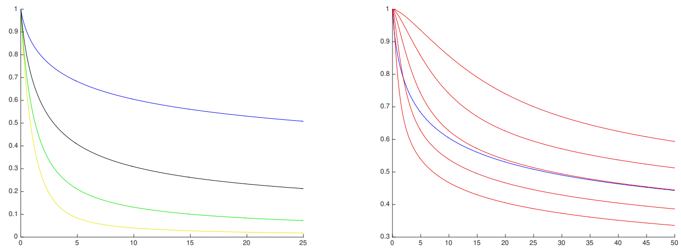


Figure 1: The left plot is the ACF based on (4.11) with $\beta = 0.1$ (yellow), $\beta = 0.2$ (green), $\beta = 0.3$ (black) and $\beta = 0.4$ (blue). With $\beta = 0.4$ fixed, the plot on the right compares the ACF based on (4.11) with $\kappa = 1$ (blue) to the ACF based on (4.12) for $\kappa = 0.125, 0.25, 0.5, 1, 2$ (red) where the ACF decreases in κ, in particular, the top curve corresponds to $\kappa = 0.125$ and the bottom to $\kappa = 2$.

Example 4.5. Suppose that η is is given by (4.1) with $\kappa = 0$, $A = -\kappa_1$, and $b = -\kappa_2$ for some $\kappa_1, \kappa_2 > 0$. In this case, $f(t) = -\kappa_2 e^{-\kappa_1 t}$ and (4.5) becomes

$$dX_t = \frac{\kappa_2}{\Gamma(1-\beta)} \int_0^\infty X_{t-u}\Big(\kappa_1 e^{-\kappa_1 u} \int_0^u e^{\kappa_1 v} v^{-\beta}\, dv - u^{-\beta}\Big) du\, dt + dL_t, \tag{4.13}$$

and since $Q(z) = z + \kappa_1$ and $R(z) = \kappa_2$ we have that

$$zQ(z) + R(z)z^\beta = z^2 + \kappa_1 z + \kappa_2 z^\beta.$$

To verify (4.3), set $z = x + iy$ for $x > 0$ and $y \in \mathbb{R}$ and note that

$$\begin{aligned} z^2 + \kappa_1 z + \kappa_2 z^\beta = &\Big(x^2 - y^2 + \kappa_1 x + \kappa_2 \cos(\beta\theta_z)|z|^\beta\Big) \\ &+ i\Big(\kappa_1 y + 2xy + \kappa_2 \sin(\beta\theta_z)|z|^\beta\Big) \end{aligned} \tag{4.14}$$

for a suitable $\theta_z \in (-\pi/2, \pi/2)$. For the imaginary part of (4.14) to be zero it must be the case that

$$(\kappa_1 + 2x)y = -\kappa_2 \sin(\beta\theta_z)|z|^\beta,$$

and this can only happen if $y = 0$, since $x, \kappa_1, \kappa_2 > 0$ and the sign of y is the same as that of $\sin(\beta\theta_z)$. However, if $y = 0$ it is easy to see that the real part of (4.14) cannot be zero for any $x > 0$, so we conclude that (4.3) holds and that there exists a stationary solution $(X_t)_{t\in\mathbb{R}}$ given through the kernel (4.4). The autocovariance function γ_X is given by

$$\gamma_X(t) = \int_{\mathbb{R}} e^{ity} \frac{y^2 + \kappa_1^2}{y^4 + 2\kappa_2\Big(\kappa_1\gamma_2|y|^{1+\beta} - \gamma_1|y|^{2+\beta}\Big) + \kappa_1^2 y^2 + \kappa_2^2|y|^{2\beta}}\, dy \tag{4.15}$$

where $\gamma_1 = \cos(\beta\pi/2)$ and $\gamma_2 = \sin(\beta\pi/2)$. The polynomials to the associated FICARMA$(2,\beta,1)$ process are given by $P(z) = z^2 + \kappa_1 z + \kappa_2$ and $Q(z) = z + \kappa_1$ (see Remark 4.3) and the autocovariance function γ_β takes the form

$$\gamma_\beta(t) = \int_{\mathbb{R}} e^{ity} \frac{y^2 + \kappa_1^2}{|y|^{4+2\beta} + (\kappa_1^2 - 2\kappa_2)|y|^{2+2\beta} + \kappa_2^2|y|^{2\beta}}\, dy. \tag{4.16}$$

In Figure 2 we have plotted the ACF based on (4.15) for $\kappa_1 = 1$ and various values of κ_2 and β. For comparison we have also plotted the ACF based on (4.16) for the same choices of κ_1, κ_2 and β.

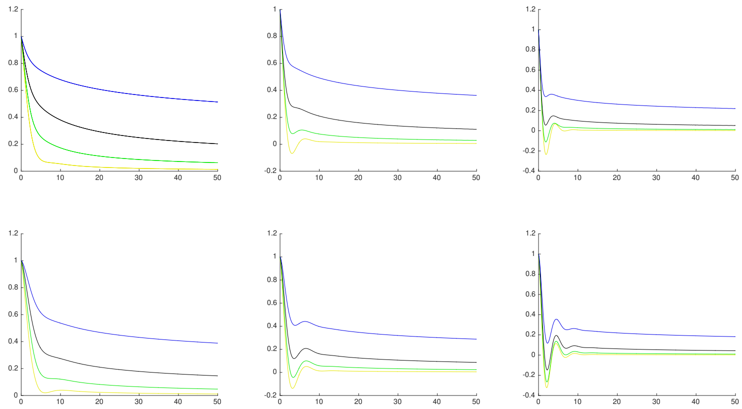


Figure 2: First row is ACF based on (4.15), second row is ACF based on (4.16), and the columns correspond to $\kappa_2 = 0.5$, $\kappa_2 = 1$ and $\kappa_2 = 2$, respectively. Within each plot, the lines correspond to $\beta = 0.1$ (yellow), $\beta = 0.2$ (green), $\beta = 0.3$ (black) and $\beta = 0.4$ (blue). In all plots, $\kappa_1 = 1$.

5 Simulation from the SFDDE

In the following we will focus on simulating from (3.2). We begin this simulation study by considering the Ornstein-Uhlenbeck type equation discussed in Example 4.4 with $\kappa = 1$ and under the assumption that $(L_t)_{t\in\mathbb{R}}$ is a standard Brownian motion. Let $c_1 = 100/\Delta$ and $c_2 = 2000/\Delta$. We generate a simulation of the solution process $(X_t)_{t\in\mathbb{R}}$ on a grid of size $\Delta = 0.01$ and with $3700/\Delta$ steps of size Δ starting from $-c_1 - c_2$ and ending at $1600/\Delta$. Initially, we set X_t equal to zero for the first c_1 points in the grid and then discretize (4.8) using the approximation

$$
\begin{aligned}
&\int_{\mathbb{R}} \Big[(n\Delta - u)_+^{-\beta} - ((n-1)\Delta - u)_+^{-\beta} \Big] X_u \, du \\
&\simeq \frac{1}{1-\beta} \Delta^{1-\beta} X_{(n-1)\Delta} \\
&\quad + \sum_{k=n-c_1}^{n-1} \frac{1}{2}(X_{k\Delta} + X_{(k-1)\Delta}) \int_{(k-1)\Delta}^{k\Delta} \Big[(n\Delta - u)_+^{-\beta} - ((n-1)\Delta - u)_+^{-\beta} \Big] du \\
&= \frac{1}{1-\beta} \Delta^{1-\beta} X_{(n-1)\Delta} + \frac{1}{1-\beta} \sum_{k=n-c_1}^{n-1} \tfrac{1}{2}(X_{k\Delta} + X_{(k-1)\Delta}) \\
&\quad \cdot \Big(2((n-k-1)\Delta)^{1-\beta} - ((n-k)\Delta)^{1-\beta} - ((n-k-2)\Delta)^{1-\beta} \Big)
\end{aligned}
$$

for $n = -c_2 + 1, \dots, 3700/\Delta - c_2 - c_1$. Next, we disregard the first $c_1 + c_2$ values of the simulated sample path to obtain an approximate sample from the stationary distribution. We assume that the process is observed on a unit grid resulting in simulated values $X_1, \dots, X_{1600}$. This is repeated 200 times, and in every repetition the sample ACF based on $X_1, \dots, X_L$ is computed for $t = 1, \dots, 25$ and $L = 100, 400, 1600$. In long-memory models, the sample mean $\bar{X}_L$ can be a poor approximation to the true mean $\mathbb{E}[X_0]$ even for large L, and this may result in considerable negative (finite sample) bias in the sample ACF (see, e.g., [22]). Due to this bias, it may be difficult to see if we succeed in simulating from (3.2), and hence we will assume that $\mathbb{E}[X_0]$ is known to be zero when computing the sample ACF. We calculate the 95% confidence interval

$$
\Big[\bar{\rho}(k) - 1.96 \tfrac{\hat{\sigma}(k)}{\sqrt{200}}, \bar{\rho}(k) + 1.96 \tfrac{\hat{\sigma}(k)}{\sqrt{200}} \Big],
$$

for the mean of the sample ACF based on L observations at lag k. Here $\bar{\rho}(k)$ is the sample mean and $\hat{\sigma}(k)$ is the sample standard deviations of the ACF at lag k based on the 200 replications. In Figure 3, the theoretical ACFs and the corresponding 95% confidence intervals for the mean of the sample ACFs are plotted for $\beta = 0.1, 0.2$ and $L = 100, 400, 1600$. We see that, when correcting for the bias induced by an unknown mean $\mathbb{E}[X_0]$, simulation from equation (4.8) results in a fairly unbiased estimator of the ACF for small values of β. When $\beta > 0.25$, in the case where the ACF of $(X_t)_{t\in\mathbb{R}}$ is not even in L^2, the results are more unstable as it requires large values of c_1 and c_2 to ensure that the simulation gives a good approximation to the stationary distribution of $(X_t)_{t\in\mathbb{R}}$. Moreover, even after correcting for the bias induced by an unknown mean of the observed process, the sample ACF for the ARFIMA process shows considerable finite sample bias when $\beta > 0.25$, see [22], and hence we may expect this to apply to solutions to (3.2) as well.

In Figure 4 we have plotted box plots for the 200 replications of the sample ACF for $\beta = 0.1, 0.2$ and $L = 100, 400, 1600$. We see that the sample ACFs have the expected convergence when L grows and that the distribution is more concentrated in the case where less memory is present.

Following the same approach as above, we simulate the solution to the equation discussed in Example 4.5. Specifically, the simulation is based on equation (3.2), restricted to the case where $\eta(dv) = -e^{-v}\, dv$ and $(L_t)_{t\in\mathbb{R}}$ is a standard Brownian

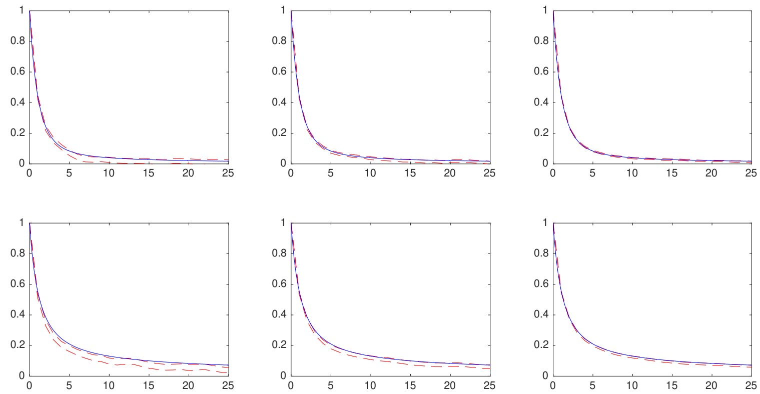


Figure 3: Theoretical ACF and 95% confidence intervals of the mean of the sample ACF based on 200 replications of $X_1,\ldots,X_L$. Columns correspond to $L = 100$, $L = 400$ and $L = 1600$, respectively, and rows correspond to $\beta = 0.1$ and $\beta = 0.2$, respectively. The model is (4.8).

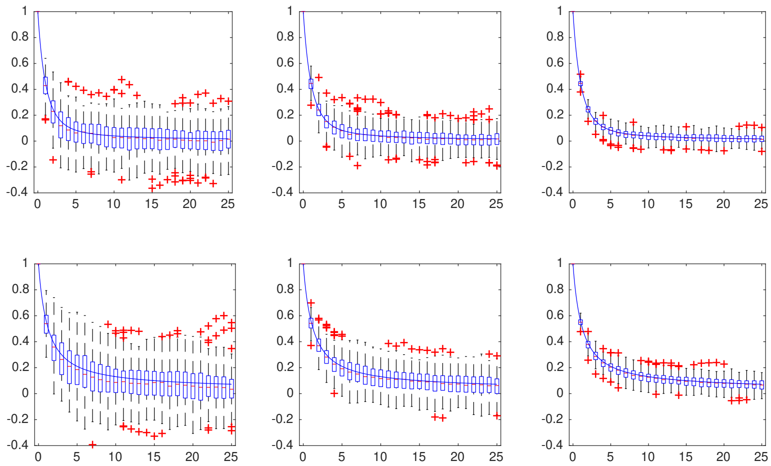


Figure 4: Box plots for the sample ACF based on 200 replications of $X_1,\ldots,X_L$ together with the theoretical ACF. Columns correspond to $L = 100$, $L = 400$ and $L = 1600$, respectively, and rows correspond to $\beta = 0.1$ and $\beta = 0.2$, respectively. The model is (4.8).

motion. In this case, we use the approximation

$$
\begin{aligned}
&\int_{\mathbb{R}} \Big[(n\Delta - u)_+^{-\beta} - ((n-1)\Delta - u)_+^{-\beta}\Big] \int_0^\infty X_{u-v}\, e^{-v}\, dv\, du \\
&= \int_0^\infty X_{n\Delta - v} \int_0^v \Big[(u-\Delta)_+^{-\beta} - u_+^{-\beta}\Big] e^{u-v}\, du\, dv \\
&\simeq \frac{1}{2}\Delta X_{(n-1)\Delta} f(\Delta) \\
&\quad + \sum_{k=2}^{c_1} \frac{1}{4}\Delta (X_{(n-k)\Delta} + X_{(n-k+1)\Delta})(\varphi(k\Delta) + \varphi((k-1)\Delta))
\end{aligned}
$$

where $\varphi : \mathbb{R} \to \mathbb{R}$ is given by

$$\varphi(v) = \int_0^v \Big[(u-\Delta)_+^{-\beta} - u^{-\beta}\Big] e^{u-v}\, dv.$$

We approximate φ recursively by noting that

$$\begin{aligned}
\varphi(k\Delta) &= \int_0^{k\Delta} \Big[(u-\Delta)_+^{-\beta} - u^{-\beta}\Big] e^{u-k\Delta}\, dv \\
&\simeq \frac{1}{2}(1+e^{-\Delta}) \int_{(k-1)\Delta}^{k\Delta} \Big[(u-\Delta)_+^{-\beta} - u_+^{-\beta}\Big] dv + e^{-\Delta}\varphi((k-1)\Delta) \\
&= \frac{1}{2(1-\beta)}\Big(1+e^{-\Delta}\Big)\Big[((k-1)\Delta)^{1-\beta} - (k\Delta)^{1-\beta})\Big] + e^{-\Delta}\varphi((k-1)\Delta)
\end{aligned}$$

for $k \geq 1$. The theoretical ACFs and corresponding 95% confidence intervals are plotted in Figure 5 and the box plots in Figure 6. The findings are consistent with first example that we considered.

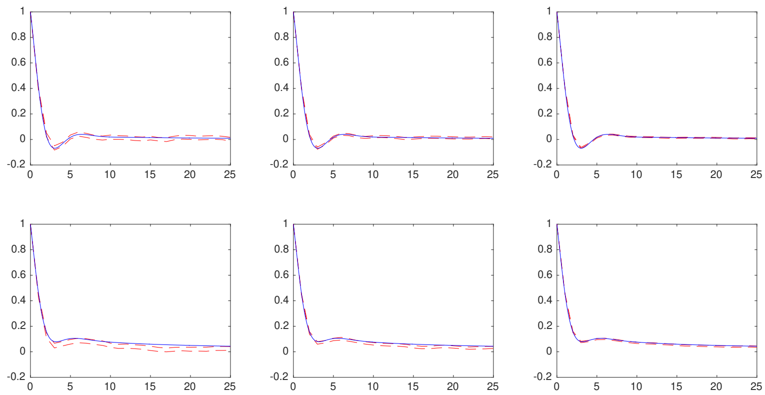


Figure 5: Theoretical ACF and 95% confidence intervals of the mean of the sample ACF sample based on 200 replications of $X_1, \dots, X_L$. Columns correspond to $L = 100$, $L = 400$ and $L = 1600$, respectively, and rows correspond to $\beta = 0.1$ and $\beta = 0.2$, respectively. The model is (4.13).

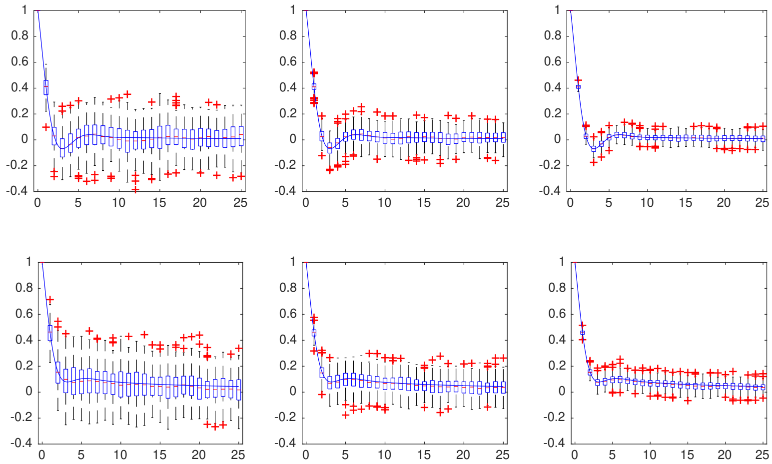


Figure 6: Box plots for the sample ACF based on 200 replications of $X_1,\ldots,X_L$ together with the theoretical ACF. Columns correspond to $L = 100$, $L = 400$ and $L = 1600$, respectively, and rows correspond to $\beta = 0.1$ and $\beta = 0.2$, respectively. The model is (4.13).

6 Supplement

6.1 Spectral representations of continuous-time stationary processes

This appendix provides an exposition of the spectral representation for continuous-time stationary, centered and square integrable processes with a continuous autocovariance function. The proofs are found in Appendix 4 For an extensive treatment we refer to [15, Section 9.4] and [19, Appendix A2.1].

Recall that if $S = \{S(t) : t \in \mathbb{R}\}$ is a (complex-valued) process such that

(i) $\mathbb{E}[|S(t)|^2] < \infty$ for all $t \in \mathbb{R}$,

(ii) $\mathbb{E}[|S(t+s) - S(t)|^2] \to 0$ as $s \downarrow 0$ for all $t \in \mathbb{R}$, and

(iii) $\mathbb{E}[(S(v) - S(u))\overline{(S(t) - S(s))}] = 0$ for all $u \le v \le s \le t$,

we may (and do) define integration of f with respect to S in the sense of [15, pp 388-390] for any $f \in L^2(G)$, where G is the control measure characterized by

$$G((s,t]) = \mathbb{E}\Big[|S(t) - S(s)|^2\Big]$$

for $s \le t$. We have the following stochastic Fubini result for this type of integral:

Proposition 6.1. *Let $S = \{S(t) : t \in \mathbb{R}\}$ be a process given as above. Let μ be a finite Borel measure on $\mathbb{R}$, and let $f : \mathbb{R}^2 \to \mathbb{C}$ be a measurable function in $L^2(\mu \times G)$. Then all the integrals below are well-defined and*

$$\int_{\mathbb{R}} \bigg(\int_{\mathbb{R}} f(x,y)\, \mu(dx) \bigg) S(dy) = \int_{\mathbb{R}} \bigg(\int_{\mathbb{R}} f(x,y)\, S(dy) \bigg) \mu(dx) \tag{6.1}$$

almost surely.

Suppose that $(X_t)_{t\in\mathbb{R}}$ is a stationary process with $\mathbb{E}[X_0^2]<\infty$ and $\mathbb{E}[X_0]=0$, and denote by γ_X its autocovariance function. Assuming that γ_X is continuous, it follows by Bochner's theorem that there exists a finite Borel measure F_X on $\mathbb{R}$ having γ_X as its Fourier transform, that is,

$$\gamma_X(t)=\int_{\mathbb{R}} e^{ity}F_X(dy),\quad t\in\mathbb{R}.$$

The measure F_X is referred to as the spectral distribution of $(X_t)_{t\in\mathbb{R}}$.

Theorem 6.2. *Let $(X_t)_{t\in\mathbb{R}}$ be given as above and let F_X be the associated spectral distribution. Then there exists a (complex-valued) process $\Lambda_X=\{\Lambda_X(y):y\in\mathbb{R}\}$ satisfying (i)-(iii) above with control measure F_X, such that*

$$X_t=\int_{\mathbb{R}} e^{ity}\,\Lambda_X(dy) \tag{6.2}$$

almost surely for each $t\in\mathbb{R}$. The process Λ_X is called the spectral process of $(X_t)_{t\in\mathbb{R}}$ and (6.2) is referred to as its spectral representation.

Remark 6.3. Let the situation be as in Theorem 6.2 and note that if there exists another process $\tilde{\Lambda}_X=\{\tilde{\Lambda}_X(y):y\in\mathbb{R}\}$ such that

$$X_t=\int_{\mathbb{R}} e^{ity}\,\tilde{\Lambda}_X(dy)$$

for all $t\in\mathbb{R}$, then its control measure is necessarily given by F_X and

$$\int_{\mathbb{R}} f(y)\,\Lambda_X(dy)=\int_{\mathbb{R}} f(y)\,\tilde{\Lambda}_X(dy)$$

almost surely for all $f\in L^2(F_X)$.

6.2 Proofs

Proof (Proof of Proposition 3.1). For a given $\gamma>0$ define $h_\gamma(z)=(-z)^\gamma/h(z)$ for $z\in\mathbb{C}\setminus\{0\}$ with $\mathrm{Re}(z)\le 0$. By continuity of h and the asymptotics $|h_\gamma(z)|\sim|\eta([0,\infty))|^{-1}|z|^\gamma$, $|z|\to 0$, and $|h_\gamma(z)|\sim|z|^{\gamma-1}$, $|z|\to\infty$, it follows that

$$\sup_{x<0}\int_{\mathbb{R}} |h_\gamma(x+iy)|^2\,dy<\infty \tag{6.3}$$

for $\gamma\in(-1/2,1/2)$. In other words, h_γ is a certain Hardy function, and thus there exists a function $f_\gamma:\mathbb{R}\to\mathbb{R}$ in L^2 which is vanishing on $(-\infty,0)$ and has $\mathcal{L}[f_\gamma](z)=h_\gamma(z)$ when $\mathrm{Re}(z)<0$, see [3, 11, 13]. Note that f_γ is indeed real-valued, since $\overline{h_\gamma(x-iy)}=h_\gamma(x+iy)$ for $y\in\mathbb{R}$ and a fixed $x<0$. We can apply [23, Proposition 2.3] to deduce that there exists a function $g\in L^2$ satisfying (3.6) and that it can be represented as the (left-sided) Riemann-Liouville fractional integral of f_0, that is,

$$g(t)=\frac{1}{\Gamma(\beta)}\int_0^t f_0(u)(t-u)^{\beta-1}\,du$$

for $t > 0$. Conversely, [23, Theorem 2.1] ensures that $D^\beta g$ given by (3.7) is a well-defined limit and that $D^\beta g = f_0$. In particular, we have shown ((ii)) and if we can argue that $f_0 \in L^1$, we have shown ((i)) as well. This follows from the assumption in (3.1), since then we have that $y \mapsto \mathcal{L}[f_0](x+iy)$ is differentiable for any $x \leq 0$ (except at 0 when $x = 0$) and

$$\begin{aligned}\mathcal{L}[u \mapsto u f_0(u)](x+iy) &= -i\frac{d}{dy}\mathcal{L}[f_0](x+iy) \\ &= \frac{\mathcal{L}[u\,\eta(du)](x+iy) - (1-\beta)(x+iy)^{-\beta}}{h(x+iy)^2}.\end{aligned} \tag{6.4}$$

The function $\mathcal{L}[u \mapsto u f_0(u)]$ is analytic on $\{z \in \mathbb{C} : \operatorname{Re}(z) < 0\}$ and from the identity (6.4) it is not too difficult to see that it also satisfies the Hardy condition (6.3). This means $u \mapsto u f_0(u)$ belongs to L^2, and hence we have that f_0 belongs to L^1. Since g is the Riemann-Liouville integral of f_0 of order β and $f_0 \in L^1 \cap L^2$, [2, Proposition 4.3] implies that $g \in L^\gamma$ for $(1-\beta)^{-1} < \gamma \leq 2$.

It is straightforward to verify (3.9) and to obtain the identity

$$\int_s^t \left(D^\beta g\right) * \eta(u-\cdot)\,du = \int_{\mathbb{R}} \left(D_-^\beta \mathbb{1}_{(s,t]}\right)(u)\, g * \eta(u-\cdot)\,du$$

almost everywhere by comparing their Fourier transforms. This establishes the relation

$$g(t-v) - g(s-v) = \int_s^t \left(D^\beta g\right) * \eta(u-v)\,du + \mathbb{1}_{(s,t]}(v)$$

By letting $s \to -\infty$, and using that $D^\beta g$ and g are both vanishing on $(-\infty, 0)$, we deduce that

$$g(t) = \mathbb{1}_{[0,\infty)}(t)\left(1 + \int_0^t (D^\beta g) * \eta(u)\,du\right),$$

for almost all $t \in \mathbb{R}$ which shows (3.8) and, thus, finishes the proof. □

Proof (Proof of Theorem 3.2). Since $g \in L^2$, according to Proposition 3.1, and $\mathbb{E}[L_1^2] < \infty$ and $\mathbb{E}[L_1] = 0$,

$$X_t = \int_{-\infty}^t g(t-u)\,dL_u, \quad t \in \mathbb{R},$$

is a well-defined process (e.g., in the sense of [25]) which is stationary with mean zero and finite second moments. By integrating both sides of (3.9) with respect to $(L_t)_{t\in\mathbb{R}}$ we obtain

$$X_t - X_s = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} \left(D_-^\beta \mathbb{1}_{(s,t]}\right)(u)\, g * \eta(u-r)\,du \right) dL_r + L_t - L_s.$$

By a stochastic Fubini result (such as [1, Theorem 3.1]) we can change the order of integration (twice) and obtain

$$\int_{\mathbb{R}} \left(\int_{\mathbb{R}} \left(D_-^\beta \mathbb{1}_{(s,t]}\right)(u)\, g * \eta(u-r)\,du \right) dL_r = \int_{\mathbb{R}} \left(D_-^\beta \mathbb{1}_{(s,t]}\right)(u)\, X * \eta(u)\,du.$$

This shows that $(X_t)_{t\in\mathbb{R}}$ is a solution to (3.2). To show uniqueness, note that the spectral process Λ_X of any purely non-deterministic solution $(X_t)_{t\in\mathbb{R}}$ satisfies

$$\int_{\mathbb{R}} \mathcal{F}[\mathbb{1}_{(s,t]}](y)(iy)^{\beta} h(-iy)\,\Lambda_X(dy) = L_t - L_s \tag{6.5}$$

almost surely for any choice of $s < t$ by Theorem 6.2 and Proposition 6.1. Using the fact that $(X_t)_{t\in\mathbb{R}}$ is purely non-deterministic, F_X is absolutely continuous with respect to the Lebesgue measure, and hence we can extend (6.5) from $\mathbb{1}_{(s,t]}$ to any function $f \in L^2$ using an approximation of f with simple functions of the form $s = \sum_{j=1}^{n} \alpha_j \mathbb{1}_{(t_{j-1},t_j]}$ for $\alpha_j \in \mathbb{C}$ and $t_0 < t_1 < \cdots < t_n$. Specifically, we establish that

$$\int_{\mathbb{R}} \mathcal{F}[f](y)(iy)^{\beta} h(-iy)\,\Lambda_X(dy) = \int_{\mathbb{R}} f(u)\,dL_u \tag{6.6}$$

almost surely for any $f \in L^2$. In particular we may take $f = g(t-\cdot)$, g being the solution kernel characterized in (3.6), so that $\mathcal{F}[g(t-\cdot)](y) = e^{ity}(iy)^{-\beta}/h(-iy)$ and (6.6) thus implies that

$$X_t = \int_{-\infty}^{t} g(t-u)\,dL_u, \qquad \square$$

which ends the proof.

Proof (Proof of Proposition 3.6). We start by arguing that the limit in (3.13) exists and is equal to $\int_{-\infty}^{t} D^{\beta} g(t-u)\,dL_u$. For a given $\delta > 0$ it follows by a stochastic Fubini result that

$$\frac{\beta}{\Gamma(1-\beta)} \int_{\delta}^{\infty} \frac{X_t - X_{t-u}}{u^{1+\beta}}\,du = \int_{\mathbb{R}} D_{\delta}^{\beta} g(t-r)\,dL_r, \tag{6.7}$$

where

$$D_{\delta}^{\beta} g(t) = \frac{\beta}{\Gamma(1-\beta)} \int_{\delta}^{\infty} \frac{g(t) - g(t-u)}{u^{1+\beta}}\,du$$

for $t > 0$ and $D_{\delta}^{\beta} g(t) = 0$ for $t \leq 0$. Suppose for the moment that $(L_t)_{t\in\mathbb{R}}$ is a Brownian motion, so that $(X_t)_{t\in\mathbb{R}}$ is γ-Hölder continuous for all $\gamma \in (0,1/2)$ by (3.2). Then, almost surely, $u \mapsto (X_t - X_{t-u})/u^{1+\beta}$ is in L^1 and the relation (6.7) thus shows that

$$\int_{\mathbb{R}} \Big[D_{\delta}^{\beta} g(t-r) - D_{\delta'}^{\beta} g(t-r)\Big]\,dL_r \xrightarrow{\mathbb{P}} 0 \quad \text{as} \quad \delta, \delta' \to 0,$$

which in turn implies that $(D_{\delta}^{\beta} g)_{\delta>0}$ has a limit in L^2. We also know that this limit must be $D^{\beta} g$, since $D_{\delta}^{\beta} g \to D^{\beta} g$ pointwise as $\delta \downarrow 0$ by (3.7). Having established this convergence, which does not rely on $(L_t)_{t\in\mathbb{R}}$ being a Brownian motion, it follows immediately from (6.7) and the isometry property of the integral map $\int_{\mathbb{R}} \cdot\, dL$ that the limit in (3.13) exists and that $D^{\beta} X_t = \int_{-\infty}^{t} D^{\beta} g(t-u)\,dL_u$. To show (3.14) we start by recalling the definition of $D_{-}^{\beta} \mathbb{1}_{(s,t]}$ in (3.3) and that $\mathcal{F}[D_{-}^{\beta} \mathbb{1}_{(s,t]}](y) = (iy)^{\beta} \mathcal{F}[\mathbb{1}_{(s,t]}](y)$.

This identity can be shown by using that the improper integral $\int_0^\infty e^{\pm iv} v^{\gamma-1}\,dv$ is equal to $\Gamma(\gamma)e^{\pm i\pi\gamma/2}$ for any $\gamma \in (0,1)$. Now observe that

$$\begin{aligned}\mathcal{F}\Big[\int_{\mathbb{R}} \big(D_-^\beta \mathbb{1}_{(s,t]}\big)(u)\, g * \eta(u-\cdot)\,du\Big](y) &= (iy)^\beta \mathcal{F}[\mathbb{1}_{(s,t]}](y)\mathcal{F}[g](-y)\mathcal{F}[\eta](-y)\\ &= \mathcal{F}[\mathbb{1}_{(s,t]}](y)\mathcal{F}\Big[\big(D^\beta g\big)*\eta\Big](-y)\\ &= \mathcal{F}\Big[\int_s^t \big(D^\beta g\big)*\eta(u-\cdot)\,du\Big](y),\end{aligned}$$

and hence $\int_{\mathbb{R}} \big(D_-^\beta \mathbb{1}_{(s,t]}\big)(u)\, g * \eta(u-\cdot)\,du = \int_s^t \big(D^\beta g\big)*\eta(u-\cdot)\,du$ almost everywhere. Consequently, using that $D^\beta X_t = \int_{-\infty}^t D^\beta g(t-u)\,dL_u$ and applying a stochastic Fubini result twice,

$$\begin{aligned}\int_s^t \big(D^\beta X\big)*\eta(u)\,du &= \int_{\mathbb{R}}\int_s^t \big(D^\beta g\big)*\eta(u-r)\,du\,dL_r\\ &= \int_{\mathbb{R}}\int_{\mathbb{R}} \big(D_-^\beta \mathbb{1}_{(s,t]}\big)(u)\, g*\eta(u-r)\,du\,dL_r\\ &= \frac{1}{\Gamma(1-\beta)}\int_{\mathbb{R}} \Big[(t-u)_+^{-\beta} - (s-u)_+^{-\beta}\Big] X*\eta(u)\,du.\end{aligned}$$

The semimartingale property of $(X_t)_{t\in\mathbb{R}}$ is now an immediate consequence of (3.2). □

Proof (Proof of Proposition 3.7). Using (3.16) and that $h(0) = -\eta([0,\infty))$, it follows that $f_X(y) \sim |y|^{-2\beta}/\eta([0,\infty))^2$ as $y \to 0$. To show the asymptotic behavior of γ_X at ∞ we start by recalling that, for $u,v \in \mathbb{R}$,

$$\int_{u\vee v}^\infty (s-u)^{\beta-1}(s-v)^{\beta-1}\,ds = \frac{\Gamma(\beta)\Gamma(1-2\beta)}{\Gamma(1-\beta)}|u-v|^{2\beta-1}$$

by [16, p. 404]. Having this relation in mind we use Proposition 3.1((ii)) and (3.15) to do the computations

$$\begin{aligned}\gamma_X(t) &= \frac{1}{\Gamma(\beta)^2}\int_{\mathbb{R}}\int_{\mathbb{R}}\int_{\mathbb{R}} D^\beta g(u) D^\beta g(v)(s+t-u)_+^{\beta-1}(s-v)_+^{\beta-1}\,dv\,du\,ds\\ &= \frac{1}{\Gamma(\beta)^2}\int_{\mathbb{R}}\int_{\mathbb{R}} D^\beta g(u) D^\beta g(v)\int_{(u-t)\vee v}^\infty (s-(u-t))^{\beta-1}(s-v)^{\beta-1}\,ds\,dv\,du\\ &= \frac{\Gamma(1-2\beta)}{\Gamma(\beta)\Gamma(1-\beta)}\int_{\mathbb{R}}\int_{\mathbb{R}} D^\beta g(u) D^\beta g(v)|u-v-t|^{2\beta-1}\,dv\,du\\ &= \frac{\Gamma(1-2\beta)}{\Gamma(\beta)\Gamma(1-\beta)}\int_{\mathbb{R}} \gamma(u)|u-t|^{2\beta-1}\,du,\end{aligned} \tag{6.8}$$

where $\gamma(u) := \int_{\mathbb{R}} D^\beta g(u+v) D^\beta g(v)\,dv$. Note that $\gamma \in L^1$ since $D^\beta g \in L^1$ by Proposition 3.1 and, using Plancherel's theorem,

$$\gamma(u) = \int_{\mathbb{R}} e^{iuy}\Big|\mathcal{F}\Big[D^\beta g\Big](y)\Big|^2\,dy = \mathcal{F}\Big[|h(i\cdot)|^{-2}\Big](u).$$

In particular $\int_{\mathbb{R}} \gamma(u)\,du = |h(0)|^{-2} = \eta([0,\infty))^{-2}$, and hence it follows from (6.8) that we have shown the result if we can argue that

$$\frac{\int_{\mathbb{R}}\gamma(u)|u-t|^{2\beta-1}\,du}{t^{2\beta-1}} = \int_{\mathbb{R}} \frac{\gamma(u)}{|\frac{u}{t}-1|^{1-2\beta}}\,du \to \int_{\mathbb{R}}\gamma(u)\,du, \quad t\to\infty. \tag{6.9}$$

It is clear by Lebesgue's theorem on dominated convergence that

$$\int_{-\infty}^{0} \frac{\gamma(u)}{|\frac{u}{t}-1|^{1-2\beta}}\,du \to \int_{-\infty}^{0} \gamma(u)\,du, \quad t \to \infty.$$

Moreover, since $|h(i\cdot)|^{-2}$ is continuous at 0 and differentiable on $(-\infty,0)$ and $(0,\infty)$ with integrable derivatives, it is absolutely continuous on $\mathbb{R}$ with a density ϕ in L^1. As a consequence, $\gamma(u) = \mathcal{F}[\phi](y)/(-iy)$ and, thus,

$$\int_{t/2}^{\infty} \frac{\gamma(u)}{|\frac{u}{t}-1|^{1-2\beta}}\,du = \int_{1/2}^{\infty} \frac{t\gamma(tu)}{|u-1|^{1-2\beta}}\,du = i\int_{1/2}^{\infty} \frac{\mathcal{F}[\phi](tu)}{u|u-1|^{1-2\beta}}\,du. \tag{6.10}$$

By the Riemann-Lebesgue lemma and Lebesgue's theorem on dominated convergence it follows that the right-hand side of expression in (6.10) tends to zero as t tends to infinity. Finally, integration by parts and the symmetry of γ yields

$$\begin{aligned}
&\int_0^{t/2} \gamma(u)\left(1-\frac{1}{|\frac{u}{t}-1|^{1-2\beta}}\right)du \\
&\quad= \int_0^{1/2} t\gamma(tu)\left(1-\frac{1}{(1-u)^{1-2\beta}}\right)du \\
&\quad= \left(2^{1-2\beta}-1\right)\int_{-\infty}^{-t/2}\gamma(u)\,du - (1-2\beta)\int_0^{1/2}\frac{1}{(1-u)^{2-2\beta}}\int_{-\infty}^{-tu}\gamma(v)\,dv\,du,
\end{aligned}$$

where both terms on the right-hand side converge to zero as t tends to infinity. Thus, we have shown (6.9), and this completes the proof. □

Proof (Proof of Proposition 3.8). Observe that it is sufficient to argue $\mathbb{E}[(X_t - X_0)^2] \sim t$ as $t \downarrow 0$. By using the spectral representation $X_t = \int_{\mathbb{R}} e^{ity}\,\Lambda_X(dy)$ and the isometry property of the integral map $\int_{\mathbb{R}} \cdot\, d\Lambda_X : L^2(F_X) \to L^2(\mathbb{P})$, see [15, p. 389], we have that

$$\begin{aligned}
\frac{\mathbb{E}\left[(X_t - X_0)^2\right]}{t} &= t^{-2}\int_{\mathbb{R}} \left|1-e^{iy}\right|^2 f_X(y/t)\,dy \\
&= \int_{\mathbb{R}} \frac{\left|1-e^{iy}\right|^2}{|y|^{2\beta}\left|(-iy)^{1-\beta} - t^{1-\beta}\mathcal{F}[\eta](y/t)\right|^2}\,dy.
\end{aligned} \tag{6.11}$$

Consider now a $y \in \mathbb{R}$ satisfying $|y| \geq C_1 t$ with $C_1 := (2|\eta|([0,\infty)))^{1/(1-\beta)}$. In this case $|y|^{1-\beta}/2 - |t^{1-\beta}\mathcal{F}[\eta](y/t)| \geq 0$, and we thus get by the reversed triangle inequality that

$$\frac{\left|1-e^{iy}\right|^2}{|y|^{2\beta}\left|(-iy)^{1-\beta} - t^{1-\beta}\mathcal{F}[\eta](y/t)\right|^2} \leq 2\frac{\left|1-e^{iy}\right|^2}{y^2}.$$

If $|y| < C_1 t$, we note that the assumption on the function in (3.5) implies that

$$C_2 := \inf_{|x|\leq C_1} \left|(-ix)^{1-\beta} - \mathcal{F}[\eta](x)\right| > 0,$$

which shows that

$$\left|(-iy)^{1-\beta} - t^{1-\beta}\mathcal{F}[\eta](y/t)\right| \geq t^{1-\beta}C_2 \geq \frac{C_2}{C_1^{1-\beta}}|y|^{1-\beta}.$$

This establishes that

$$\frac{\left|1-e^{iy}\right|^2}{|y|^{2\beta}\left|(-iy)^{1-\beta}-t^{1-\beta}\mathcal{F}[\eta](y/t)\right|^2} \le \frac{C_1^{2(1-\beta)}}{C_2^2}\frac{\left|1-e^{iy}\right|^2}{y^2}.$$

Consequently, it follows from (6.11) and Lebesgue's theorem on dominated convergence that

$$\frac{\mathbb{E}\left[(X_t-X_0)^2\right]}{t} \to \int_{\mathbb{R}} \frac{\left|1-e^{iy}\right|^2}{y^2}\,dy = \int_{\mathbb{R}} |\mathcal{F}[\mathbb{1}_{(0,1]}](y)|^2\,dy = 1$$

as $h \downarrow 0$, which was to be shown. □

Proof (Proof of Proposition 3.11). We start by arguing that the first term on the right-hand side of the formula is well-defined. In order to do so it suffices to argue that

$$\begin{aligned}\mathbb{E}\Big[\int_0^{t-s}\int_{-\infty}^s |X_w| \int_{[0,\infty)} \left|\left(D_-^\beta \mathbb{1}_{(s,t-u]}\right)(v+w)\right| |\eta|(dv)\,dw\,|g|(du)\Big]& \\ \le \mathbb{E}[|X_0|]\int_0^{t-s}\int_{[0,\infty)}\int_{-\infty}^s \left|\left(D_-^\beta \mathbb{1}_{(s,t-u]}\right)(v+w)\right| dw\,|\eta|(dv)\,|g|(du)&\end{aligned} \tag{6.12}$$

is finite. This is implied by the facts that

$$\begin{aligned}\Gamma(1-\beta)&\int_{-\infty}^s \left|\left(D_-^\beta \mathbb{1}_{(s,t-u]}\right)(v+w)\right| dw \\ &\le \int_{u+s-t}^0 (t-s-u+w)^{-\beta}\,dw + \int_0^1 w^{-\beta}-(t-s-u+w)^{-\beta}\,dw \\ &\quad + (1+\beta)\int_1^\infty w^{-1-\beta}(t-s-u)\,dw \\ &= \frac{1}{1-\beta}\left(2(t-s-u)^{1-\beta}+1-(t-s-u+1)^{1-\beta}\right)+\frac{(1+\beta)}{\beta}(t-s-u) \\ &\le \frac{2}{1-\beta}(t-s)^{1-\beta}+\frac{(1+\beta)}{\beta}(t-s)\end{aligned}$$

for $u \in [0,t-s]$ and $g(du)$ is a finite measure (since $D^\beta g \in L^1$ by Proposition 3.1). Now fix an arbitrary $z \in \mathbb{C}$ with $\mathrm{Re}(z) < 0$. It follows from (3.2) that

$$\begin{aligned}\mathcal{L}[X\mathbb{1}_{(s,\infty)}](z) =& X_s\mathcal{L}[\mathbb{1}_{(s,\infty}](z)+\mathcal{L}[\mathbb{1}_{(s,\infty)}(L_\cdot - L_s)](z) \\ &+\mathcal{L}\Big[\mathbb{1}_{(s,\infty)}\int_{\mathbb{R}} X_u \int_{[0,\infty)}\left(D_-^\beta\mathbb{1}_{(s,\cdot]}\right)(u+v)\,\eta(dv)\,du\Big](z).\end{aligned} \tag{6.13}$$

By noting that $(D_-^\beta \mathbb{1}_{(s,t]})(u) = 0$ when $t \le s < u$ we obtain

$$\begin{aligned}\mathcal{L}\Big[\mathbb{1}_{(s,\infty)}&\int_s^\infty X_u\int_{[0,\infty)}\left(D_-^\beta\mathbb{1}_{(s,\cdot]}\right)(u+v)\,\eta(dv)\,du\Big](z) \\ &= \frac{1}{\Gamma(1-\beta)}\mathcal{L}\Big[\int_s^\infty X_u\int_{[0,\infty)}(\cdot-u-v)_+^{-\beta}\,\eta(dv)\,du\Big](z) \\ &= \mathcal{L}[\mathbb{1}_{(s,\infty)}X](z)\mathcal{L}[\eta](z)(-z)^{\beta-1}.\end{aligned}$$

Combining this observation with (6.13) we get the relation

$$\begin{aligned}&\Big(-z-(-z)^{\beta}\mathcal{L}[\eta](z)\Big)\mathcal{L}[\mathbb{1}_{(s,\infty)}X](z)\\&\quad=X_s(-z)\mathcal{L}[\mathbb{1}_{(s,\infty)}](z)+(-z)\mathcal{L}[\mathbb{1}_{(s,\infty)}(L-L_s)](z)\\&\qquad+(-z)\mathcal{L}\bigg[\mathbb{1}_{(s,\infty)}\int_{-\infty}^{s}X_u\int_{[0,\infty)}\Big(D_-^{\beta}\mathbb{1}_{(s,\cdot]}\Big)(u+v)\,\eta(dv)\,du\bigg](z),\end{aligned}$$

which implies

$$\begin{aligned}&\mathcal{L}[\mathbb{1}_{(s,\infty)}X](z)\\&\quad=\mathcal{L}[g](z)\mathcal{L}[X_s\delta_0(s-\cdot)](z)+\mathcal{L}[g](z)(-z)\mathcal{L}[\mathbb{1}_{(s,\infty)}(L-L_s)](z)\\&\qquad+\mathcal{L}[g](z)(-z)\mathcal{L}\bigg[\mathbb{1}_{(s,\infty)}\int_{-\infty}^{s}X_u\int_{[0,\infty)}\Big(D_-^{\beta}\mathbb{1}_{(s,\cdot]}\Big)(u+v)\,\eta(dv)\,du\bigg](z)\\&\quad=\mathcal{L}[g(\cdot-s)X_s](z)+\mathcal{L}\bigg[\int_s^{\cdot}g(\cdot-u)\,dL_u\bigg](z)\\&\qquad+\mathcal{L}\bigg[\int_0^{\cdot-s}\int_{-\infty}^{s}X_w\int_{[0,\infty)}\Big(D_-^{\beta}\mathbb{1}_{(s,\cdot-u]}\Big)(v+w)\,\eta(dv)\,dw\,g(du)\bigg](z).\end{aligned}$$

This establishes the identity

$$\begin{aligned}X_t=&\,g(t-s)X_s+\int_s^t g(t-u)\,dL_u\\&+\int_0^{t-s}\int_{-\infty}^{s}X_w\int_{[0,\infty)}\Big(D_-^{\beta}\mathbb{1}_{(s,t-u]}\Big)(v+w)\,\eta(dv)\,dw\,g(du)\end{aligned}\tag{6.14}$$

almost surely for Lebesgue almost all $t>s$. Since both sides of (6.14) are continuous in $L^1(\mathbb{P})$, the identity holds for each fixed pair $s<t$ almost surely as well. By applying the conditional mean $\mathbb{E}[\cdot\mid X_u,\,u\le s]$ on both sides of (6.14) we obtain the result. □

Proof (Proof of Corollary 4.1). In this setup it follows that the function h in (3.5) is given by

$$h(z)=(-z)^{1-\beta}+\kappa+\frac{R(-z)}{Q(-z)},$$

where $Q(z)\neq 0$ whenever $\mathrm{Re}(z)\geq 0$ by the assumption on A. This shows that h is non-zero (on $\{z\in\mathbb{C}:\mathrm{Re}(z)\leq 0\}$) if and only if

$$Q(z)\Big[z^{1-\beta}+\kappa\Big]+R(z)\neq 0\quad\text{for all } z\in\mathbb{C}\text{ with }\mathrm{Re}(z)\geq 0.\tag{6.15}$$

Condition (6.15) may equivalently be formulated as $Q(z)[z+\kappa z^{\beta}]+R(z)z^{\beta}\neq 0$ for all $z\in\mathbb{C}\setminus\{0\}$ with $\mathrm{Re}(z)\geq 0$ and $h(0)=\kappa+b^TA^{-1}e_1\neq 0$, which by Theorem 3.2 shows that a unique solution to (4.5) exists. It also provides the form of the solution, namely (3.11) with

$$\mathcal{F}[g](y)=\frac{(-iy)^{-\beta}}{(-iy)^{1-\beta}+\kappa+\frac{R(-iy)}{Q(-iy)}}=\frac{Q(-iy)}{Q(-iy)\Big[-iy+\kappa(-iy)^{\beta}\Big]+R(-iy)(-iy)^{\beta}}$$

for $y\in\mathbb{R}$. This finishes the proof. □

Proof (Proof of Proposition 4.2). We will first show that $D^\beta f \in L^1$. By using that $\int_0^\infty e^{Au}\,du = -A^{-1}$ we can rewrite $D^\beta f$ as

$$D^\beta f(t) = \frac{1}{\Gamma(1-\beta)} b^T A\biggl(\int_0^t e^{Au}\Bigl[(t-u)^{-\beta} - t^{-\beta}\Bigr]\,du - \int_t^\infty e^{Au} t^{-\beta}\,du\biggr)e_1$$

for $t > 0$, from which we see that it suffices to argue that (each entry of)

$$t \mapsto \int_0^t e^{Au}\Bigl[(t-u)^{-\beta} - t^{-\beta}\Bigr]\,du$$

belongs to L^1. Since $u \mapsto e^{Au}$ is continuous and with all entries decaying exponentially fast as $u \to \infty$, this follows from the fact that, for a given $\gamma > 0$,

$$\begin{aligned}\int_0^\infty \int_0^t & e^{-\gamma u}\Bigl|(t-u)^{-\beta} - t^{-\beta}\Bigr|\,du\,dt \\ &\leq \int_0^\infty e^{-\gamma u}\biggl(\int_u^{u+1}\Bigl((t-u)^{-\beta} + t^{-\beta}\Bigr)dt + \beta u \int_1^\infty t^{-\beta-1}\,dt\biggr)du < \infty.\end{aligned}$$

Here we have used the mean value theorem to establish the inequality

$$\Bigl|(t-u)^{-\beta} - t^{-\beta}\Bigr| \leq \beta u (t-u)^{-\beta-1}$$

for $0 < u < t$. To show that $D^\beta f \in L^2$, note that it is the left-sided Riemann-Liouville fractional derivative of f, that is,

$$D^\beta f(t) = \frac{1}{\Gamma(1-\beta)} \frac{d}{dt} \int_0^t f(t-u) u^{-\beta}\,du$$

for $t > 0$. Consequently, it follows by [26, Theorem 7.1] that the Fourier transform $\mathcal{F}[D^\beta f]$ of f is given by

$$\mathcal{F}\Bigl[D^\beta f\Bigr](y) = (-iy)^\beta \mathcal{F}[f](y) = -(-iy)^\beta b^T (A+iy)^{-1} e_1, \quad y \in \mathbb{R},$$

in particular it belongs to L^2 (e.g., by Cramer's rule), and thus $D^\beta f \in L^2$. By comparing Fourier transforms we establish that $(D^\beta g) * f = g * (D^\beta f)$, and hence it holds that

$$\int_0^\infty D^\beta X_{t-u} f(u)\,du = \int_{\mathbb{R}} \Bigl(D^\beta g\Bigr) * f(t-r)\,dL_r = \int_0^\infty X_{t-u} D^\beta f(u)\,du$$

using Proposition 3.6 and a stochastic Fubini result. This finishes the proof. □

Proof (Proof of Proposition 6.1). First, note that (6.1) is trivially true when f is of the form

$$f(x,y) = \sum_{j=1}^n \alpha_j \mathbb{1}_{A_j}(x)\mathbb{1}_{B_j}(y) \tag{6.16}$$

for $\alpha_1,\dots,\alpha_n \in \mathbb{C}$ and Borel sets $A_1, B_1, \dots, A_n, B_n \subseteq \mathbb{R}$. Now consider a general $f \in L^2(\mu \times G)$ and choose a sequence of functions $(f_n)_{n\in\mathbb{N}}$ of the form (6.16) such that $f_n \to f$ in $L^2(\mu\times G)$ as $n \to \infty$. Set

$$X_n = \int_{\mathbb{R}}\biggl(\int_{\mathbb{R}} f_n(x,y)\,\mu(dx)\biggr)S(dy), \quad X = \int_{\mathbb{R}}\biggl(\int_{\mathbb{R}} f(x,y)\,\mu(dx)\biggr)S(dy)$$

$$\text{and} \quad Y = \int_{\mathbb{R}}\biggl(\int_{\mathbb{R}} f(x,y)\,S(dy)\biggr)\mu(dx).$$

Observe that X and Y are indeed well-defined, since $x \mapsto f(x,y)$ is in $L^1(\mu)$ for G-almost all y, $y \mapsto f(x,y)$ is in $L^2(G)$ for μ-almost all x,

$$\int_{\mathbb{R}} \left| \int_{\mathbb{R}} f(x,y)\,\mu(dx) \right|^2 G(dy) \leq \mu(\mathbb{R}) \int_{\mathbb{R}^2} |f(x,y)|^2 \,(\mu \times G)(dx,dy) < \infty$$

$$\text{and} \quad \mathbb{E}\left[\int_{\mathbb{R}} \left| \int_{\mathbb{R}} f(x,y)\,S(dy) \right|^2 \mu(dx) \right] = \int_{\mathbb{R}^2} |f(x,y)|^2 \,(\mu \times G)(dx,dy) < \infty.$$

Next, we find that

$$\begin{aligned} \mathbb{E}[|X - X_n|^2] &= \int_{\mathbb{R}} \left| \int_{\mathbb{R}} (f(x,y) - f_n(x,y)\,\mu(dx) \right|^2 G(dy) \\ &\leq \mu(\mathbb{R}) \int_{\mathbb{R}^2} |f(x,y) - f_n(x,y)|^2 \,(\mu \times G)(dx,dy) \end{aligned}$$

which tends to zero by the choice of $(f_n)_{n\in\mathbb{N}}$. Similarly, using that $X_n = \int_{\mathbb{R}} \big(\int_{\mathbb{R}} f_n(x,y)\,S(dy) \big)\,\mu(dx)$, one shows that $X_n \to Y$ in $L^2(\mathbb{P})$, and hence we conclude that $X = Y$ almost surely. □

Proof (Proof of Theorem 6.2). For any given $t \in \mathbb{R}$ set $f_t(y) = e^{ity}$, $y \in \mathbb{R}$, and let H_F and H_X be the set of all (complex) linear combinations of $\{f_t : t \in \mathbb{R}\}$ and $\{X_t : t \in \mathbb{R}\}$, respectively. By equipping H_F and H_X with the usual inner products on $L^2(F_X)$ and $L^2(\mathbb{P})$, their closures $\overline{H_F}$ and $\overline{H_X}$ are Hilbert spaces. Due to the fact that

$$\langle X_s, X_t \rangle_{L^2(\mathbb{P})} = \mathbb{E}[X_s X_t] = \int_{\mathbb{R}} e^{i(t-s)x}\, F_X(dy) = \langle f_s, f_t \rangle_{L^2(F_X)}, \quad s,t \in \mathbb{R},$$

we can define a linear isometric isomorphism $\mu : \overline{H_F} \to \overline{H_X}$ as the one satisfying

$$\mu\bigg(\sum_{j=1}^n \alpha_j f_{t_j} \bigg) = \sum_{j=1}^n \alpha_j X_{t_j}$$

for any given $n \in \mathbb{N}$, $\alpha_1, \dots, \alpha_n \in \mathbb{C}$ and $t_1 < \cdots < t_n$. Since $\mathbb{1}_{(-\infty,y]} \in \overline{H_F}$ for each $y \in \mathbb{R}$, cf. [31, p. 150], we can associate a (complex-valued) process $\Lambda_X = \{\Lambda_X(y) : y \in \mathbb{R}\}$ to $(X_t)_{t\in\mathbb{R}}$ through the relation

$$\Lambda_X(y) = \mu(\mathbb{1}_{(-\infty,y]}).$$

It is straight-forward to check from the isometry property that Λ_X is right-continuous in $L^2(\mathbb{P})$, has orthogonal increments and satisfies

$$\mathbb{E}\Big[|\Lambda_X(y_2) - \Lambda_X(y_1)|^2 \Big] = F_X((y_1, y_2])$$

for $y_1 < y_2$. Consequently, integration with respect to Λ_X of any function $f \in L^2(F_X)$ can be defined in the sense of [15, pp 388-390]. For any $n \in \mathbb{N}$, $\alpha_1, \dots, \alpha_n \in \mathbb{C}$ and $t_0 < t_1 < \cdots < t_n$, we have

$$\int_{\mathbb{R}} \bigg(\sum_{j=1}^n \alpha_j \mathbb{1}_{(t_{j-1},t_j]}(y) \bigg) \Lambda_X(dy) = \sum_{j=1}^n \alpha_j \mu(\mathbb{1}_{(t_{j-1},t_j]}) = \mu\bigg(\sum_{j=1}^n \alpha_j \mathbb{1}_{(t_{j-1},t_j]} \bigg).$$

Since $f \mapsto \int_{\mathbb{R}} f(y)\,\Lambda_X(dy)$ is a continuous map (from $L^2(F_X)$ into $L^2(\mathbb{P})$), it follows by approximation with simple functions and from the relation above that

$$\int_{\mathbb{R}} f(y)\,\Lambda_X(dy) = \mu(f)$$

almost surely for any $f \in \overline{H_F}$. In particular, it shows that

$$X_t = \mu(f_t) = \int_{\mathbb{R}} e^{ity}\,\Lambda_X(dy), \quad t \in \mathbb{R},$$

which is the spectral representation of $(X_t)_{t\in\mathbb{R}}$. □

Acknowledgments

The authors thank Andreas Basse-O'Connor and Jan Pedersen for helpful comments. The research of Mikkel Slot Nielsen and Victor Rohde was supported by the Danish Council for Independent Research grant DFF-4002-00003. The research of Richard Davis was supported in part by ARO MURI grant W911NF-12-1-0385.

References

[1] Barndorff-Nielsen, O.E. and A. Basse-O'Connor (2011). Quasi Ornstein–Uhlenbeck processes. *Bernoulli* 17(3), 916–941. DOI: 10.3150/10-BEJ311.

[2] Basse-O'Connor, A., M.S. Nielsen, J. Pedersen and V. Rohde (2018). Multivariate stochastic delay differential equations and CAR representations of CARMA processes. *arXiv preprint arXiv:1801.04091*.

[3] Basse-O'Connor, A., M.S. Nielsen, J. Pedersen and V. Rohde (2019). Stochastic delay differential equations and related autoregressive models. *Stochastics.* Forthcoming. DOI: 10.1080/17442508.2019.1635601.

[4] Bennedsen, M. (2015). Rough electricity: a new fractal multi-factor model of electricity spot prices. *CREATES Research Paper* 42.

[5] Bennedsen, M., A. Lunde and M.S. Pakkanen (2016). Decoupling the short- and long-term behavior of stochastic volatility. arXiv: 1610.00332.

[6] Beran, J., Y. Feng, S. Ghosh and R. Kulik (2016). *Long-Memory Processes*. Springer.

[7] Bichteler, K. (1981). Stochastic integration and L^p-theory of semimartingales. *Ann. Probab.* 9(1), 49–89.

[8] Brockwell, P.J. and R.A. Davis (2006). *Time series: theory and methods*. Springer Series in Statistics. Reprint of the second (1991) edition. Springer, New York.

[9] Brockwell, P.J. and T. Marquardt (2005). Lévy-driven and fractionally integrated ARMA processes with continuous time parameter. *Statist. Sinica* 15(2), 477–494.

[10] Delbaen, F. and W. Schachermayer (1994). A general version of the fundamental theorem of asset pricing. *Math. Ann.* 300(3), 463–520.

[11] Doetsch, G. (1937). Bedingungen für die Darstellbarkeit einer Funktion als Laplace-integral und eine Umkehrformel für die Laplace-Transformation. *Math. Z.* 42(1), 263–286. DOI: 10.1007/BF01160078.

[12] Doukhan, P., G. Oppenheim and M.S. Taqqu, eds. (2003). *Theory and applications of long-range dependence*. Boston, MA: Birkhäuser Boston Inc.

[13] Dym, H. and H.P McKean (1976). *Gaussian processes, function theory, and the inverse spectral problem*. Probability and Mathematical Statistics, Vol. 31. New York: Academic Press [Harcourt Brace Jovanovich Publishers].

[14] Granger, C.W. and R. Joyeux (1980). An introduction to long-memory time series models and fractional differencing. *J. Time Series Anal.* 1(1), 15–29.

[15] Grimmett, G. and D. Stirzaker (2001). *Probability and random processes*. Oxford University Press.

[16] Gripenberg, G. and I. Norros (1996). On the prediction of fractional Brownian motion. *J. Appl. Probab.* 33(2), 400–410.

[17] Gushchin, A.A. and U. Küchler (2000). On stationary solutions of delay differential equations driven by a Lévy process. *Stochastic Process. Appl.* 88(2), 195–211. DOI: 10.1016/S0304-4149(99)00126-X.

[18] Hosking, J.R. (1981). Fractional differencing. *Biometrika* 68(1), 165–176.

[19] Koopmans, L.H. (1995). *The spectral analysis of time series*. Academic Press.

[20] Marquardt, T. (2006). Fractional Lévy processes with an application to long memory moving average processes. *Bernoulli* 12(6), 1099–1126.

[21] Mohammed, S.E.A. and M.K.R. Scheutzow (1990). Lyapunov exponents and stationary solutions for affine stochastic delay equations. *Stochastics Stochastics Rep.* 29(2), 259–283.

[22] Newbold, P. and C. Agiakloglou (1993). Bias in the sample autocorrelations of fractional noise. *Biometrika* 80(3), 698–702.

[23] Pipiras, V. and M.S. Taqqu (2003). Fractional calculus and its connections to fractional Brownian motion. *Theory and applications of long-range dependence*, 165–201.

[24] Pipiras, V. and M.S. Taqqu (2017). *Long-range dependence and self-similarity*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.

[25] Rajput, B.S. and J. Rosiński (1989). Spectral representations of infinitely divisible processes. *Probab. Theory Related Fields* 82(3), 451–487.

[26] Samko, S.G., A.A. Kilbas, O.I. Marichev, et al. (1993). Fractional integrals and derivatives. *Theory and Applications, Gordon and Breach, Yverdon* 1993.

[27] Samorodnitsky, G. (2016). *Stochastic processes and long range dependence*. Vol. 26. Springer.

[28] Samorodnitsky, G. et al. (2007). Long range dependence. *Foundations and Trends® in Stochastic Systems* 1(3), 163–257.

[29] Samorodnitsky, G. and M.S. Taqqu (1994). *Stable Non-Gaussian Random Processes*. Stochastic Modeling. Stochastic models with infinite variance. New York: Chapman & Hall.

[30] Sato, K. (1999). *Lévy Processes and Infinitely Divisible Distributions*. Vol. 68. Cambridge Studies in Advanced Mathematics. Translated from the 1990 Japanese original, Revised by the author. Cambridge University Press.

[31] Yaglom, A.M (1987). *Correlation theory of stationary and related random functions. Vol. I*. Springer Series in Statistics. Basic results. New York: Springer-Verlag.

Paper F

Recovering The Background Noise Of A Lévy-Driven CARMA Process Using An SDDE Approach

Mikkel Slot Nielsen and Victor Rohde

Abstract

Based on a vast amount of literature on continuous-time ARMA processes, the so-called CARMA processes, we exploit their relation to stochastic delay differential equations (SDDEs) and provide a simple and transparent way of estimating the background driving noise. An estimation technique for CARMA processes, which is particularly tailored for the SDDE specification, is given along with an alternative and (for the purpose) suitable state-space representation. Through a simulation study of the celebrated CARMA$(2,1)$ process we check the ability of the approach to recover the distribution.

Keywords: continuous-time ARMA process; Lévy processes; noise estimation; stochastic volatility

1 Introduction

Continuous-time ARMA processes, specifically the class of CARMA processes, have been studied extensively and found several applications. The most basic CARMA process is the CAR(1) process, which corresponds to the Ornstein-Uhlenbeck process. This process serves as the building block in stochastic modeling, e.g., [1] use it as the stochastic volatility component in option pricing modeling and [13] models (log) spot price of many different commodities through an Ornstein-Uhlenbeck specification. More recently, several researchers have paid attention to higher order CARMA processes. To give a few examples, [8] model turbulent wind speed data as a CAR(2) process, [11] and [3] fit a CARMA(2, 1) process to electricity spot prices, and [4] find a good fit of the CAR(3) to daily temperature observations (and thus, suggests a suitable model for the OTC market for temperature derivatives). In addition, as for the CAR(1) process, several studies have concerned the use of CARMA processes in the modeling of stochastic volatility (see, e.g., [7, 14, 16]).

From a statistical point of view, as noted in the above references, the ability to recover the underlying noise of the CARMA process is important. However, while it is possible to recover the driving noise process, it is a subtle task. Due to the non-trivial nature of the typical algorithm, see [7], implementation is not straightforward and approximation errors may be difficult to locate. The recent study of [2] on processes of ARMA structure relates CARMA processes to certain stochastic (delay) differential equations, and this leads to an alternative way of backing out the noise from the observed process which is transparent and easy to implement. The contribution of this paper is exploiting this result to get a simple way to recover the driving noise. The study both relies and supports the related work of [7].

Section 2 recalls a few central definitions and gives a dynamic interpretation of CARMA processes by relating them to solutions of stochastic differential equations. Section 3 briefly discusses how to do (consistent) estimation and inference in the dynamic model and, finally, in Section 4 we investigate through a simulation study the ability of the approach to recover the distribution of the underlying noise for two sample frequencies.

2 CARMA processes and their dynamic SDDE representation

Recall that a Lévy process is interpreted as the continuous-time analogue to the (discrete-time) random walk. More precisely, a (one-sided) Lévy process $(L_t)_{t\geq 0}$, $L_0 = 0$, is a stochastic process having stationary independent increments and cádlág sample paths. From these properties it follows that the distribution of L_1 is infinitely divisible, and the distribution of $(L_t)_{t\geq 0}$ is determined by the one of L_1 according to the relation

$$\mathbb{E}[e^{iyL_t}] = \mathbb{E}[e^{iyL_1}]^t$$

for $y \in \mathbb{R}$ and $t \geq 0$. The definition is extended to a two-sided Lévy process $(L_t)_{t\in\mathbb{R}}$, $L_0 = 0$, which can be constructed from a one-sided Lévy process $(L_t^1)_{t\geq 0}$ by taking an independent copy $(L_t^2)_{t\geq 0}$ and setting $L_t = L_t^1$ if $t \geq 0$ and $L_t = -L^2_{(-t)-}$ if $t < 0$. Throughout, $(L_t)_{t\in\mathbb{R}}$ denotes a two-sided Lévy process, which is assumed to be square integrable.

Next, we will give a brief recap of Lévy-driven CARMA processes. (For an extensive treatment, see [5, 7, 9].) Let $p \in \mathbb{N}$ and set

$$P(z) = z^p + a_1 z^{p-1} + \cdots + a_p \quad \text{and} \quad Q(z) = b_0 + b_1 z + \cdots + b_{p-1} z^{p-1} \tag{2.1}$$

for $z \in \mathbb{C}$ and $a_1, \ldots, a_p, b_0, \ldots, b_{p-1} \in \mathbb{R}$. Define

$$\tilde{A}_p = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ -a_p & -a_{p-1} & -a_{p-2} & \cdots & -a_1 \end{bmatrix},$$

$e_p = \begin{bmatrix} 0 & 0 & \cdots & 0 & 1 \end{bmatrix}' \in \mathbb{R}^p$, and $b = \begin{bmatrix} b_0 & b_1 & \cdots & b_{p-2} & b_{p-1} \end{bmatrix}'$. In order to ensure the existence of a casual CARMA process we will assume that the eigenvalues of $\tilde{A}_p$ or, equivalently, the zeroes of P all have negative real parts. Then there is a unique (strictly) stationary $\mathbb{R}^p$-valued process $(X_t)_{t\in\mathbb{R}}$ satisfying

$$dX_t = \tilde{A}_p X_t \, dt + e_p \, dL_t, \tag{2.2}$$

and it is explicitly given by $X_t = \int_{-\infty}^{t} e^{\tilde{A}_p(t-u)} e_p \, dL_u$ for $t \in \mathbb{R}$. For a given $q \in \mathbb{N}_0$ with $q < p$, we set $b_q = 1$ and $b_j = 0$ for $q < j < p$. A CARMA(p,q) process $(Y_t)_{t\in\mathbb{R}}$ is then the strictly stationary process defined by

$$Y_t = b' X_t \tag{2.3}$$

for $t \in \mathbb{R}$. This is the state-space representation of the formal stochastic differential equation

$$P(D)Y_t = Q(D)DL_t, \tag{2.4}$$

where D denotes differentiation with respect to time. One says that $(Y_t)_{t\in\mathbb{R}}$ is causal, since Y_t is independent of $(L_s - L_t)_{s>t}$ for all $t \in \mathbb{R}$. We will say that $(Y_t)_{t\in\mathbb{R}}$ is invertible if all the zeroes of Q have negative real parts. The word "invertible" is justified by Theorem 2.1 below and the fact that this is the assumption imposed in [7] in order to make the recovery of the increments of the Lévy process possible. In Figure 1 we have simulated a CARMA$(2,1)$ process driven by a gamma (Lévy) process and by a Brownian motion, respectively.

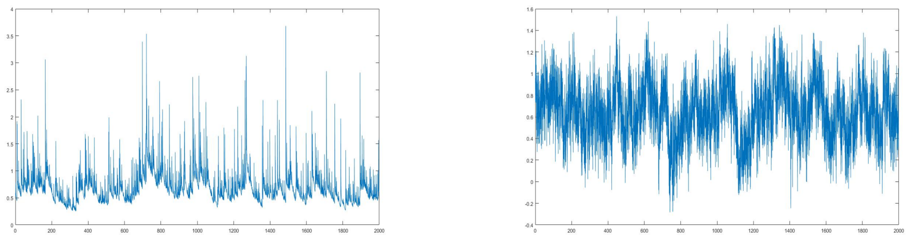


Figure 1: A simulation of a CARMA$(2,1)$ process with parameters $a_1 = 1.3619$, $a_2 = 0.0443$, and $b_0 = 0.2061$. It is driven by a gamma (Lévy) process with parameters $\lambda = 0.2488$ and $\xi = 0.5792$ on the left and a Brownian motion with mean $\mu = 0.1441$ and standard deviation $\sigma = 0.2889$ on the right.

For a given finite (signed) measure η concentrated on $[0, \infty)$ we will adopt a definition from [2] and say that an integrable measurable process $(Y_t)_{t\in\mathbb{R}}$ is a solution to the associated Lévy-driven stochastic delay differential equation (SDDE) if it is stationary and satisfies

$$dY_t = \int_{[0,\infty)} Y_{t-v}\,\eta(dv)\,dt + dL_t. \tag{2.5}$$

In the formulation of the next result we denote by δ_0 the Dirac measure at 0 and use the convention $\prod_\emptyset = 1$ and $\sum_\emptyset = 0$. Furthermore, we introduce the finite measure $\eta_\beta = \mathbb{1}_{[0,\infty)}(v)e^{\beta v}\,dv$ for $\beta \in \mathbb{C}$ with $\mathrm{Re}(\beta) < 0$, and let $\eta_0 = \delta_0$ and $\eta_j = \eta_{j-1} * \eta_{\beta_j}$ for $j = 1,\ldots,p-1$. By relying on [2, Theorem 3.12] we get the following dynamic SDDE representation of an invertible CARMA$(p,p-1)$ process:

Theorem 2.1. *Let $(Y_t)_{t\in\mathbb{R}}$ be an invertible CARMA$(p,p-1)$ process and let $\beta_1,\ldots,\beta_{p-1}$ be the roots of Q. Then $(Y_t)_{t\in\mathbb{R}}$ is the (up to modification) unique stationary solution to (2.5) with the real-valued measure η given by*

$$\eta = \sum_{j=0}^{p-1} \alpha_j \eta_j, \tag{2.6}$$

where $\alpha_0,\ldots,\alpha_{p-1} \in \mathbb{C}$ are chosen such that the relation

$$P(z) = z\prod_{k=1}^{p-1}(z-\beta_k) - \sum_{j=0}^{p-1}\alpha_j \prod_{k=j+1}^{p-1}(z-\beta_k) \tag{2.7}$$

holds for all $z \in \mathbb{C}$. In particular, if $\beta_1,\ldots,\beta_{p-1}$ are distinct,

$$\eta(dv) = \gamma_0\delta_0(dv) + \Bigl(\mathbb{1}_{[0,\infty)}(v)\sum_{i=1}^{p-1}\gamma_i e^{\beta_i v}\Bigr)dv \tag{2.8}$$

where

$$\gamma_0 = -\Bigl(a_1 + \sum_{j=1}^{p-1}\beta_j\Bigr) \quad \text{and} \quad \gamma_i = -\frac{P(\beta_i)}{Q'(\beta_i)} \quad \text{for} \quad i = 1,\ldots,p-1.$$

Proof. It follows immediately from [2, Theorem 3.12] that $(Y_t)_{t\in\mathbb{R}}$ is the unique stationary solution to (2.5) with η given by (2.6). Assume now that the roots of Q are distinct. Then relation (2.7) implies in particular that $\gamma_0 = \alpha_0 = -(a_1 + \sum_{j=1}^{p-1}\beta_j)$. Moreover, an induction argument shows that

$$\eta_j(dv) = \mathbb{1}_{[0,\infty)}(v)\sum_{i=1}^{j} e^{\beta_i v}\prod_{k=1,k\neq i}^{j}(\beta_i-\beta_k)^{-1}\,dv,$$

from which it follows that

$$\begin{aligned}\eta(dv) - \alpha_0\delta_0(dv) &= \sum_{j=1}^{p-1}\alpha_j\Bigl(\mathbb{1}_{[0,\infty)}(v)\sum_{i=1}^{j}e^{\beta_i v}\prod_{k=1,k\neq i}^{j}(\beta_i-\beta_k)^{-1}\,dv\Bigr)\\ &= \mathbb{1}_{[0,\infty)}(v)\sum_{i=1}^{p-1}e^{\beta_i v}\sum_{j=i}^{p-1}\alpha_j\prod_{k=1,k\neq i}^{j}(\beta_i-\beta_k)^{-1}\,dv.\end{aligned}$$

Finally, observe that the definition of $\alpha_0, \alpha_1, \dots, \alpha_{p-1}$ implies that

$$\gamma_i = \frac{\sum_{j=i}^{p-1} \alpha_j \prod_{k=j+1}^{p-1}(\beta_i - \beta_k)}{\prod_{k=1,k\neq i}^{p-1}(\beta_i - \beta_k)} = \sum_{j=i}^{p-1} \alpha_j \prod_{k=1,k\neq i}^{j} (\beta_i - \beta_k)^{-1}, \ i = 1, \dots, p-1,$$

which concludes the proof. □

Remark 2.2. In [7] they assume that the roots of P are distinct. This makes it possible to write $(Y_t)_{t\in\mathbb{R}}$ as a sum of dependent Ornstein-Uhlenbeck processes, which they in turn use to recover the driving Lévy process. In Theorem 2.1 above we invert the CARMA process by using that it is a solution to an SDDE and thereby circumvent the assumption of distinct roots. On the other hand, when $q \geq 2$, the roots of Q may complex-valued and this would make an estimation procedure that is parametrized by these roots (such as the one given in Section 3) more complicated in practice.

Theorem 2.1 gives an insightful intuition about inverting CARMA processes as well. Let $\mathcal{F}$ be the Fourier transform where $\mathcal{F}[f](y) = \int_{\mathbb{R}} e^{iyx} f(x)\,\mathrm{d}x$ for $f \in L^1$. If we then heuristically take this Fourier transform on both sides of (2.4) we get

$$P(-iy)\mathcal{F}[Y](y) = Q(-iy)\mathcal{F}[DL](y).$$

For $\gamma_0 \in \mathbb{R}$, this can be rewritten as

$$\mathcal{F}[DL](y) = \left(\frac{P(-iy) + (iy + \gamma_0)Q(-iy)}{Q(-iy)} - \gamma_0\right)\mathcal{F}[Y](y) + \mathcal{F}[DY](y).$$

If we let $\gamma_0 = -\left(a_1 + \sum_{j=1}^{p-1} \beta_j\right)$ then

$$y \mapsto \frac{P(-iy) + (iy + \gamma_0)Q(-iy)}{Q(-iy)} \in L^2,$$

and consequently, there exists $f \in L^2$ such that

$$\left(\frac{P(-iy) + (iy + \gamma_0)Q(-iy)}{Q(-iy)} - \gamma_0\right)\mathcal{F}[Y](y) = \mathcal{F}[-f * Y - \gamma_0 Y](y).$$

We conclude that $(Y_t)_{t\in\mathbb{R}}$ satisfy the formal equation $DY_t = f * Y_t + \gamma_0 Y_t + DL_t$. By integrating this equation we get an equation of the form (2.5), and in the case where Q has distinct roots, contour integration and Cauchy's residue theorem imply that

$$f(v) = \mathbb{1}_{[0,\infty)}(v) \sum_{i=1}^{p-1} -\frac{P(\beta_i)}{Q'(\beta_i)} e^{\beta_i v}$$

in line with Theorem 2.1.

The simplest example beyond the (Lévy-driven) Ornstein-Uhlenbeck process is the invertible CARMA$(2,1)$ process:

Example 2.3. Suppose that $a_0, a_1 \in \mathbb{R}$ are chosen such that the zeroes of $P(z) = z^2 + a_1 z + a_2$ have negative real parts and let $b_0 > 0$ so that the same holds for $Q(z) =$

$b_0 + z$. Then there exists an associated invertible CARMA(2,1) process $(Y_t)_{t\in\mathbb{R}}$, and Theorem 2.1 implies that

$$dY_t = \alpha_0 Y_t\,dt + \alpha_1 \int_0^\infty e^{\beta_1 v} Y_{t-v}\,dv\,dt + dL_t,$$

where $\beta_1 = -b_0$, $\alpha_0 = b_0 - a_1$, and $\alpha_1 = (a_1 - b_0)b_0 - a_2$. Note that, in this particular case, we have $\gamma_0 = \alpha_0$ and $\gamma_1 = \alpha_1$.

We end this section by giving the mean and the autocovariance function of the invertible CARMA$(p,p-1)$ process. To formulate the result we introduce the $p\times p$-matrix

$$A_p = \begin{bmatrix} \beta_1 & 0 & 0 & \cdots & 0 & 1 \\ 1 & \beta_2 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \beta_3 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & \beta_{p-1} & 0 \\ \alpha_1 & \alpha_2 & \cdots & \alpha_{p-2} & \alpha_{p-1} & \alpha_0 \end{bmatrix}, \tag{2.9}$$

where $\alpha_0, \alpha_1, \beta_1, \ldots, \alpha_{p-1}, \beta_{p-1} \in \mathbb{C}$ are given as in Theorem 2.1. In case $p = 1$, respectively $p = 2$, the matrix in (2.9) reduces to $A_1 = \alpha_0$, respectively

$$A_2 = \begin{bmatrix} \beta_1 & 1 \\ \alpha_1 & \alpha_0 \end{bmatrix}.$$

Proposition 2.4. *Let $(Y_t)_{t\in\mathbb{R}}$ be an invertible CARMA$(p,p-1)$ process and let η be the associated measure introduced in Theorem 2.1. Then*

$$\mathbb{E}[Y_t] = -\frac{\mu}{\eta([0,\infty))} \quad \textit{and} \quad \gamma(t) := Cov(Y_t, Y_0) = \sigma^2 e_p' e^{A_p|t|}\Sigma e_p, \quad t \in \mathbb{R},$$

where

$$\mu = \mathbb{E}[L_1], \quad \sigma^2 = \mathrm{Var}(L_1), \quad \textit{and} \quad \Sigma = \int_0^\infty e^{A_p y} e_p e_p' e^{A_p' y}\,dy.$$

In particular, $(Y_t)_{t\in\mathbb{R}}$ is centered if and only if $(L_t)_{t\in\mathbb{R}}$ is centered.

Proof. The mean of Y_t is obtained from (2.5) using the stationarity of $(Y_t)_{t\in\mathbb{R}}$. The autocovariance of $(Y_t)_{t\in\mathbb{R}}$ function is given in [2, p. 14]. □

3 Estimation of the SDDE parameters

Fix $\Delta > 0$ and $n \in \mathbb{N}$, and suppose that we have $n+1$ equidistant observations $Y_0, Y_\Delta, Y_{2\Delta}, \ldots, Y_{n\Delta}$ of an invertible CARMA$(p,p-1)$ process $(Y_t)_{t\in\mathbb{R}}$. Our interest will be on estimating the vector of parameters

$$\theta_0 = (\alpha_0, \alpha_1, \beta_1, \alpha_2, \beta_2, \ldots, \alpha_{p-1}, \beta_{p-1})'$$

of η in (2.6). We will restrict our attention to the case where $\theta_0 \in \mathbb{R}^{2p-1}$. For simplicity, we will also assume that $(Y_t)_{t\in\mathbb{R}}$ or, equivalently, $(L_t)_{t\in\mathbb{R}}$ is centered. For any given θ let $P_{k-1}(Y_{k\Delta} \mid \theta)$ be the $L^2(\mathbb{P}_\theta)$-projection of $Y_{k\Delta}$ onto the linear span of $Y_0, Y_\Delta, Y_{2\Delta}, \dots, Y_{(k-1)\Delta}$ and set $\epsilon_k(\theta) = Y_{k\Delta} - P_{k-1}(Y_{k\Delta} \mid \theta)$. Then the least squares estimator $\hat{\theta}_n$ of θ_0 is the point that minimizes

$$\theta \mapsto \sum_{k=1}^{n} \epsilon_k(\theta)^2.$$

In practice, the projections $P_{k-1}(Y_{k\Delta} \mid \theta)$, $k = 1, \dots, n$, can be computed using the Kalman recursions (see, e.g., [6, Proposition 12.2.2]) together with the state-space representation given in Proposition 3.1 below. We stress that one can compute the projections without a state-space representation, e.g., using the Durbin-Levinson algorithm (see [6, p. 169]), but this approach will be very time-consuming for large n and a cut-off is necessary in practice. (This technique is used by [12] in the SDDE framework (2.5) when η is compactly supported and $(L_t)_{t\in\mathbb{R}}$ is a Brownian motion.) Under weak regularity assumptions, following the arguments in [7, Proposition 4-5] that rely on [10], one can show that the estimator $\hat{\theta}_n$ of θ_0 is (strongly) consistent and asymptotically normal.

Proposition 3.1 provides a convenient state-space representation of $(Y_{k\Delta})_{k\in\mathbb{N}_0}$ in terms of $\alpha_0, \alpha_1, \beta_1, \dots, \alpha_{p-1}, \beta_{p-1}$ (rather than the one from the definition of $(Y_t)_{t\in\mathbb{R}}$ in terms of the coefficients of P and Q).

Proposition 3.1. *Let the setup be as above and let A_p be the matrix given in (2.9). Then $(Y_{k\Delta})_{k\in\mathbb{N}_0}$ has the state-space representation $Y_{k\Delta} = e_p' Z_k$, $k \in \mathbb{N}_0$, with $(Z_k)_{k\in\mathbb{N}_0}$ satisfying the state-equation*

$$Z_k = e^{A_p\Delta} Z_{k-1} + U_k, \quad k \in \mathbb{N},$$

where $(U_k)_{k\in\mathbb{N}}$ is a sequence of i.i.d. random vectors with mean 0 and covariance matrix $\int_0^\Delta e^{A_p u} e_p e_p' e^{A_p' u}\, du$.

Proof. It follows by [2, Proposition 3.13] that $Y_t = e_p' \tilde{Z}_t$, $t \in \mathbb{R}$, where $(\tilde{Z}_t)_{t\in\mathbb{R}}$ is the $\mathbb{R}^p$-valued Ornstein-Uhlenbeck process given by

$$\tilde{Z}_t = \int_{-\infty}^{t} e^{A_p(t-u)} e_p \, dL_u$$

for $t \in \mathbb{R}$. Thus, by defining $Z_k = \tilde{Z}_{k\Delta}$ so that $Y_{k\Delta} = e_p' Z_k$, $k \in \mathbb{N}_0$, and observing that

$$Z_k = \int_{-\infty}^{(k-1)\Delta} e^{A_p(k\Delta-u)} e_p \, dL_u + \int_{(k-1)\Delta}^{k\Delta} e^{A_p(k\Delta-u)} e_p \, dL_u = e^{A_p\Delta} Z_{k-1} + U_k$$

with $U_k := \int_{(k-1)\Delta}^{k\Delta} e^{A_p(k\Delta-u)} e_p \, dL_u$ for $k \in \mathbb{N}$, the result is immediate. □

4 A simulation study, $p = 2$

The simulation of the invertible CARMA(2,1) is done in a straightforward manner by the (defining) state-space representation of $(Y_t)_{t\in\mathbb{R}}$ and an Euler discretization of (2.2). In order to ensure that X_0 is a realization of the stationary distribution we take $20{,}000$ steps of size 0.01 before time 0. Given X_0 the simulation is based on

200,000 steps each of size 0.01, and then it is assumed that we have $n+1 = 2{,}000$, respectively $n+1 = 20{,}000$, observations of the process $Y_0, Y_\Delta, Y_{2\Delta}, \ldots, Y_{(n-1)\Delta}$ on a grid with distance $\Delta = 1$, respectively $\Delta = 0.1$, between adjacent points. We will be considering the case where the background noise $(L_t)_{t\in\mathbb{R}}$ is a gamma (Lévy) process with shape parameter $\lambda > 0$ and scale parameter $\xi > 0$. Recall that the gamma process with parameters λ and ξ is a pure jump process with infinite activity, and the density f (at time 1) is given by

$$f(x) = \frac{1}{\Gamma(\lambda)\xi^\lambda} x^{\lambda-1} e^{-\frac{x}{\xi}}, \quad x > 0,$$

where Γ is the gamma function. In line with [7] we will choose the parameters to be $\lambda = 0.2488$ and $\xi = 0.5792$. For comparison we will also study the situation where $(L_t)_{t\in\mathbb{R}}$ is Brownian motion with mean $\mu = \lambda\xi = 0.1441$ and standard deviation $\sigma = \xi\sqrt{\lambda} = 0.2889$ (these parameters are chosen so that the Brownian motion matches the mean and standard deviation of the gamma process). After subtracting the sample mean $\bar{Y}_n = n^{-1}\sum_{k=0}^{n-1} Y_{k\Delta}$ from the observations, the vector of true parameters $\theta_0 = (\alpha_0, \alpha_1, \beta_1)$ is estimated as outlined in Section 3. We will choose $\theta_0 = (-1.1558, 0.1939, -0.2061)$ as in [7] (this choice corresponds to $a_1 = 1.3619$, $a_2 = 0.0443$, and $b_0 = 0.2061$, which are certain estimated values of a stochastic volatility model by [15]). We repeat the experiment 100 times and the estimated parameters are given in Table 1.

Noise	Spacing	Parameter	Sample mean	Bias	Sample std deviation
Gamma	$\Delta = 1$	α_0	-1.2075	- 0.0517	0.1155
		α_1	0.2157	0.0218	0.0501
		β_1	-0.2190	-0.0129	0.0366
	$\Delta = 0.1$	α_0	-1.1688	-0.0130	0.0466
		α_1	0.1934	-0.0005	0.0315
		β_1	-0.2053	0.0008	0.0296
Gaussian	$\Delta = 1$	α_0	-1.1967	- 0.0409	0.1147
		α_1	0.2117	0.0178	0.0524
		β_1	-0.2201	-0.0140	0.0358
	$\Delta = 0.1$	α_0	-1.1653	-0.0095	0.0469
		α_1	0.2002	0.0062	0.0353
		β_1	-0.2121	-0.0060	0.0324

Table 1: Estimated SDDE parameters based on 100 simulations of the CARMA(2,1) process on $[0, 2000]$ with true parameters $\alpha_0 = -1.1558$, $\alpha_1 = 0.1939$, and $\beta_1 = -0.2061$.

It appears that the (absolute value of the) bias of $(\alpha_0, \alpha_1, \beta_1)$ is very small when $\Delta = 0.1$. The general picture is that the bias is largest for α_0, and it is also consistently negative. This observations should, however, be seen in light of the relative size of α_0 compared to α_1 and β_1.

Once we have estimated θ_0 we can estimate the driving Lévy process by exploiting the relation presented in Theorem 2.1 and using the trapezoidal rule. Note that, as in the estimation, we use the relation in Theorem 2.1 on the demeaned data so that we in turn recover the centered version of the Lévy process. Finally, to obtain an estimate of the true Lévy process we estimate $\mu = \mathbb{E}[L_1]$ using Proposition 2.4. In order to get a proper approximation of the integral $\int_0^\infty e^{\beta_1 v}(Y_{t-v} - \mathbb{E}_{\theta_0}[Y_0])\,dv$ we will only be estimating $L_{k\Delta} - L_{(k-1)\Delta}$ for $m := 50\Delta^{-1} \leq k \leq n$. If one is interested in estimating the entire path $L_{(m+1)\Delta} - L_{m\Delta}, L_{(m+2)\Delta} - L_{m\Delta}, \ldots, L_{n\Delta} - L_{m\Delta}$, one will need data observed at a high frequency, that is, small Δ, since the approximation errors accumulate over time. Typically, one is more interested in estimating the distribution of L_1, which is less sensitive to these approximation errors, and this is our focus in the following. For this reason, we have in Figure 2 plotted five estimations of the distribution function of L_1 in dashed lines against the true distribution function (represented by a solid line) in the low frequency case ($\Delta = 1$). The left, respectively right, figure corresponds to the gamma, respectively Gaussian, case. Due to the above conventions, each estimated distribution function is based on $1,950$ estimated realizations of L_1. Generally, the estimated distribution functions in the figures seem to capture the true structure and give a fairly precise estimate, however, there is a slight tendency to over-estimate small values and under-estimate large values.

Due to the high degree of precision of the estimated distribution functions, we plot an associated histogram, based on $1,950$ realizations of L_1 and a sampling frequency of $\Delta = 1$, against the theoretical probability density function in order to detect potential (smaller) biases. We compare this to a histogram of the actual increments. For simplicity, we have restricted ourselves to the Gaussian case as the gamma case is difficult to analyze close to zero (specifically, this will require more observations). The plots are found in Figure 3. We see that the two histograms have very similar appearances, but the histogram based on estimated parameters has a slightly smaller mean.

5 Conclusion and further research

In this paper we have studied the ability to recover the underlying Lévy process from an observed invertible CARMA process using the SDDE relation presented in Theorem 2.1. In particular, after discussing the theoretical foundations, we did a simulation study similar to the one in the classical approach presented in [7] and estimated the underlying Lévy noise. Our findings supported the theory and it seemed possible to (visually) detect the distribution of the underlying Lévy process.

Future research could include a further study of the performance of the presented SDDE inversion technique compared to the classical approach in [7]. Specifically, in light of Remark 2.2, a suggestion could be to consider a situation where P has a root of multiplicity strictly greater than one or where $q \geq 2$ and some of the roots of Q are not real numbers. Such situations may complicate the analysis in one approach relative to the other. Furthermore, it may be interesting to study inversion formulas for invertible CARMA(p,q) processes when $p > q+1$. In particular, a manipulation of the equation (2.4) yields

$$dL_t = \left(\frac{P(D)}{Q(D)} Y_t\right) dt. \tag{5.1}$$

The content of Theorem 2.1 is that the right-hand side of (5.1) is meaningful when $p = q+1$ and it should be interpreted as $dY_t - \int_{[0,\infty)} Y_{t-v}\,\eta(dv)\,dt$. It seems that this statement continues to hold when $p > q+1$ as well when dY_t is replaced by a suitable linear combination of $dY_t, dY_t^{(1)}, \ldots, dY_t^{(p-q-1)}$.

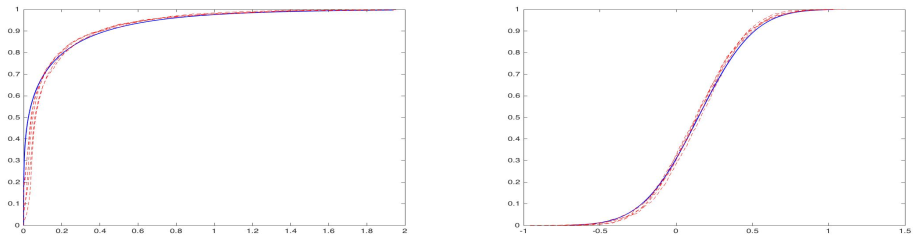


Figure 2: Five estimations of the distribution function of L_1, based on estimates of α_0, α_1, and β_1, plotted against the true distribution function for a sampling frequency of $\Delta = 1$. The left corresponds to gamma noise and the right to Gaussian noise.

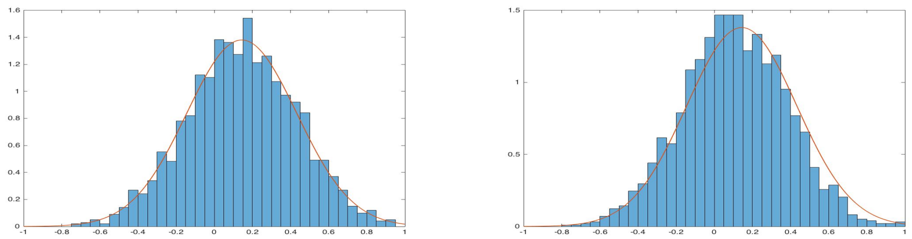


Figure 3: Histograms of the true increments on the left and estimated increments, based on estimates of α_0, α_1, and β_1 for a sampling frequency of $\Delta = 1$, on the right plotted against the theoretical (Gaussian) probability density function.

References

[1] Barndorff-Nielsen, O.E. and N. Shephard (2001). Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial economics. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 63(2), 167–241.

[2] Basse-O'Connor, A., M.S. Nielsen, J. Pedersen and V. Rohde (2019). Stochastic delay differential equations and related autoregressive models. *Stochastics.* Forthcoming. DOI: 10.1080/17442508.2019.1635601.

[3] Benth, F.E., C. Klüppelberg, G. Müller and L. Vos (2014). Futures pricing in electricity markets based on stable CARMA spot models. *Energy Econ.* 44, 392–406.

[4] Benth, F.E., J. Šaltytė-Benth and S. Koekebakker (2007). Putting a price on temperature. *Scand. J. Statist.* 34(4), 746–767. DOI: 10.1111/j.1467-9469.2007.00564.x.

[5] Brockwell, P.J. (2001). Lévy-driven CARMA processes. *Ann. Inst. Statist. Math.* 53(1). Nonlinear non-Gaussian models and related filtering methods (Tokyo, 2000), 113–124. DOI: 10.1023/A:1017972605872.

[6] Brockwell, P.J. and R.A. Davis (2006). *Time series: theory and methods.* Springer Series in Statistics. Reprint of the second (1991) edition. Springer, New York.

[7] Brockwell, P.J., R.A. Davis and Y. Yang (2011). Estimation for non-negative Lévy-driven CARMA processes. *J. Bus. Econom. Statist.* 29(2), 250–259. DOI: 10.1198/jbes.2010.08165.

[8] Brockwell, P.J., V. Ferrazzano and C. Klüppelberg (2013). High-frequency sampling and kernel estimation for continuous-time moving average processes. *J. Time Series Anal.* 34(3), 385–404. DOI: 10.1111/jtsa.12022.

[9] Brockwell, P.J. and A. Lindner (2009). Existence and uniqueness of stationary Lévy-driven CARMA processes. *Stochastic Process. Appl.* 119(8), 2660–2681. DOI: 10.1016/j.spa.2009.01.006.

[10] Francq, C. and J.-M. Zakoïan (1998). Estimating linear representations of nonlinear processes. *J. Statist. Plann. Inference* 68(1), 145–165.

[11] García, I., C. Klüppelberg and G. Müller (2011). Estimation of stable CARMA models with an application to electricity spot prices. *Stat. Model.* 11(5), 447–470.

[12] Küchler, U. and M. Sørensen (2013). Statistical inference for discrete-time samples from affine stochastic delay differential equations. *Bernoulli* 19(2), 409–425. DOI: 10.3150/11-BEJ411.

[13] Schwartz, E.S. (1997). The stochastic behavior of commodity prices: Implications for valuation and hedging. *J. Finance* 52(3), 923–973.

[14] Todorov, V. (2009). Estimation of continuous-time stochastic volatility models with jumps using high-frequency data. *J. Econometrics* 148(2), 131–148.

[15] Todorov, V. (2011). Econometric analysis of jump-driven stochastic volatility models. *J. Econometrics* 160(1), 12–21.

[16] Todorov, V. and G. Tauchen (2006). Simulation methods for Lévy-driven continuous-time autoregressive moving average (CARMA) stochastic volatility models. *J. Bus. Econom. Statist.* 24(4), 455–469. DOI: 10.1198/073500106000000260.

PAPER G

On non-negative modeling with CARMA processes

Fred Espen Benth and Victor Rohde

Abstract

Two stationary and non-negative processes that are based on continuous-time autoregressive moving average (CARMA) processes are discussed. First, we consider a generalization of Cox-Ingersoll-Ross (CIR) processes. Next, we consider CARMA processes driven by compound Poisson processes with exponential jumps which are generalizations of Ornstein-Uhlenbeck (OU) processes driven by the same noise. The way in which the two processes generalize CIR and OU processes and the relation between them will be discussed. Furthermore, the stationary distribution, the autocorrelation function, and pricing of zero-coupon bonds are considered.

MSC 2010: 60G10, 91G30

Keywords: continuous-time ARMA processes; Lévy processes; Ornstein-Uhlenbeck processes; stationary processes; square-root process

1 Introduction

The CIR process has been studied extensively and found many applications, in particular, as a model for the volatility of a financial asset (known as the Heston volatility model [24]) and as a model for spot interest rates (see [19]). More recently, a CIR process has been proposed to model wind speeds in [5] where a significant reduction in the mean-square prediction error is obtained compared with more classical static models. The OU process is another extensively studied process that is popular in modeling volatility (see [35]), interest rates (known as the Vasicek model [38]), and wind (see [7]). Both the CIR and OU processes have the desirable trait of being continuous-time processes that are relatively simple, analytically tractable, stationary and, when the OU process is driven by a subordinator, non-negative. On the other hand, a drawback of CIR and OU processes is that the correlation structure is not very flexible; the autocorrelation function (ACF) is simply an exponential function. This often results in models that do not capture some of the correlation structure. For example, in their application of a CIR process to wind, we see in [5, Figure 1] (see also Figures 1 and 2 in this paper) that the model ACF overshoots the population ACF for small lags and undershoots it for large lags. This motivates studying non-negative and stationary processes that generalize CIR and OU processes but with a correlation structure that is more flexible.

A natural generalization of OU processes is CARMA processes. CARMA processes are stationary, there are conditions for non-negativity (see [13, 37]), and they have a more flexible correlation structure (cf. Proposition 4.1). Many results about CARMA processes are given in the literature, for example, a prediction formula, a multivariate extension, noise recovery, an extension to Hilbert space valued processes, an extension to incorporate long memory, and a CAR(∞) representation (see [4, 9, 13, 15, 16, 30]). Moreover, there exists a large body of literature on the application of CARMA processes. For example, CARMA processes have been used to model realized volatility, interest rates, electricity prices, and temperature (see [1, 6, 8, 13, 17, 22, 36]).

CIR and OU processes are closely related in two ways. First, sums of independent copies of squared Gaussian OU processes constitute special cases of CIR processes. Second, both a CIR process and an OU process driven by a compound Poisson process with exponentially distributed jumps have a gamma distribution as their stationary distribution and an exponential function as their ACF. In this paper we discuss two stationary and non-negative models based on CARMA processes that build on these two connections between CIR and OU processes.

Firstly, we consider a sum of independent copies of squared Gaussian CARMA processes which, motivated by the discussion above, we will call a CIR-CARMA process. Using a connection between the sum of independent copies of squared CARMA processes and Wishart processes, we are able to extend this class of non-negative stationary processes that naturally generalizes CIR processes.

Secondly, we will discuss a CARMA process driven by a compound Poisson process with exponential jumps. This model is very tractable but its stationary distribution is not well understood. We show that, under assumptions frequently satisfied in practice, the stationary distribution is an infinite sum of independent gamma distributed random variables.

In both models we also discuss the ACF, which we will compare with the ACF of a CIR process (or, equivalently, an OU process), and we price zero-coupon bonds when the spot interest rate is governed by the models to show their analytical tractability.

We start with Section 2 where we briefly introduce the setup. Section 3 introduces the CIR-CARMA model and Section 4 discusses CARMA processes with a focus on when the Lévy process is a compound Poisson process with exponential jumps.

2 Preliminaries

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a complete probability space, equipped with a filtration $(\mathcal{F}_t)_{t\geq 0}$ satisfying the usual hypotheses (complete and right-continuous).

We now introduce CARMA processes. We will only consider causal CARMA processes that is, where the auto-regressive polynomial P has roots with negative real part, and we therefore limit our discussion to those. For a more thorough discussion see for example [12]. For $p \in \mathbb{N}, q \in \mathbb{N} \cup \{0\}$, consider two polynomials P and Q given by

$$\begin{aligned} P(z) &= z^p + a_1 z^{p-1} + \cdots + a_p \quad \text{and} \\ Q(z) &= b_0 + b_1 z + \cdots + b_{q-1} z^{q-1} + z^q, \end{aligned}$$

where $p > q$, and assume that the roots of P have negative real part. Then a causal CARMA(p,q) process $(Y(t))_{t\in\mathbb{R}}$ satisfies the formal differential equation

$$P(D)Y(t) = Q(D)DL(t) \tag{2.1}$$

where $(L(t))_{t\in\mathbb{R}}$ is a two-sided square integrable Lévy process. In the case $(L(t))_{t\in\mathbb{R}}$ is a standard Brownian motion, we denote it by $(B(t))_{t\in\mathbb{R}}$. CARMA processes may be defined for Lévy processes possessing only finite log-moments (see [14]) but, for convenience, we restrict ourselves to the square integrable case. The representation (2.1) can be made rigorous by considering the state-space representation. In particular, let $(X(t))_{t\in\mathbb{R}}$ be a p-dimensional Ornstein-Uhlenbeck process that satisfies

$$dX(t) = AX(t)\,dt + e_p dL(t),$$

where e_p is the p-th standard basis vector in $\mathbb{R}^p$ and

$$A = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 1 \\ -a_p & -a_{p-1} & \cdots & -a_2 & -a_1 \end{bmatrix}. \tag{2.2}$$

Then $(Y(t))_{t\in\mathbb{R}}$ is defined by $Y(t) = b^\top X(t)$ where

$$b = (b_0, \ldots, b_{q-1}, 1, 0, \ldots, 0)^\top \in \mathbb{R}^p, \tag{2.3}$$

and $x^\top$ denotes the transpose of x. We call $(X(t))_{t\in\mathbb{R}}$ the state-space process associated to the CARMA process $(Y(t))_{t\in\mathbb{R}}$.

CARMA processes have a moving average representation, that is,

$$Y(t) = \int_{\mathbb{R}} g(t-u)\,dL(u), \quad t \in \mathbb{R}, \tag{2.4}$$

where $g : \mathbb{R} \to \mathbb{R}$ is characterized by

$$\mathcal{F}[g](y) := \int_{\mathbb{R}} e^{-iyx} g(x)\,dx = \frac{Q(iy)}{P(iy)}, \quad y \in \mathbb{R}.$$

If $(Y(t))_{t\in\mathbb{R}}$ is causal then $g(x) = b^\top e^{Ax} e_p$ for $x \geq 0$ and zero otherwise.

If the polynomial P has distinct roots with negative real part then the spectral decomposition of g takes the simple form

$$g(x) = \sum_{i=1}^{p} \frac{Q(\alpha_i)}{P'(\alpha_i)} e^{\alpha_i x}, \quad x \geq 0, \tag{2.5}$$

where $\alpha_1, \ldots, \alpha_p$ are the roots of P and P' denotes the differential of P. In general, if P does not have distinct roots, g still has a spectral decomposition but, to remain brief, we will not give the details here (see [12] for more on this case).

We will say a random variable has an exponential distribution with parameter $\kappa > 0$ if it has density

$$x \mapsto \kappa e^{-\kappa x}, \quad x > 0,$$

and say it has a gamma distribution with shape $\alpha > 0$ and rate $\beta > 0$ if it has density

$$x \mapsto \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, \quad x > 0.$$

Here, Γ denotes the gamma function.

3 The CIR-CARMA process

A sum of independent copies of squared OU processes constitutes a special case of a CIR process. In this section we will investigate what happens if the OU processes are replaced by CARMA processes. In particular, we will connect a sum of independent copies of squared CARMA processes to the Wishart process, which is an extension of CIR processes to the matrix valued case. As we will see, the Wishart process will play a similar role to a sum of independent copies of squared CARMA processes as the state-space process does for CARMA processes.

We start by introducing CIR processes. A process $(r(t))_{t\geq 0}$ is said to be a CIR process if it has dynamics given by

$$dr(t) = a(b - r(t))\,dt + \sigma\sqrt{r(t)}\,dB(t), \qquad r(0) = r_0 \geq 0, \tag{3.1}$$

where $a, b, \sigma > 0$ are constants and $(B(t))_{t\in\mathbb{R}}$ is a standard Brownian motion. The process $(r(t))_{t\geq 0}$ is mean-reverting and non-negative (even positive under the Feller condition $2ab \geq \sigma^2$).

Another extensively studied mean-reverting stochastic process is the OU process (see, for example, [2, 34]). We say $(X(t))_{t\geq 0}$ is a Gaussian OU process with mean-reversion $\lambda > 0$ and volatility $\sigma_{OU}^2 > 0$ if

$$dX(t) = -\lambda X(t)\,dt + \sigma_{OU}\,dB(t), \qquad X(0) = X_0 \in \mathbb{R}.$$

The stationary distribution of $(X(t))_{t\geq 0}$ is a mean zero Gaussian distribution with variance $\sigma_{OU}^2/2\lambda$. The following connection between CIR and Gaussian OU processes is well-known (see, for example, [23, Proposition 4] or [26, Chapter 6]).

Proposition 3.1. *Let $(X_i(t))_{t\in\mathbb{R}}$, $i = 1,\dots,n$, be OU processes driven by independent standard Brownian motions $(B_i(t))_{t\in\mathbb{R}}$, $i = 1,\dots,n$, all with mean-reversion $\lambda > 0$ and volatility $\sigma_{OU}^2 > 0$. Then $r(t) = \sum_{i=1}^n X_i^2(t)$ is a CIR process of the form in* (3.1) *with*

$$a = 2\lambda, \quad b = \frac{n\sigma_{OU}^2}{2\lambda}, \quad \textit{and} \quad \sigma = 2\sigma_{OU}.$$

Remark 3.2. From Proposition 3.1 we see that not all CIR processes are squared OU processes. In particular, for a and σ fixed, λ and σ_{OU} are determined. Then b needs to be an integer times $\frac{\sigma_{OU}^2}{2\lambda}$ for the CIR process to be a squared OU process. A CIR process makes sense for general b, and, in this way, CIR processes generalize sums of squared OU processes.

Instead of summing independent copies of squared Gaussian OU processes as in Proposition 3.1 we will now consider a sum of independent copies of squared causal Gaussian CARMA processes. Let $n \in \mathbb{N}$ and $(Y_i(t))_{t\in\mathbb{R}}$, $i = 1,\dots,n$, be independent copies of a causal CARMA(p,q) process driven by $(\sigma B_i(t))_{t\in\mathbb{R}}$, $i = 1,\dots,n$, for $\sigma > 0$, where $(B_i(t))_{t\in\mathbb{R}}$, $i = 1,\dots,n$ are independent standard Brownian motions, and let $(X_i(t))_{t\in\mathbb{R}}$, $i = 1,\dots,n$ be the associated state-space processes. Finally, we define the process $(C(t))_{t\in\mathbb{R}}$ by

$$C(t) = \sum_{i=1}^n Y_i^2(t) = b^\top Z(t) b \tag{3.2}$$

where $(Z(t))_{t\in\mathbb{R}}$ is given by

$$Z(t) = \sum_{i=1}^n X_i(t) X_i^\top(t). \tag{3.3}$$

We will say that $(C(t))_{t\in\mathbb{R}}$ is a CIR-CARMA(p,q) process.

The following proposition is a direct consequence of the definition of $(C(t))_{t\in\mathbb{R}}$ but we still highlight the result since it is an important point for the model.

Proposition 3.3. *Let $(C(t))_{t\in\mathbb{R}}$ be given by* (3.2)*. Then $C(t)$ is gamma distributed with scale $n/2$ and rate $(2\sigma^2\|g\|_2^2)^{-1}$, where $g : \mathbb{R} \to \mathbb{R}$ is given by*

$$g(x) = b^\top e^{Ax} e_p, \quad x \geq 0,$$

and zero otherwise, and $\|\cdot\|_2$ is the $L^2(\mathbb{R})$-norm.

Proof. It follows by the moving average representation of a CARMA process in (2.4) that $Y_i(t)/(\sigma\|g\|_2)$ has a standard Gaussian distribution. Since

$$C(t) = \sigma^2\|g\|_2^2 \sum_{i=1}^n \left(\frac{Y_i(t)}{\sigma\|g\|_2}\right)^2$$

we therefore have that $C(t)$ is the product of $\sigma^2\|g\|_2^2$ and a chi-square distributed variable with n degrees of freedom, or, equivalently, $C(t)$ is gamma distributed with scale $n/2$ and rate $(2\sigma^2\|g\|_2^2)^{-1}$. □

We now state the ACF of a CIR-CARMA process.

Proposition 3.4. *Let $(C(t))_{t\in\mathbb{R}}$ be given by* (3.2)*. Then, for $s \leq t$,*

$$\operatorname{corr}(C(t), C(s)) = \left(\frac{b^\top e^{A(t-s)} \int_0^\infty e^{Au} e_p e_p^\top e^{A^\top u} \, du \, b}{\int_0^\infty (b^\top e^{Au} e_p)^2 \, du} \right)^2. \tag{3.4}$$

Proof. Initially, we note that for any mean zero two-dimensional Gaussian variable (G_1, G_2) it follows from Isserlis' Theorem (cf. [25]) that

$$\mathbb{E}[G_1^2 G_2^2] = \mathbb{E}[G_1^2]\mathbb{E}[G_2^2] + 2(\mathbb{E}[G_1 G_2])^2,$$

and therefore

$$\begin{aligned} \operatorname{cov}(G_1^2, G_2^2) &= \mathbb{E}[G_1^2 G_2^2] - \mathbb{E}[G_1^2]\mathbb{E}[G_2^2] \\ &= 2(\mathbb{E}[G_1 G_2])^2. \end{aligned}$$

Next note that we may assume $n = 1$ in (3.2) since $C(t)$ is a sum of independent copies. Then, since $(Y_1(t))_{t\in\mathbb{R}}$ is a mean zero Gaussian process and

$$\begin{aligned} \mathbb{E}[Y_1(t)Y_1(s)] &= \mathbb{E}\left[\int_{-\infty}^t g(t-u)\, dB(u) \int_{-\infty}^s g(s-u)\, dB(u) \right] \\ &= \int_{\mathbb{R}} g(t-u) g(s-u)\, du \\ &= b^\top \int_0^\infty e^{A(t-s+u)} e_p e_p^\top e^{A^\top u} \, du \, b, \end{aligned}$$

the result follows. □

If the roots of P are distinct, then it is argued in [12, section 3] that

$$\begin{aligned} &b^\top e^{A(t-s)} \int_0^\infty e^{Au} e_p e_p^\top e^{A^\top u} \, du \, b \\ &\qquad = \sum_{i=1}^p \frac{Q(\alpha_i) Q(-\alpha_i)}{P'(\alpha_i) P(-\alpha_i)} e^{\alpha_i (t-s)} \end{aligned} \tag{3.5}$$

where $\alpha_1, \ldots, \alpha_p$ are the roots of P, which simplifies calculating the correlation given in Proposition 3.4.

In Figure 1 the population ACF for the wind speeds at the turbine rotor height every 10 minutes over 365 days in 2005 (ID 24500 located 43.48N and 107.29W in Wyoming, USA) is plotted.[1] We also plot the ACFs of the CIR and CIR-CARMA(2,1) that minimize the squared distance between the 10 minutes lags for a total of 4 days. A similar plot can be found in [5, Figure 1] where the population ACF is compared with the CIR ACF calibrated using maximum likelihood estimation on the same data. The CIR process given by (3.1) has ACF

$$t \mapsto e^{-at}, \quad t \geq 0,$$

[1] These data have been made available to us by Alexandre Brouste and corresponds to the same data used in [5]. Unfortunately, these data are not anymore available on wind.nrel.gov as referred to in [5].

(see, for example, [5]). Here, t is measured in units of ten minutes. For the CIR process we find $a = 0.0069$ as the calibrated parameter and for the CIR-CARMA(2,1) process we find the calibrated parameters

$$\alpha_1 = -0.0024, \quad \alpha_2 = -0.0296, \quad \text{and} \quad \beta_1 = -0.0165,$$

where α_1 and α_2 are the roots of P and β_1 is the root of Q. The CIR-CARMA process is seen to capture the correlation structure very well and, in particular, overcome the difficulties of the CIR process in both capturing the rapid decay for small lags and slow decay for large lags of the population ACF.

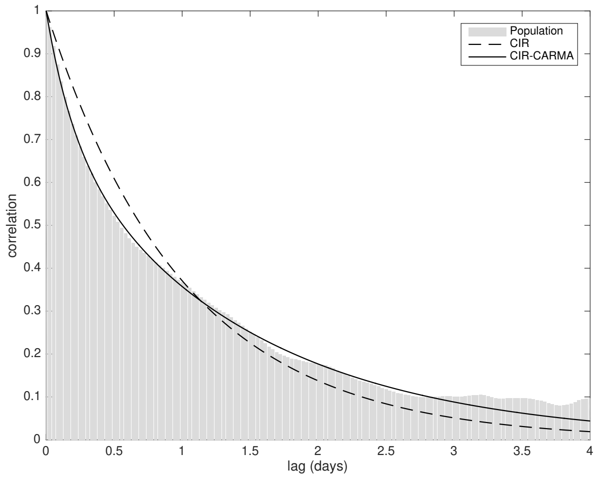


Figure 1: Population ACF in gray, CIR ACF (dotted line), and CIR-CARMA(2,1) ACF (solid line).

Next, to show the analytical tractability of the CIR-CARMA process, we apply it to zero-coupon bond pricing. We assume that the short rate is given by the CIR-CARMA process in (3.2), that is, $r(t) = C(t)$, and where we use $\mathbb{P}$ as the pricing measure. Following the standard theory for pricing based on short rate processes, see e.g. [21], we define the zero-coupon bond price $P(t,T)$ at time $t \geq 0$ for a contract maturing at time $T \geq t$ by

$$P(t,T) = \mathbb{E}\left[\exp\left\{-\int_t^T r(s)\,ds\right\}\middle|\mathcal{F}_t\right]. \tag{3.6}$$

We find the following result:

Theorem 3.5. *Let* $(C(t))_{t\in\mathbb{R}}$ *and* $(Z(t))_{t\in\mathbb{R}}$ *be given by* (3.2) *and* (3.3), *respectively. Then, for* $0 \leq t \leq T < \infty$,

$$P(t,T) = \exp\{n\phi(T-t) + \mathrm{tr}(\Phi(T-t)Z(t))\}, \tag{3.7}$$

where tr is the trace operator,

$$\Phi(t) = \Phi_1(t)\Phi_2(t)^{-1} \tag{3.8}$$

with $\Phi_1 : \mathbb{R} \to \mathbb{R}^{p\times p}, \Phi_2 : \mathbb{R} \to \mathbb{R}^{p\times p}$, *and* $\phi : \mathbb{R} \to \mathbb{R}$ *given by*

$$\begin{pmatrix}\Phi_1(t)\\ \Phi_2(t)\end{pmatrix} = \exp\left(\begin{pmatrix} A^\top & -bb^\top \\ -2\sigma^2 e_p e_p^\top & -A\end{pmatrix} t\right)\begin{pmatrix}0_p\\ I_p\end{pmatrix} \tag{3.9}$$

and

$$\phi(t) = \sigma^2 \int_0^t e_p^\top \Phi(s) e_p \, ds.$$

Here, I_p *is the* $p \times p$ *identity matrix and* 0_p *is the matrix in* $\mathbb{R}^{p\times p}$ *of zeros.*

Proof. Initially, we note that, by independence, we may assume $n = 1$ in (3.2). For $u : [0,\infty) \times \mathbb{R}^p \to \mathbb{R}$, consider the parabolic partial differential equation

$$\begin{aligned} \partial_t u(t,x) &= \tfrac{1}{2}\sigma^2 \partial_p^2 u(t,x) + \nabla_x u(t,x) A x - (b^\top x)^2 u(t,x) \\ u(0,x) &= 1. \end{aligned} \tag{3.10}$$

Here, $x \in \mathbb{R}^p$, $t \geq 0$, and ∂_p denotes the differentiation with respect to the p'th entry of x. Assume that the solution u is given by

$$u(t,x) = \exp\{\phi(t) + x^\top \Phi(t) x\}. \tag{3.11}$$

Then

$$\begin{aligned} \partial_t u(t,x) &= (\phi'(t) + x^\top \Phi'(t) x) u(t,x) \\ \nabla_x u(t,x) &= x^\top (\Phi(t) + \Phi^\top(t)) u(t,x) \\ \partial_p^2 u(t,x) &= (2e_p^\top \Phi(t) e_p + x^\top (\Phi(t) + \Phi^\top(t)) e_p e_p^\top (\Phi(t) + \Phi^\top(t)) x) u(t,x), \end{aligned}$$

and it therefore follows that u solves (3.10) if Φ is symmetric, and ϕ and Φ solves

$$\begin{aligned} \phi'(t) &= \sigma^2 e_p^\top \Phi(t) e_p, \\ \phi(0) &= 0, \\ \Phi'(t) &= 2\sigma^2 \Phi(t) e_p e_p^\top \Phi(t) + \Phi(t) A + A^\top \Phi(t) - bb^\top, \\ \Phi(0) &= 0_p, \quad t \geq 0. \end{aligned} \tag{3.12}$$

Here, we have used that $x^\top 2\Phi^\top A x = x^\top A^\top \Phi x + x^\top \Phi^\top A x$. Given Φ, ϕ is readily found by integration. To find Φ we recognize the equation as a matrix Riccati equation (see for example [28] where a simple change of variable is necessary since the Riccati equation is considered backward). It follows from [28, Theorem 8] that (3.12) has a unique solution on $[0,\infty)$. Since the transpose of a solution again is a solution, it is symmetric. By [28, Theorem 9] we have that the solution is negative semi-definite and therefore that u of the form in (3.11) is bounded and, in particular, of at most polynomial growth. Furthermore, by [29], the solution is given by (3.8) and (3.9).

By the Feynman-Kac formula (see for example [27, Theorem 7.6, Chapter 5]), the solution u is unique with representation

$$\mathbb{E}\left[\exp\left\{-\int_t^T (b^\top X(s))^2 \, ds\right\} \middle| X(t) = x\right] = u(T-t, x).$$

Since

$$\begin{aligned}\mathbb{E}\left[\exp\left\{-\int_t^T r(s)\,ds\right\}\middle|\mathcal{F}_t\right] &= \mathbb{E}\left[\exp\left\{-\int_t^T (b^\top X(s))^2\,ds\right\}\middle|\mathcal{F}_t\right] \\ &= \mathbb{E}[\exp\{\phi(t)+X(t)^\top\Phi(t)X(t))\}|\mathcal{F}_t] \\ &= \exp\{\phi(t)+\mathrm{tr}(\Phi(t)Z(t))\},\end{aligned}\qquad\square$$

the result now follows.

Some comments are in place. Firstly, by considering the case $p = 1$ where we are back to the classical CIR process, it is simple to see that Theorem 3.5 reduces to the known zero-coupon bond price for CIR short rate models (see [19]). Next, in mathematical finance, one usually differentiates between the market probability and pricing measures, sometimes called equivalent martingale measures, denoted $\mathbb{Q} \sim \mathbb{P}$. In the theorem above, we simply used $\mathbb{Q} = \mathbb{P}$. We may, however, assume that the CIR-CARMA model is formulated under some $\mathbb{Q} \sim \mathbb{P}$ directly, and obtain the price for $P(t,T)$ under a conditional expectation with respect to $\mathbb{Q}$. We obtain, of course, the price as in Theorem 3.5. The market prices of $P(t,T)$ is, on the other hand, observed under the market probability $\mathbb{P}$, and to have its dynamics under $\mathbb{P}$ we must specify the $\mathbb{P}$-dynamics of $(Z(t))_{t\in\mathbb{R}}$. As we have a Brownian-based model for $(Z(t))_{t\in\mathbb{R}}$, this can be done by referring to Girsanov's Theorem. Further, we note that the zero-coupon bond price is represented in terms of the process $(Z(t))_{t\in\mathbb{R}}$, which is not directly observable. This is contrary to the CIR model, where the price is explicit in terms of the current state of the interest rate $r(t)$. To make inference, say filtering techniques are called for in the general CIR-CARMA case.

We now connect the CIR-CARMA model to Wishart processes. This will allow us do define more general processes that still are generalizations of CIR processes. The main idea of the proof is to connect multivariate Gaussian OU processes to a Wishart process. This is also discussed in [18].

Theorem 3.6. *Let $(C(t))_{t\in\mathbb{R}}$ be defined as in* (3.2) *and assume $n \geq p$. Then*

$$C(t) = b^\top Z(t)b, \tag{3.13}$$

where $(Z(t))_{t\in\mathbb{R}}$ is an $p\times p$ dimensional process with dynamics

$$\begin{aligned} dZ(t) &= (n\sigma^2 e_p e_p^\top + AZ(t) + Z(t)A^\top)dt \\ &\quad + Z(t)^{1/2}dW(t)\sigma e_p e_p^\top + \sigma e_p e_p^\top dW(t)^\top Z(t)^{1/2},\end{aligned}\tag{3.14}$$

and $(W(t))_{t\in\mathbb{R}}$ is a $p\times p$ dimensional standard matrix Brownian motion, that is, the entries consist of independent standard Brownian motions.

Proof. We have

$$\sum_{i=1}^n Y_i(t)^2 = b^\top\left(\sum_{i=1}^n X_i(t)X_i(t)^\top\right)b.$$

By the multivariate version of Ito's lemma (see for example [32, Theorem 4.2.1])

$$\begin{aligned} d(X_i(t)X_i(t)^\top) &= X_i(t)dX_i(t)^\top + (dX_i(t))X_i(t)^\top + dX_i(t)dX_i(t)^\top \\ &= (X_i(t)X_i(t)^\top A^\top + AX_i(t)X_i(t)^\top + \sigma^2 e_p e_p^\top)dt \\ &\quad + (X_i(t)e_p^\top + e_p X_i(t)^\top)\sigma dB_i(t). \end{aligned}$$

Since $n \geq p$, $\sum_{i=1}^n X_i(s)X_i(s)^\top$ is positive definite almost surely. Let

$$\tilde{W}(t) = \int_0^t \left(\sum_{i=1}^n X_i(s)X_i(s)^\top\right)^{-1/2} \sum_{i=1}^n X_i(s)e_p^\top dB_i(s),$$

where we have used the notation $A^{-1/2} = (A^{1/2})^{-1}$ for a positive definite matrix A. Denote by $[\cdot,\cdot]$ the quadratic covariation and by $B_{i,j}$ the (i,j)'th entry of a $p \times p$ matrix B. Consider $1 \leq i_1, i_2, j_1, j_2 \leq p$. Then

$$\begin{aligned} \left[\tilde{W}_{i_1,j_1}, \tilde{W}_{i_2,j_2}\right](t) = \sum_{i=1}^n \int_0^t &\left(\left(\sum_{i=1}^n X_i(s)X_i(s)^\top\right)^{-1/2} X_i(s)e_p^\top\right)_{i_1,j_1} \\ &\times \left(\left(\sum_{i=1}^n X_i(s)X_i(s)^\top\right)^{-1/2} X_i(s)e_p^\top\right)_{i_2,j_2} ds. \end{aligned} \tag{3.15}$$

If either j_1 or j_2 is not equal to p then (3.15) is zero. Furthermore,

$$\begin{aligned} &\sum_{i=1}^n \left(\left(\sum_{i=1}^n X_i(s)X_i(s)^\top\right)^{-1/2} X_i(s)e_p^\top\right)_{i_1,p} \\ &\quad \times \left(\left(\sum_{i=1}^n X_i(s)X_i(s)^\top\right)^{-1/2} X_i(s)e_p^\top\right)_{i_2,p} \\ &\quad = \sum_{i=1}^n \left(\left(\sum_{i=1}^n X_i(s)X_i(s)^\top\right)^{-1/2} X_i(s)X_i(s)\top \left(\sum_{i=1}^n X_i(s)X_i(s)^\top\right)^{-1/2}\right)_{i_1,i_2} \\ &\quad = (I_p)_{i_1,i_2}. \end{aligned}$$

It follows that $(\tilde{W}(t))_{t\in\mathbb{R}}$ is a $p \times p$ dimensional Brownian motion by Lévy's characterization of Brownian motion (see for example [32, Theorem 8.6.1]) of the form $W(t)e_p e_p^\top$, where $(W(t))_{t\in\mathbb{R}}$ is a $p \times p$ standard matrix Brownian motion, that is, the entries of $(W(t))_{t\in\mathbb{R}}$ are independent standard Brownian motions. Let $Z(t)$ be given by (3.3). We then conclude that

$$\begin{aligned} dZ(t) &= (n\sigma^2 e_p e_p^\top + AZ(t) + Z(t)A^\top)dt \\ &\quad + Z(t)^{1/2} dW(t)\sigma e_p e_p^\top + \sigma e_p e_p^\top dW(t)^\top Z(t)^{1/2} \end{aligned}$$

and the proof is complete. □

It is natural to generalize the level $n\sigma^2 e_p e_p^\top$ in the dynamics of $(Z(t))_{t\geq 0}$ given in (3.14) to a general level as a direct parallel to the way sums of squared OU processes can be generalized by a CIR process to accommodate a general level. Thus, we are interested in finding a solution to

$$\begin{aligned} dZ(t) &= (\kappa e_p e_p^\top + AZ(t) + Z(t)A^\top)dt \\ &\quad + Z(t)^{1/2} dW(t)\sigma e_p e_p^\top + \sigma e_p e_p^\top dW(t)^\top Z(t)^{1/2}, \quad Z(0) = Z_0 \end{aligned} \tag{3.16}$$

for some general κ and positive semi-definite $p \times p$ matrix Z_0. Such processes are Wishart processes and have been studied in [18, 20, 31]. In particular, in [20], it is argued that whenever $\kappa \geq (p-1)\sigma^2$ then there exists an affine Markov process on the space of positive semi-definite matrices with the dynamics in (3.16). We are led to extending the class of CIR-CARMA(p,q) processes to be a process $(C(t))_{t\in\mathbb{R}}$ of the form

$$C(t) = b^\top Z(t) b \tag{3.17}$$

where b is given in (2.3), A is given in (2.2), and where $(Z_t)_{t\geq 0}$ is the affine Markov process given in [20] with dynamics given by (3.16).

The definition of a CIR-CARMA process has a similar structure to that of a CARMA process. In particular, it is defined as a linear transformation of a higher dimensional process. It therefore seems natural to consider the process $(Z(t))_{t\geq 0}$ as a state-space process associated to the CIR-CARMA process $(C(t))_{t\geq 0}$.

It follows immediately from the definition of b and the CIR-CARMA(p,q) process that its paths are $p-q-1$ times continuously differentiable. This is a direct parallel to CARMA processes where a similar statement holds.

If we have $\kappa \geq (p+1)\sigma^2$ in (3.16) then [31] shows that there is a strong unique solution, and it is argued that if, additionally, Z_0 is positive definite then $Z(t)$ remains positive definite for all $t \geq 0$. This result may be regarded as a version of the Feller condition for Wishart processes. We summarize in the following Proposition:

Proposition 3.7 ([31]). *Let $(C(t))_{t\geq 0}$ be a CIR-CARMA(p,q) process and assume that Z_0 is positive definite and $\kappa \geq (p+1)\sigma^2$. Then $C(t) > 0$ for all $t \geq 0$.*

4 CARMA processes driven by a subordinator

In this section we will consider causal CARMA processes as an alternative to CIR-CARMA. We are mainly interested in the case where the background driving Lévy process is a compound Poisson process with exponentially distributed jumps and specifications of the CARMA parameters such that the kernel function is non-negative. In this case, the CARMA process will itself become non-negative.

Modeling with a non-negative CARMA process as opposed to the CIR-CARMA discussed above has the advantage of, in many respects, being a simpler non-negative model that still maintains a large class of possible correlation structures. A downside is, however, that the stationary distribution is more complicated. We address the question of characterizing the stationary distribution under some specific assumptions, but start by considering the ACF and pricing of zero-coupon bonds when the spot interest is modeled by the CARMA process.

The following proposition is well-known (see for example [12, Section 3]).

Proposition 4.1. *Let $(Y(t))_{t\in\mathbb{R}}$ be a causal CARMA(p,q) process driven by a square integrable Lévy process. Then*

$$\mathrm{corr}(Y(t),Y(s)) = \frac{b^\top e^{A(t-s)}\int_0^\infty e^{Au}e_pe_p^\top e^{A^\top u}\,du\,b}{\int_0^\infty (b^\top e^{Au}e_p)^2\,du}, \quad s\le t.$$

Remark 4.2. The above result should be compared with Proposition 3.4. Furthermore, (3.5) may be applied to easily calculate the ACF if the roots of P are distinct.

We again compare the correlation structure of a CARMA(2,1) with that of a CIR process when calibrated to the wind data described in Section 3. The resulting correlations are given in Figure 2. The calibrated CARMA parameters are

$$\alpha_1 = -0.0049, \quad \alpha_2 = -0.0339, \quad \text{and} \quad \beta_1 = -0.0179, \tag{4.1}$$

where α_1 and α_2 are the roots of P and β_1 is the root of Q. Appealing to [37], these parameters corresponds to a non-negative CARMA kernel.

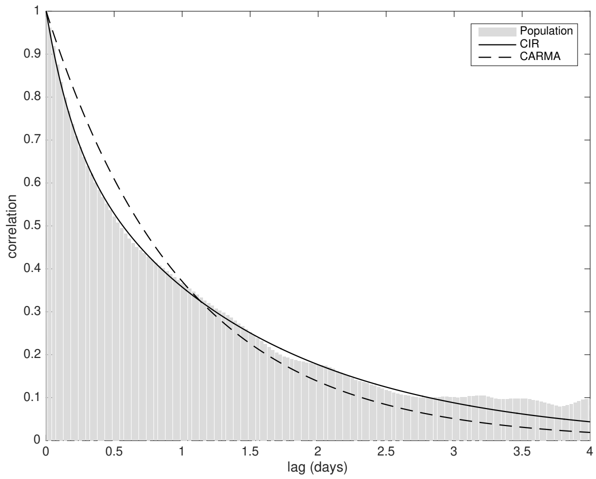


Figure 2: Population acf in gray, CIR ACF (dotted line), and CARMA(2,1) ACF (solid line).

For comparison with the CIR-CARMA case, we give the zero-coupon bond price when the short rate is modeled as a CARMA(p,q) process.

Proposition 4.3. *Let $(Y(t))_{t\in\mathbb{R}}$ be a causal CARMA(p,q) process driven by a compound Poisson process with intensity $\mu>0$ and exponentially distributed jumps with parameter $\kappa>0$. Assume that the kernel of the CARMA process is non-negative. Then*

$$\begin{aligned} P(t,T) :=& \mathbb{E}\left[\exp\left\{-\int_t^T Y(s)\,ds\right\}\middle| L(t)-L(u),\, u<t\right] \\ =& \exp\left\{ b^\top A^{-1}\left(I_p - e^{A(T-t)}\right)X(t) \right. \\ & \left. + \mu\int_0^{T-t} \frac{b^\top A^{-1}\left(I_p-e^{Au}\right)e_p}{\kappa + b^\top A^{-1}(e^{Au}-I_p)e_p}\,du \right\} \end{aligned}$$

where $(X(t))_{t\in\mathbb{R}}$ *is the associated state-space process.*

Before proving Proposition 4.3 we first give the following expression for the moment generating function of a moving average process, which also will be useful in the proof of Theorem 4.8:

Lemma 4.4. *Let* $f \in L^1(\mathbb{R})$ *be a right-continuous and non-negative bounded function with support on* $[0,\infty)$ *and define* $d = \sup_{u\geq 0}|f(u)|$*. Let* $(L(t))_{t\in\mathbb{R}}$ *be a compound Poisson process with intensity* $\mu > 0$ *and exponentially distributed jumps with parameter* $\kappa > 0$*. Then*

$$\log \mathbb{E}\left[\exp\left\{x\int_{-\infty}^{t} f(t-u)\,dL(u)\right\}\right] = \mu\int_0^\infty \frac{xf(u)}{\kappa - xf(u)}\,du$$

for $x \in (-\infty, \kappa/d)$*, and*

$$\int_0^\infty \frac{xf(u)}{\kappa - xf(u)}\,du = \sum_{n=1}^{\infty}\left(\frac{x}{\kappa}\right)^n \int_0^\infty f(u)^n\,du$$

for $x \in (-\kappa/d, \kappa/d)$*.*

Proof. Initially we note that for $s < t$ and any $x \in (-\infty, \kappa)$,

$$\log \mathbb{E}[\exp\{x(L(t)-L(s))\}] = (t-s)\mu\frac{x}{\kappa - x}. \tag{4.2}$$

Since f is right-continuous and non-negative there exists an increasing sequence of piece-wise constant functions $(f_n)_{n\in\mathbb{N}}$ of the form

$$f_n(t) = \sum_{m=0}^{\infty} a_{m,n}\mathbb{1}_{[m/2^n,(m+1)/2^n)}(t) \tag{4.3}$$

with $f_n \to f$ which implies that $\int_{\mathbb{R}} f_n(t-u)\,dL(u) \to \int_{\mathbb{R}} f(t-u)\,dL(u)$ in probability as $n\to\infty$ (cf. [33]). Since $(L(t))_{t\in\mathbb{R}}$ is non-decreasing and f_n is given by (4.3), $\int_{\mathbb{R}} f_n(t-u)\,dL(u)$ is non-decreasing in n. We therefore also have almost sure convergence. Let $x \in (-\infty, \kappa/d)$, and note that

$$|\kappa - xf(u)| \geq \kappa - |x|d > 0.$$

Now it follows from (4.2) and (4.3) that

$$\begin{aligned}
&\log \mathbb{E}\left[\exp\left\{x\int_{-\infty}^{t} f(t-u)\,dL(u)\right\}\right]\\
&= \lim_{n\to\infty}\log \mathbb{E}\left[\exp\left\{x\int_{-\infty}^{t} f_n(t-u)\,dL(u)\right\}\right]\\
&= \mu\lim_{n\to\infty}\int_0^\infty \frac{xf_n(u)}{\kappa - xf_n(u)}\,du\\
&= \mu\int_0^\infty \frac{xf(u)}{\kappa - xf(u)}\,du,
\end{aligned} \tag{4.4}$$

□

where we have used both monotone and dominated convergence. The last statement of the lemma is straightforward.

Proof (Proof of Proposition 4.3). Since $(X(t))_{t\in\mathbb{R}}$ is a multivariate OU process,

$$\int_t^T X(s)\,ds = A^{-1}(X(T)-X(t)-e_p(L(T)-L(t)),$$

and therefore

$$\int_t^T Y(s)\,ds = \int_t^T b^\top X(s)\,ds = b^\top A^{-1}(X(T)-X(t)-e_p(L(T)-L(t))).$$

Note that

$$X(T) = \int_t^T e^{A(T-u)} e_p\,dL(u) + e^{A(T-t)}X(t)$$

and, since the kernel is non-negative,

$$b^\top A^{-1}\left(e^{At}-I_p\right)e_p = \int_0^t b^\top e^{Au} e_p\,du \geq 0.$$

Consequently, by Lemma 4.4, with $f(t) = b^\top A^{-1}(e^{At}-I_p)e_p$ and $x=-1$, we conclude that

$$\begin{aligned}
&\mathbb{E}\left[\exp\left\{-\int_t^T Y(s)\,ds\right\}\middle| L(t)-L(u),\, u\leq t\right] \\
&\quad = \exp\{b^\top A^{-1}(I_p - e^{A(T-t)})X(t)\} \\
&\qquad \times \mathbb{E}\left[\exp\left\{-b^\top A^{-1}\left(\int_t^T e^{A(T-u)}e_p\,dL(u) - e_p(L(T)-L(t))\right)\right\}\right] \\
&\quad = \exp\left\{b^\top A^{-1}(I_p - e^{A(T-t)})X(t) + \mu\int_0^{T-t} \frac{b^\top A^{-1}(I_p - e^{Au})e_p}{\kappa + b^\top A^{-1}(e^{Au}-I_p)e_p}\,du\right\}.
\end{aligned}$$

The proof is complete. □

The following Corollary follows from Lemma 4.4, and gives a series representation of the cumulant function of Y.

Corollary 4.5. *Let $(Y(t))_{t\in\mathbb{R}}$ be a causal CARMA(p,q) process driven by a compound Poisson process with intensity $\mu>0$ and exponentially distributed jumps with parameter $\kappa>0$. Let $d=\sup_{u\geq 0}|g(u)|$. Then*

$$\log \mathbb{E}[e^{xY(t)}] = \mu\sum_{n=1}^{\infty}\frac{a_n}{n}\left(\frac{x}{\kappa}\right)^n, \quad x\in(-\kappa/d,\kappa/d), \tag{4.5}$$

where

$$a_n = n\int_0^\infty \left(b^\top e^{Au}e_p\right)^n du. \tag{4.6}$$

Let us consider the case of an OU process:

Example 4.6. Let $(Y(t))_{t\in\mathbb{R}}$ be an OU process (corresponding to the CARMA$(1,0)$ case). Then A is a negative scalar and therefore

$$a_n = n\int_0^\infty e^{nAu}\,du = -\frac{1}{A}.$$

Consequently, the expression in (4.5) reduces to

$$\log \mathbb{E}[e^{xY(t)}] = -\frac{\mu}{A}\sum_{n=1}^\infty \frac{1}{n}\left(\frac{x}{\kappa}\right)^n = \frac{\mu}{A}\log\left(1-\frac{x}{\kappa}\right),$$

which we recognize as the cumulant of the gamma distribution with shape $-\mu/A$ and rate κ. In particular, by Corollary 4.5, we recover the well-known result that an OU process driven by a compound Poisson process with exponential jumps has the gamma distribution as its stationary distribution (see for example [3]).

We now constrain our analysis to the case of CARMA$(p,p-1)$ processes.

Lemma 4.7. *Let $(Y(t))_{t\in\mathbb{R}}$ be a causal CARMA$(p,p-1)$ process driven by a compound Poisson process with intensity $\mu>0$ and exponentially distributed jumps with parameter $\kappa>0$. Let $(a_n)_{n\in\mathbb{N}}$ be given by (4.6) and assume*

$$\frac{d}{du}b^\top e^{Au}e_p = b^\top A e^{Au}e_p < 0, \quad u>0.$$

Then $(a_n)_{n\in\mathbb{N}}$ is a convergent sequence with

$$a_1 = -b^\top A^{-1}e_p \quad \textit{and} \quad a_\infty := \lim_{n\to\infty} a_n = -\frac{1}{b^\top A e_p}.$$

Proof. Using integration by parts we find that

$$\begin{aligned} a_n &= n\int_0^\infty (b^\top e^{Au}e_p)^n\,du \\ &= \int_0^\infty \frac{d}{du}\Big((b^\top e^{Au}e_p)^n\Big)\frac{b^\top e^{Au}e_p}{b^\top A e^{Au}e_p}\,du \\ &= \left[\frac{(b^\top e^{Au}e_p)^{n+1}}{b^\top A e^{Au}e_p}\right]_0^\infty - \int_0^\infty (b^\top e^{Au}e_p)^n \frac{d}{du}\frac{b^\top e^{Au}e_p}{b^\top A e^{Au}e_p}\,du \\ &= -\frac{1}{b^\top A e_p} - \int_0^\infty (b^\top e^{Au}e_p)^n\left(1-\frac{b^\top e^{Au}e_p b^\top A^2 e^{Au}e_p}{(b^\top A e^{Au}e_p)^2}\right)du. \end{aligned} \tag{4.7}$$

Since the function $u\mapsto b^\top e^{Au}e_p$ is continuous, converges to zero when $u\to\infty$, is 1 at $u=0$, and is strictly decreasing on $(0,\infty)$ by assumption, we conclude that $0<b^\top e^{Au}e_p<1$ for all $u>0$. Thus, by monotone convergence, $(a_n)_{n\in\mathbb{N}}$ is convergent with limit $-\frac{1}{b^\top A e_p}$. The statement about a_1 is a simple calculation. □

Let $\Gamma(\eta,\theta)$ denote the gamma distribution with rate η and scale θ. The following is the main result of this section.

Theorem 4.8. *For $p \geq 2$, let $(Y(t))_{t\in\mathbb{R}}$ be a causal CARMA$(p,p-1)$ process driven by a compound Poisson process with intensity $\mu > 0$ and exponentially distributed jumps with parameter $\kappa > 0$. Let $\alpha_1,\dots,\alpha_p$ be the roots of P and assume that they are distinct and real. Furthermore, assume*

$$\frac{Q(\alpha_i)}{P'(\alpha_i)} > 0, \quad i = 1,\dots,p. \tag{4.8}$$

Then $G + \sum_{i=1}^m G_{i,m}$ converges to $Y(t)$ in distribution as $m \to \infty$ where

$$G \sim \Gamma(\mu\eta,\kappa) \quad \text{and} \quad G_{i,m} \sim \Gamma(\mu\eta_{i,m},\kappa\theta_{i,m})$$

are independent. Here $\eta = -1/(b^\top A e_p)$, and $\eta_{i,m} > 0$ and $\theta_{i,m} > 1$, for $i = 1,\dots,m$, are such that

$$\sum_{i=1}^m \frac{\eta_{i,m}}{\theta_{i,m}^n} \to -\int_0^\infty (b^\top e^{Au} e_p)^n \frac{d}{du}\frac{b^\top e^{Au} e_p}{b^\top A e^{Au} e_p}\, du \tag{4.9}$$

as $m \to \infty$.

Proof. By the spectral decomposition,

$$b^\top e^{At} e_p = \sum_{i=1}^p c_i e^{\alpha_i t}$$

where $c_i = \frac{Q(\alpha_i)}{P'(\alpha_i)} > 0$. Since the CARMA process is causal, $\alpha_i < 0$, $i = 1,\dots,p$, and we conclude that the derivative of $t \mapsto b^\top e^{At} e_p$ is negative. It follows that the assumptions of Lemma 4.7 are satisfied. Furthermore,

$$\begin{aligned}\Delta \sum_{i=0}^\infty (b^\top e^{A\Delta(i+1/2)} e_p)^n \left(\frac{b^\top e^{A\Delta i} e_p}{b^\top A e^{A\Delta i} e_p} - \frac{b^\top e^{A\Delta(i+1)} e_p}{b^\top A e^{A\Delta(i+1)} e_p}\right)&\\ \to -\int_0^\infty (b^\top e^{Au} e_p)^n \frac{d}{du}\frac{b^\top e^{Au} e_p}{b^\top A e^{Au} e_p}\, du&\end{aligned} \tag{4.10}$$

as $\Delta \to 0$. From the assumption in (4.8) we have $0 < b^\top e^{A\Delta(i+1/2)} e_p < 1$, and we find that

$$\begin{aligned}(b^\top A e^{Au} e_p)^2 &= \sum_{i=1}^p c_i^2 \alpha_i^2 e^{2\alpha_i u} + \sum_{i=1}^p \sum_{j>i}^p c_j c_i 2\alpha_j \alpha_i e^{(\alpha_j+\alpha_i)u}\\ &< \sum_{i=1}^p c_i^2 \alpha_i^2 e^{2\alpha_i u} + \sum_{i=1}^p \sum_{j>i}^p c_j c_i (\alpha_j^2 + \alpha_i^2) e^{(\alpha_j+\alpha_i)u}\\ &= b^\top e^{Au} e_p b^\top A^2 e^{Au} e_p,\end{aligned}$$

which implies $\frac{d}{du}\frac{b^\top e^{Au} e_p}{b^\top A e^{Au} e_p} < 0$. Let

$$\theta_{i,m} = \frac{1}{b^\top e^{A\Delta_m(i+1/2)} e_p} \tag{4.11}$$

and

$$\eta_{i,m} = \Delta_m \Big(\frac{b^\top e^{A\Delta_m i} e_p}{b^\top A e^{A\Delta_m i} e_p} - \frac{b^\top e^{A\Delta_m (i+1)} e_p}{b^\top A e^{A\Delta_m (i+1)} e_p} \Big), \tag{4.12}$$

for $i = 1, \dots, m$ and a sequence $(\Delta_m)_{m\in\mathbb{N}}$ such that $\Delta_m \to 0$ and $m\Delta_m \to \infty$ as $m \to \infty$. Then we conclude that (4.9) is satisfied. Now, since convergence of the moment generating function implies convergence in distribution (see, for example, [10, Example 5.5]), the result follows from Corollary 4.5 and (4.7). □

Remark 4.9. The condition in (4.8) ensures that the kernel function of the CARMA process is a sum of positive scalars times exponential function by the spectral decomposition in (2.5). This is often the case when CARMA($p, p-1$) processes are calibrated to data, for example, it holds for all the applications of CARMA($p, p-1$) processes discussed in the introduction in which the calibrated parameters are reported, that is, for [13, 17, 22, 36]. Additionally, it also holds for the parameters in (4.1).

Remark 4.10. The distributional result in Theorem 4.8 is closely related to generalized gamma convolutions albeit slightly different since the gamma distributed variables are dependent on m. A distribution μ is a generalized gamma convolution if there exists a sequence of independent gamma distributed random variables $(G_i)_{i\in\mathbb{N}}$ such that $\sum_{i=1}^m G_i$ converges to μ in distribution as $m \to \infty$. For more on gamma convolutions see for example [11].

Remark 4.11. From (4.11) and (4.12) in the proof of Theorem 4.8 we have available an explicit expression of shapes and rates of gamma distributions where the sum approximates the stationary distribution of the CARMA($p, p-1$). In practice however, it may be more natural to find coefficients that minimize

$$\sum_{n=1}^{M} |\eta + \sum_{i=1}^{m} \frac{\eta_{i,m}}{\theta_{i,m}^n} - a_n|^2.$$

for some $M \in \mathbb{N}$, where $(a_n)_{n\in\mathbb{N}}$ is given by (4.7). This works well for small m, but we quickly face an identification problem for large m. In such cases, the explicit expression of the approximating coefficients have been found to work well.

We now investigate numerically the approximating sequence of the gamma distribution given in Theorem 4.8. We will consider two CARMA(2, 1) processes, the first with

$$\alpha_1 = -0.5, \quad \alpha_2 = -1, \quad \text{and} \quad \beta_1 = -0.75,$$

and the second with

$$\alpha_1 = -0.5, \quad \alpha_2 = -2, \quad \text{and} \quad \beta_1 = -0.75,$$

where α_1 and α_2 are the roots of P and β_1 is the root of Q. We will also consider a CARMA(3, 2) process with

$$\alpha_1 = -0.5, \quad \alpha_2 = -1, \quad \alpha_3 = -1.2, \quad \beta_1 = -0.75, \quad \text{and} \quad \beta_2 = -1.1,$$

where α_1, α_2, and α_3 are the roots of P, and β_1 and β_2 are the roots of Q. The compound Poisson process has intensity $\mu = 1$ and the exponential jumps have parameter $\kappa = 1$ in all three CARMA models. A simulation of the stationary distributions together with the approximating sum of gamma distributions are given in Figure 3, 4, and 5. We have also plotted the gamma distribution with a probability density function (pdf) that minimizes the squared error to the empirical pdf for comparison. The shape and rate in the approximating gamma distribution are calculated using the approximation in (4.11) and (4.12) with $\Delta = 1$ and where we truncate the sum at $i = 3$, resulting in a sum of 4 gamma distributed variables that approximate the distribution of the CARMA process. The stationary distribution of the CARMA processes is simulated by discretizing the state-space equation over an interval of length $1,000,050$. We simulate on a grid with grid-size 1/10 but only consider every tenth simulated value, and we disregard the first 50 observations to approximately be in the stationary distribution. The simulation of the stationary distribution is therefore based on $1,000,000$ observations.

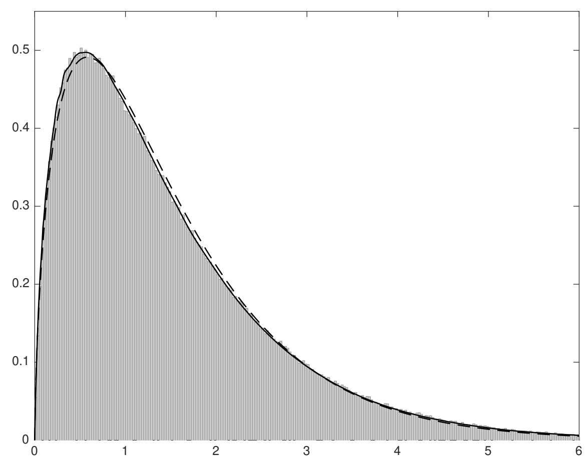


Figure 3: Histogram for simulated stationary distribution of a CARMA(2,1) process in grey, pdf of sum of 4 independent gamma distributed variables (solid line), and pdf of the gamma distribution minimizing the squared error to the histogram (dotted line). Here, $\alpha_2 = -1$.

We see that the sum of independent gamma distributed variables approximates the stationary distribution of the CARMA process well and that there is a significantly better fit than the gamma distribution.

Acknowledgements

We would like to thank Alexandre Brouste for providing access to the wind data analyzed in this paper. FEB gratefully acknowledges the financial support from FINEWSTOCH, funded by the Norwegian Research Council. VR gratefully acknowledges the financial support from the Danish Council for Independent Research, grant DFF-4002-00003. We would also like to thank two referees for many valuable suggestions and comments.

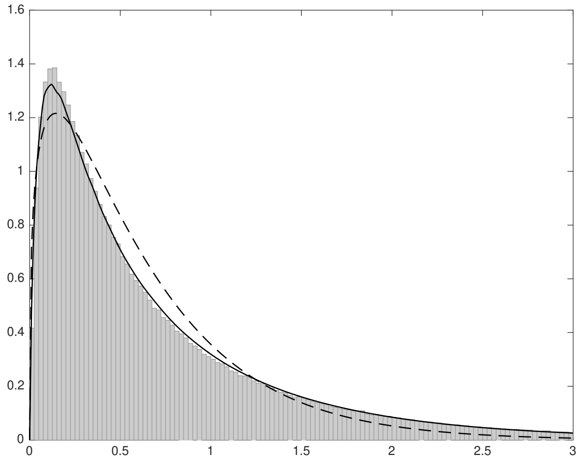


Figure 4: Histogram for simulated stationary distribution of a CARMA(2, 1) process in grey, pdf of sum of 4 independent gamma distributed variables (solid line), and pdf of the gamma distribution minimizing the squared error to the histogram (dotted line). Here, $\alpha_2 = -2$.

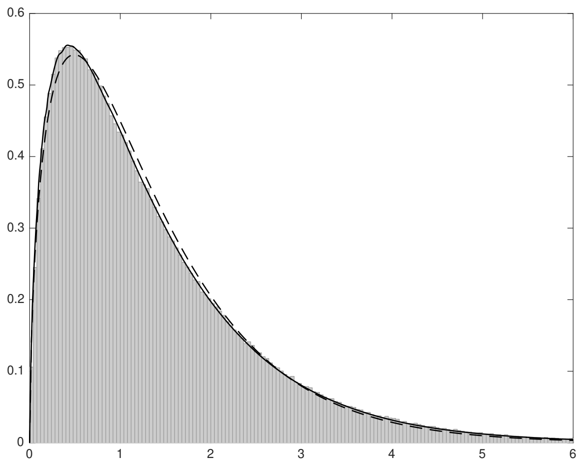


Figure 5: Histogram for simulated stationary distribution of a CARMA(3, 2) process in grey, pdf of sum of 4 independent gamma distributed variables (solid line), and pdf of the gamma distribution minimizing the squared error to the histogram (dotted line).

References

[1] Andresen, A., F. E. Benth, S. Koekebakker and V. Zakamulin (2014). The CARMA interest rate model. *International Journal of Theoretical and Applied Finance* 17(02), 1450008.

[2] Barndorff-Nielsen, O. E. and A. Basse-O'Connor (2011). Quasi Ornstein–Uhlenbeck processes. *Bernoulli* 17(3), 916–941. DOI: 10.3150/10-BEJ311.

[3] Barndorff-Nielsen, O.E. and N. Shephard (2001). Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial economics. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 63(2), 167–241.

[4] Basse-O'Connor, A., M.S. Nielsen, J. Pedersen and V. Rohde (2018). Multivariate stochastic delay differential equations and CAR representations of CARMA processes. *arXiv preprint arXiv:1801.04091*.

[5] Bensoussan, A. and A. Brouste (2016). Cox–Ingersoll–Ross model for wind speed modeling and forecasting. *Wind Energy* 19(7), 1355–1365.

[6] Benth, F.E., C. Klüppelberg, G. Müller and L. Vos (2014). Futures pricing in electricity markets based on stable CARMA spot models. *Energy Economics* 44, 392–406.

[7] Benth, F.E. and A. Pircalabu (2018). A non-Gaussian Ornstein–Uhlenbeck model for pricing wind power futures. *Applied Mathematical Finance* 25(1), 36–65.

[8] Benth, F.E., J. Šaltytė-Benth and S. Koekebakker (2007). Putting a price on temperature. *Scand. J. Statist.* 34(4), 746–767. DOI: 10.1111/j.1467-9469.2007.00564.x.

[9] Benth, F.E. and A. Suess (2017). Continuous-time autoregressive moving-average processes in Hilbert space. *To appear in Abel Proceedings 2016, Springer Verlag*.

[10] Billingsley, P. (1999). *Convergence of Probability Measures*. Second. Wiley Series in Probability and Statistics: Probability and Statistics. A Wiley-Interscience Publication. New York: John Wiley & Sons Inc., x+277.

[11] Bondesson, L. (2012). *Generalized gamma convolutions and related classes of distributions and densities*. Vol. 76. Springer Science & Business Media.

[12] Brockwell, P.J. (2014). Recent results in the theory and applications of CARMA processes. *Ann. Inst. Statist. Math.* 66(4), 647–685. DOI: 10.1007/s10463-014-0468-7.

[13] Brockwell, P.J., R.A. Davis and Y. Yang (2011). Estimation for non-negative Lévy-driven CARMA processes. *J. Bus. Econom. Statist.* 29(2), 250–259. DOI: 10.1198/jbes.2010.08165.

[14] Brockwell, P.J. and A. Lindner (2009). Existence and uniqueness of stationary Lévy-driven CARMA processes. *Stochastic Process. Appl.* 119(8), 2660–2681. DOI: 10.1016/j.spa.2009.01.006.

[15] Brockwell, P.J. and A. Lindner (2015). Prediction of Lévy-driven CARMA processes. *J. Econometrics* 189(2), 263–271.

[16] Brockwell, P.J. and T. Marquardt (2005). Lévy-driven and fractionally integrated ARMA processes with continuous time parameter. *Statist. Sinica* 15(2), 477–494.

[17] Brockwell, P. and A. Lindner (2013). Integration of CARMA processes and spot volatility modelling. *Journal of Time Series Analysis* 34(2), 156–167.

[18] Bru, M.-F. (1991). Wishart processes. *Journal of Theoretical Probability* 4(4), 725–751.

[19] Cox, J.C., J.E. Ingersoll Jr and S.A. Ross (1985). A theory of the term structure of interest rates. *Econometrica* 53(2), 385–408.

[20] Cuchiero, C., D. Filipović, E. Mayerhofer and J. Teichmann (2011). Affine processes on positive semidefinite matrices. *The Annals of Applied Probability* 21(2), 397–463.

[21] Filipovic, D. (2009). *Term-Structure Models. A Graduate Course.* Springer.

[22] García, I., C. Klüppelberg and G. Müller (2011). Estimation of stable CARMA models with an application to electricity spot prices. *Stat. Model.* 11(5), 447–470. DOI: 10.1177/1471082X1001100504.

[23] Gouriéroux, C. (2006). Continuous time Wishart process for stochastic risk. *Econometric Reviews* 25(2-3), 177–217.

[24] Heston, S.L. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *The review of financial studies* 6(2), 327–343.

[25] Isserlis, L. (1918). On a formula for the product-moment coefficient of any order of a normal frequency distribution in any number of variables. *Biometrika* 12(1/2), 134–139.

[26] Jeanblanc, M., M. Yor and M. Chesney (2009). *Mathematical methods for financial markets*. Springer Science & Business Media.

[27] Karatzas, I. and S.E. Shreve (1991). *Brownian motion and stochastic calculus*. Second. Vol. 113. Graduate Texts in Mathematics. New York: Springer-Verlag, xxiv+470.

[28] Kučera, V. (1973). A review of the matrix Riccati equation. *Kybernetika* 9(1), 42–61.

[29] Levin, J. (1959). On the matrix Riccati equation. *Proceedings of the American Mathematical Society* 10(4), 519–524.

[30] Marquardt, T. and R. Stelzer (2007). Multivariate CARMA processes. *Stochastic Process. Appl.* 117(1), 96–120. DOI: 10.1016/j.spa.2006.05.014.

[31] Mayerhofer, E., O. Pfaffel and R. Stelzer (2011). On strong solutions for positive definite jump diffusions. *Stochastic Processes and Their Applications* 121(9), 2072–2086.

[32] Øksendal, B. (2003). *Stochastic Differential Equations*. Springer.

[33] Rajput, B.S. and J. Rosiński (1989). Spectral representations of infinitely divisible processes. *Probab. Theory Related Fields* 82(3), 451–487.

[34] Rogers, L.C.G. and D. Williams (2000). *Diffusions, Markov processes, and martingales. Vol. 1*. Cambridge Mathematical Library. Foundations, Reprint of the second (1994) edition. Cambridge: Cambridge University Press, xx+386.

[35] Stein, E.M and J.C. Stein (1991). Stock price distributions with stochastic volatility: an analytic approach. *The Review of Financial Studies* 4(4), 727–752.

[36] Todorov, V. (2009). Estimation of continuous-time stochastic volatility models with jumps using high-frequency data. *J. Econometrics* 148(2), 131–148.

[37] Tsai, H. and K. Chan (2005). A note on non-negative continuous time processes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 67(4), 589–597.

[38] Vasicek, O. (1977). An equilibrium characterization of the term structure. *Journal of Financial Economics* 5(2), 177–188.

PAPER H

Multivariate Continuous-Time Modeling Of Wind Indexes And Hedging Of Wind Risk

Fred E. Benth, Troels S. Christensen, and Victor Rohde

Abstract

With the introduction of the exchange-traded German wind power futures, opportunities for German wind power producers to hedge their volumetric risk are present. We propose two continuous-time multivariate models for the wind power utilization at different wind sites, and discuss the properties and estimation procedures for the models. Employing the models to wind index data for wind sites in Germany and the underlying wind index of exchange-traded wind power futures contracts, the estimation results of both models suggest that they capture key statistical features of the data. We argue how these models can be used to find optimal hedging strategies using exchange-traded wind power futures for the owner of a portfolio of so-called tailor-made wind power futures. Both in-sample and out-of-sample hedging scenarios are considered, and, in both cases, significant variance reductions are achieved. Additionally, the risk premium of the German wind power futures is analysed, leading to an indication of the risk premium of tailor-made wind power futures.

MSC 2010: 60G10, 60G22, 60H10, 60H20

Keywords: multivariate Ornstein-Uhlenbeck process; wind power futures; hedg- ing; risk premium

1 Introduction

In the power market, producers in general face market risk in the sense of uncertainty of the prices at which they can sell their generated power. The intermittent nature of renewable energy sources such as wind and photo voltaic power production adds yet another layer of risk, known as volumetric risk in the sense that the produced amount of electricity is uncertain due to the dependence on weather. Globally, so-called power purchase agreements and subsidies from governments have minimized the market risk for renewable power producers. In contrast, the volumetric risk has only recently been addressed in Germany—and only for wind power producers (WPPs)—by the introduction of the exchange-traded wind power futures (WPF) contracts. The underlying of a WPF contract is an index between zero and one representing the overall utilization of the installed German wind power production. By taking an appropriate position in WPF contracts, the lost income of the WPPs implied by low wind scenarios is (partially) offset by the position in WPF contracts, hence minimizing the volumetric risk. Due to the prioritization of the cheapest power producers in Germany, the opposite part of the WPF market is typically conventional power producers (CPPs) such as gas-fired power plants. By taking an appropriate position in exchange-traded WPF contracts, CPPs can hedge their exposure to the cheap electricity generated by WPPs.

The WPF market is considered in detail in [10], where the authors propose an equilibrium pricing model. They find that the willingness to engage in the WPF market is greater for the WPPs compared to CPPs. In other words, the hedging benefits of the WPF contracts are greater for WPPs than CPPs. This is supported by the results in [8] who employ an ARMA-GARCH copula framework to the joint modelling of one site-specific wind index and the underlying WPF index. In [5] modeling of the underlying WPF index is considered and closed-form formulas for the WPF price and the price of European options written on the WPF index are derived.

Continuous-time modeling using univariate Ornstein-Uhlenbeck (OU) processes driven by non-decreasing Lévy processes, like the compound Poisson process with exponential jumps, have been studied extensively, and used to model, for example, stochastic volatility of financial assets, wind, electicity prices, and temperature (see [3, 7, 4, 5]). A detailed treatment of Lévy processes can be found in [15]. The multivariate modeling of more than two stochastic processes using multidimensional non-Gaussian Lévy processes is, however, more limited. Here we mention the work of [12] and [16] that introduce the multivariate construction by subordination of Brownian motions, and the work of [2] using linear transformations of Lévy processes.

Our contribution to the literature is twofold. Firstly, we propose a joint model for the simultaneous behaviour of wind indexes that allows for a parsimonious representation of the correlation structure. This model can be seen as the multivariate version of the model presented in [5]. As a consequence of the scarce literature on such multivariate models, we propose an alternative model for comparison reasons. Secondly, we suggest the idea of so-called tailor-made WPF contracts to eliminate the volumetric risk of WPPs completely. Employing our proposed joint model of wind indexes, we investigate the hedging benefits of exchange-traded WPF contracts for a owner of a portfolio of tailor-made WPF contracts, and comment on the risk premium of tailor-made WPF contracts. We show that this construction is beneficial for both parties of the tailor-made WPF contracts.

The rest of the paper is organized as follows. Sec. 2 presents the data of wind indexes we later analyze in greater detail and also serves as motivation for the proposed model. In Sec. 3 we present models for the joint behaviour of wind indexes and corresponding estimation procedures. In Sec. 4 we present the estimation results of the two models. Sec. 5 discusses the hedging of wind power production using WPF contracts implied by the proposed models. Lastly, Sec. 6 concludes.

2 Data presentation

The empirical observation period spans from 1 July 2016 to 30 June 2019, which corresponds to 1095 daily observations for each considered wind index. The data set consists of

1. A daily wind index at three wind sites provided by Centrica Energy Trading. The wind index at wind site i is calculated by

$$\frac{Q_i(t)}{h(t)C_i},$$

   where $h(t)$ is the number of hours in day t, $Q_i(t)$ is the power production at day t at site i, and C_i is the installed capacity at site i. Figure 1 shows the approximate geographical locations for the three wind sites.[1]

2. A daily German wind index provided by Nasdaq, representing the German utilization of wind power plants. The acronym used for it is NAREX WIDE (NAsdaq REnewable indeX WInd DE (Germany)) and is used as the underlying for WPF contracts traded on Nasdaq. We will simply denote it as the German wind index in the remaining part of the paper.

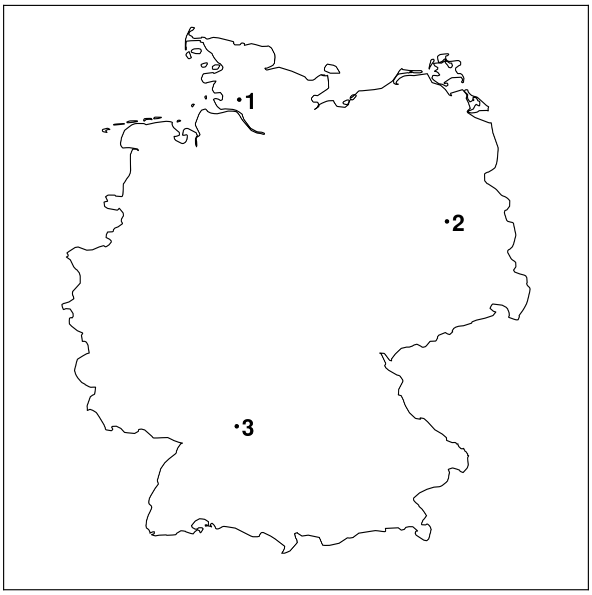


Figure 1: Locations of wind sites with site ID in Germany.

The wind indexes are bounded between zero and one. Fig. 2 shows all four wind indexes, and the corresponding autocorrelation function for each wind index. In all four cases, the wind index displays a pronounced yearly cycle consistent with the

[1] The locations are approximate due to confidentially issues.

observations made in [5] and [8]. Since the German wind index is by construction made up of all wind power production in Germany, the behaviour of the German wind index is less extreme compared to the individual wind indexes. To concretize, a value of zero for the German wind index is not observed in our observation period, whereas it is for all three wind sites. Also the maximum attained value for each site wind index is higher than the maximum value of the German index; however, it does not reach one in any of the cases.

3 Model description

3.1 General model considerations

Let n denote the number of wind sites and $P_i(t)$ the ith wind index. We assume that the ith wind index can be described by

$$P_i(t) = 1 - e^{-S_i(t)X_i(t)}, \quad i = 1, \dots, n, \tag{3.1}$$

where $S_i(t) : \mathbb{R} \to \mathbb{R}^+$ is a deterministic function intended to filter out potential seasonal effects, and $X_i(t)$ is a mean-reverting stochastic process satisfying $X_i(t) \geq 0$ for all t. The intention of $X(t) = (X_1(t), ..., X_n(t))^\top$ is to capture the short-term uncertainty and the dependence between the n wind indexes. By this specification we are ensured that $P_i(t) \in [0, 1)$.

The proposed model in Eq. (3.1) distinguishes itself from the specification in [5], where the natural extension of their univariate setup to the present multivariate setup would be

$$P_i(t) = S_i(t)e^{-X_i(t)}. \tag{3.2}$$

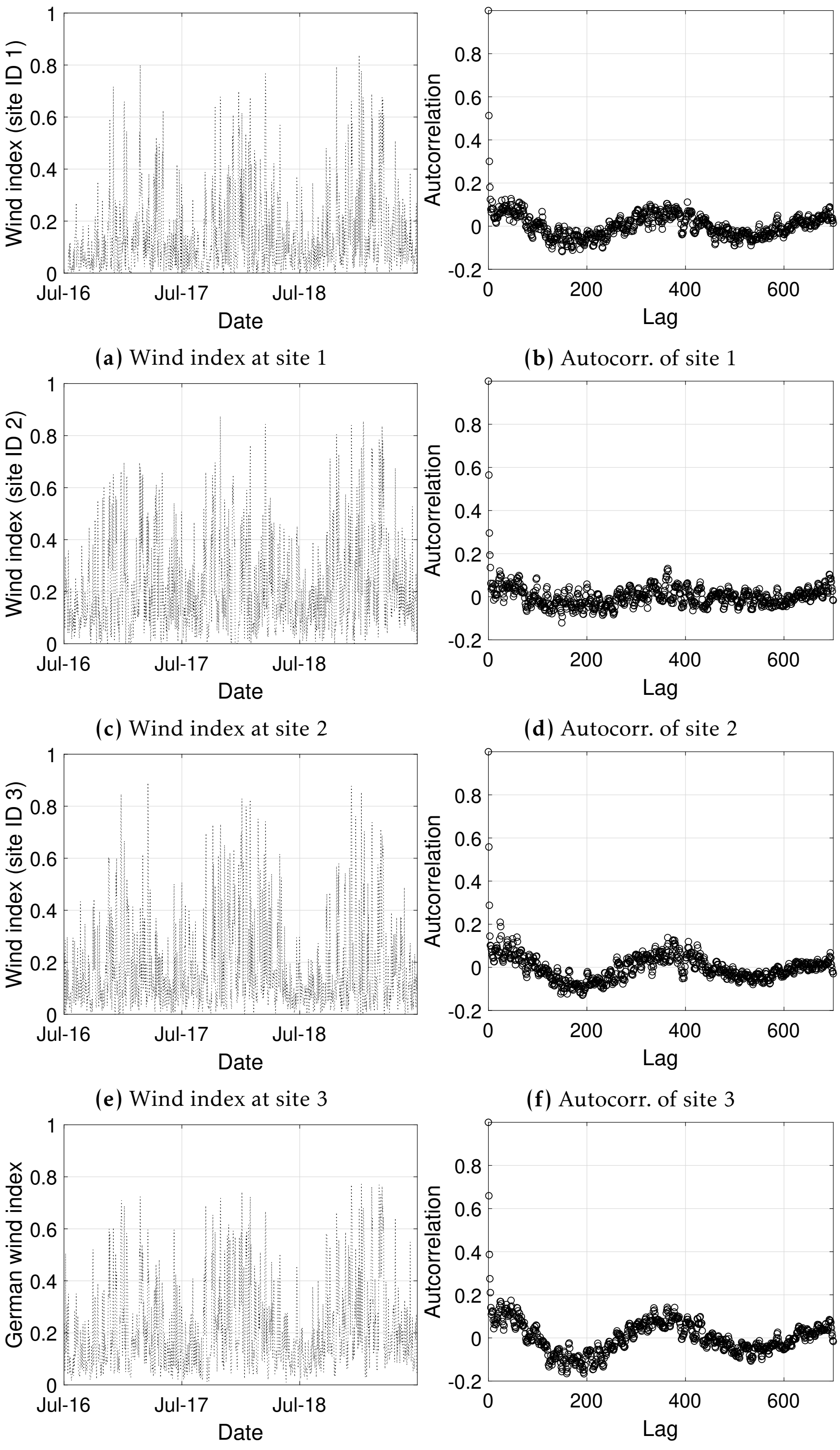
with appropriate choices of $X_i(t)$ and $S_i(t)$. We do, however, prefer Eq. (3.1) over Eq. (3.2). Due to our specification with regard to the deterministic part $S_i(t)$ of the model, we do not face any potential model inconsistencies as is the case of Eq. (3.2). We refer the interested reader to [5] more information and discussion. Additionally, as discussed in Sec. 2, the wind index at a given site can attain a value of zero, whereas, on the other hand, we have not observed full utilization of the capacity at a single wind site. Since [5] consider the German wind index separately, which is never zero or one due to the construction of it, Eq. (3.2) is applicable without any modifications. Lastly, Eq. (3.1) implies that increased values of $S_i(t)$ and $X_i(t)$ translate to an increased value of $P_i(t)$, which is more intuitively appealing.

Moving on to the seasonal components of the model, we include the following yearly seasonality motivated by the observations made in Fig. 2,

$$S_i(t) = a_i + b_i \sin(2\pi t/365) + c_i \cos(2\pi t/365), \quad i = 1, \dots, n,$$

where $a_i, b_i, c_i \in \mathbb{R}$. With N being the number of observations, the coefficients are determined by

$$\min_{a_i, b_i, c_i} \sum_{t=1}^{N} \left[-\log(1 - P_i(t)) - S_i(t) \right]^2, \quad i = 1, \dots, n.$$



(a) Wind index at site 1 **(b)** Autocorr. of site 1

(c) Wind index at site 2 **(d)** Autocorr. of site 2

(e) Wind index at site 3 **(f)** Autocorr. of site 3

(g) German wind index **(h)** Autocorr. of German

Figure 2: All four wind indexes with corresponding empirical auto-correlation function.

Having obtained the estimated seasonal functions, the observed values of $X_i(t)$ implied by the estimated seasonal function $\hat{S}_i(t)$ can then be calculated as

$$X_i(t) = \frac{-\log(1 - P_i(t))}{\hat{S}_i(t)}. \tag{3.3}$$

We will in the following discuss two approaches for modeling $X_i(t)$.

3.2 A gamma model

In this section a multivariate model for $n-1$ wind indexes and the German wind index is discussed, which we refer to as the gamma model in the sequel. In Sec. 4 we will consider the case $n = 4$. We start by introducing the noise process. In particular, we say a Lévy process L is a compound Poisson process with exponential jumps and parameters $\alpha > 0$ and $\beta > 0$ if

$$L(t) = \sum_{i=1}^{N(t)} J_i$$

where $(N(t))_{t\in\mathbb{R}}$ is a Poisson process with intensity α and J_i, $i \in \mathbb{N}$, are independent exponentially distributed random variables with parameter β. We say a random variable has an exponential distribution with parameter β if it has density $x \mapsto \mathbb{1}_{[0,\infty)}(x)\beta e^{-\beta x}$.

The gamma model assumes X is a multidimensional Lévy-driven Ornstein-Uhlenbeck (OU) process,

$$dX(t) = -\Lambda X(t)dt + \Sigma_L dL(t). \tag{3.4}$$

Here, L is an n-dimensional Lévy process where the i'th entry is an independent compound Poisson process with exponential jumps, variance equal to one, and parameters α_i and β_i for $i = 1,\dots,n$. Furthermore, Λ is a diagonal matrix, $\mathrm{diag}(\lambda_1,\dots,\lambda_n)$ with $\lambda_i > 0$ for $i = 1,\dots,n$. We assume Σ_L is given by

$$\Sigma_L = \begin{pmatrix} \sigma_{1,1} & 0 & \dots & 0 & \sigma_{1,n} \\ 0 & \sigma_{2,2} & \dots & 0 & \sigma_{2,n} \\ \vdots & \vdots & \ddots & 0 & \vdots \\ 0 & 0 & 0 & \sigma_{n-1,n-1} & \sigma_{n-1,n} \\ 0 & 0 & 0 & 0 & \sigma_{n,n} \end{pmatrix} \tag{3.5}$$

and that all entries of Σ_L are non-negative. Due to the form of Σ_L, each individual wind index has an idiosyncratic risk associated to it through one of the first $n-1$ compound Poisson process $L_1,\dots,L_{n-1}$. Furthermore, all sites and the German index share a systematic risk through the n'th compound Poisson process L_n. A similar construction is also considered in [2] where a multivariate model is proposed for modeling financial products written on more than one underlying asset.

Distribution of $P_n(t)$ in the gamma model

Since X_n, the process associated with the German wind index, is an OU process driven by one compound Poisson process with exponential jumps, it has a gamma distribution as its stationary distribution. Thus, we conclude the following:

Proposition 3.1. *The stationary distribution of $P_n(t)$ in the gamma model has density*

$$f_{P_n(t)}(x) = \frac{(-\log(1-x))^{\alpha_n - 1}(1-x)^{\beta_n/S_n(t)-1}}{S_n(t)^{\alpha_n}} \qquad x \in (0,1) \tag{3.6}$$

Proof. This is a direct consequence of $X_n(t)$ being gamma distributed with shape α_n and rate β_n (see for example [3]). □

In Figure 3 the density of P_n is depicted for different α_n and β_n (with $S_n(t) = 1$). We see that the densities implied by the gamma model are rather flexible and able to cover both low and high utilization scenarios.

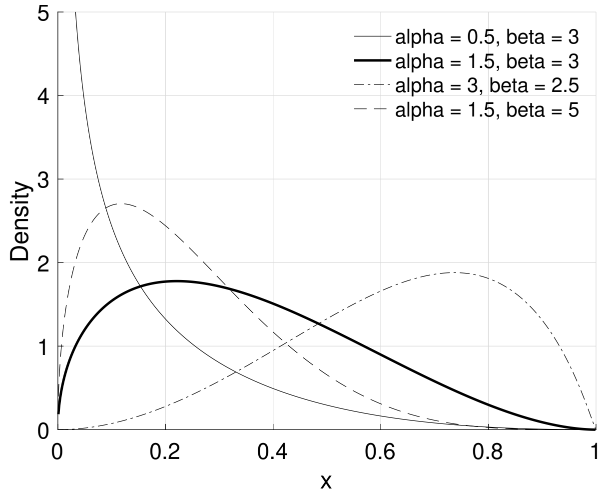


Figure 3: Different variations of the density in (3.6)

The processes $X_1, \ldots, X_{n-1}$ are sums of two independent gamma distributions, and thus, there does not, in general, exist simple expressions for the densities of the individual site index similar to the one for the German index stated in Prop. 3.1.

Covariance between wind indexes in the gamma model

We now give (semi-)analytical expressions of the covariances implied by the gamma model, which will be useful for fast calculation of the minimum variance hedges discussed in Section 5. The integral in Eq. (3.7) is the only non-analytical part of the expression, but we argue in Remark 7.3 that this integral is small and can be coarsely approximated without significant effect. We thereby maintain fast computation time.

Before we state the result, let us introduce some notation to help making a concise statement. To this end define the $n \times (n+1)$ matrix $\tilde{\Sigma}_L$ by

$$\tilde{\Sigma}_L = \begin{pmatrix} \sigma_{1,1} & 0 & \ldots & 0 & 0 & \sigma_{1,n} \\ 0 & \sigma_{2,2} & \ldots & 0 & 0 & \sigma_{2,n} \\ \vdots & \vdots & \ddots & 0 & 0 & \vdots \\ 0 & 0 & 0 & \sigma_{n-1,n-1} & 0 & \sigma_{n-1,n} \\ 0 & 0 & 0 & 0 & 0 & \sigma_{n,n} \end{pmatrix}$$

where $\sigma_{i,j}$ is the (i,j)'th entry of Σ_L. Let $\tilde{\sigma}_{i,j}$ denote the (i,j)'th entry of $\tilde{\Sigma}_L$. Furthermore, define $\tilde{\alpha}, \tilde{\beta} \in \mathbb{R}^{n+1}$ by

$$\tilde{\alpha} = (\alpha_1, \dots, \alpha_{n-1}, 0, \alpha_n)^\top \quad \text{and} \quad \tilde{\beta} = (\beta_1, \dots, \beta_{n-1}, 1, \beta_n)^\top,$$

and denote the i'th entry of $\tilde{\alpha}$ and $\tilde{\beta}$ by $\tilde{\alpha}_i$ and $\tilde{\beta}_i$. We now give the expressions of the covariances of the gamma model. The proof is relegated to Section 7.

Proposition 3.2. *Consider $s \le t$ and define*

$$f_{i,j}(u) = S_i(t)\tilde{\sigma}_{i,n+1}e^{-\lambda_i(t-s+u)} + S_j(s)\tilde{\sigma}_{j,n+1}e^{-\lambda_j u}, \quad i,j = 1,\dots,n.$$

Then

$$\begin{aligned}
&\operatorname{cov}(P_i(t), P_j(s)) \\
&= \left(\frac{\tilde{\beta}_i}{\tilde{\beta}_i + \tilde{\sigma}_{i,i}S_i(t)}\right)^{\tilde{\alpha}_i/\lambda_i} \left(\frac{\tilde{\beta}_{n+1} + \tilde{\sigma}_{i,n+1}S_i(t)e^{-\lambda_i(t-s)}}{\tilde{\beta}_{n+1} + \tilde{\sigma}_{i,n+1}S_i(t)}\right)^{\tilde{\alpha}_{n+1}/\lambda_i} \\
&\times \left(\frac{\tilde{\beta}_j}{\tilde{\beta}_j + \tilde{\sigma}_{j,j}S_j(s)}\right)^{\tilde{\alpha}_j/\lambda_j} \left[\left(1 + \frac{f_{i,j}(0)}{\tilde{\beta}_{n+1}}\right)^{\tilde{\alpha}_{n+1}f_{i,j}(0)/f'_{i,j}(0)}\right. \\
&\times \exp\left\{\tilde{\alpha}_{n+1}\int_0^\infty \frac{d}{du}\left(\frac{f_{i,j}(u)}{\frac{d}{du}f_{i,j}(u)}\right)\log\left(1 + \frac{f_{i,j}(u)}{\tilde{\beta}_{n+1}}\right)du\right\} \\
&\left. - \left(\frac{\tilde{\beta}_{n+1}}{\tilde{\beta}_{n+1} + \tilde{\sigma}_{i,n+1}S_4(t)e^{-\lambda_i(t-s)}}\right)^{\tilde{\alpha}_{n+1}/\lambda_i} \left(\frac{\tilde{\beta}_{n+1}}{\tilde{\beta}_{n+1} + \tilde{\sigma}_{j,n+1}S_j(s)}\right)^{\tilde{\alpha}_{n+1}/\lambda_j}\right]
\end{aligned} \tag{3.7}$$

for $i,j = 1,\dots,n$, $i \ne j$, and

$$\begin{aligned}
&\operatorname{cov}(P_i(t), P_i(s)) \\
&= \left(\frac{\tilde{\beta}i + \tilde{\sigma}_{i,i}S_i(t)e^{-\lambda_i(t-s)}}{\tilde{\beta}_i + \tilde{\sigma}_{i,i}S_i(t)}\right)^{\tilde{\alpha}_i/\lambda_i} \left(\frac{\tilde{\beta}_{n+1} + \tilde{\sigma}_{i,n+1}S_i(t)e^{-\lambda_i(t-s)}}{\tilde{\beta}_{n+1} + \tilde{\sigma}_{i,n+1}S_i(t)}\right)^{\tilde{\alpha}_{n+1}/\lambda_i} \\
&\times \left[\left(\frac{\tilde{\beta}_i}{\tilde{\beta}_i + \left(S_i(t)e^{-\lambda_i(t-s)} + S_i(s)\right)\tilde{\sigma}_{i,i}}\right)^{\tilde{\alpha}_i/\lambda_i}\right. \\
&\times \left(\frac{\tilde{\beta}_{n+1}}{\tilde{\beta}_{n+1} + \left(S_i(t)e^{-\lambda_i(t-s)} + S_i(s)\right)\tilde{\sigma}_{i,n+1}}\right)^{\tilde{\alpha}_{n+1}/\lambda_i} \\
&- \left(\frac{\tilde{\beta}_i^2}{\left(\tilde{\beta}_i + \tilde{\sigma}_{i,i}S_i(t)e^{-\lambda_i(t-s)}\right)\left(\tilde{\beta}_i + \tilde{\sigma}_{i,i}S_i(s)\right)}\right)^{\tilde{\alpha}_i/\lambda_i} \\
&\left. \times \left(\frac{\tilde{\beta}_{n+1}^2}{\left(\tilde{\beta}_{n+1} + \tilde{\sigma}_{i,n+1}S_i(t)e^{-\lambda_i(t-s)}\right)\left(\tilde{\beta}_{n+1} + \tilde{\sigma}_{i,n+1}S_i(s)\right)}\right)^{\tilde{\alpha}_{n+1}/\lambda_i}\right]
\end{aligned} \tag{3.8}$$

for $i = 1,\dots,n$.

Identification of parameters in the gamma model

Let Λ_{var} be the $n \times n$ matrix given by

$$(\Lambda_{var})_{i,j} = \frac{1}{\lambda_i + \lambda_j}.$$

Furthermore, denote by $\circ$ the Hadamard product. Then the following result will be used to estimate the parameters of the gamma model. Again, we relegate the proof of Proposition 3.3 to Section 7.

Proposition 3.3. *The mean of X is*

$$\mathbb{E}[X(0)] = \Lambda^{-1}\Sigma_L\beta/2 \tag{3.9}$$

and the auto-covariance of X is

$$\operatorname{cov}(X(0), X(t)) = \left(\Lambda_{var} \circ \left(\Sigma_L\Sigma_L^\top\right)\right)e^{-\Lambda t} \tag{3.10}$$

for $t \geq 0$.

The parameters of the gamma model will be estimated in three steps. First, the mean-reversion matrix Λ will be fitted to the empirical auto-correlation function based on the first 25 lags. From (3.10), it follows that the model auto-correlation function of X_i is $t \mapsto e^{-\lambda_i t}$. To find $\hat{\lambda}_i$, the estimate of λ_i, we therefore minimize

$$\sum_{t=1}^{25}\left(\hat{\rho}_i(t) - e^{-\hat{\lambda}_i t}\right)^2$$

such that $\hat{\lambda}_i > 0$ for $i = 1, \ldots, n$, where $\hat{\rho}_i(t)$ is the empirical auto-correlation function of X_i.

Next, $\hat{\Sigma}_L$ is chosen such that the model matches the empirical covariances. In particular, we choose $\hat{\Sigma}_L$ to minimize

$$\left\|\hat{\Sigma}_X - \Lambda_{var} \circ \left(\hat{\Sigma}_L\hat{\Sigma}_L^\top\right)\right\|^2$$

where $\hat{\Sigma}_X$ is the sample covariance of X, $\|\cdot\|$ is the Frobenius norm and the minimization is done over matrices $\hat{\Sigma}_L$ with non-negative entries of the form in (3.5).

Finally, we discuss how the parameters α and β are estimated. First, we choose $\hat{\beta} = (\hat{\beta}_1, \ldots, \hat{\beta}_n)$ to minimize

$$\left\|\hat{\mu}_X - \Lambda^{-1}\hat{\Sigma}_L\hat{\beta}/2\right\|^2$$

such that $\hat{\beta}_i > 0$, where $\hat{\mu}_X$ is the empirical mean of X.

It is not too difficult to show that $\operatorname{var}(L_i(1)) = 2\alpha_i/\beta_i^2$ and, since the compound Poisson processes are assumed to have unit variance, it therefore follows that

$$1 = \operatorname{var}(L_i(1)) = \frac{2\alpha_i}{\beta_i^2}$$

Consequently, we take $\hat{\alpha}_i = \hat{\beta}_i^2/2$.

3.3 A lognormal model

In this section we present a lognormal model relying on different assumptions than the gamma model. We assume that $G(t) := \log X(t)$ can be modelled as a multidimensional Gaussian Ornstein-Uhlenbeck process,

$$dG(t) = -\Upsilon(G(t) - \Theta)dt + \Sigma dB(t), \tag{3.11}$$

where $(B(t))_{t\in\mathbb{R}}$ is an n-dimensional Brownian motion, $\Upsilon \in \mathbb{R}^{n\times n}$ is a diagonal matrix, $\Sigma \in \mathbb{R}^{n\times n}$ is a lower triangular matrix, and $\Theta \in \mathbb{R}^n$.

It is well-known (see e.g. [13]) that the stationary distribution of $G(t)$, when the diagonal elements of Υ all are positive, is normal with mean Θ. The autocovariance of $G(t)$ is given in, for example, [13], and it is the same as for the gamma model found in Prop. 3.3. In particular, we find

$$\Sigma_G(t) := \operatorname{cov}(G(0), G(t))) = \left(\Upsilon_{var} \circ \left(\Sigma\Sigma^\top\right)\right) e^{-\Upsilon t}, \quad t \geq 0. \tag{3.12}$$

Here, Υ_{var} is the $n \times n$ matrix given by

$$(\Upsilon_{var})_{i,j} = \frac{1}{v_i + v_j},$$

where v_i is the i'th entry of Υ, $i = 1, \ldots, n$.

Consequently, the stationary distribution of $X(t)$ is multivariate lognormal with expected value of $X_i(0)$ being

$$\mathbb{E}[X_i(0)] = \exp\left(\Theta_i + \frac{\Sigma_G(0)_{ii}}{2}\right),$$

while the autocovariance is

$$\operatorname{cov}(X_i(0), X_j(t)) = \mathbb{E}[X_i(0)]\mathbb{E}[X_j(0)](e^{\Sigma_G(t)_{ij}} - 1) \tag{3.13}$$

for $i, j = 1, \ldots, n$ (see e.g. [11] for more information on the multivariate lognormal distribution). This implies that the autocorrelation of $X(t)$ is

$$\operatorname{corr}(X_i(0), X_j(t)) = \frac{\exp\left(\Sigma_G(0)_{ij} e^{-t v_j}\right) - 1}{\sqrt{(e^{\Sigma_G(0)_{ii}} - 1)(e^{\Sigma_G(0)_{jj}} - 1)}}, \tag{3.14}$$

for $i, j = 1, \ldots, n$.

Distribution of $P_i(t)$ in the lognormal model

Having the results for $X(t)$ from the previous section in mind, the stationary distribution of $P_i(t)$ follows and is given in Prop. 3.4.

Proposition 3.4. *The stationary distribution of $P_i(t)$ is characterized by the density*

$$f_{P_i(t)}(x) = \frac{-1}{(1-x)\log(1-x)\sqrt{\Sigma_G(0)_{i,i}}}\,\phi\left(\frac{\log\left(-\frac{\log(1-x)}{S_i(t)}\right) - \Theta_i}{\sqrt{\Sigma_G(0)_{i,i}}}\right), \tag{3.15}$$

where $\Sigma_G(0)_{i,i}$ is the i'th element of the diagonal of $\Sigma_G(0)$, Θ_i is the i'th element of Θ, and $\phi(\cdot)$ is the density of the standard normal distribution.

To investigate the density of $P_i(t)$ in more detail, consider for a moment a more generic version of Eq. (3.15), given by

$$f(x|\mu, \sigma) = \frac{-1}{(1-x)\log(1-x)\sigma}\,\phi\left(\frac{\log(-\log(1-x)) - \mu}{\sigma}\right). \tag{3.16}$$

As can be seen in Fig. 4, showing examples of the density given different values of μ and σ in Eq. (3.16), the distribution is rather flexible and capable of attaining quite different forms.

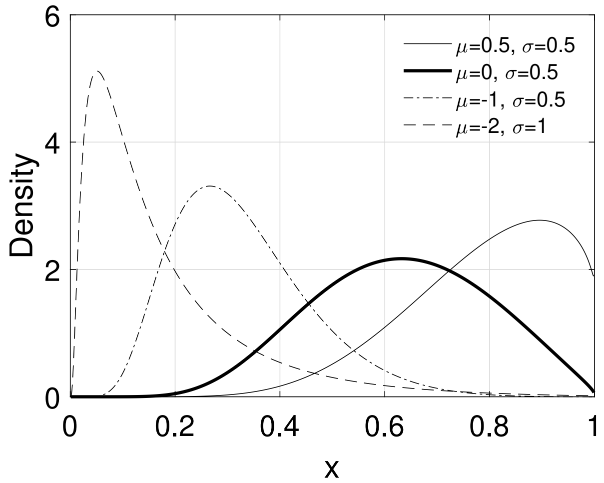


Figure 4: Different variations of the density in Eq. (3.16).

Covariance between wind indexes in the lognormal model

Deriving the covariances between wind indexes in the lognormal model is closely related to the derivation of the Laplace transform of the lognormal distribution. To the best of our knowledge, no closed-form has been derived for the Laplace transform of the lognormal distribution, but there exist approximations, see e.g. [1]. With regard to this paper, we refer the interested reader to [1] and the references herein for further information, and employ numerical integration by exploiting our knowledge of the distribution of $G(t)$ to determine the covariances between the wind indexes.

Identification of parameters in the lognormal model

To identify the parameters of the model, we employ the method of moments as in the gamma model case. We first identify $\Sigma_G(0)$ by exploiting Eq. (3.13),

$$\Sigma_G(0)_{ij} = \log\left(\frac{\hat{\Sigma}_{X,ij}}{\hat{\mu}_i\hat{\mu}_j} + 1\right), \tag{3.17}$$

with $\hat{\mu}_i$ being the empirical mean of $X_i(t)$, and $\hat{\Sigma}_{X,ij}$ is the (i,j)th entry of the empirical covariance between $X_i(0)$ and $X_j(0)$.

Having obtained an estimate of $\Sigma_G(0)$ and remembering the model implied auto-correlation in Eq. (3.14), we identify υ_i by minimizing

$$\sum_{t=1}^{25}\left(\hat{\rho}_i(t) - \frac{\exp\left(\hat{\Sigma}_G(0)_{ii}e^{-t\upsilon_i}\right) - 1}{\sqrt{(e^{\hat{\Sigma}_G(0)_{ii}} - 1)(e^{\hat{\Sigma}_G(0)_{ii}} - 1)}}\right)^2,$$

where $\hat{\rho}_i(t)$ is the empirical autocorrelation function of $X_i(0)$ and $X_i(t)$. Here, as in the gamma model, we use the first 25 lags of the empirical auto-covariance function to estimate λ_i. With $\hat{\Upsilon}$, consisting of the estimated υ_i for $i = 1, \dots, n$ in the diagonal, at hand, we identify $\Sigma\Sigma^\top$ by

$$\Sigma\Sigma^\top = \hat{\Sigma}_G(0) \oslash \hat{\Upsilon}_{var},$$

where $\oslash$ is the Hadamard division defined for two matrices A and B by $A \oslash B = A_{ij}/B_{ij}$. Lastly, we determine Θ by

$$\Theta_i = \log(\mu_i) - \frac{\hat{\Sigma}_G(0)_{ii}}{2}, \quad i = 1, \dots, n. \tag{3.18}$$

3.4 Comparison of the gamma and lognormal model

The covariances between indexes in the gamma model can be calculated fast using Proposition 3.2 to find the optimal hedging strategy (see Sec. 5). The noise in the gamma model also has a compelling interpretation, where an idiosyncratic risk is associated to each site index and a systematic risk is associated to all site indexes and the German index. On the other hand, the lognormal model gives rise to closed-form expressions of the densities of all indexes as opposed to only the German index in the gamma model. The lognormal model is simple in the sense that the underlying process is a Gaussian driven OU process. This makes it possible to do numerical analysis based on Gaussian theory.

Both the gamma and lognormal model have straightforward and fast estimation procedures, making them easy to implement. Furthermore, as we will see in Sec. 4, both models capture the autocorrelation of X_i, the cross-autocorrelations between X_i and X_j, and the stationary distribution of X_i well.

4 Estimation results

In this section we summarize and discuss the estimation results. As a starting point, we consider the fitted seasonal functions. In Table 1 we report the fitted parameters for all four wind indexes.

	$\hat{a}_i$	$\hat{b}_i$	$\hat{c}_i$
Site 1	0.1721	-0.0491	-0.0804
Site 2	0.2848	-0.0405	-0.0956
Site 3	0.2294	-0.0322	-0.1226
German	0.2732	-0.0298	-0.1285

Table 1: Fitted seasonal parameters for the four wind indexes.

4.1 Gamma model

Fig. 5 shows the theoretical autocorrelation implied by the estimated gamma model compared to the empirical autocorrelation. The fit to the empirical autocorrelation is convincing, and it is worth noticing that the cross-autocorrelations match well even though the model has only been estimated to the marginal autocorrelation functions.

In Fig. 6 the histogram of X_i and the model density based on a simulation are shown. We see that the distribution of the data is captured well by the model.

We report the estimated parameters in Table 2. The parameters α_4 and β_4 are considerably larger than α_i and β_i for $i = 1, 2, 3$. This implies that the systematic risk

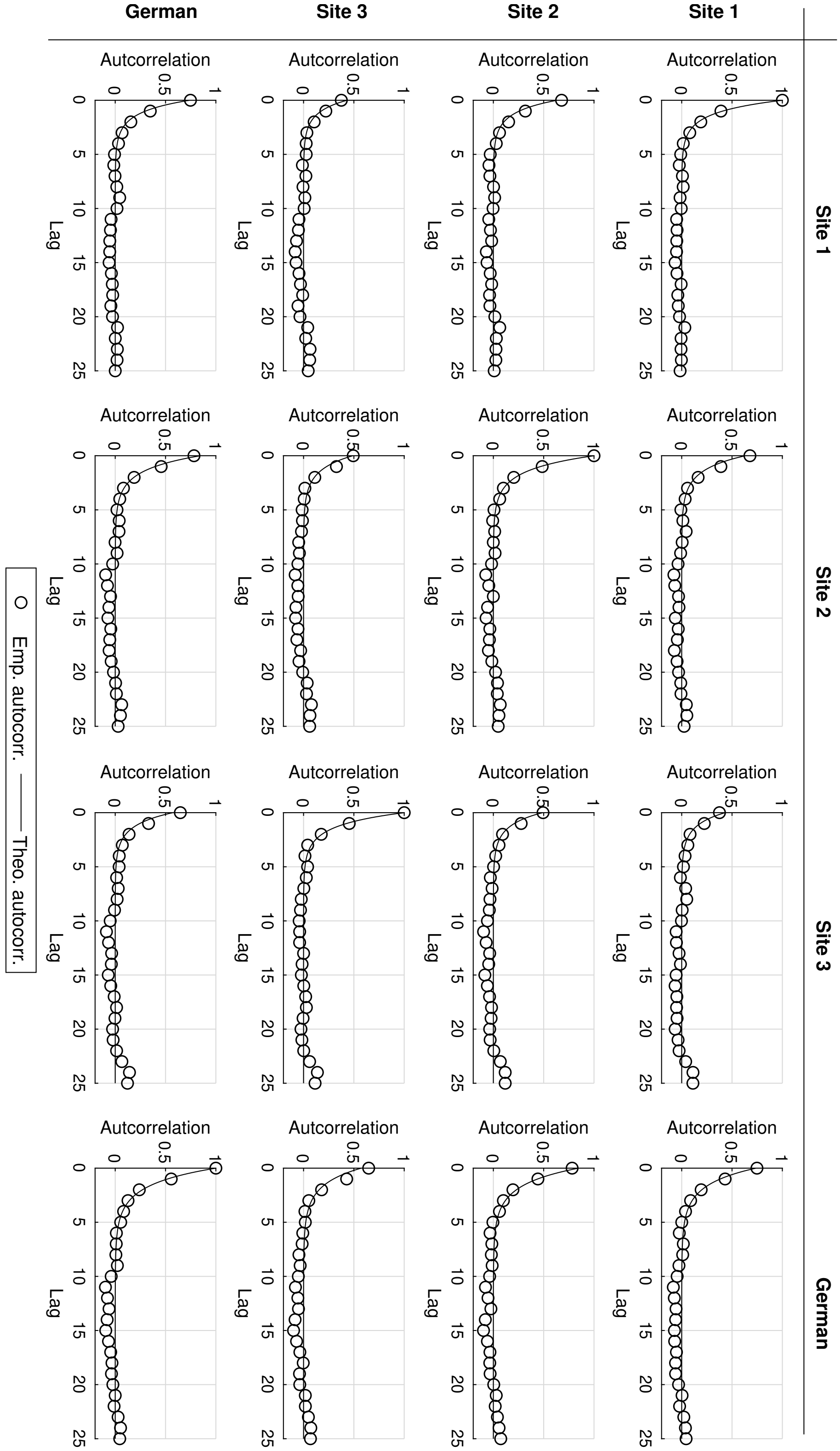
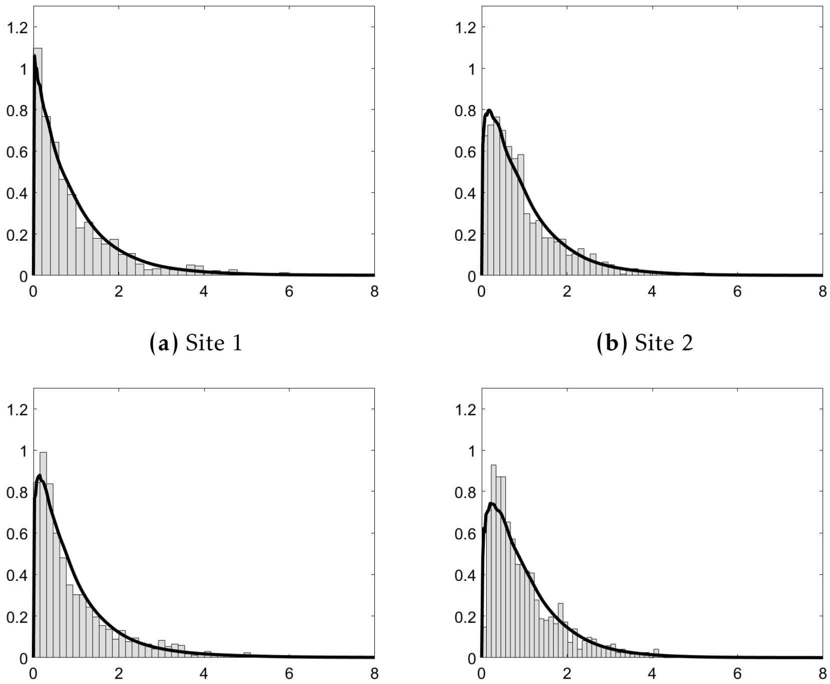


Figure 5: Empirical autocorrelation and theoretical autocorrelation implied by the fitted gamma model. The (i, j)'th panel shows $\mathrm{cor}(X_i(0), X_j(t))$ for $t = 0, 1, \ldots, 25$.



(**a**) Site 1 (**b**) Site 2

(**c**) Site 3 (**d**) German

Figure 6: Histograms of $X_i(t)$ with the fitted densities of the gamma model.

factor L_4 jumps much more frequently than L_i, $i = 1, 2, 3$, but that the jumps of L_4 are relatively small compared to the jumps of L_i, $i = 1, 2, 3$. This aligns well with the intuition that the systematic risk is associated to the wind utilization of the whole of Germany.

	$\hat{\alpha}_i$	$\hat{\beta}_i$	$\hat{\lambda}_i$	$\hat{\sigma}_{i,i}$	$\hat{\sigma}_{i,4}$
Site 1	0.0271	0.2328	0.8977	1.0305	1.1593
Site 2	0.0538	0.3282	0.7589	0.6101	0.9792
Site 3	0.1383	0.5260	0.8513	1.1674	0.8247
German	0.8960	1.3387	0.6539	(-)	0.9781

Table 2: Estimated parameters in the gamma model.

To further assess the model, we report in Table 3 the mean, variance, skewness, and kurtosis of the gamma model along with the empirical and lognormal equivalents for the German wind index[2]. The gamma model captures the first two moments very well as expected from the estimation procedure, where we match the gamma model to the empirical mean and variance. Further, the empirical skewness and kurtosis are also captured very well by the gamma model.

[2]Since the same quantities for the site wind indexes are not relevant in the remaining part of the paper, we have chosen to omit them.

	Mean	Variance	Skewness	Kurtosis
Gamma	1.00	0.73	1.71	7.38
Lognormal	1.00	0.73	3.17	24.98
Empiricial	1.00	0.73	1.67	6.27

Table 3: Mean, variance, skewness, and kurtosis of the German wind index in the gamma and lognormal model together with the empirical values for the German wind index.

4.2 Lognormal model

Fig. 7 shows the theoretical autocorrelation implied by the estimated lognormal model compared to the empirical autocorrelation. As in the gamma model case, the lognormal model captures the autocorrelation and cross-autocorrelation well, in particular taking into account that only the autocorrelation is used to estimate the parameters affecting both the autocorrelations and the cross-autocorrelations.

Fig. 8 shows histograms of the marginal distributions and the fitted lognormal densities. The lognormal model provides overall a decent fit, but seems to capture the distribution of the German wind index better than the site indexes. The estimated Θ and Υ for the lognormal model is reported in Table 4 and the estimated Σ is

$$\hat{\Sigma} = \begin{bmatrix} 1.0987 & 0 & 0 & 0 \\ 0.6763 & 0.5886 & 0 & 0 \\ 0.4902 & 0.3505 & 0.8376 & 0 \\ 0.6381 & 0.2539 & 0.2162 & 0.3035 \end{bmatrix}.$$

Although the speed of mean reversion parameters $\hat{\upsilon}_i$ differ in the lognormal model compared to the gamma model, the same pattern is observed, with the German wind index being the most persistent.

	$\hat{\Theta}_i$	$\hat{\upsilon}_i$
Site 1	-0.4282	0.7080
Site 2	-0.3215	0.6341
Site 3	-0.3803	0.6837
German	-0.2711	0.5607

Table 4: Estimated parameters in the lognormal model.

Returning to Table 3, the lognormal model matches the empirical mean and variance as a results of the estimation procedure, but it does not capture the higher order standardized moments. This indicates that the lognormal model does not capture the whole distribution of the data as well as the gamma model.

5 Hedging wind power production

In the following we denote the German wind index at day t by $P_n(t)$. An exchange-traded WPF contract is written on the underlying daily wind index, $P_n(t)$. The payoff of a long position in such a contract is

$$H(\bar{P}_n(S,T) - P_n(t_0,S,T))X, \tag{5.1}$$

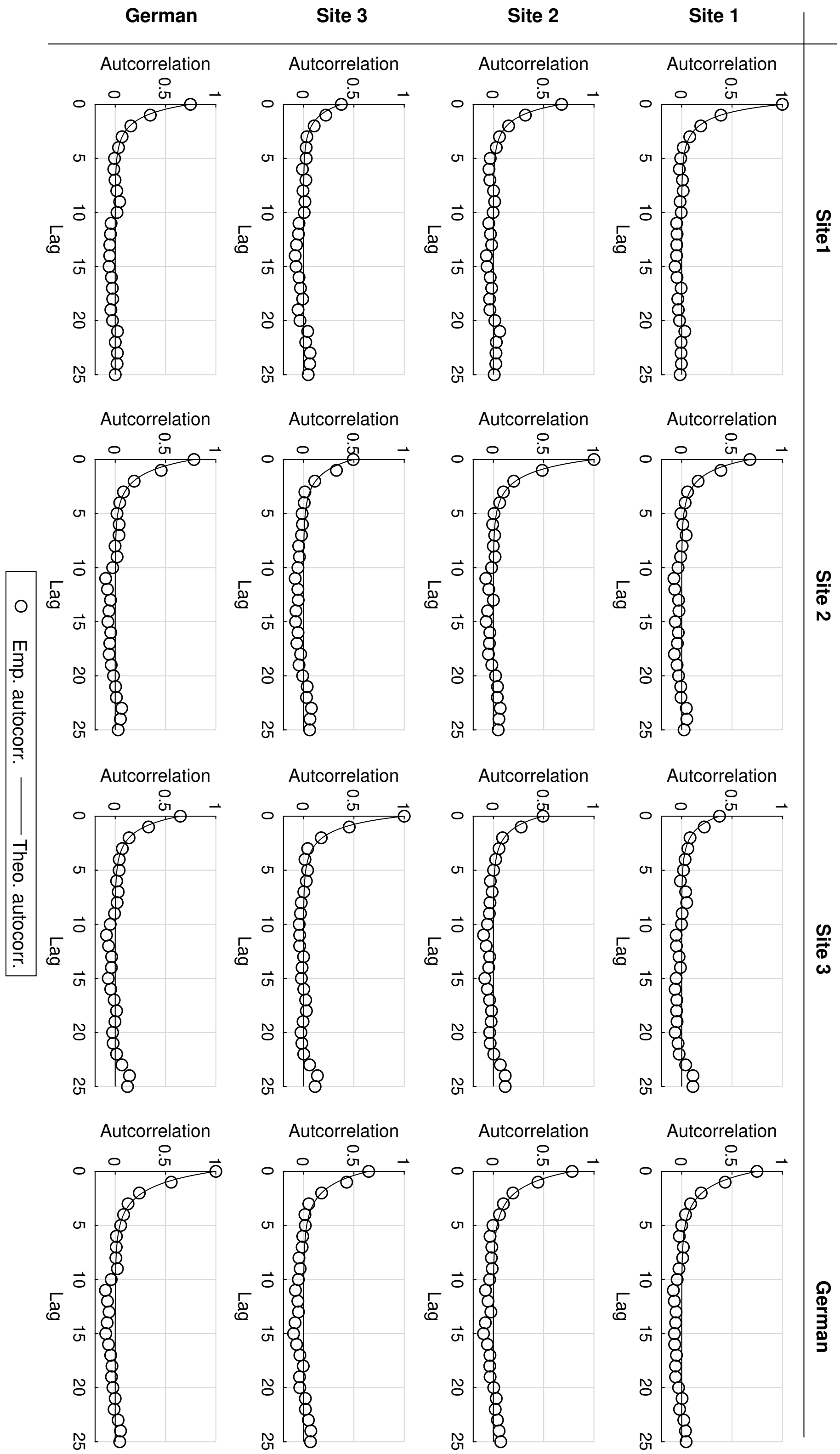
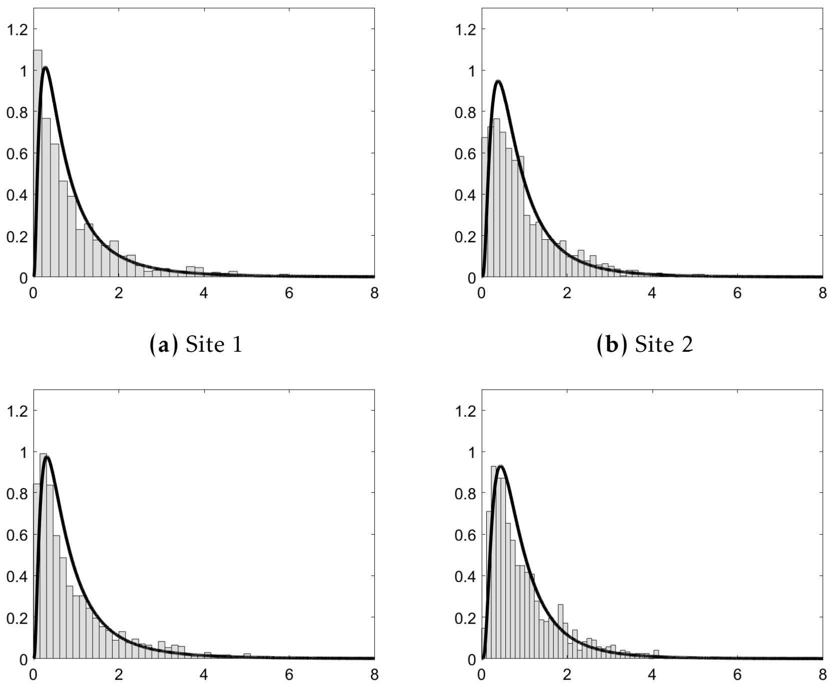


Figure 7: Empirical autocorrelation and theoretical autocorrelation implied by the fitted lognormal model. The (i, j)'th panel shows $\mathrm{cor}(X_i(0), X_j(t))$ for $t = 0, 1, \ldots, 25$.



(a) Site 1 **(b)** Site 2

(c) Site 3 **(d)** German

Figure 8: Histograms of $X_i(t)$ with fitted lognormal densities.

where H is the number of hours during the delivery period $[S,T]$, $P_n(t_0,S,T)$ is the WPF price agreed on at time $t_0 < S < T$, X is a specified tick size, and

$$\bar{P}_n(S,T) = \frac{1}{T-S+1}\sum_{t=S}^{T} P_n(t).$$

From Eq. (5.1) it is apparent that a short position results in a positive payoff in low wind scenarios according to the short position equivalent to Eq. (5.1),

$$H(P_n(t_0,S,T) - \bar{P}_n(S,T))X.$$

That is, if the realization of $\bar{P}_n(S,T)$ is lower than $P_n(t_0,S,T)$. This is favourable for a WPP, since this payoff will offset the loss in income from the long position in wind power production.

To be more specific, let C_i denote the capacity of WPP i, and let $P_i(t)$ denote the daily wind index/utilization of WPP i such that $C_iP_i(t)$ is the actual production of power. Further assume that the WPP receives a fixed price Q_i per produced MWh. The long position in wind power production for WPP i doing the period $[S,T]$ is therefore

$$\bar{P}_i(S,T)C_iHQ_i, \tag{5.2}$$

where

$$\bar{P}_i(S,T) = \frac{1}{T-S+1}\sum_{t=S}^{T} P_i(t). \tag{5.3}$$

Assume that the WPP takes a position $\gamma_i \in \mathbb{Z}$ in WPF contracts with delivery period being $[S,T]$. The payoff from taking this position and the long position in wind power production results in a portfolio with payoff

$$H\bar{P}_i(S,T)C_iQ_i + \gamma_i H(\bar{P}_n(S,T) - P_n(t_0,S,T))X. \tag{5.4}$$

From Eq. (5.4) it is clear that perfectly hedging the volumetric risk would mean to choose γ_i such that $H\bar{P}_i(S,T)C_iQ_i = -\gamma_i H\bar{P}_n(S,T)X$, resulting in the deterministic payoff $\gamma_i HP_n(t_0,S,T)X$. However, the problem for the WPP is that the stochastic terms $\bar{P}_i(S,T)$ and $\bar{P}_n(S,T)$ are not perfectly dependent, and hence obtaining the deterministic payoff $\gamma_i HP_n(t_0,S,T)X$ is not possible. In fact, as shown in [8], it is far from optimal using the exchange-traded WPF contracts for hedging purposes for a single WPP, depending on the dependence structure between the site-specific wind index and the underlying index of the WPF contract.

5.1 Perfect hedging of volumetric risk using tailor-made wind power futures

Tailor-made over-the-counter WPF contracts is a way of perfectly hedging the volumetric risk. Instead of going short the exchange-traded WPF contract, the WPP could instead go short an over-the-counter WPF contract with the underlying being $P_i(t)$ instead of $P_n(t)$. In the following we therefore consider the situation of an energy management company (EMC) acting as counterparty of these tailor-made WPF contracts from $n-1$ different WPPs in Germany. Let $H(\bar{P}_i(S,T) - P_i(t_0,S,T))C_iQ_i$ be the payoff of a long position in a tailor-made WPF contract for WPP i. Thus, from the point of view of the EMC, the payoff of acting as counterparty for $n-1$ different WPPs and taking a position $\gamma \in \mathbb{Z}$ in the exchange-traded WPF contract is

$$R_C(\gamma) = \sum_{i=1}^{n-1} H(\bar{P}_i(S,T) - P_i(t_0,S,T))C_iQ_i + \gamma H(\bar{P}_n(S,T) - P_n(t_0,S,T))X, \tag{5.5}$$

while the payoff from the point of view of the ith WPP is

$$\begin{aligned} R_{WPP,i} &= H\bar{P}_i(S,T)C_iQ_i + H(P_i(t_0,S,T) - \bar{P}_i(S,T))C_iQ_i \\ &= HP_i(t_0,S,T)C_iQ_i. \end{aligned}$$

We argue that this construction can be beneficial for both the individual WPPs and the EMC: Firstly, the individual WPPs obtain a perfect hedge of their volumetric risk, and secondly, with an appropriate number of WPPs and distribution of the WPPs geographically, the portfolio of tailor-made WPF contracts approximately replicates the exchange-traded WPF contract.

5.2 Minimum variance hedge of a tailor-made WPF contracts portfolio

In this section we discuss a minimum variance hedge of a portfolio consisting of tailor-made WPF contracts for the EMC. I.e., from Eq. (5.5) we define the objective to

$$\min_{\gamma} \operatorname{var}(R_C(\gamma)).$$

The variance is

$$\begin{aligned}
\operatorname{var}(R_C(\gamma)) &= \operatorname{var}\Bigg[\sum_{i=1}^{n-1} H\left(\frac{1}{T-S+1}\sum_{t=S}^{T} P_i(t) - P_i(t_0,S,T)\right) C_i Q_i \\
&\quad + \gamma H\left(\frac{1}{T-S+1}\sum_{t=S}^{T} P_n(t) - P_n(t_0,S,T)\right) X\Bigg] \\
&= \sum_{i=1}^{n-1}\sum_{j=1}^{n-1}\left(\frac{H}{T-S+1}\right)^2 C_i Q_i C_j Q_j \sum_{t=S}^{T}\sum_{s=S}^{T} \operatorname{cov}(P_i(t), P_j(s)) \\
&\quad + \left(\gamma \frac{H}{T-S+1} X\right)^2 \sum_{t=S}^{T}\sum_{s=S}^{T} \operatorname{cov}(P_n(t), P_n(s)) \\
&\quad + 2\sum_{i=1}^{n-1}\left(\frac{H}{T-S+1}\right)^2 \gamma X C_i Q_i \sum_{t=S}^{T}\sum_{s=S}^{T} \operatorname{cov}(P_n(t), P_i(s)).
\end{aligned} \tag{5.6}$$

It follows from Eq. (5.6) that the optimal position of WPF contracts is

$$\gamma = -\frac{\sum_{i=1}^{n-1} C_i Q_i \sum_{t=S}^{T}\sum_{s=S}^{T} \operatorname{cov}(P_n(t), P_i(s))}{X \sum_{t=S}^{T}\sum_{s=S}^{T} \operatorname{cov}(P_n(t), P_n(s))}. \tag{5.7}$$

Besides the fact that the dependencies between the stochastic variables impact the size of γ, the size of each wind site measured by C_i and the price paid for each MWh to each wind site measured by Q_i both translate linearly to the size of γ. Therefore, the larger the wind site or the higher the price paid for each MWh, the larger γ will be in absolute terms (all other things being equal).

In-sample hedging effectiveness

In the following we consider the case of an EMC that needs to hedge its portfolio of tailor-made WPF from one year ahead to two years ahead. The considered wind sites are the ones depicted in Fig. 2. We assume that the contract specifications for each site is as shown in Table 5. Further, we assume that $X = 100$ EUR. The estimated parameters of the gamma and lognormal model are the ones reported in Sec. 4.

Site ID	Capacity in MW, C_i	Price in EUR/MWh, Q_i
1	100	30
2	100	30
3	100	30

Table 5: Fictional contract specifications for the sites in Fig. 2.

In Table 6 we present the hedging results for the gamma and lognormal model. We include the case with all three sites and the German WPF in the portfolio, and then three cases where we only include one of the wind sites and the German WPF. In each case, we report the model-implied optimal position of exchange-traded WPF contracts, $\hat{\gamma}$, and the variance reduction (in percentage) implied by the model calculated by $[\mathrm{var}(R_C(0)) - \mathrm{var}(R_C(\hat{\gamma}))]/\mathrm{var}(R_C(0))$. It is apparent that the portfolio with all three sites outperforms the three other cases, confirming the diversification approach of the EMC discussed in Sec. 5.1.

The fact that the difference regarding $\hat{\gamma}$ is small indicate that both models could be used interchangeably to determine an appropriate hedge, though the difference in variance reduction will mislead in a risk management context. In other words, if the wind indexes are driven by the gamma (lognormal) model, and one uses the lognormal (gamma) model to determine hedges, the variance reduction implied by the used model is wrong, while the hedging quantity is relatively close to the optimal hedge.

Case	Sites in portfolio	$\hat{\gamma}$	$\mathrm{var}(R_C(0))$	$\mathrm{var}(R_C(\hat{\gamma}))$	Variance reduction (%)
			Gamma		
1	1,2,3	-63.60	$8.12 \cdot 10^{11}$	$1.31 \cdot 10^{11}$	83.83
2	1	-18.87	$8.55 \cdot 10^{10}$	$2.56 \cdot 10^{10}$	70.12
3	2	-26.79	$1.55 \cdot 10^{11}$	$3.41 \cdot 10^{10}$	77.96
4	3	-17.94	$1.25 \cdot 10^{11}$	$7.07 \cdot 10^{10}$	43.42
			Lognormal		
1	1,2,3	-64.00	$7.06 \cdot 10^{11}$	$1.21 \cdot 10^{11}$	82.89
2	1	-18.97	$8.12 \cdot 10^{10}$	$2.98 \cdot 10^{10}$	63.33
3	2	-25.10	$1.34 \cdot 10^{11}$	$4.42 \cdot 10^{10}$	67.04
4	3	-19.92	$1.13 \cdot 10^{11}$	$5.66 \cdot 10^{10}$	50.05

Table 6: Optimal hedging quantity γ implied by the gamma and lognormal model for different portfolios consisting of different wind sites, and the corresponding variance of the portfolio excluding the exchange-traded WPF contract, and the variance of the portfolio when the optimal hedge is employed. Additionally, we show in all cases the associated variance reduction in percentage.

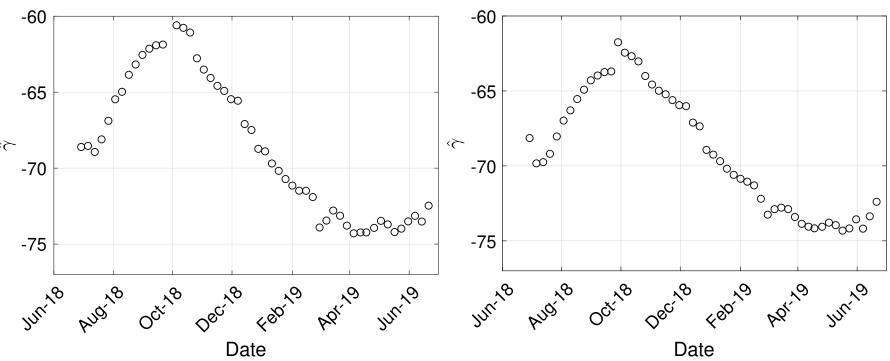
Comparing Eq. (5.4) to Eq. (5.5), the cases 2, 3, and 4 represent the variance reduction implied by the model if the individual wind sites were to hedge their power production themselves by using the exchange-traded WPF contract. From a social welfare point of view, the sum of variances of case 2, 3, and 4 is approximately 8% larger for both models. So not only does the model suggest that tailor-made WPF contracts constitute an obvious way of mitigating uncertainty for wind power producers, but also as a way of optimizing the integration of wind power penetration in the electricity grid.

Out-of-sample hedging effectiveness

In this section we consider the same portfolio of wind sites as in the previous case study, and the specifications of the sites are therefore specified in Table 5. However, here we assess the model on out-of-sample observations. We assume that an EMC has bought tailor-made WPF contracts at the three sites for the period from 2 July 2018 to 30 June 2019, corresponding to 364 days or 52 weeks. We employ a weekly minimum variance hedging strategy, meaning that the EMC has a naked position in a portfolio of tailor-made WPF contracts for the entire period with the exception of the front week. To concretize, the first position taken in exchange-traded WPF contracts is the contract with a weekly delivery period from 2 July 2018 to 8 July 2018. The position is taken based on a model that is estimated by using two years of observations ending the last trading day before the delivery period of the weekly exchange-traded WPF contract. With the delivery period starting the 2 July 2018, the last trading day turns out to be 29 June 2018. Then we step one week ahead and determine the appropriate hedge for the week starting 9 July 2018 and ending 15 July 2018, but again only by employing two years of in-sample observations to estimate the model (the estimation period again ends on the last day where one can trade the weekly exchange-traded WPF contract). In this way we end up with 52 hedging quantities, where each quantity is calculated using different estimated parameters of the model due to the moving two-year observation period.

A comment on the model specifications is in place. The stationarity of the models in Sec. 3 might seem unreasonable in the present context, given the short period of time between an estimation date and the corresponding start date of delivery of the exchange-traded WPF contract. However, we also implemented the models that take the conditional distribution into account, resulting in similar results. For the sake of keeping the presentation as clear as possible, we have therefore only chosen to present the stationary versions of the models.

The resulting optimal hedge quantities are depicted in Fig. 9, indicating a seasonal pattern with more exchange-traded WPF contracts needed during spring compared to autumn. Considering Eq. (5.7), this is the result of the fact that the difference between the sum of the autocovariances of the German wind index and the sum of the autocovariances between the German wind index and the site indexes increases. To assess the hedging effectiveness, we calculate the corresponding implied weekly

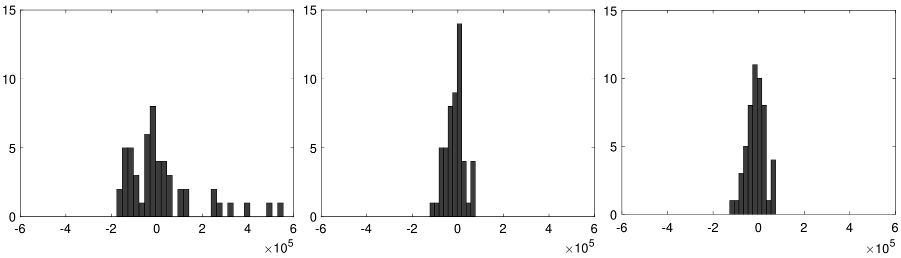


(a) Gamma model **(b)** Lognormal model

Figure 9: Variance minimizing hedge quantitiy, $\hat{\gamma}$, implied by (a) the gamma model and (b) the lognormal model for the 52 weeks covering the period from 2 July 2018 to 30 June 2019.

payoff, $R_C(\hat{\gamma})$, for each weekly hedge quantity, $\hat{\gamma}$. Since we have a variance minimizing perspective, we force a simplistic view on $P_i(t_0, S, T)$ for all wind indexes. Specifically, we assume that for each i, $P_i(t_0, S, T)$ for all weeks during the out-of-sample period from 2 July 2018 to 30 June 2019 is the mean of $P_i(t)$ over the first estimation period spanning 1 July 2016 to 29 June 2018,

Fig. 10 shows a histogram of the payoffs of the portfolio of tailor-made WPF contracts and the exchange-traded WPF contract acting as hedging instrument. Compared to Fig. 10(a), the variances in Fig. 10(b) and Fig. 10(c) are clearly reduced. In fact, the variance reduction in percentage of using the exchange-traded WPF contracts as hedging instrument is 93.64% for the gamma model and 93.62% for the lognormal model.



(a) No hedge **(b)** With hedge (gamma) **(c)** With hedge (lognormal)

Figure 10: Histograms of (a) the payoff for EMC by not hedging the portfolio of tailor-made WPF contracts with exchange-traded WPF contracts, and (b)-(c) using the gamma and lognormal model to find the position in exchange-traded WPF contracts used as a hedging instrument for the portfolio of tailor-made WPF contracts.

Risk premium of wind power futures

Since tailor-made WPF contracts are, by construction, traded over-the-counter, it is worth to discuss the risk premium of such contracts. As a reference point, we consider the risk premium of the exchange-traded WPF contracts. We define the risk premium as the model implied WPF contract price under the physical measure subtracted from the observed exchange-traded WPF contract price. The model implied price is defined by $\mathbb{E}[\bar{P}_n(S,T)]$, meaning that the risk premium $RP(t_o,S,T)$ is

$$RP(t_0,S,T) = \bar{P}_n(t_0,S,T) - \mathbb{E}[\bar{P}_n(S,T)] \tag{5.8}$$

on day t_0 for the delivery period $[S,T]$. The observed quoted exchange-traded WPF prices are obtained from NASDAQ OMX. As concluded in Sec. 5.2, the stationarity of the models does not imply different results compared to the conditional versions of the models for such long time periods, so to ease the presentation, we only consider the unconditional expected value here[3].

We limit ourselves to yearly and quarterly exchange-traded WPF contracts for two reasons. First, it is unlikely that the tailor-made WPF contracts in general will be specified for a short delivery such as a week as a result of such non-standardized instrument. Secondly, as concluded in [5], fundamentals impact the information premium of exchange-traded WPF contracts with a short delivery period (e.g. a week) and a short period of time to delivery, which we would like to avoid. Thirdly, to assess the seasonal differences we also consider quarterly contracts.

For the period from 1 July 2016 to 30 June 2019, we show $\mathbb{E}[\bar{P}_n(S,T)]$ and $\bar{P}_n(t_0,S,T)$ in Fig. 11 for the front year (that is, for a given date, the front year denotes the next

[3]Despite the fact that $RP(t_0,S,T)$ still depends on t_0 thorugh $\bar{P}_n(t_0,S,T)$, the assumption of stationarity is to some degree confirmed by the constant pattern of $\bar{P}_n(t_0,S,T)$ observed in Figs. 11 and 12(a).

year). The quoted prices are fairly constant throughout the entire period, which could be a consequence of illiquidity of exchange-traded WPF contracts. The risk premium

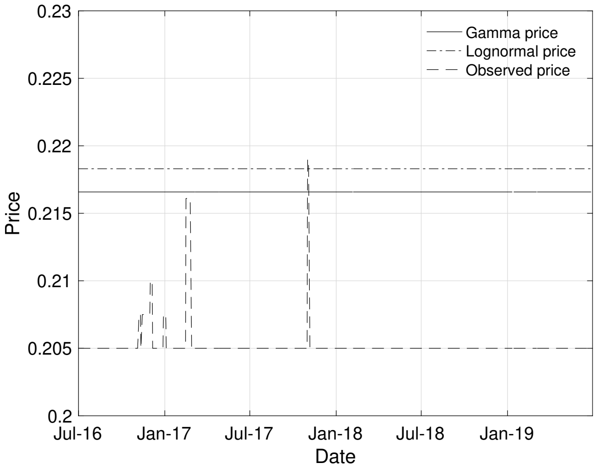
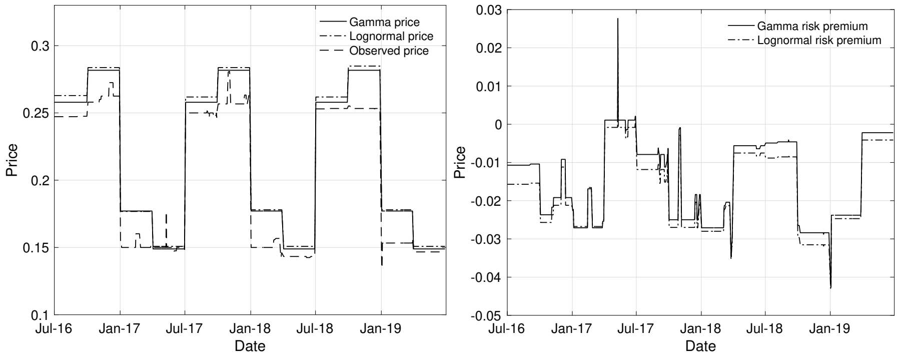


Figure 11: The model implied and quoted price for the front year for the period from 1 July 2016 to 30 June 2019. Notice that the date refers to the observation date; i.e., the date where the contract is quoted.

is -0.011 for the gamma model and -0.013 for the lognormal model on average. Since we are considering a yearly WPF contract we can ignore the seasonality and use the empirical mean to assess the risk premium. The empirical risk premium is -0.011, agreeing with the gamma model. This is likely a consequence of the gamma model having a better fit to the distribution of the German wind index as discussed in Sec. 4 (see also Table 3).

Fig. 12(a) shows the model implied and quoted prices for the front quarter, and Fig. 12(b) shows the corresponding risk premium. The mean of the risk premium in this case is -0.014 for the gamma model and -0.016 for the lognormal model. The seasonal variation in the prices peaks for contracts with delivery during Q4 and Q1, simply since more wind is present during these quarters. This is also reflected in the model-implied prices. The peaks in the risk premium are observed for contracts with delivery during Q1 and Q2. One explanation of this could be non-aligned incentives to engage in the WPF market throughout the year for the buying and selling side. [9] shows that the hedging benefits are greater for CPPs during Q3 and Q4 compared to Q1 and Q2; hence, during Q3 and Q4, CPPs are more interested in WPF contracts and thus willing to pay more.

A negative risk premium is in line with the findings in [5] and [10]. One might argue that this is expected from a hedging benefit perspective, since the hedging benefits in general are greater for the selling side than the buying side (see [8], [9], and [10]). Continuing this argument, the risk premium is likely to be even more negative in the tailor-made WPF contracts market as a result of the perfect hedge implied by the tailor-made WPF contracts for WPPs. However, from the perspective of the individual WPP, this extra risk premium associated with the tailor-made WPF contract compared to the exchange-traded WPF contract has to be weighted against the deterministic payoff implied by the tailor-made WPF contract.



(a) Model and quoted price

(b) Risk premium

Figure 12: (a) the model implied and quoted price for the front quarter, and (b) and the risk premium for the front quarter. The observations period is from 1 July 2016 to 30 June 2019. Notice that the date refers to the observation date; i.e., the date where the contract is quoted.

6 Conclusion

In this paper, we propose and compare two multivariate continuous-time models, the gamma and lognormal model, for the joint behaviour of wind indexes. We discuss the properties of the models, and propose estimation procedures. Empirically, we employ the models to a joint model for the wind indexes at three different wind sites in Germany, and the German wind index that represents the overall utilization of wind power production in Germany. We find that both models are able to capture the autocorrelation structure well. However, the gamma model captures the skewness and kurtosis of the German wind index better than the lognormal model.

The models are applied to a variance-minimizing hedging strategy of a portfolio consisting of long positions in so-called tailor-made wind power futures contracts at the three wind sites, and a short position in the exchange-traded wind power futures contract. The hedging effectiveness is assessed in an in-sample and out-of-sample context. Both models indicate that a significant variance reduction can be obtained by hedging the portfolio with the exchange-traded wind power futures contracts in-sample as well as out-of-sample. Further, the hedging benefits are greater for the portfolio of tailor-made wind power futures compared to hedging each individual wind site with exchange-traded wind power futures contracts.

The risk premium of the exchange-traded wind power futures contracts is examined, where we find that the gamma model implies a more reliable estimate of the risk premium. A negative risk premium is observed in line with other findings in the literature for both yearly and quarterly contracts. Even though the tailor-made wind power futures contracts give each wind power producer a perfect volumetric hedge of her wind power production, we argue that it is likely that the risk premium for a tailor-made wind power futures contract is even more negative compared to the exchange-traded contract.

Acknowledgements

Troels Sønderby Christensen is supported by the Innovation Fund Denmark under Grant 5189-00117B. Victor Rohde is supported by the Danish Council for Independent Research under Grant DFF-4002-00003.

7 Theoretical results for the gamma model

This section is dedicated to proving Prop. 3.2 and Proposition 3.3. We start by proving Prop. 3.3, which the lemma below is a first step towards. We will use some standard results about continuous-time moving averages, all of which can be found in [14].

The following Lemma is well-known, but we give a proof for the sake of completeness.

Lemma 7.1. *Let $t \geq 0$ and consider the two one-dimensional processes*

$$Y_1(t) = \int_{-\infty}^{t} f_1(t-u)dZ(u) \quad \text{and} \quad Y_2(t) = \int_{-\infty}^{t} f_2(t-u)dZ(u) \tag{7.1}$$

for functions f_1 and f_2 in $L^1(\mathbb{R}) \cap L^2(\mathbb{R})$, and where Z is a one-dimensional Lévy process with second moment. Then

$$\mathbb{E}[Y_1(0)] = \int_0^{\infty} f_1(u)du \mathbb{E}[Z(1)]$$

and

$$\mathbb{E}[(Y_1(0) - \mathbb{E}[Y_1(0)])(Y_2(t) - \mathbb{E}[Y_2(t)])] = \int_0^{\infty} f_1(u)f_2(t+u)du \operatorname{var}(Z(1)).$$

Proof. Let $\psi_{Y_1(0),Y_2(t)}$ be the cumulant generating function of $(Y_1(0), Y_2(t))$ and ψ_Z be the cumulant generating function of Z. Then

$$\begin{aligned}\psi_{Y_1(0),Y_2(t)}(x) &= \log \mathbb{E}[\exp\{x_1 Y_1(0) + x_2 Y_2(t)\}] \\ &= \int_0^t \psi_Z(x_2 f_2(u))du + \int_0^{\infty} \psi_Z(x_1 f_1(u) + x_2 f_2(t+u))du.\end{aligned}$$

It follows that for $n_1, n_2 \in \mathbb{N}_0$ with $n_1 + n_2 \leq 2$,

$$\begin{aligned}\frac{d^{n_1+n_2}}{dx_1^{n_1} dx_2^{n_2}} \psi_{Y_1,Y_2}(x) = &\int_0^t f_2^{n_2}(u)\psi_Z^{(n_2)}(x_2 f_2(u))du \\ &+ \int_0^{\infty} f_1^{n_1}(u) f_2^{n_2}(u) \psi_Z^{(n_1+n_2)}(x_1 f_1(u) + x_2 f_2(u))du.\end{aligned}$$

where $\psi_Z^{(n_1+n_2)}$ denotes the $n_1 + n_2$ times derivative of ψ_Z. We conclude that

$$\mathbb{E}[Y_1(0)] = \frac{d}{dx_1}\psi_{Y_1(0),Y_2(t)}(0) = \int_0^{\infty} f_1(u)du \mathbb{E}[Z(1)].$$

Assume now, without loss of generality, $\mathbb{E}[Z(1)] = 0$. Then

$$\begin{aligned}&\mathbb{E}[(Y_1(0) - \mathbb{E}[Y_1(0)])(Y_2(t) - \mathbb{E}[Y_2(t)])] \\ &= \frac{d^2}{dx_1 dx_2}\psi_{Y_1(0),Y_2(t)}(0) \\ &= \int_0^{\infty} f_1(u) f_2(t+u)du \operatorname{var}[Z(1)]\end{aligned}$$

□

Proof (Proof of Prop. 3.3). Let $\sigma_{i,k}$ denote the (i,k)'th entry of Σ_L. Then, using Lemma 7.1,

$$\begin{aligned}\mathbb{E}[X_i(t)] &= \sum_{k=1}^{n} \mathbb{E}\left[\int_{-\infty}^{t} e^{-\lambda_i(t-u)} \sigma_{i,k} dL_k(u)\right] \\ &= \sum_{k=1}^{n} \frac{1}{\lambda_i} \sigma_{i,k} \beta_k / 2 \\ &= \left(\Lambda^{-1} \Sigma_L \beta / 2\right)_i.\end{aligned}$$

This gives (3.9). Assume now, without loss of generality, $\mathbb{E}[L(1)] = 0$. Then, using Lemma 7.1 again,

$$\begin{aligned}&\mathbb{E}[X_i(0)X_j(t)] \\ &\quad = \mathbb{E}\left[\left(\sum_{k=1}^{n} \int_{-\infty}^{0} e^{-\lambda_i(-u)} \sigma_{i,k} dL_k(u)\right)\left(\sum_{k=1}^{n} \int_{-\infty}^{t} e^{-\lambda_j(t-u)} \sigma_{j,k} dL_k(u)\right)\right] \\ &\quad = \sum_{k=1}^{n} \sigma_{i,k}\sigma_{j,k} \int_{0}^{\infty} e^{-\lambda_j(t+u)} e^{-\lambda_i u} du \\ &\quad = \frac{e^{-\lambda_j t}}{\lambda_i + \lambda_j} \sum_{k=1}^{n} \sigma_{i,k}\sigma_{j,k} \\ &\quad = ((\Lambda_{var} \circ \Sigma_L \Sigma_L^{\top}) e^{-\Lambda t})_{i,j}\end{aligned}$$

from which (3.10) follows. □

We now turn to proving Prop. 3.2. Initially, we give the next result which is a special case of [6, Theorem 4.8], but again, we give a proof for the sake of completeness.

Proposition 7.2. *Let L be a compound Poisson process with intensity $\alpha > 0$ and exponential jumps with parameter $\beta > 0$. Consider $t \in \mathbb{R}$, $\lambda, \mu > 0$ and $x_1, x_2 \in \mathbb{R}$ with $x_1 + x_2 < \beta$. Furthermore, assume $x_1 x_2 \geq 0$ and $x_1 \neq 0$, and let $f(t) = x_1 e^{-\lambda t} + x_2 e^{-\mu t}$. Then*

$$\begin{aligned}&\log \mathbb{E}\left[\exp\left\{\int_{-\infty}^{t} f(t-u) dL(u)\right\}\right] \\ &\quad = \alpha \frac{f(0)}{f'(0)} \log\left(1 - \frac{f(0)}{\beta}\right) + \alpha \int_{0}^{\infty} \left(\frac{f(u)}{f'(u)}\right)' \log\left(1 - \frac{f(u)}{\beta}\right) du,\end{aligned} \tag{7.2}$$

where

$$\left|\left(\frac{f(u)}{f'(u)}\right)'\right| \leq \frac{(\lambda-\mu)^2}{2\lambda\mu} \tag{7.3}$$

for all $u \geq 0$.

Proof. Initially, note that f/f' is bounded and

$$\begin{aligned}\left|\left(\frac{f(u)}{f'(u)}\right)'\right| &= \frac{|f'(u)^2 - f(u)f''(u)|}{f'(u)^2} \\ &= \frac{x_1 x_2 (\lambda-\mu)^2 e^{-(\lambda+\mu)u}}{x_1^2\lambda^2 e^{-2\lambda u} + x_2^2\mu^2 e^{-2\mu u} + 2x_1x_2\lambda\mu e^{-(\lambda+\mu)u}} \\ &\leq \frac{(\lambda-\mu)^2}{2\lambda\mu}.\end{aligned}$$

This gives the bound on $(f/f')'$. Additionally, we find that

$$\begin{aligned}\left|\left(\frac{f(u)}{f'(u)}\right)'\right| &= \frac{x_1 x_2 (\lambda-\mu)^2 e^{-(\lambda+\mu)u}}{x_1^2\lambda^2 e^{-2\lambda u} + x_2^2\mu^2 e^{-2\mu u} + 2x_1x_2\lambda\mu e^{-(\lambda+\mu)u}} \\ &= \frac{x_1 x_2 (\lambda-\mu)^2}{x_1^2\lambda^2 e^{-(\lambda-\mu)u} + x_2^2\mu^2 e^{-(\mu-\lambda)u} + 2x_1x_2\lambda\mu}.\end{aligned}$$

We conclude that $(f(u)/f'(u))' = O(e^{-|\lambda-\mu|u})$ as $u \to \infty$. Thus, all integrals below are convergent and the integration by parts is justified. Next let

$$\psi(u) = \log \mathbb{E}[\exp(uL(1))] = \alpha \frac{u}{\beta - u}$$

be the cumulant-generating function of $L(1)$ and let $\phi(u) = -\alpha \log(1 - u/\beta)$ be the cumulant-generating funciton of a gamma distribution with shape α and rate β (see for example [5]). Note that $\psi(u) = u\phi'(u)$. Then, using integration by parts,

$$\begin{aligned}&\log \mathbb{E}\left[\exp\left\{\int_{-\infty}^{t} f(t-u)dL(u)\right\}\right] \\ &= \int_0^\infty \psi(f(u))du \\ &= \int_0^\infty \frac{f(u)}{f'(u)}(\phi(f(u)))'du \\ &= -\frac{f(0)}{f'(0)}\phi(f(0)) - \int_0^\infty \left(\frac{f(u)}{f'(u)}\right)' \phi(f(u))du.\end{aligned}$$

□

Remark 7.3. Considering the proof of Prop. 7.2 there are two approaches to calculate

$$\log \mathbb{E}\left[\exp\left\{\int_{-\infty}^{t} f(t-u)dL(u)\right\}\right]. \tag{7.4}$$

Either by calculating

$$\int_0^\infty \psi(f(u))du \tag{7.5}$$

or

$$-\frac{f(0)}{f'(0)}\phi(f(0)) - \int_0^\infty \left(\frac{f(u)}{f'(u)}\right)' \phi(f(u))du. \tag{7.6}$$

Here, ψ and ϕ are the cumulant-generating function of $L(1)$ and of a gamma distribution with shape α and rate β as defined in the proof of Prop. 7.2. By (7.3), the integral in (7.6) will be small whenever $(\lambda-\mu)^2/(2\lambda\mu)$ is small. In the application we consider we are concerned with the case where $\lambda = \hat{\lambda}_i$ and $\mu = \hat{\lambda}_j$ for some $i,j = 1,2,3,4$, where $\hat{\lambda}_i$ and $\hat{\lambda}_j$ are given in Table 2. We have

$$\max_{i,j} \frac{(\hat{\lambda}_i - \hat{\lambda}_j)^2}{2\hat{\lambda}_i\hat{\lambda}_j} = 0.0695,$$

and therefore, indeed, that $(\lambda-\mu)^2/(2\lambda\mu)$ is small in the case relevant to us. The integral in (7.6) has ϕ in the kernel whereas (7.5) has ψ, making a direct comparison more difficult. We do, however, have

$$\phi(u) = \alpha u + O(u^2) \text{ and } \psi(u) = \alpha u + O(u^2) \text{ as } u \to 0$$

(by a Taylor approximation argument), indicating that ϕ and ψ are of comparable size, at least for small values. Furthermore, by numerical comparison, we have found them to be of similar size. We conclude that the kernel of (7.6) is expected to be considerably smaller than the kernel of (7.5). We therefore prefer to do the calculation in (7.6) instead of (7.5) since we can do a much more coarse approximation for a desired precision of a approximation of (7.4).

Proposition 7.4. *Let L be a compound Poisson process with intensity $\alpha > 0$ and exponential jumps with parameter $\beta > 0$. Consider $s < t$, $\lambda > 0$ and $x < \beta$. Then*

$$\mathbb{E}\left[\exp\left\{x\int_s^t e^{-\lambda(t-u)}dL(u)\right\}\right] = \left(\frac{\beta - xe^{-\lambda(t-s)}}{\beta - x}\right)^{\alpha/\lambda}$$

and

$$\mathbb{E}\left[\exp\left\{x\int_{-\infty}^t e^{-\lambda(t-u)}dL(u)\right\}\right] = \left(\frac{\beta}{\beta - x}\right)^{\alpha/\lambda} \tag{7.7}$$

Proof. Let

$$\psi(t) = \log\mathbb{E}[\exp(tL(1))] = \alpha\frac{t}{\beta - t}$$

be the cumulant-generating function of L. Then

$$\begin{aligned}\log\mathbb{E}&\left[\exp\left\{\int_s^t f(t-u)dL(u)\right\}\right]\\ &= \int_0^{t-s}\psi(e^{-\lambda u})du\\ &= \frac{\alpha}{\lambda}\Big(\log(\beta - xe^{-\lambda(t-s)}) - \log(\beta - x)\Big).\end{aligned}$$

A similar calculation gives (7.7). □

Proof (Proof of Theorem 3.2). For notional convenience, let

$$\tilde{L}(t) = (L_1(t),\ldots,L_{n-1}(t),0,L_n(t))^\top \in \mathbb{R}^{n+1}.$$

First consider (3.7) and assume $i \neq j$. We have

$$X_i(t) = \int_{-\infty}^{t} e^{-\lambda_i(t-u)}\tilde{\sigma}_{i,i}\, d\tilde{L}_i(u) + \int_{s}^{t} e^{-\lambda_i(t-u)}\tilde{\sigma}_{i,n+1}\, d\tilde{L}_{n+1}(u) + \int_{-\infty}^{s} e^{-\lambda_i(t-u)}\tilde{\sigma}_{i,n+1}\, d\tilde{L}_{n+1}(u)$$

and

$$X_j(s) = \int_{-\infty}^{s} e^{-\lambda_j(s-u)}\tilde{\sigma}_{j,j}\, d\tilde{L}_j(u) + \int_{-\infty}^{s} e^{-\lambda_j(s-u)}\tilde{\sigma}_{j,n+1}\, d\tilde{L}_{n+1}(u).$$

Next, note that $\operatorname{cov}(UV, W) = \operatorname{cov}(V, UW) = \mathbb{E}[U]\operatorname{cov}(V, W)$ for a random variable U independent of the random variables V and W. Applying this, and the above, we conclude that

$$\begin{aligned}
&\operatorname{cov}(P_i(t), P_j(s)) \\
&= \operatorname{cov}\left(e^{-S_i(t)X_i(t)}, e^{-S_j(s)X_j(s)}\right) \\
&= \mathbb{E}\left[\exp\left\{-S_i(t)\int_{-\infty}^{t} e^{-\lambda_i(t-u)}\tilde{\sigma}_{i,i}\, d\tilde{L}_i(u)\right\}\right] \\
&\times \mathbb{E}\left[\exp\left\{-S_i(t)\int_{s}^{t} e^{-\lambda_i(t-u)}\tilde{\sigma}_{i,n+1}\, d\tilde{L}_{n+1}(u)\right\}\right] \\
&\times \mathbb{E}\left[\exp\left\{-S_j(s)\int_{-\infty}^{s} e^{-\lambda_j(t-u)}\tilde{\sigma}_{j,j}\, d\tilde{L}_j(u)\right\}\right] \\
&\times \operatorname{cov}\left(\exp\left\{-S_i(t)\int_{-\infty}^{s} e^{-\lambda_i(t-u)}\tilde{\sigma}_{i,n+1}\, d\tilde{L}_i(u)\right\},\right. \\
&\qquad \left.\exp\left\{-S_j(s)\int_{-\infty}^{s} e^{-\lambda_j(s-u)}\sigma_{j,n+1}\, d\tilde{L}_{n+1}(u)\right\}\right)
\end{aligned} \tag{7.8}$$

Expressions of the three expectations in Eq. (7.8) are given in Prop. 7.4. Furthermore,

$$\begin{aligned}
&\operatorname{cov}\left(\exp\left\{-S_i(t)\int_{-\infty}^{s} e^{-\lambda_i(t-u)}\tilde{\sigma}_{i,n+1}\, d\tilde{L}_i(u)\right\},\right. \\
&\qquad \left.\exp\left\{-S_j(s)\int_{-\infty}^{s} e^{-\lambda_j(s-u)}\tilde{\sigma}_{j,n+1}\, d\tilde{L}_{n+1}(u)\right\}\right) \\
&= \mathbb{E}\left[\exp\left\{-\int_{-\infty}^{s} f_{i,j}(s-u)\, d\tilde{L}_{n+1}(u)\right\}\right] \\
&- \mathbb{E}\left[\exp\left\{-S_i(t)\int_{-\infty}^{s} e^{-\lambda_i(t-u)}\tilde{\sigma}_{i,n+1}\, d\tilde{L}_{n+1}(u)\right\}\right] \\
&\times \mathbb{E}\left[\exp\left\{-S_j(s)\int_{-\infty}^{s} e^{-\lambda_j(s-u)}\tilde{\sigma}_{j,n+1}\, d\tilde{L}_{n+1}(u)\right\}\right]
\end{aligned}$$

for which expressions are given in Prop. 7.2 and Prop. 7.4.

Next, consider (3.8). We write

$$X_i(t) = \int_{s}^{t} e^{-\lambda_i(t-u)}\left(\tilde{\sigma}_{i,i}\, d\tilde{L}_i(u) + \tilde{\sigma}_{i,n+1}\, d\tilde{L}_{n+1}(u)\right) + \int_{-\infty}^{s} e^{-\lambda_i(t-u)}\left(\tilde{\sigma}_{i,i}\, d\tilde{L}_i(u) + \tilde{\sigma}_{i,n+1}\, d\tilde{L}_{n+1}(u)\right)$$

and

$$X_i(s) = \int_{-\infty}^{s} e^{-\lambda_i(s-u)} \left(\tilde{\sigma}_{i,i} d\tilde{L}_i(u) + \tilde{\sigma}_{i,n+1} d\tilde{L}_{n+1}(u)\right).$$

Consequently,

$$\begin{aligned}
&\operatorname{cov}(P_i(t), P_i(s)) \\
&\quad = \operatorname{cov}\left(e^{-S_i(t)X_i(t)}, e^{-S_i(s)X_i(s)}\right) \\
&\quad = \mathbb{E}\left[\exp\left\{-S_i(t)\int_s^t e^{-\lambda_i(t-u)}\left(\tilde{\sigma}_{i,i} d\tilde{L}_i(u) + \tilde{\sigma}_{i,n+1} d\tilde{L}_{n+1}(u)\right)\right\}\right] \\
&\quad \times \operatorname{cov}\left(\exp\left\{-S_i(t)\int_{-\infty}^s e^{-\lambda_i(t-u)}\left(\tilde{\sigma}_{i,i} d\tilde{L}_i(u) + \tilde{\sigma}_{i,n+1} d\tilde{L}_{n+1}(u)\right)\right\},\right. \\
&\qquad \left.\exp\left\{-S_i(s)\int_{-\infty}^s e^{-\lambda_i(s-u)}\left(\tilde{\sigma}_{i,i} d\tilde{L}_j(u) + \tilde{\sigma}_{i,n+1} d\tilde{L}_{n+1}(u)\right)\right\}\right)
\end{aligned}$$

Again, expressions for the expectation in (7.8) can be found using Prop. 7.4. Finally,

$$\begin{aligned}
&\operatorname{cov}\left(\exp\left\{-S_i(t)\int_{-\infty}^s e^{-\lambda_i(t-u)}\left(\tilde{\sigma}_{i,i} d\tilde{L}_i(u) + \tilde{\sigma}_{i,n+1} d\tilde{L}_{n+1}(u)\right)\right\},\right. \\
&\qquad \left.\exp\left\{-S_i(s)\int_{-\infty}^s e^{-\lambda_i(s-u)}\left(\tilde{\sigma}_{i,i} d\tilde{L}_i(u) + \tilde{\sigma}_{i,n+1} d\tilde{L}_{n+1}(u)\right)\right\}\right) \\
&\quad = \mathbb{E}\left[\exp\left\{-\int_{-\infty}^s \tilde{\sigma}_{i,i}\left(S_i(t)e^{-\lambda_i(t-s)} + S_i(s)\right)e^{-\lambda_i(s-u)} d\tilde{L}_i(u)\right\}\right] \\
&\quad \times \mathbb{E}\left[\exp\left\{-\int_{-\infty}^s \tilde{\sigma}_{i,n+1}\left(S_i(t)e^{-\lambda_i(t-s)} + S_i(s)\right)e^{-\lambda_i(s-u)} d\tilde{L}_{n+1}(u)\right\}\right] \\
&\quad - \mathbb{E}\left[\exp\left\{-\int_{-\infty}^s \tilde{\sigma}_{i,i} S_i(t) e^{-\lambda_i(t-u)} d\tilde{L}_i(u)\right\}\right] \\
&\quad \times \mathbb{E}\left[\exp\left\{-\int_{-\infty}^s \tilde{\sigma}_{i,n+1} S_i(t) e^{-\lambda_i(t-u)} d\tilde{L}_{n+1}(u)\right\}\right] \\
&\quad \times \mathbb{E}\left[\exp\left\{-\int_{-\infty}^s \tilde{\sigma}_{i,i} S_i(s) e^{-\lambda_i(s-u)} d\tilde{L}_i(u)\right\}\right] \\
&\quad \times \mathbb{E}\left[\exp\left\{-\int_{-\infty}^s \tilde{\sigma}_{i,n+1} S_i(s) e^{-\lambda_i(s-u)} d\tilde{L}_{n+1}(u)\right\}\right]
\end{aligned}$$

where expressions are given in Prop. 7.4. □

References

[1] Asmussen, S., J.L. Jensen and L. Rojas-Nandayapa (2016). On the Laplace transform of the lognormal distribution. *Methodology and Computing in Applied Probability* 18(2), 441–458.

[2] Ballotta, L. and E. Bonfiglioli (2016). Multivariate asset models using Lévy processes and applications. *The European Journal of Finance* 22(13), 1320–1350.

[3] Barndorff-Nielsen, O.E. and N. Shephard (2001). Non-Gaussian Ornstein–Uhlenbeck-based models and some of their uses in financial economics. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 63(2), 167–241.

[4] Benth, F.E., J. Kallsen and T. Meyer-Brandis (2007a). A non-Gaussian Ornstein–Uhlenbeck process for electricity spot price modeling and derivatives pricing. *Applied Mathematical Finance* 14(2), 153–169.

[5] Benth, F.E. and A. Pircalabu (2018). A non-Gaussian Ornstein-Uhlenbeck model for pricing wind power futures. *Applied Mathematical Finance* 25(1), 36–65.

[6] Benth, F.E. and V. Rohde (2019). On non-negative modeling with CARMA processes. *Journal of Mathematical Analysis and Applications* 476(1), 196–214.

[7] Benth, F.E., J. Šaltytė-Benth and S. Koekebakker (2007b). Putting a price on temperature. *Scand. J. Statist.* 34(4), 746–767. DOI: 10.1111/j.1467-9469.2007.00564.x.

[8] Christensen, T.S. and A. Pircalabu (2018). On the spatial hedging effectiveness of German wind power futures for wind power generators. *Journal of Energy Markets*.

[9] Christensen, T.S., A. Pircalabu and E. Høg (2019). A seasonal copula mixture for hedging the clean spark spread with wind power futures. *Energy Economics* 78, 64–80.

[10] Gersema, G. and D. Wozabal (2017). An equilibrium pricing model for wind power futures. *Energy Economics* 65, 64–74.

[11] Halliwell, L.J. (2015). "The lognormal random multivariate". *Casualty Actuarial Society E-Forum, Spring*, 5.

[12] Leoni, P. and W. Schoutens (2007). Multivariate smiling.

[13] Meucci, A. (2009). Review of statistical arbitrage, cointegration, and multivariate Ornstein-Uhlenbeck. *Cointegration, and Multivariate Ornstein-Uhlenbeck (May 14, 2009)*.

[14] Rajput, B.S. and J. Rosiński (1989). Spectral representations of infinitely divisible processes. *Probab. Theory Related Fields* 82(3), 451–487.

[15] Sato, K. (1999). *Lévy Processes and Infinitely Divisible Distributions*. Vol. 68. Cambridge Studies in Advanced Mathematics. Translated from the 1990 Japanese original, Revised by the author. Cambridge University Press.

[16] Semeraro, P. (2008). A multivariate variance gamma model for financial applications. *International Journal of Theoretical and Applied Finance* 11(01), 1–18.

Paper I

A Statistical View On A Surrogate Model For Estimating Extreme Events With An Application To Wind Turbines

Mikkel Slot Nielsen and Victor Rohde

Abstract

In the present paper we propose a surrogate model, which particularly aims at estimating extreme events from a vector of covariates and a suitable simulation environment. The first part introduces the model rigorously and discusses the flexibility of each of its components by drawing relations to literature within fields such as incomplete data, statistical matching, outlier detection and conditional probability estimation. In the second part of the paper we study the performance of the model in the estimation of extreme loads on an operating wind turbine from its operational statistics.

AMS 2010 subject classifications: 62P30; 65C20; 91B68

Keywords: extreme event estimation; wind turbines; surrogate model

1 Introduction

Suppose that we want to assess the distributional properties of a certain one-dimensional random variable Y. For instance, one could be interested in knowing the probability of the occurrence of large values of Y as they may be associated with a large risk such as system failure or a company default. One way to evaluate such risks would be to collect observations $y_1,\dots,y_n$ of Y and then fit a suitable distribution (e.g., the generalized Pareto distribution) to the largest of them. Extreme event estimation is a huge area and there exists a vast amount on literature of both methodology and applications; a few references are [4, 5, 12, 17]. This is one example where knowledge of the empirical distribution of Y,

$$\widehat{\mathbb{P}_Y}(\delta_{y_1},\ldots,\delta_{y_n}) = \frac{1}{n}\sum_{i=1}^{n}\delta_{y_i}, \tag{1.1}$$

is valuable. (Here δ_y denotes the Dirac measure at the point y.) If one is interested in the entire distribution of Y, one may use the estimator (1.1) directly or a smoothed version, e.g., replacing δ_{y_i} by the Gaussian distribution with mean y_i and variance $\sigma^2 > 0$ (the latter usually referred to as the bandwidth). The problem in determining (1.1) arises if Y is not observable. Such a situation can happen for several reasons, e.g., it may be that Y is difficult or expensive to measure or that its importance has just recently been recognized, and hence one have not collected the historic data that is needed. Sometimes, a solution to the problem of having a latent variable could be to set up a suitable simulation environment and, by varying the conditions of the system, obtain various realizations of Y. Since we cannot be sure that the variations in the simulation environment correspond to the variations in the physical environment, the realizations of Y are not necessarily drawn from the true distribution. This is essentially similar to any experimental study and one will have to rely on the existence of control variables.

By assuming the existence of an observable d-dimensional vector X of covariates carrying information about the environment, a typical way to proceed would be regression/matching which in turn would form a surrogate model. To be concrete, given a realization x of X, a surrogate model is expected to output (approximately) $f(x) = \mathbb{E}[Y \mid X = x]$, the conditional mean of Y given $X = x$. Consequently, given inputs $x_1,\ldots,x_n$, the model would produce $f(x_1),\ldots,f(x_n)$ as stand-ins for the missing values $y_1,\ldots,y_n$ of Y. Building a surrogate for the distribution of Y on top of this could now be done by replacing y_i by $f(x_i)$ in (1.1) to obtain an estimate $\widehat{\mathbb{P}_Y}(\delta_{f(x_1)},\ldots,\delta_{f(x_n)})$ of the distribution of Y. This surrogate model for the distribution of Y can thus be seen as a composition of two maps:

$$(x_1,\ldots,x_n) \longrightarrow (\delta_{f(x_1)},\ldots,\delta_{f(x_n)}) \longrightarrow \widehat{\mathbb{P}_Y}(\delta_{f(x_1)},\ldots,\delta_{f(x_n)}). \tag{1.2}$$

In the context of an incomplete data problem, the strategy of replacing unobserved quantities by the corresponding conditional means is called regression imputation and will generally not provide a good estimate of the distribution of Y. For instance, while the (unobtainable) estimate in (1.1) converges weakly to the distribution of Y as the sample size n increases, the one provided by (1.2) converges weakly to the distribution of the conditional expectation $\mathbb{E}[Y \mid X]$ of Y given X. In fact, any of the so-called single imputation approaches, including regression imputation, usually results in proxies $\hat{y}_1,\ldots,\hat{y}_n$ which exhibit less variance than the original values $y_1,\ldots,y_n$, and in this case $\widehat{\mathbb{P}_Y}(\delta_{\hat{y}_1},\ldots,\delta_{\hat{y}_n})$ will provide a poor estimate of the distribution of Y (see [15] for details).

The reason that the approach (1.2) works unsatisfactory is that $\delta_{f(X)}$ is an (unbiased) estimator for the distribution of $\mathbb{E}[Y \mid X]$ rather than of Y. For this reason we will replace $\delta_{f(x)}$ by an estimator for the conditional distribution μ_x of Y given $X = x$ and maintain the overall structure of (1.2):

$$(x_1,\ldots,x_n) \longrightarrow (\mu_{x_1},\ldots,\mu_{x_n}) \longrightarrow \widehat{\mathbb{P}_Y}(\mu_{x_1},\ldots,\mu_{x_n}). \tag{1.3}$$

In Section 2 we introduce the model (1.3) rigorously and relate the assumptions on the simulation environment needed to estimate μ_x to the classical strong ignorability (or unconfoundedness) assumption within a matching framework. Given a simulation environment that satisfies this assumption, an important step in order to apply the surrogate model (1.3) is of course to decide how to estimate μ_x, and hence we discuss in Section 2.1 some methods that are suitable for conditional probability estimation. In Section 2.2 we address the issue of checking if the simulation environment meets the imposed assumptions. Finally, in Section 3 we apply the surrogate model to real-world data as we estimate extreme tower loads on a wind turbine from its operational statistics.

2 The model

Let $\mathbb{P}$ be the physical probability measure. Recall that Y is the one-dimensional random variable of interest, X is a d-dimensional vector of covariates and $x_1, \ldots, x_n$ are realizations of X under $\mathbb{P}$. We are interested in a surrogate model that delivers an estimate of $\mathbb{P}(Y \in B)$ for every measurable set B. The model is given by

$$\widehat{\mathbb{P}_Y} = \frac{1}{n} \sum_{i=1}^{n} \widehat{\mu}_{x_i}, \tag{2.1}$$

where $\widehat{\mu}_x$ is an estimator for the conditional distribution μ_x of Y given $X = x$. Since each x_i is drawn independently of μ_x under $\mathbb{P}$, each $\widehat{\mu}_{x_i}$ provides an estimator of $\mathbb{P}_Y$, and the averaging in (2.1) may be expected to force the variance of the estimator $\widehat{\mathbb{P}_Y}$ to zero as n tends to infinity. In order to obtain $\widehat{\mu}_x$ we need to assume the existence of a valid simulation tool:

Condition 2.1. *Realizations of (X, Y) can be obtained under an artificial probability measure $\mathbb{Q}$ which satisfies*

(i) *The support of $\mathbb{P}(X \in \cdot)$ is contained in the support of $\mathbb{Q}(X \in \cdot)$.*

(ii) *The conditional distribution of Y given $X = x$ is the same under both $\mathbb{P}$ and $\mathbb{Q}$, that is,*

$$\mathbb{Q}(Y \in \cdot \mid X = x) = \mu_x$$

for all x in the support of $\mathbb{P}(X \in \cdot)$.

In words, Condition 2.1 says that any outcome of X that can happen in the real world can also happen in the simulation environment and, given an outcome of X, the probabilistic structure of Y in the real world is perfectly mimicked by the simulation tool. Note that, while this is a rather strict assumption, it may of course be relaxed to $\mathbb{Q}(Y \in B \mid X = x) = \mu_x(B)$ for all x in the support of $\mathbb{P}(X \in \cdot)$ and any set B of interest. For instance, in Section 3 we will primarily be interested in $B = (\tau, \infty)$ for a large threshold τ.

Remark 2.2. We can assume, possibly by modifying the sample space, the existence of a random variable $Z \in \{0, 1\}$ and a probability measure $\tilde{\mathbb{P}}$ such that

$$\mathbb{P} = \tilde{\mathbb{P}}(\cdot \mid Z = 0) \quad \text{and} \quad \mathbb{Q} = \tilde{\mathbb{P}}(\cdot \mid Z = 1).$$

Effectively, Z indicates whether we are using the simulation tool or not, and $\tilde{\mathbb{P}}(Z = 1) \in (0, 1)$ defines the probability of drawing (X, Y) from the simulation environment (as opposed to drawing X from the measurement environment). In this case, according to Bayes' rule, Condition 2.1 is equivalent to

$$\tilde{\mathbb{P}}(Z = 1 \mid X, Y) = \tilde{\mathbb{P}}(Z = 1 \mid X). \tag{2.2}$$

In words, (2.2) means that Y and Z are conditionally independent under $\tilde{\mathbb{P}}$ given X. The assumption (2.2) was introduced in [13] as the strong ignorability assumption in relation to estimating heterogeneous treatment effects. In the literature on incomplete data, where Z indicates whether Y is observed or not, (2.2) is usually known as the Missing at Random (in short, MAR) mechanism, referring to the pattern of which Y is missing. This assumption is often imposed and viewed as necessary in order to do inference. See [9, 14, 15] for details about the incomplete data problem and the MAR mechanism.

Remark 2.3. Usually, to meet Condition 2.1((ii)), one will search for a high-dimensional X (large d) to control for as many factors as possible. However, as this complicates the estimation of μ_x, one may be interested in finding a function $b : \mathbb{R}^d \to \mathbb{R}^m$, $m < d$, maintaining the property

$$\mathbb{P}(Y \in \cdot \mid b(X) = b(x)) = \mathbb{Q}(Y \in \cdot \mid b(X) = b(x)) \tag{2.3}$$

for all x in the support of $\mathbb{P}(X \in \cdot)$. This is a well-studied problem in statistical matching with the main reference being [13], who referred to any such b as a balancing function. They characterized the class of balancing functions by first showing that (2.3) holds if b is chosen to be the propensity score under $\tilde{\mathbb{P}}$ (cf. Remark 2.2), $\pi(x) = \tilde{\mathbb{P}}(Z = 1 \mid X = x)$, and next arguing that a general function b is a balancing function if and only if

$$f(b(x)) = \pi(x) \quad \text{for some function } f. \tag{2.4}$$

2.1 Estimation of the conditional probability

The ultimate goal is to estimate $\mu_x = \mathbb{Q}(Y \in \cdot \mid X = x)$, e.g., in terms of the cumulative distribution function (CDF) or density function, from a sample $(x_1^s, y_1^s), \dots, (x_m^s, y_m^s)$ of (X, Y) under the artificial measure $\mathbb{Q}$. (We use the notation x_i^s rather than x_i to emphasize that the quantities are simulated values and should not be confused with x_i in (2.1).) The literature on conditional probability estimation is fairly large and includes both parametric and non-parametric approaches varying from simple nearest neighbors matching to sophisticated deep learning techniques. A few references are [7, 8, 10, 18]. In Section 3 we have chosen to use two simple but robust techniques in order to estimate μ_x:

(i) *Smoothed k-nearest neighbors*: for a given $k \in \mathbb{N}$, $k \le m$, let $I_k(x) \subseteq \{1, \dots, m\}$ denote the k indices corresponding to the k points in $\{x_1^s, \dots, x_m^s\}$ which are closest to x with respect to some distance measure. Then μ_x is estimated by

$$\widehat{\mu}_x = \frac{1}{k} \sum_{i \in I_k(x)} \mathcal{N}(y_i^s, \sigma),$$

where $\mathcal{N}(\xi,\sigma)$ denotes the Gaussian distribution with mean ξ and standard deviation $\sigma \geq 0$ (using the convention $\mathcal{N}(\xi,0) = \delta_\xi$).

(ii) *Smoothed random forest classification*: suppose that one is interested in the CDF of μ_x at certain points $\alpha_1 < \alpha_2 < \cdots < \alpha_k$ and consider the random variable $C \in \{0,1,\ldots,k\}$ defined by $C = \sum_{j=1}^k \mathbb{1}_{\{Y>\alpha_j\}}$. From $y_1^s,\ldots,y_m^s$ one obtains realizations $c_1,\ldots,c_m$ of C under $\mathbb{Q}$ and, next, random forest classification (as described in [2]) can be used to obtain estimates of the functions

$$p_j(x) = \mathbb{Q}(C = j \mid X = x), \quad j = 0,1,\ldots,k-1.$$

Given these estimates, say $\widehat{p_0},\widehat{p_1},\ldots,\widehat{p_{k-1}}$, the CDF of μ_x is estimated by

$$\widehat{\mu_x}((-\infty,\alpha_i]) = \sum_{j=1}^{k} \widehat{p_{j-1}}(x)\Phi\Big(\frac{\alpha_i - \alpha_j}{\sigma}\Big), \quad i = 1,\ldots,k,$$

where Φ is the CDF of a standard Gaussian distribution (using the convention $\Phi(\cdot/0) = \mathbb{1}_{[0,\infty)}$).

Both techniques are easily implemented in Python using modules from the scikit-learn library (see [11]). The distance measure d, referred to in ((i)), would usually be of the form

$$d(x,y) = \sqrt{(x-y)^T M(x-y)}, \quad x,y \in \mathbb{R}^d,$$

for some positive definite $d \times d$ matrix M. If M is the identity matrix, d is the Euclidean distance, and if M is the inverse sample covariance matrix of the covariates, d is the Mahalanobis distance. Note that, since the k-nearest neighbors (kNN) approach suffers from the curse of dimensionality, one would either require that X is low-dimensional, reduce the dimension by applying dimensionality reduction techniques or use another balancing function than the identity function (i.e., finding an alternative function b satisfying (2.4)).

2.2 Validation of the simulation environment

The validation of the simulation environment concerns how to evaluate whether or not Condition 2.1 is satisfied. Part ((i)) of the condition boils down to checking whether it is plausible that a realization x of X under the physical measure $\mathbb{P}$ could also happen under the artificial measure $\mathbb{Q}$ or, by negation, whether x is an outlier relative to the simulations of X. Outlier detection methods have received a lot of attention over decades and, according to [6], they generally fall into one of three classes: unsupervised clustering (pinpoints most remote points to be considered as potential outliers), supervised classification (based on both normal and abnormal training data, an observation is classified either as an outlier or not) and semi-supervised detection (based on normal training data, a boundary defining the set of normal observations is formed). We will be using a kNN outlier detection method, which belongs to the first class, and which bases the conclusion of whether x is an outlier or not on the average distance from x to its k nearest neighbors. The motivation for applying this method is two-fold: (i) an extensive empirical study [3] of the unsupervised outlier detection methods concluded that the kNN method, despite its simplicity, is a robust method that remains the state of the art when compared across various datasets, and (ii) given that we already compute the distances to the k nearest neighbors to estimate μ_x, the additional computational burden induced by using the kNN outlier detection method is minimal. For more on outlier detection methods, see [1, 3, 6, 19] and references therein.

Following the setup of Section 2.1, let $x_1^s,\dots,x_m^s$ be realizations of X under $\mathbb{Q}$ and denote by $I_k(x)$ the set of indices corresponding to the k realizations closest to x with respect to some metric d (e.g., the Euclidean or Mahalanobis distance). Then, for observations $x_1,\dots,x_n$ under $\mathbb{P}$, the algorithm goes as follows:

(1) For $i = 1,\dots,n$ compute the average distance from x_i to its k-nearest neighbors

$$\bar{d}_i = \frac{1}{k}\sum_{j\in I_k(x_i)} d(x_i, x_j^s).$$

(2) Obtain a sorted list $\bar{d}_{(1)} \le \cdots \le \bar{d}_{(n)}$ of $\bar{d}_1,\dots,\bar{d}_n$ and detect, e.g., by visual inspection, a point j at which the structure of the function $i \mapsto \bar{d}_{(i)}$ changes significantly.

(3) Regard any x_i with $\bar{d}_i \ge \bar{d}_{(j)}$ as an outlier.

Part ((ii)) of Condition 2.1 can usually not be checked, since we do not have any realizations of Y under $\mathbb{P}$; this is similar to the issue of verifying the MAR assumption in an incomplete data problem. Of course, if such realizations are available we can estimate the conditional distribution of Y given $X = x$ under both $\mathbb{P}$ and $\mathbb{Q}$ and compare the results.

3 Application to extreme event estimation for wind turbines

In this section we will consider the possibility of estimating the distribution of the 10-minute maximum down-wind bending moment (load) on the tower top, middle and base on an operating wind turbine from its 10-minute operational statistics. The data consists of 19,976 10-minute statistics from the turbine under normal operation over a period from February 17 to September 30, 2017. Since this particular turbine is part of a measurement campaign, load measurements are available, and these will be used to assess the performance of the surrogate model (see Figure 1 for the histogram and CDF of measured loads).

To complement the measurements, a simulation tool is used to obtain 50,606 simulations of both the operational statistics and the corresponding tower loads. We choose to use the following eight operational statistics as covariates:

- Electrical power (maximum and standard deviation)
- Generator speed (maximum)
- Tower top down-wind acceleration (standard deviation)
- Blade flap bending moment (maximum, standard deviation and mean)
- Blade pitch angle (minimum)

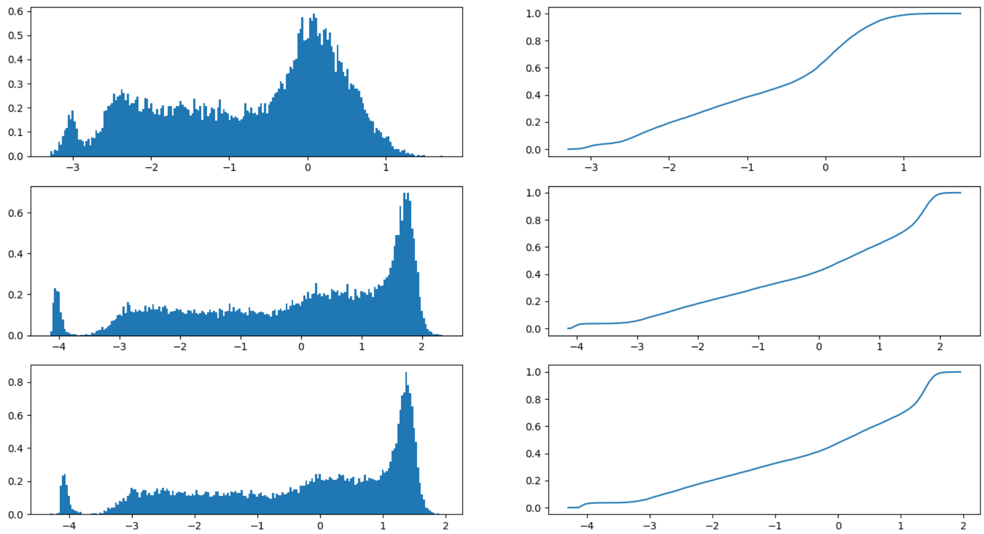


Figure 1: Measured load distributions. Left and right plots correspond to histograms and CDFs, respectively, based on 19,976 observations of the tower top (first row), middle (second row) and base (third row) down-wind bending moments.

The selection of covariates is based on a physical interpretation of the problem and by leaving out covariates which from a visual inspection (i.e., plots of the two-dimensional coordinate projections) seem to violate the support assumption imposed in Condition 2.1((i)). The loads and each of the covariates are standardized by subtracting the sample mean and dividing by the sample standard deviation (both of these statistics are computed from the simulated values). In the setup of Section 2, this means that we have realizations of $X \in \mathbb{R}^8$ and $Y \in \mathbb{R}$ under both $\mathbb{P}$ and $\mathbb{Q}$ (although the typical case would be that Y is not realized under $\mathbb{P}$). This gives us the opportunity to compare the results of our surrogate model with the, otherwise unobtainable, estimate (1.1) of $\mathbb{P}(Y \in \cdot)$.

In order to sharpen the estimate of μ_x for covariates x close to the measured ones, we discard simulations which are far from the domain of the measured covariates. Effectively, this is done by reversing the kNN approach explained in Section 2.2 as we compute average distances from simulated covariates to the k nearest measured covariates, sort them and, eventually, choosing a threshold that defines the relevant simulations. We will use $k = 1$ and compute the sorted average distances in terms of the Mahalanobis distance. The selection of threshold is not a trivial task and, as suggested in Section 2.2, the best strategy may be to inspect visually if there is a certain point, at which the structure of the sorted average distances changes significantly. To obtain a slightly less subjective selection rule, we use the following ad hoc rule: the threshold is defined to be $d_{(\tau)}$, the τ-th smallest average distance, where τ is the point that minimizes the L^1 distance

$$d_1(f, f_\tau) := \int_1^m |f(x) - f_\tau(x)|\, dx \tag{3.1}$$

between the function f that linearly interpolates $(1, d_{(1)}), \ldots, (m, d_{(m)})$ and f_τ that linearly interpolates $(1, d_{(1)}), (\tau, d_{(\tau)}), (m, d_{(m)})$ over the interval $[1, m]$ (see the left plot of Figure 2). This selection rule implies a threshold of 6.62 with $\tau = 46,100$, which in turn implies that $4,506$ (8.90 %) of the simulations are discarded before estimating the conditional load distributions. See the right plot of Figure 2 for a visual illustration of the threshold selection. Of course, a more (or less) conservative selection rule can be obtained by using another distance measure than (3.1).

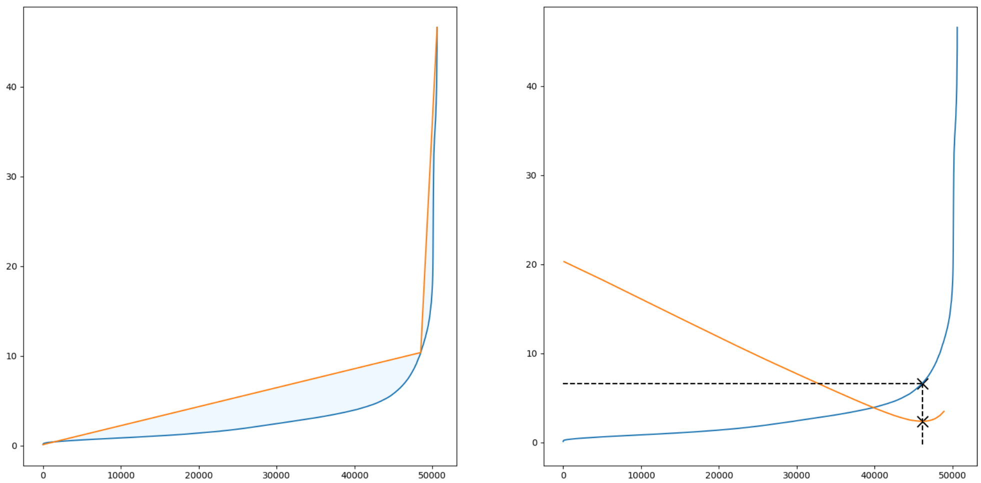


Figure 2: Blue curve: sorted distance from simulated covariates to nearest measured covariates. Left: linear interpolation of $(1, d_{(1)}), (\tau, d_{(\tau)}), (m, d_{(m)})$ with shaded region representing the corresponding L^1 error for $\tau = 48,500$. Right: the orange curve is the normalised L^1 error as a function of τ and the dashed black lines indicate the corresponding minimum and selected threshold.

The same procedure is repeated, now precisely as described in Section 2.2, to detect potential outliers in the measurements. In this case, $k = 10$ is used since this will be the same number of neighbors used to estimate μ_x. The threshold is 2.45 with $\tau = 18,400$, and hence $1,577$ (8.57 %) of the measurements are found to be potential outliers (see also Figure 3).

To assess which points that have been labeled as potential outliers, two-dimensional projections of the outliers, inliers and simulations are plotted in Figure 4 (if a point seems to be an outlier in the projection plot the original eight-dimensional vector should also be labeled an outlier). To restrict the number of plots we only provide 18 (out of 28) of the projection plots corresponding to plotting electrical power (maximum), blade flap bending moment (maximum) and generator speed (maximum) against each other and all the remaining five covariates. The overall picture of Figure 4 is that a significant part of the observations that seem to be outliers is indeed labeled as such. Moreover, some of the labeled outliers seem to form a horizontal or vertical line, which could indicate a period of time where one of the inputs was measured to be constant. Since this is probably caused by a logging error, such measurements should indeed be declared invalid (outliers).

Next, we would need to check if the distributional properties of the load can be expected to change by removing outliers. In an incomplete data setup, the outliers may be treated as the missing observations, and hence we want to assess whether the Missing (Completely) at Random mechanism is in force (recall the discussion in Remark 2.2). If the operation of removing outliers causes a significant change in the

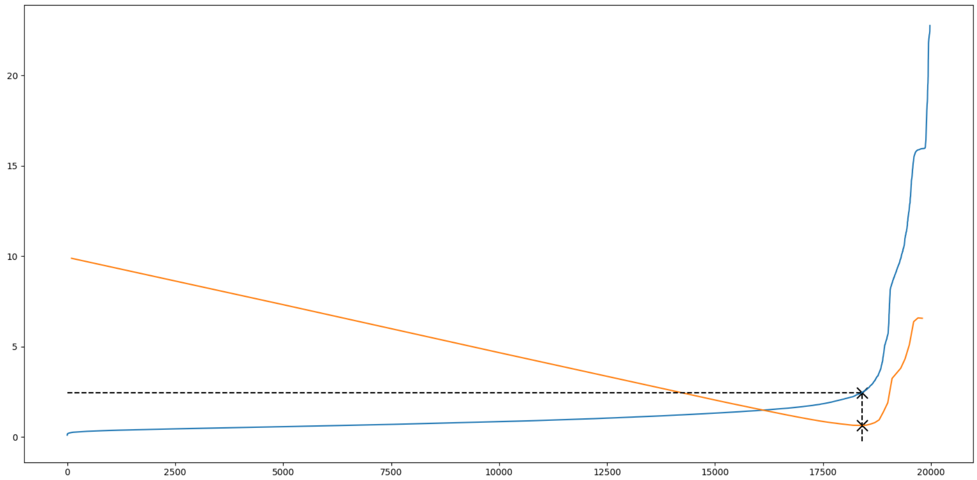


Figure 3: The blue curve is the sorted distance from measured covariates to the 10 nearest simulated covariates, the orange curve is the L^1 error as a function of τ, and the dashed black lines indicate the corresponding minimum and selected threshold. All points with average distance larger than the threshold are labeled possible outliers.

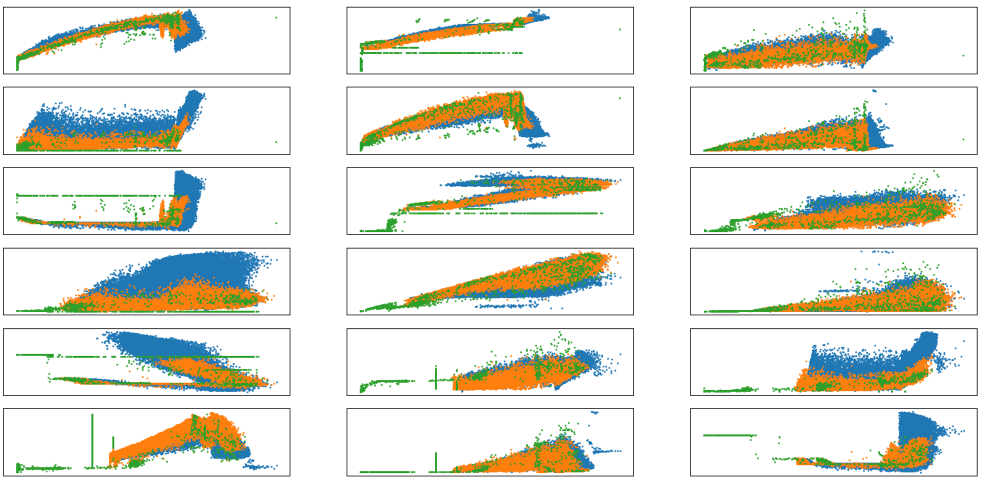


Figure 4: Some of the two-dimensional projections of the covariates. Blue dots are simulations, orange dots are inliers and green dots are potential outliers.

load distribution, then the outliers cannot be ignored and would need to be handled separately. In Figure 5 the histograms of tower top, middle and base load obtained from all measurements (the same as those in the three rows of Figure 1) are compared to those where the outliers have been removed. It becomes immediately clear that the distributions are not unchanged, since most of the outliers correspond to the smallest loads of all measurements. However, it seems reasonable to believe that the conditional distribution of the load given that it exceeds a certain (even fairly small) threshold is not seriously affected by the exclusion of outliers. Since the interest is on the estimation of extreme events, i.e., one often focuses only on large loads, it may be sufficient to match these conditional excess distributions. Hence, we choose to exclude the outliers without paying further attention to them. It should be noted that, since the outlier detection method only focuses on covariates, it does not take into account their explanatory power on the loads. For instance, it might be that a declared outlier only differs from the simulations with respect to covariates that do not significantly help explaining the load level. While this could suggest using other distance measures, this is not a direction that we will pursue here.

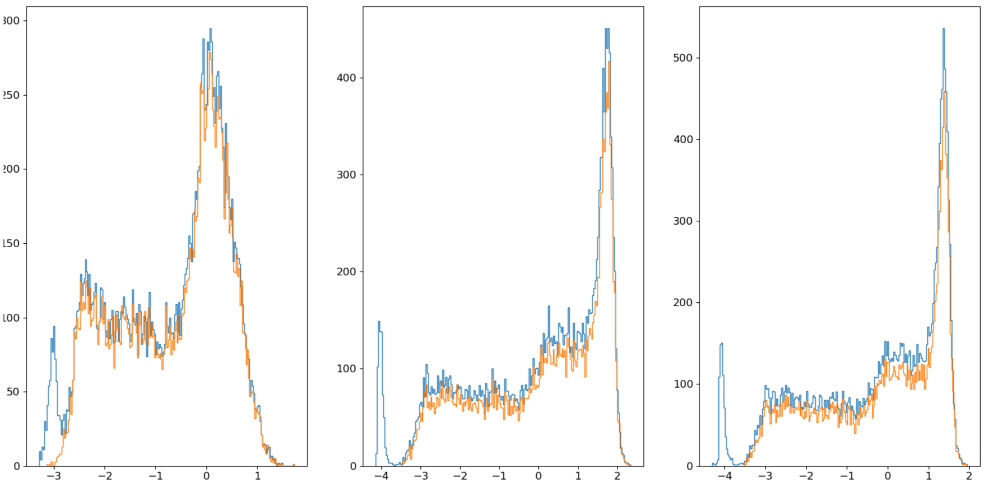


Figure 5: Histograms of measurements on tower top (left), middle (mid) and base (right) down-wind bending moments. Measurements including and excluding outliers are represented in blue and orange, respectively.

We will rely on (2.1) together with the two methods presented in Section 2.1 to estimate the load distributions. The unsmoothed version of both methods (i.e., $\sigma = 0$) will be used, and for the kNN method we will choose $k = 10$. There are at least two reasons for initially choosing the bandwidth σ to be zero: (i) it can be a subtle task to select the optimal bandwidth as there is no universally accepted approach, and (ii) given that we have a fairly large dataset, most of the estimated values of the CDFs should be fairly insensitive to the choice of bandwidth. In Figure 6 we have plotted the empirical CDF of the loads (i.e., the CDF of (1.1) based on measured loads) together with the estimates provided by the kNN and random forest approach. Since the loads are 10-minute maxima, it is natural to compare the CDFs to those of GEV type (cf. the Fisher-Tippett-Gnedenko theorem). For this reason, and in order to put attention on the estimation of the tail, we have also plotted the $-\log(-\log(\cdot))$ transform of the

CDFs. Recall that, when applying such a transformation to the CDF, the Gumbel, Weibull and Fréchet distributions would produce straight lines, convex curves and concave curves, respectively. From the plots it follows that, generally, the estimated CDFs are closest to the empirical CDF for small and large quantiles. Estimated α-quantiles tend to be smaller than the true ones for moderate values of α. One would expect that, given only the eight covariates as considered here, a significant part of errors would be due to differences between the simulation environment and the real-world environment. From an extreme event estimation perspective, the most important part of the curve would be the last 10-20 % corresponding to quantiles above 0.8 or 0.9. On this matter, the $-\log(-\log(\cdot))$ transform of the CDFs reveals that the estimated CDFs have some difficulties in replicating the tail of the distribution for middle and base load. However, since there are few extreme observations, this is also the part where a potential smoothing (positive bandwidth) would have an effect. To test the smoothing effect, we choose σ according to Silverman's rule of thumb, that is, $\sigma = 1.06(kn)^{-1/5}\hat{\sigma}_s$, where $n = 18,399$ is the number of measurements (without outliers) and $\hat{\sigma}_s$ is the sample standard deviation of the kn load simulations (top, middle or base) used for obtaining the kNN estimate of the given load distribution. For details about this choice of bandwidth, and bandwidth selection in general, see [16]. In Figure 7 we have compared the $-\log(-\log(\cdot))$ transforms of the smoothed estimates of the CDFs and the empirical CDF.

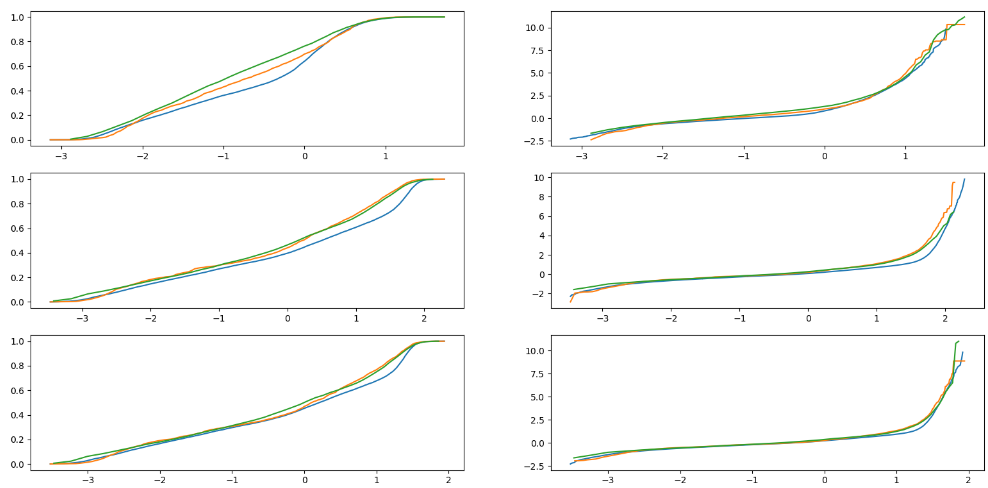


Figure 6: Plots of CDFs (first column) and the corresponding $-\log(-\log(\cdot))$ transforms (second column) of tower top (first row), middle (second row) and base (third row) down-wind bending moments. The blue curve is the empirical distribution of the measurements, and the orange and green curves are the kNN and random forest predictions, respectively.

It seems that the smoothed versions of the estimated curves generally fit the tail better for the tower top and middle loads, but tend overestimate the larger quantiles for the tower base load. This emphasizes that the smoothing should be used with caution; when smoothing the curve, one would need to decide from which point the estimate of the CDF is not reliable (as the Gaussian smoothing always will dominate the picture sufficiently far out in the tail). When no smoothing was used, the uncertainty of the estimates was somewhat reflected in the roughness of

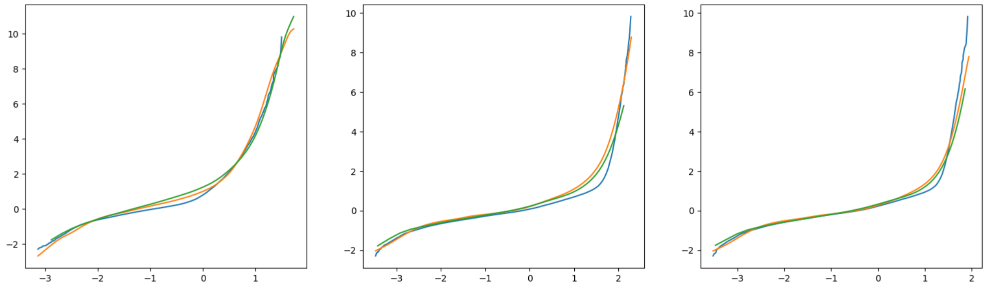


Figure 7: Plots of $-\log(-\log(\cdot))$ transforms of CDFs of tower top (left), middle (center) and base (right) down-wind bending moments. The blue curve is the empirical distribution of the measurements, and the orange and green curves are the smoothed kNN and random forest predictions, respectively, using Silverman's rule of thumb.

the curves. We end this study with Table 1 which compares some of the estimated quantiles with the true (empirical) ones. From this table we see that the errors tend to be largest for the 25 %, 50 % and 75 % quantiles and fairly small for the 95 %, 99 % and 99.5 % quantiles, which is in line with the conclusion based on Figure 6. Moreover, it also appears that no consistent improvements of the tail estimates are obtained by using the smoothed CDF estimates.

Quantile (%)		kNN	kNN (smoothed)	Random forest	Random forest (smoothed)	Empirical
25	Top	−1.7349	−1.7344	−1.7941	−1.7315	−1.5528
	Mid	−1.4252	−1.4434	−1.3607	−1.2773	−1.1427
	Base	−1.4689	−1.4794	−1.4474	−1.3653	−1.3576
50	—	−0.7111	−0.7106	−0.9544	−0.8928	−0.3204
		0.2181	0.2114	0.1587	0.2147	0.5002
		0.1018	0.1152	0.0047	0.0547	0.2087
75	—	0.1643	0.1626	−0.0501	−0.0055	0.1991
		1.1114	1.1076	1.1819	1.2192	1.5460
		0.9407	0.9366	0.9978	1.0247	1.2192
95	—	0.6936	0.7122	0.6951	0.7414	0.7161
		1.6855	1.7090	1.7283	1.7913	1.8670
		1.6782	1.4653	1.4651	1.5184	1.4957
99	—	0.9611	0.9815	1.0068	1.0631	1.0271
		1.8583	1.9383	1.9386	2.0385	1.9917
		1.5877	1.6676	1.6245	1.7240	1.6179
99.5	—	1.0313	1.0687	1.0944	1.1522	1.1155
		1.9180	2.0113	2.0195	2.1213	2.0418
		1.6341	1.7337	1.6716	1.7910	1.6594

Table 1: Some quantiles of the empirical load distributions and of the corresponding kNN and random forest estimates.

4 Conclusion

In this paper we presented a surrogate model for the purpose of estimating extreme events. The key assumption was the existence of a simulation environment which produces realizations of the vector (X,Y) in such a way that the conditional distribution of the variable of interest Y equals the true one given a suitable set of observable covariates X. It was noted that this corresponds to the Missing at Random assumption in an incomplete data problem. Next, we briefly reviewed the literature on conditional probability estimation as this is the critical step in order to translate valid simulations into an estimate of the true unconditional distribution of Y. Finally, we checked the performance of the surrogate model on real data as we used an appropriate simulation environment to estimate the distribution of the tower top, middle and base down-wind loads on an operating wind turbine from its operational statistics. The surrogate model seemed to succeed in estimating the tail of the load distributions, but it tended to underestimate loads of normal size.

Acknowledgments

We thank James Alexander Nichols from Vestas (Load & Control) and Jan Pedersen for fruitful discussions. This work was supported by the Danish Council for Independent Research (Grant DFF - 4002 - 00003).

References

[1] Ben-Gal, I. (2005). “Outlier detection”. *Data mining and knowledge discovery handbook*. Springer, 131–146.

[2] Breiman, L. (2001). Random forests. *Machine learning* 45(1), 5–32.

[3] Campos, G.O., A. Zimek, J. Sander, R.J.G.B. Campello, B. Micenková, E. Schubert, I. Assent and M.E. Houle (2016). On the evaluation of unsupervised outlier detection: measures, datasets, and an empirical study. *Data Min. Knowl. Discov.* 30(4), 891–927. DOI: 10.1007/s10618-015-0444-8.

[4] De Haan, L. and A. Ferreira (2007). *Extreme value theory: an introduction*. Springer Science & Business Media.

[5] Gilli, M. et al. (2006). An application of extreme value theory for measuring financial risk. *Comput. Econ.* 27(2-3), 207–228.

[6] Hodge, V. and J. Austin (2004). A survey of outlier detection methodologies. *Artif. Intell. Rev.* 22(2), 85–126.

[7] Husmeier, D. (2012). *Neural networks for conditional probability estimation: Forecasting beyond point predictions*. Springer Science & Business Media.

[8] Hyndman, R.J., D.M Bashtannyk and G.K. Grunwald (1996). Estimating and visualizing conditional densities. *J. Comput. Graph. Statist.* 5(4), 315–336. DOI: 10.2307/1390887.

[9] Little, R.J. and D.B. Rubin (2019). *Statistical analysis with missing data*. Vol. 793. Wiley.

[10] Neuneier, R., F. Hergert, W. Finnoff and D. Ormoneit (1994). "Estimation of conditional densities: A comparison of neural network approaches". *International Conference on Artificial Neural Networks*. Springer, 689–692.

[11] Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, et al. (2011). Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* 12(Oct), 2825–2830.

[12] Ragan, P. and L. Manuel (2008). Statistical extrapolation methods for estimating wind turbine extreme loads. *Journal of Solar Energy Engineering* 130(3), 031011.

[13] Rosenbaum, P.R. and D.B. Rubin (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika* 70(1), 41–55. DOI: 10.1093/biomet/70.1.41.

[14] Rubin, D.B. (1976). Inference and missing data. *Biometrika* 63(3). With comments by R. J. A. Little and a reply by the author, 581–592. DOI: 10.1093/biomet/63.3.581.

[15] Scheffer, J. (2002). Dealing with missing data. *Res. Lett. Inf. Math. Sci.* 3, 153–160.

[16] Silverman, B.W. (1986). *Density estimation for statistics and data analysis*. Monographs on Statistics and Applied Probability. Chapman & Hall, London. DOI: 10.1007/978-1-4899-3324-9.

[17] Smith, R.L. (1990). Extreme value theory. *Handbook of applicable mathematics* 7, 437–471.

[18] Sugiyama, M., I. Takeuchi, T. Suzuki, T. Kanamori, H. Hachiya and D. Okanohara (2010). Least-squares conditional density estimation. *IEICE T. Inf. Syst.* 93(3), 583–594.

[19] Zimek, A., E. Schubert and H.-P. Kriegel (2012). A survey on unsupervised outlier detection in high-dimensional numerical data. *Stat. Anal. Data Min.* 5(5), 363–387. DOI: 10.1002/sam.11161.