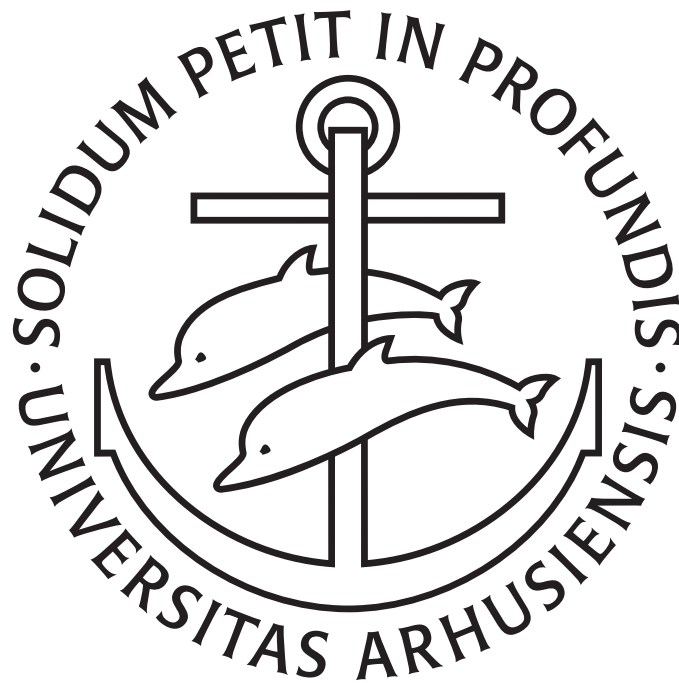# PhD Dissertation

Continuous-Time Modeling Using Lévy-Driven Moving Averages

Representations, Limit Theorems and Other Properties

Mikkel Slot Nielsen

Department of Mathematics
Aarhus University

2019

Continuous-time modeling using Lévy-driven moving averages
Representations, limit theorems and other properties

PhD dissertation by
Mikkel Slot Nielsen

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, July 3, 2019

AARHUS
UNIVERSITY
DEPARTMENT OF MATHEMATICS

Contents

Preface

This dissertation is the result of my PhD studies carried out from May 1, 2015 to July 3, 2019 at Department of Mathematics, Aarhus University, under supervision of Andreas Basse-O'Connor (main supervisor) and Jan Pedersen (co-supervisor). My studies were fully funded by Andreas' grant (DFF–4002–00003) from the Danish Council for Independent Research.

The dissertation consists of the following nine self-contained papers:

Paper A   Equivalent martingale measures for Lévy-driven moving averages and related processes. *Stochastic Processes and their Applications* 128(8), 2538–2556.

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

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

Paper D   Multivariate stochastic delay differential equations and CAR representations of CARMA processes. *Stochastic Processes and their Applications* (forthcoming), 25 pages.

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

Paper F   Limit theorems for quadratic forms and related quantities of discretely sampled continuous-time moving averages. *ESAIM: Probability and Statistics* (forthcoming), 20 pages.

Paper G   On non-stationary solutions to MSDDEs: representations and the cointegration space. Submitted.

Paper H   Low frequency estimation of Lévy-driven moving averages. Submitted.

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

Up to notation and minor adjustments, Papers A–H align with their published or submitted version. Main parts of Papers A–C are written during the first two years of my PhD studies, and thus these were also included in my progress report used for the qualifying examination, after which I obtained a master's degree in mathematical economics. While a few of the ideas of Papers D and F were briefly discussed in the

progress report as well, Papers D–I are primarily a result of the last two years of my studies. I have contributed comprehensively in both the writing as well as the research phase of Papers A–B and D–H. Together with Victor Rohde I have written Papers C and I, and to these we have contributed equally.

The first chapter of the dissertation is an introduction, which motivates the use of Lévy-driven moving averages in the modeling of continuous-time stochastic systems and discusses the importance of obtaining knowledge of their representations, limit theorems and certain other properties. The findings of Papers A–H deliver answers to many of the questions raised in this discussion, and hence the main results of these papers will also be highlighted in this chapter. Paper I, however, is an industrial collaboration with Vestas Wind Systems A/S and concerns estimation of extreme loads on wind turbines using covariates. Since the details are carefully explained in the included paper and the overall aim differs from that of Papers A–H, I have chosen not to address its findings in the introductory chapter.

My four years of PhD studies have been both challenging and rewarding, and I owe several people huge thanks for making the journey joyful. First of all, I thank my main supervisor Andreas Basse-O'Connor for giving me the unique opportunity of pursuing a PhD degree in a truly inspiring and intellectually stimulating research environment and for our many fruitful discussions. His support, enthusiasm and high ambitions have definitely pushed my limits as a researcher. A special thanks goes to my co-supervisor Jan Pedersen, with whom I have had uncountably many conversations spanning from technical details in proofs and general probabilistic and statistical considerations to an analysis of the outcome of yesterday's hockey match. Due to his extraordinary guidance, his trust in my abilities and his positive mindset, Jan has had a significant impact on my development and well-being during my studies. I feel honored that Andreas and Jan have invested this much time and effort in me—it exceeds by far what could be expected of a supervisor, and for this I am deeply grateful.

I would also like to thank my co-author Richard A. Davis from Department of Statistics, Columbia University, for letting me visit him and his group in New York and for his interest in my research. Our frequent meetings and his generous hospitality ensured that I had a constructive and pleasant stay. I thank as well my office mate Victor Rohde for numerous fruitful collaborations, and the local LaTeX expert Lars 'daleif' Madsen and my office mate Mathias Ljungdahl for helping me with the technical typesetting. Furthermore, I want to thank my colleagues at Department of Mathematics, Aarhus University, for giving me a perfect working environment, which I have enjoyed being a part of throughout my studies. A particular thanks goes to Claudio Heinrich, Julie Thøgersen, Mads Stehr, Mathias Ljungdahl, Patrick Laub, Thorbjørn Grønbæk and Victor Rohde for all the (non-)mathematical discussions and social activities.

Finally, my family and friends deserve an abundance of gratitude for their endless support and encouragement. I conclude with a very special thanks to my fiancée Marianne, since none of this would have been possible without her.

Mikkel Slot Nielsen
Aarhus, July 2019

Summary

Similarly to the discrete-time framework, moving averages driven by white noise processes play a crucial role in the modeling of continuous-time stochastic processes. The main purpose of this dissertation is to address various aspects of Lévy-driven moving averages. The existence of equivalent martingale measures, autoregressive representations and limit theorems will be of particular interest.

Based on earlier literature on the semimartingale property for Lévy-driven moving averages, and under rather general conditions on the Lévy process, we give necessary and sufficient conditions on the driving kernel for an equivalent martingale measure to exist. In particular, these conditions extend previous results for Gaussian moving averages to the symmetric α-stable case with an arbitrary $\alpha \in (1, 2]$.

A significant part of the dissertation concerns various properties of solutions to a range of stochastic delay differential equations (SDDEs). Among other things, we obtain sufficient conditions for existence and uniqueness of solutions to univariate, multivariate, higher order and fractional SDDEs, provide moving average representations of the solutions and discuss its memory properties. A few implications of the obtained results are that (i) invertible continuous-time ARMA processes can be viewed as unique solutions to SDDEs, (ii) solutions can be semimartingales and exhibit long memory at the same time, and (iii) cointegration can be embedded in multivariate SDDEs in a straightforward manner. From the properties that we prove for SDDEs we draw several parallels to classical results for autoregressive representations in the discrete-time literature and, hence, indicate that it may be reasonable to think of SDDEs as the continuous-time counterpart.

We also study the limiting behavior of quadratic forms and related quantities of discretely sampled Lévy-driven moving averages. The linear nature of Lévy-driven moving averages and their tractable probabilistic structure allow us to obtain rather explicit conditions on the driving kernel and the coefficients of the quadratic form ensuring asymptotic normality. The result differs from those obtained in related literature due to the quite delicate interplay between discrete-time sampling and continuous-time convolutions. The applications of these asymptotic results are many; in particular, we demonstrate how they can be used to obtain central limit theorems when estimating the driving kernel parametrically using least squares.

The last part of the dissertation is related to an industrial collaboration, where we consider prediction of extreme loads on wind turbines using only a number of covariates and a simulation tool. In particular, we discuss how to set up a statistical model in this situation, address some of the key assumptions and, finally, check its performance on real-world data.

Resumé

Præcis som for modeller i diskret tid spiller glidende gennemsnit drevet af hvid støj en fundamental rolle i modelleringen af stokastiske processer i kontinuert tid. Hovedformålet med denne afhandling er at undersøge forskellige aspekter af Lévy-drevne glidende gennemsnit. Vi vil være særligt interesserede i eksistens af ækvivalente martingalmål, autoregressive repræsentationer og grænseværdisætninger.

Baseret på tidligere litteratur om semimartingal-egenskaben for Lévy-drevne glidende gennemsnit, og under ret svage antagelser på Lévy-processen, giver vi nødvendige og tilstrækkelige betingelser på den drivende kerne, der sikrer, at et ækvivalent martingalmål eksisterer. Som et specialtilfælde af dette resultat opnår vi en generalisering af resultater for gaussiske glidende gennemsnit til det symmetrisk α-stabile tilfælde for et vilkårligt $\alpha \in (1, 2]$.

En stor del af afhandlingen omhandler forskellige egenskaber ved løsninger til en række stokastiske differentialligninger, som involverer processens egen fortid (disse vil herfra kort refereres til som SDDEer). Vi giver tilstrækkelige betingelser til at sikre eksistens og entydighed af løsninger til en- og flerdimensionale SDDEer, SDDEer af højere orden og fraktionelle SDDEer. Desuden repræsenterer vi løsningerne som glidende gennemsnit samt studerer deres afhængighedsstruktur. Umiddelbare konsekvenser af disse resultater er, at (i) invertible ARMA processer i kontinuert tid er entydige løsninger til SDDEer, (ii) løsninger kan være semimartingaler og have lang hukommelse på samme tid, og (iii) kointegration kan nemt indlejres i de flerdimensionale SDDEer. Fra de beviste egenskaber for SDDEer trækker vi adskillige paralleller til klassiske resultater for autoregressive modeller i diskret tid og indikerer på den måde, at SDDEer kan opfattes som modstykket i kontinuert tid.

Vi studerer også den asymptotiske opførsel af kvadratiske former og relaterede størrelser af diskrete observationer fra Lévy-drevne glidende gennemsnit. De glidende gennemsnits lineære struktur samt transparente fordelingsmæssige egenskaber gør det muligt for os at udlede eksplicitte betingelser på den drivende kerne og koefficienterne i den kvadratiske form, som sikrer asymptotisk normalitet. På grund af det udfordrende samspil mellem diskrete observationer og foldninger i kontinuert tid adskiller resultatet sig fra dem, der er udledt i lignende litteratur. Sådanne asymptotiske resultater har mange anvendelser: f.eks. viser vi, hvordan de kan bruges til at udlede centrale grænseværdisætninger ifm. parametrisk estimation af den drivende kerne ved brug af mindste kvadraters metode.

Afhandlingens sidste del er relateret til et industrielt samarbejde, hvor vi studerer estimation af ekstreme belastninger på vindmøller ved brug af en række kovariater samt et simuleringsværktøj. Her diskuterer vi, hvordan man kan formulere en fornuftig statistisk model samt belyser de væsentligste antagelser. Endeligt undersøger vi, hvordan modellen klarer sig på data fra virkeligheden.

ix

Introduction

This chapter motivates the study of Lévy-driven moving average processes, highlights key results obtained in the included papers and addresses their relation to existing literature. In Section 1 we discuss why Lévy-driven moving averages constitute a convenient class for modeling a wide range of stochastic systems in time by relying on a Wold–Karhunen type decomposition, and we review some of their properties. This leads naturally to a discussion of the key findings of Paper A. Section 2 concerns the specification of the deterministic kernel driving the moving average. Specifically, by drawing parallels to the discrete-time literature on ARMA type equations, we motivate the continuous-time ARMA processes as well as solutions to certain stochastic delay differential equations. These are all special cases of moving averages, which have formed the foundations of Papers B–E and G, and hence we end the section by giving an overview of the main contributions of each of these papers. Finally, in Section 3 we discuss the relevance of limit theorems for quadratic forms and related quantities of Lévy-driven moving averages and relate it to Papers F and H.

1 A Wold–Karhunen type decomposition and the Lévy-driven moving averages

There may be many reasons for modeling stochastic processes continuously in time. To give an example, financial data are nowadays sampled at both very high and irregular frequencies, and the continuous-time specification is a way to model this type of observations in a consistent manner. Another reason is due to the remarkable result of Delbaen and Schachermayer [18], which essentially characterizes arbitrage opportunities in a financial market driven by semimartingales in terms of the existence of a so-called equivalent martingale measure (cf. Paper A). For further examples on the use of continuous-time models, see [1, 8] and [22, Section 1.2].

Suppose now that $(X_t)_{t \in \mathbb{R}}$ is a centered and weakly stationary (continuous-time) process, that is,

$$\mathbb{E}[X_t^2] < \infty, \quad \mathbb{E}[X_t] = 0 \quad \text{and} \quad \left(h \longmapsto \mathbb{E}[X_{t+h}X_t]\right) = \left(h \longmapsto \mathbb{E}[X_h X_0]\right) =: \gamma_X \qquad (1.1)$$

for all $t \in \mathbb{R}$. While some phenomena may be reasonably described by such $(X_t)_{t \in \mathbb{R}}$, one may often need to transform, deseasonalize and/or detrend observations to align with such assumptions (see [11, Section 1.4] for details on this). A classical example is the evolution of a stock price $(S_t)_{t \in \mathbb{R}}$ exhibiting the random walk behavior $\lim_{t \to \infty} \mathrm{Var}(S_t) = \infty$, while its Δ-period log-returns $X_t = \log S_{t+\Delta} - \log S_t$, $t \in \mathbb{R}$, might approximately meet (1.1). A related example is the (log-)prices of two stocks $(S_t^1)_{t \in \mathbb{R}}$ and $(S_t^2)_{t \in \mathbb{R}}$, which individually may wander widely, but the spread $X_t = S_t^1 - S_t^2$, $t \in \mathbb{R}$,

behaves in a stationary manner. Such a situation can happen if the two stocks are very similar by nature and one would in this case refer to them as being *cointegrated* (see also Paper G). Despite the fact that the class of processes satisfying (1.1) is large and general, Theorem 1.1 shows that these conditions are not far from ensuring that, up to a term that can be perfectly predicted from the remote past (in an $L^2(\mathbb{P})$ sense), they correspond to moving averages driven by white noise processes. In the result it will be required that $(X_t)_{t\in\mathbb{R}}$ is continuous in $L^2(\mathbb{P})$ or, equivalently, γ_X is continuous at 0. Under this assumption it follows by Bochner's theorem that there exists a finite and symmetric Borel measure F_X, usually referred to as the spectral distribution of $(X_t)_{t\in\mathbb{R}}$, which has characteristic function γ_X:

$$\gamma_X(h) = \int_{\mathbb{R}} e^{ihy} F_X(\mathrm{d}y), \qquad h \in \mathbb{R}. \tag{1.2}$$

In the formulation, f_X refers to the density of the absolutely continuous part of F_X and $\overline{\mathrm{sp}}$ denotes the $L^2(\mathbb{P})$ closure of the linear span.

Theorem 1.1 (Karhunen [28]). *Suppose that $(X_t)_{t\in\mathbb{R}}$ is a centered and weakly stationary process, which is continuous in $L^2(\mathbb{P})$. Moreover, suppose that the Paley–Wiener condition*

$$\int_{\mathbb{R}} \frac{|\log f_X(y)|}{1+y^2}\, \mathrm{d}y < \infty \tag{1.3}$$

is satisfied. Then there exists a unique decomposition of $(X_t)_{t\in\mathbb{R}}$ as

$$X_t = \int_{-\infty}^{t} g(t-u)\, \mathrm{d}Z_u + V_t, \qquad t \in \mathbb{R}, \tag{1.4}$$

where $g\colon \mathbb{R} \to \mathbb{R}$ belongs to L^2, $(Z_t)_{t\in\mathbb{R}}$ is a process with weakly stationary and orthogonal increments satisfying $\mathbb{E}[(Z_t - Z_s)^2] = t - s$ for all $s < t$, and $(V_t)_{t\in\mathbb{R}}$ is a weakly stationary process with $V_t \in \bigcap_{s\in\mathbb{R}} \overline{\mathrm{sp}}\{X_u : u \le s\}$ for $t \in \mathbb{R}$. Moreover, if F_X is absolutely continuous with a density f_X satisfying (1.3) then $V_t = 0$ for all $t \in \mathbb{R}$.

The stochastic integral in (1.4) is defined as an $L^2(\mathbb{P})$ limit of integrals of simple functions. While the proof of Theorem 1.1 can be found in [28, Satz 5–6], the formulation of the result is borrowed from [5, Theorem 4.1]. It is straightforward to verify that a converse of Theorem 1.1 is also true: if $g \in L^2$ and $(Z_t)_{t\in\mathbb{R}}$ is a process with weakly stationary and orthogonal increments satisfying $\mathbb{E}[(Z_t - Z_s)^2] = t - s$ for $s < t$, then

$$X_t = \int_{-\infty}^{t} g(t-u)\, \mathrm{d}Z_u, \qquad t \in \mathbb{R}, \tag{1.5}$$

satisfies (1.1), and γ_X can be represented as (1.2) with $F_X(\mathrm{d}y) = (2\pi)^{-1}|\mathcal{F}[g](y)|^2\, \mathrm{d}y$. Here $\mathcal{F}[g]$ denotes the Fourier transform of g; we define it as $\mathcal{F}[g](y) = \int_{\mathbb{R}} e^{-ity} g(t)\, \mathrm{d}t$ for $g \in L^1$, $y \in \mathbb{R}$, and extend it to functions in $L^1 \cup L^2$ by Plancherel's theorem.

Loosely speaking, the above considerations show that weakly stationary processes correspond to causal moving averages of the form (1.5) and, thus, it would be natural to focus on modeling g and $(Z_t)_{t\in\mathbb{R}}$. Note that, unless $(X_t)_{t\in\mathbb{R}}$ can be assumed to be Gaussian, in which case $(Z_t)_{t\in\mathbb{R}}$ is a standard Brownian motion, (1.5) does not reveal anything about $(X_t)_{t\in\mathbb{R}}$ beyond its second order properties. In particular, for a general noise process $(Z_t)_{t\in\mathbb{R}}$, the relation (1.5) leaves us with no insight about the path

properties and the probabilistic structure of $(X_t)_{t\in\mathbb{R}}$. For instance, to assess properties of estimators based on $(X_t)_{t\in\mathbb{R}}$, it is necessary to have a better understanding of its dependence structure. This should indicate that, while the overall moving average (convolution) structure can possibly produce a wide class of interesting processes, we should require that $(Z_t)_{t\in\mathbb{R}}$ is a particularly nice process. Natural candidates are provided by the extensively studied class of Lévy processes ([6, 9, 34]), since these will allow us to keep track of the entire distribution of the process while maintaining the same second order properties. Since Lévy processes have stationary and independent increments, the use of these can be seen as the continuous-time equivalent of using i.i.d. noise rather than just uncorrelated noise in a discrete-time setting.

Recall that a one-sided Lévy process $(L_t)_{t\geq 0}$, $L_0 = 0$, is a stochastic process with càdlàg sample paths having stationary and independent increments. These properties imply that $\log\mathbb{E}[\exp\{iyL_t\}] = t\log\mathbb{E}[\exp\{iyL_1\}]$ for $y \in \mathbb{R}$. Consequently, since

$$\psi_L(y) := \log\mathbb{E}\big[e^{iyL_1}\big] = iyb - \frac{1}{2}c^2y^2 + \int_{\mathbb{R}} (e^{iyx} - 1 - iyx\mathbb{1}_{\{|x|\leq 1\}})F(\mathrm{d}x), \qquad y \in \mathbb{R},$$

for some $b \in \mathbb{R}$, $c^2 \geq 0$ and Lévy measure F by the Lévy–Khintchine formula, the distribution of $(L_t)_{t\geq 0}$ may be summarized as a triplet (b, c^2, F). We extend $(L_t)_{t\geq 0}$ to a two-sided Lévy process $(L_t)_{t\in\mathbb{R}}$ by setting $L_t = -\tilde{L}_{(-t)-}$ for $t < 0$, where $(\tilde{L}_t)_{t\geq 0}$ is an independent copy of $(L_t)_{t\geq 0}$. When $\mathbb{E}[|L_1|] < \infty$, or equivalently $\int_{|x|>1} |x|\,F(\mathrm{d}x) < \infty$, we let $\bar{L}_t = L_t - t\mathbb{E}[L_1]$, $t \in \mathbb{R}$, denote the centered version of $(L_t)_{t\in\mathbb{R}}$.

For a measurable function $g \colon \mathbb{R} \to \mathbb{R}$, which vanishes on $(-\infty, 0)$, necessary and sufficient conditions on (b, c^2, F, g) for the Lévy-driven moving average

$$X_t = \int_{-\infty}^{t} g(t - u)\,\mathrm{d}L_u, \qquad t \in \mathbb{R}, \tag{1.6}$$

to exist (as limits in probability of integrals of simple functions) are given in [31, Theorem 2.7]. It follows as well from [31] that the finite dimensional distributions of $(X_t)_{t\in\mathbb{R}}$ are characterized in terms of (b, c^2, F, g) by the relation

$$\log\mathbb{E}\big[e^{i(y_1 X_{t_1} + \cdots + y_n X_{t_n})}\big] = \int_{\mathbb{R}} \psi_L(y_1 g(t_1 + u) + \cdots + y_n g(t_n + u))\,\mathrm{d}u,$$

which holds for any $n \in \mathbb{N}$ and $t_1, y_1, \ldots, t_n, y_n \in \mathbb{R}$. One immediate consequence of this relation is that $(X_t)_{t\in\mathbb{R}}$ is a stationary and infinitely divisible stochastic process (the finite dimensional distributions of $(X_{t+h})_{t\in\mathbb{R}}$ are infinitely divisible and do not depend on h). Note that, in contrast to (1.5), $(X_t)_{t\in\mathbb{R}}$ given by (1.6) needs not satisfy (1.1), e.g., it may allow for a heavy-tailed marginal distribution. For instance, if L_1 has a symmetric α-stable distribution for some $\alpha \in (0, 2)$, then (1.6) is well-defined if and only if $g \in L^\alpha$, in which case the distribution of X_0 is also symmetric α-stable ([33, Propositions 6.2.1–6.2.2]). In particular, for $p \in (0, \infty)$ it holds that $\mathbb{E}[|X_0|^p] < \infty$ if and only if $p < \alpha$ ([33, Property 1.2.16]). While the class of Lévy-driven moving averages is rather large, it should be pointed out that more general specifications of stationary infinitely divisible processes, such as mixed moving averages (in particular, superpositions of Ornstein–Uhlenbeck processes) and Lévy semistationary processes have also received some attention in the literature; see [3, 4] for details.

The path properties of $(X_t)_{t\in\mathbb{R}}$ are very much related to those of g, and for details beyond the following discussion we refer to [32]. A fundamental question to ask

is when $(X_t)_{t\geq 0}$ is a semimartingale (with respect to a suitable filtration). Indeed, Delbaen and Schachermayer [18] argue that the semimartingale property is desirable when modeling financial markets, and by Bichteler–Dellacherie theorem it is necessary and sufficient that $(X_t)_{t\geq 0}$ is a semimartingale if it is supposed to serve as a "good" integrator (see [10, Theorem 7.6] and [19] for precise statements). Under rather mild conditions on the driving Lévy process $(L_t)_{t\in\mathbb{R}}$, [7, Corollary 4.8] provides a complete characterization of the semimartingale property within the moving average framework (1.6):

Theorem 1.2 (Basse-O'Connor and Rosiński [7]). *Suppose that $(L_t)_{t\in\mathbb{R}}$ has sample paths of locally unbounded variation and that either $x \mapsto F((-x,x)^c)$ is regularly varying at ∞ of index $\beta \in [-2,-1)$ or $\int_{|x|>1} x^2 F(\mathrm{d}x) < \infty$. Then $(X_t)_{t\geq 0}$ defined as in (1.6) is a semimartingale with respect to the least filtration $(\mathscr{F}_t)_{t\geq 0}$ satisfying the usual conditions and*

$$\sigma(L_s : s \leq t) \subseteq \mathscr{F}_t, \qquad t \geq 0,$$

if and only if g is absolutely continuous on $[0,\infty)$ with a density g' satisfying

$$\int_0^\infty \left(c^2 g'(t)^2 + \int_\mathbb{R} \left(|xg'(t)| \wedge |xg'(t)|^2 \right) F(\mathrm{d}x) \right) \mathrm{d}t < \infty. \tag{1.7}$$

Furthermore, if (1.7) is satisfied, $(X_t)_{t\geq 0}$ admits the semimartingale decomposition

$$X_t = X_0 + g(0)\bar{L}_t + \int_0^t \left(\int_{-\infty}^s g'(s-u)\,\mathrm{d}\bar{L}_u \right) \mathrm{d}s, \qquad t \geq 0. \tag{1.8}$$

If Theorem 1.2 is applicable we have that $\mathbb{E}[|L_1|] < \infty$, and it follows that $(X_t)_{t\geq 0}$ can be decomposed into a sum of a martingale and an absolutely continuous stochastic process (in fact, this implies that $(X_t)_{t\geq 0}$ is a so-called special semimartingale as defined in [26, Definition 4.21]). Sometimes, such as when pricing derivatives or fixed income securities in a financial market driven by semimartingales, it might be important to know if the latter term can be absorbed by a suitable equivalent change of measure. To be precise, for a given $T \in (0,\infty)$ one asks if there is a probability measure $\mathbb{Q}$ on $\mathscr{F}_T$ such that:

  (i) For all $A \in \mathcal{F}_T$, $\mathbb{Q}(A) > 0$ if and only if $\mathbb{P}(A) > 0$.

  (ii) Under $\mathbb{Q}$, $(X_t)_{t\in[0,T]}$ is a local martingale with respect to $(\mathscr{F}_t)_{t\in[0,T]}$.

Such $\mathbb{Q}$ is referred to as an *equivalent local martingale measure* (ELMM) for $(X_t)_{t\in[0,T]}$. It should be mentioned that equivalent or, more generally, absolutely continuous change of measure for some stochastic processes (such as Markov processes and solutions to certain stochastic differential equations) is well-studied; see the introduction of Paper A for references. While it might be tempting to require that $T = \infty$ (with $\mathscr{F}_\infty = \bigvee_{t\geq 0} \mathscr{F}_t$), this is a rather serious restriction. As an example, the probability measures induced by two homogeneous Poisson processes with different intensities (on the space $D([0,\infty))$ equipped with the Skorokhod topology) are equivalent on $\mathscr{F}_T$ for any $T \in (0,\infty)$ but singular on $\mathscr{F}_\infty$, cf. [17, Remark 9.2]. The intuition is that when one has an infinite horizon, the intensity can be estimated almost surely from the Poisson process. Consequently, we will return to the question of the existence of an ELMM for $(X_t)_{t\in[0,T]}$ when fixing $T \in (0,\infty)$.

Recall that it is a prerequisite that $(X_t)_{t \in [0,T]}$ is a semimartingale in order to admit an ELMM ([26, Theorem 3.13 (Chapter III)]). This means that if $(L_t)_{t \in \mathbb{R}}$ is a Lévy process satisfying the assumptions of Theorem 1.2, the conditions imposed on g in this theorem are necessary. Except in trivial cases it must also be the case that $g(0) \neq 0$; indeed, if $g(0) = 0$ and an ELMM $\mathbb{Q}$ exists, the representation (1.8) shows that

$$\int_0^t \left(\int_{-\infty}^s g'(s-u)\, d\bar{L}_u \right) ds, \qquad t \in [0,T],$$

is a local martingale under $\mathbb{Q}$, and hence it must be identically equal to zero ([20, Theorem 3.3 (Section 2)]). If the distribution of L_1 is not degenerate, this happens only if g is vanishing almost everywhere. On the other hand, Cheridito [14] showed that if $(L_t)_{t \in \mathbb{R}}$ is a Brownian motion (that is, $c^2 > 0$ and $F \equiv 0$), the condition $g(0) \neq 0$ combined with the assumptions of Theorem 1.2 are also sufficient for the existence of an ELMM. The main purpose of Paper A has been to establish conditions ensuring that $(X_t)_{t \in [0,T]}$ admits an ELMM beyond the Gaussian setting.

1.1 Paper A

Inspired by the structure of $(X_t)_{t \in [0,T]}$ in (1.8), this paper investigates when an ELMM exists for semimartingales of the form

$$\tilde{X}_t = L_t + \int_0^t Y_s\, ds, \qquad t \in [0,T],$$

under the assumption that $(Y_t)_{t \in [0,T]}$ is a predictable process such that $\int_0^T |Y_t|\, dt < \infty$ almost surely and $\mathbb{E}[|L_1|] < \infty$. In Theorem 2.1 (Paper A) we give rather explicit sufficient conditions for $(\tilde{X}_t)_{t \in [0,T]}$ to admit an ELMM. Specifically, each of the following two statements is sufficient:

(i) The collection $(Y_t)_{t \in [0,T]}$ is tight, each Y_t is infinitely divisible and the corresponding Lévy measures $(F_t)_{t \in [0,T]}$ meet $\sup_{t \in [0,T]} F_t([-K,K]^c) = 0$ for some $K > 0$. Moreover, the Lévy measure F of $(L_t)_{t \in [0,T]}$ satisfies $F((-\infty,0)), F((0,\infty)) > 0$.

(ii) The Lévy measure F of $(L_t)_{t \in [0,T]}$ satisfies $F((-\infty,-K]), F([K,\infty)) > 0$ for all $K > 0$.

The somewhat canonical example of a process $(Y_t)_{t \in [0,T]}$ satisfying (i) is a stationary and infinitely divisible process where the Lévy measure of Y_0 is compactly supported. More concretely, it could be a moving average with a bounded kernel driven by a Lévy process with a compactly supported Lévy measure. Loosely speaking, (ii) states that no further assumptions on $(Y_t)_{t \in [0,T]}$ are needed as long as $(L_t)_{t \in [0,T]}$ can have jumps of arbitrarily large positive and negative size. As an almost immediate consequence of these findings and Theorem 1.2 above, we obtain a quite general result on the existence of an ELMM for $(X_t)_{t \in [0,T]}$ given by (1.6); see Theorem 1.2 of Paper A for details. Among other things, this result implies that if $(L_t)_{t \in \mathbb{R}}$ is a symmetric α-stable Lévy process for some $\alpha \in (1,2]$, then there exists an ELMM for $(X_t)_{t \in [0,T]}$ if and only if $g(0) \neq 0$ and g is absolutely continuous on $[0,\infty)$ with a density g' which belongs L^α (cf. Corollary 1.3 of Paper A). Consequently, this result provides a natural extension of the Gaussian setup studied in [14].

It should be stressed that the techniques used in [14] cannot be transferred into the non-Gaussian setting that we consider in this paper. Specifically, his proof relies on a localized version of the Novikov condition by showing that

$$\mathbb{E}\left[\exp\left\{\frac{1}{2}\int_s^t Y_u^2\,\mathrm{d}u\right\}\right] < \infty \tag{1.9}$$

as long as $t - s \in (0, \delta)$ for a $\delta > 0$ sufficiently small. While this can be verified in a Gaussian setup, such a requirement is rarely satisfied in other situations. In fact, if $\int_s^t Y_u\,\mathrm{d}u$ is infinitely divisible with a non-trivial Lévy measure, (1.9) will never be satisfied ([34, Theorem 26.1]). The conditions (i)–(ii) above are instead results of two alternative and very different techniques. Indeed, (i) makes use of a general predictable criterion of Lépingle and Mémin [29], and (ii) is obtained by carefully constructing $\mathbb{Q}$ so that it changes the distribution of the large jumps of $(L_t)_{t\in[0,T]}$, but leaves the jump intensity constant and thereby avoiding finite explosion times.

2 Dynamic models for Lévy-driven moving averages

While the Lévy-driven moving averages define a rather flexible and tractable class of stationary continuous-time processes, we are still left with the question: *What are reasonable choices of the kernel g?* It may be desirable to choose g so that $(X_t)_{t\in\mathbb{R}}$ exhibits a certain autoregressive (dynamic) behavior. Since autoregressive and moving average representations have different advantages, one would often aim at getting parsimonious representations in both domains without losing too much flexibility—e.g., in terms of possible autocovariances or, equivalently, spectral distributions that can be generated by the model.

Motivation: To make the above discussion more concrete, let us take a step back and consider the discrete-time equations

$$Y_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} \quad \text{and} \quad \sum_{j=0}^{\infty} \pi_j Y_{t-j} = \varepsilon_t, \qquad t \in \mathbb{Z}, \tag{2.1}$$

for suitable sequences of coefficients $(\psi_t)_{t\in\mathbb{N}_0}$ and $(\pi_t)_{t\in\mathbb{N}_0}$, and an i.i.d. noise $(\varepsilon_t)_{t\in\mathbb{Z}}$. Some choices of $(\psi_t)_{t\in\mathbb{N}_0}$ lead to a stationary moving average $(Y_t)_{t\in\mathbb{Z}}$, defined by the first equation of (2.1), which satisfies the second equation of (2.1) for a suitable choice of $(\pi_t)_{t\in\mathbb{N}_0}$. Conversely, for some choices of $(\pi_t)_{t\in\mathbb{N}_0}$ the second equation of (2.1) has a unique stationary solution given by the first equation of (2.1) with a suitably chosen sequence $(\psi_t)_{t\in\mathbb{N}_0}$. We will refer to the first and second equation of (2.1) as a *moving average representation* and an *autoregressive representation* of $(Y_t)_{t\in\mathbb{Z}}$, respectively. While a moving average representation is convenient for assessing several distributional properties of $(Y_t)_{t\in\mathbb{Z}}$, an autoregressive representation provides a lot of valuable insight concerning the dynamic behavior of $(Y_t)_{t\in\mathbb{Z}}$; e.g., it can be used for prediction and estimation purposes, to simulate sample paths or to filter out the noise $(\varepsilon_t)_{t\in\mathbb{Z}}$ from the observed process $(Y_t)_{t\in\mathbb{Z}}$.

There is no guarantee that a simple moving average representation leads to a particularly simple autoregressive representation and vice versa. However, an extremely popular modeling class in discrete time, which allows for rather tractable

representations in both domains, consists of the causal and invertible ARMA processes. Specifically, given two real polynomials P and Q with no zeroes on the unit disc $\mathcal{D} := \{z \in \mathbb{C} : |z| \leq 1\}$, the corresponding ARMA process $(Y_t)_{t \in \mathbb{Z}}$ is the unique stationary solution to the linear difference equation

$$P(B)Y_t = Q(B)\varepsilon_t, \qquad t \in \mathbb{Z}. \tag{2.2}$$

Here B denotes the backward shift operator. In this case, $(\psi_t)_{t \in \mathbb{N}_0}$ and $(\pi_t)_{t \in \mathbb{N}_0}$ correspond to the coefficients in the power series expansion on $\mathcal{D}$ of the rational functions Q/P and P/Q, respectively. The difficulty of computing the coefficients depends ultimately on the denominator polynomial, and hence there is a tradeoff between the simplicity of the moving average and the autoregressive specification. One advantage of the ARMA framework, however, is that the coefficients can always be obtained by relying on simple properties of the geometric series and, possibly, a partial fraction decomposition. An easy example is the AR(1) process where $P(z) = 1 - \alpha z$ for some $\alpha \in (-1, 1)$ and $Q \equiv 1$. In this case $\pi_0 = 1$, $\pi_1 = -\alpha$ and $\pi_j = 0$ for $j \geq 2$, while $\psi_j = \alpha^j$ for all $j \geq 0$. There exists a vast amount of literature related to ARMA processes and various extensions. For further details, see [11, 25].

Continuous-time ARMA equations: Since the coefficients in the moving average representation of the discrete-time AR(1) process take a geometric form, the continuous-time equivalent is naturally $g(t) = e^{-\lambda t}$ for $t \geq 0$ and a given $\lambda > 0$. The corresponding process $(X_t)_{t \in \mathbb{R}}$ given by (1.6), known as the Ornstein–Uhlenbeck process, is perhaps the most well-studied Lévy-driven moving average of all time, and it can be characterized as the unique stationary solution to the stochastic differential equation

$$X_t - X_s = -\lambda \int_s^t X_u \, \mathrm{d}u + L_t - L_s, \qquad s < t. \tag{2.3}$$

Ornstein–Uhlenbeck processes enjoy many properties: they are Markovian, their possible one-dimensional marginal laws coincide with the self-decomposable distributions and a sampled Ornstein–Uhlenbeck process $(X_{t\Delta})_{t \in \mathbb{Z}}$ is an AR(1) process for any $\Delta > 0$. For details about Ornstein–Uhlenbeck processes and further references, see Section 1 of Paper B.

Defining formally the derivatives $(DX_t)_{t \in \mathbb{R}}$ and $(DL_t)_{t \in \mathbb{R}}$ of $(X_t)_{t \in \mathbb{R}}$ and $(L_t)_{t \in \mathbb{R}}$, respectively, (2.3) reads $(D + \lambda)X_t = DL_t$ for $t \in \mathbb{R}$. In light of this equation and (2.2) it makes sense to view a process $(X_t)_{t \in \mathbb{R}}$ as a continuous-time ARMA (CARMA) process if it is stationary and satisfies the formal equation

$$P(D)X_t = Q(D)DL_t, \qquad t \in \mathbb{R}, \tag{2.4}$$

for two real polynomials P and Q. Although the derivatives on the right-hand side will not be well-defined in the usual sense (except in trivial cases), $(X_t)_{t \in \mathbb{R}}$ is defined rigorously through its corresponding moving average representation. Specifically, by assuming that $p := \deg(P)$ and $q := \deg(Q)$ satisfy $p > q$ and that P has no zeroes on $\{z \in \mathbb{C} : \mathrm{Re}(z) \geq 0\}$, there exists a function $g \colon \mathbb{R} \to \mathbb{R}$ which vanishes on $(-\infty, 0)$ and has Fourier transform

$$\mathcal{F}[g](y) = \frac{Q(iy)}{P(iy)}, \qquad y \in \mathbb{R}.$$

As for the ARMA processes, the rational form of the Fourier transform ensures that one can compute g explicitly by relying on the fact that $t \mapsto \mathbb{1}_{[0,\infty)}(t)e^{-\lambda t}$ has Fourier transform $y \mapsto (iy + \lambda)^{-1}$ for any $\lambda > 0$. This construction ensures that g is absolutely continuous on $[0,\infty)$ and decays exponentially fast at ∞, and hence the *causal* CARMA(p,q) process with polynomials P and Q can be rigorously defined as the moving average (1.6) with kernel g as long as $\mathbb{E}[\log_+|L_1|] < \infty$. On a heuristic level, one can apply the Fourier transform on both sides of the equation (2.4) and rearrange terms in order to reach the conclusion that a CARMA process should have such a moving average representation. For applications and properties of the CARMA process as well as details about its definition, see Sections 1 and 4.3 of Paper D and references therein.

Continuous-time autoregressive representations: To sum up, the continuous-time version of the moving average representation in (2.1) is the Lévy-driven moving average (1.6), and the ARMA equation (2.2) may naturally be interpreted as (2.4), which in turn leads to the CARMA processes that have a fairly tractable kernel g. Still, when comparing to the discrete-time setup, some questions arise immediately:

   (i) *What is an autoregressive representation in continuous time?*

   (ii) *Which types of moving averages admit such a representation?*

   (iii) *Does the CARMA process admit an autoregressive representation and is it particularly simple?*

Suppose that $\mathbb{E}[L_1] = 0$ and $\mathbb{E}[L_1^2] < \infty$. For a process $(X_t)_{t \in \mathbb{R}}$ with $\mathbb{E}[X_0] = 0$ and $\mathbb{E}[X_t^2] < \infty$ to admit an autoregressive representation it seems reasonable to require that

$$\overline{\mathrm{sp}}\{X_u : u \le t\} \supseteq \overline{\mathrm{sp}}\{L_t - L_s : s \le t\}, \qquad t \in \mathbb{R}. \tag{2.5}$$

When $(X_t)_{t \in \mathbb{R}}$ is of the moving average form (1.6) for some $g \in L^2$ which is vanishing on $(-\infty, 0)$, the reverse inclusion of (2.5) is always satisfied and equality holds if and only if $\mathcal{F}[g]$ is a so-called outer function ([21, pp. 94–95]). While there exist conditions ensuring that a function is outer, these are often not easy to check and, more importantly, in many situations the recipe for going from $(X_u)_{u \le t}$ to $L_t - L_s$ is not clear. Instead, we take the opposite standpoint and define a class of processes by an autoregressive type of equations, such that this transition is simple and transparent. Of course, then we need to argue that it contains a sufficiently wide class of moving averages—ideally, to align with the discrete-time representations, the invertible CARMA processes should form a particularly nice subclass. The relation between this class of autoregressions and moving averages should be somewhat as depicted in Figure 1.

    The class of interest will be solutions to the so-called *stochastic delay differential equations* (SDDEs), which in the simplest case (univariate, first order and non-fractional) are of the form

$$X_t - X_s = \int_s^t \int_{[0,\infty)} X_{u-v}\, \eta(\mathrm{d}v)\, \mathrm{d}u + L_t - L_s, \qquad s < t. \tag{2.6}$$

Here η is a finite signed measure and $(X_t)_{t \in \mathbb{R}}$ is a measurable process such that the integral in (2.6) is well-defined almost surely for each $s < t$. Among other things, the
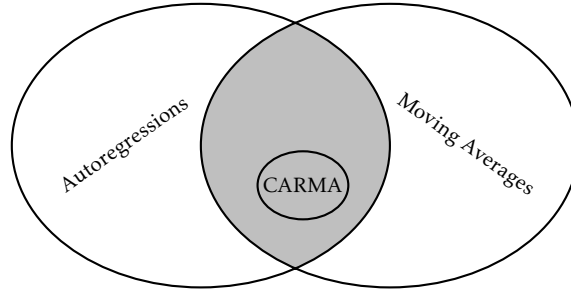
Figure 1: Invertible and causal CARMA processes being a strict subset of processes which both admit an autoregressive representation and a moving average representation.

purpose of Papers B–E and G has been to address each of the questions (i)–(iii) in frameworks related to (2.6) and show that many properties of the solutions are akin to those of discrete-time autoregressions. Depending on the paper, different assumptions are put on $(X_t)_{t\in\mathbb{R}}$ in order to ensure that the integral in (2.6) is well-defined. For now let us just remark that each of the following three conditions is sufficient: (i) η is compactly supported and $t \mapsto X_t$ is càdlàg, (ii) $(X_t)_{t\in\mathbb{R}}$ is stationary and $\mathbb{E}[|X_0|] < \infty$, and (iii) $(X_t)_{t\in\mathbb{R}}$ has stationary increments, $\mathbb{E}[|X_t|] < \infty$ for all t and $\int_{[0,\infty)} t\,|\eta|(\mathrm{d}t) < \infty$ (the latter condition is due to [5, Corollary A.3]). One of the simplest SDDEs is the Ornstein–Uhlenbeck equation (2.3), which corresponds to $\eta = -\lambda\delta_0$ with δ_0 being the Dirac measure at 0. The literature has primarily focused on the case where η is compactly supported (cf. [24, 30]), but as we shall see in Paper D, this restriction unfortunately rules out the possibility of representing CARMA processes with a non-trivial moving average polynomial as solutions to SDDEs. To the best of our knowledge, SDDEs have historically not been viewed as continuous-time equivalents to discrete-time autoregressive representations, and hence questions such as (i)–(iii) have not been raised.

Before jumping into technical descriptions of the attached papers on SDDEs, we will briefly comment on their scopes. Papers B and D address existence and uniqueness of stationary solutions to (2.6), also when the noise is much more general than $(L_t)_{t\in\mathbb{R}}$, and in Paper D the results are shown to hold true in a multidimensional and higher order setting as well. Moreover, Paper E defines a large class of fractional delays which all give rise to stationary solutions that are semimartingales and have hyperbolically decaying autocovariance functions. While the equations considered in this paper do indeed take the form (2.6) in special cases, the general framework is different and specifically tailored for producing long-memory processes. Finally, Paper G studies existence and uniqueness of solutions which are not necessarily stationary, but have stationary increments, in the same type of multivariate setting as in Paper D, and it characterizes the space of the corresponding cointegration vectors. In general, the papers draw clear parallels to well-known discrete-time models such as the fractionally integrated ARMA model and the cointegrated VAR model.

Besides whether we consider a univariate or multivariate version of (2.6), there is another factor discriminating the papers: to find solutions to (2.6) using Papers B–D we *must* have that $\eta([0,\infty)) \neq 0$, while Papers E and G sometimes apply in cases where $\eta([0,\infty)) = 0$. The condition $\eta([0,\infty)) = 0$ corresponds to the autoregressive polynomial having a zero at $z = 1$ in a discrete-time setting, and it is closely related

to memory and stationarity properties of the solution. Table 1 gives an overview of the focus in each of the five papers on SDDEs.

Table 1: An overview of the five papers on SDDEs.

	Univariate	Multivariate
$\eta([0,\infty)) \neq 0$	B, C	D
$\eta([0,\infty)) = 0$	E	G

2.1 Papers B and D

Papers B and D are very much related in the sense that the latter extends the former to a multivariate framework, and questions such as existence and uniqueness of stationary solutions are addressed in both papers. Despite this, they still have fairly different aims:

(i) Paper B also contains a study of an alternative type of autoregressive representation than the SDDE and many examples are provided.

(ii) Paper D is generally more technical and is also concerned with representations of solutions, prediction formulas, higher order SDDEs and their relation to invertible CARMA processes.

Here we will briefly discuss the main findings of the two papers, but only formulate them in the univariate setting. The multivariate extension is more demanding from a notational point of view and, thus, we refer to Paper D for further details. The majority of the proofs in Papers B and D rely on the idea of rephrasing the problems in the frequency domain and then exploiting key results from harmonic analysis, such as certain Paley–Wiener theorems and characterizations of Hardy spaces, to establish the existence of the appropriate functions.

The equation of interest is (2.6) with a more general noise and of higher order, namely

$$X_t^{(m-1)} - X_s^{(m-1)} = \sum_{j=0}^{m-1} \int_s^t \int_{[0,\infty)} X_{u-v}^{(j)} \, \eta_j(\mathrm{d}v) \, \mathrm{d}u + Z_t - Z_s, \qquad s < t. \qquad (2.7)$$

where $(Z_t)_{t\in\mathbb{R}}$ is a measurable process with stationary increments, $Z_0 = 0$ and $\mathbb{E}[|Z_t|] < \infty$ for all $t \in \mathbb{R}$. Here $m \in \mathbb{N}$, the measures $\eta_0, \eta_1, \ldots, \eta_{m-1}$ are finite and signed, and $(X_t^{(j)})_{t\in\mathbb{R}}$ denotes the jth derivative of $(X_t)_{t\in\mathbb{R}}$ with respect to t. For convenience, we will assume that $(Z_t)_{t\in\mathbb{R}}$ is a *regular integrator* in the sense of Proposition 4.1 (Paper D). For now it suffices to know that a regular integrator ensures that the solutions we construct can be expressed as moving averages and that Lévy processes, fractional Lévy processes and many semimartingales with stationary increments are regular integrators. It should be stressed that existence and uniqueness of solutions to (2.7) can still be obtained when $(Z_t)_{t\in\mathbb{R}}$ is not a regular integrator; see Theorem 2.5 of Paper B and Theorem 3.1 of Paper D for the case $m = 1$.

As discussed in relation to Table 1, we need to impose conditions ensuring that $\sum_{j=0}^{m-1} \eta_j([0,\infty)) \neq 0$ in order to prove existence and uniqueness of stationary solutions

to (2.7). Specifically, it is assumed that $\int_{[0,\infty)} t^2 |\eta_j|(\mathrm{d}t) < \infty$ for $j = 0, 1, \ldots, m-1$ and that the equation

$$h_\eta(z) := z^m - \sum_{j=0}^{m-1} z^j \int_{[0,\infty)} e^{-zt} \eta_j(\mathrm{d}t) = 0 \tag{2.8}$$

has no solutions on the imaginary axis $\{z \in \mathbb{C} : \mathrm{Re}(z) = 0\}$. Here $|\eta_j|$ denotes the variation of η_j. Theorem 4.5 (Paper D) states that, under these assumptions, the unique stationary solution to (2.7) is given by

$$X_t = \int_{\mathbb{R}} g(t-u) \, \mathrm{d}Z_u, \qquad t \in \mathbb{R}. \tag{2.9}$$

where $g \colon \mathbb{R} \to \mathbb{R}$ can be characterized through its Fourier transform as $\mathcal{F}[g](y) = h_\eta(iy)^{-1}$ for $y \in \mathbb{R}$. Note that $\mathcal{F}[g]$ is well-defined due to the imposed assumption on h_η. Here uniqueness means that for any other measurable and stationary process $(X_t)_{t \in \mathbb{R}}$ which has $\mathbb{E}[|X_0|] < \infty$ and satisfies (2.7), the equality in (2.9) holds true almost surely for each $t \in \mathbb{R}$. It follows that $(X_t)_{t \in \mathbb{R}}$ is a backward moving average of the form (1.5) if g is vanishing on $(-\infty, 0)$ almost everywhere, and this is the case if the equation in (2.8) has no solutions on $\{z \in \mathbb{C} : \mathrm{Re}(z) \geq 0\}$.

The last result addressed here concerns the possibility of representing CARMA processes as unique solutions to certain SDDEs. Hence, we consider any two real and monic polynomials P and Q with corresponding degrees $p > q$, and we assume that P has no zeroes in $\{z \in \mathbb{C} : \mathrm{Re}(z) = 0\}$ and does not share any zeroes with Q. Moreover, we let $(X_t)_{t \in \mathbb{R}}$ be given by (2.9) with $\mathcal{F}[g](y) = Q(iy)/P(iy)$ for $y \in \mathbb{R}$. This setup covers in particular the causal Lévy-driven CARMA process introduced above when $\mathbb{E}[|L_1|] < \infty$, but also more general CARMA frameworks as discussed in Section 4.3 (Paper D). In line with discrete-time ARMA processes we need an invertibility assumption in order to obtain an autoregressive representation, and this amounts in turn to assuming that the zeroes of Q do not belong to $\{z \in \mathbb{C} : \mathrm{Re}(z) \geq 0\}$. Note that this is exactly what is needed for g to be outer (see [21, Exercise 2 (Section 2.7)]), which is necessary and sufficient for (2.5) to hold when $\mathbb{E}[L_1] = 0$ and $\mathbb{E}[L_1^2] < \infty$. While the rational function P/Q was the key ingredient in order to obtain an autoregressive representation of ARMA processes in a discrete-time setup, the continuous-time SDDE setup requires a decomposition of P. Specifically, we decompose P as

$$P = QR + S,$$

where R and S are polynomials such that $\deg(R) = p - q$ and $\deg(S) < q$ ($S \equiv 0$ if $q = 0$). Such a decomposition is unique and can be obtained using polynomial long division. Set $m = p - q$ and write

$$R(z) = z^m - c_{m-1} z^{m-1} - \cdots - c_0, \qquad z \in \mathbb{C},$$

for suitable $c_0, \ldots, c_{m-1} \in \mathbb{R}$. The essence of Theorem 4.8 (Paper D) is that $(X_t)_{t \in \mathbb{R}}$ is the unique stationary solution to (2.7) when

$$\eta_0(\mathrm{d}t) = c_0 \delta_0(\mathrm{d}t) + f(t) \, \mathrm{d}t \quad \text{and} \quad \eta_j = c_j \delta_0, \qquad j = 1, \ldots, m-1, \tag{2.10}$$

where $f \colon \mathbb{R} \to \mathbb{R}$ is vanishing on $(-\infty, 0)$ and characterized by $\mathcal{F}[f](y) = S(iy)/Q(iy)$ for $y \in \mathbb{R}$. One should notice here that, similarly to computing the coefficients in the

autoregressive representation of an ARMA process, writing up the SDDE associated to a particular CARMA process reduces to finding a function with a certain rational Fourier transform.

2.2 Paper C

Inspired by the study of Brockwell et al. [12], the purpose of this paper is to carry out a simulation study, which is designed to check the possibility of using SDDEs to filter out (or recover) the noise process from an observed invertible Lévy-driven CARMA(2, 1) process $(X_t)_{t\in\mathbb{R}}$. Specifically, the results of Papers B and D ensure the existence of $\alpha, \beta \in \mathbb{R}$ and $\gamma > 0$ such that

$$\mathrm{d}X_t = \alpha X_t \, \mathrm{d}t + \beta \int_0^\infty e^{-\gamma u} X_{t-u} \, \mathrm{d}u \, \mathrm{d}t + \mathrm{d}L_t, \qquad t \in \mathbb{R}, \tag{2.11}$$

so by observing $(X_t)_{t\in\mathbb{R}}$ on a sufficiently fine grid the distribution of L_1 is estimated by discretizing (2.11). Before this step we estimate the vector (α, β, γ) of parameters by a least squares approach. We refer to Sections 3 and 4 (in particular, Figures 2 and 3) of Paper C for further details.

2.3 Paper E

This paper is concerned with the question of incorporating long memory into the solutions of equations of a similar type as the SDDE in (2.6) when $\mathbb{E}[L_1] = 0$ and $\mathbb{E}[L_1^2] = 1$. The notion of long memory refers in this context to a certain asymptotic behavior of either the autocovariance function γ_X or, if it exists, the spectral density f_X of the solution $(X_t)_{t\in\mathbb{R}}$, namely that

$$\gamma_X(h) \sim \alpha h^{2\beta-1} \quad \text{as } h \to \infty \qquad \text{or} \qquad f_X(y) \sim \alpha|y|^{-2\beta} \quad \text{as } y \to 0 \tag{2.12}$$

for some $\alpha > 0$ and $\beta \in (0, 1/2)$. Here, and in what follows, we use the notation $f(t) \sim g(t)$ for two functions $f, g\colon \mathbb{R} \to \mathbb{C}$ to indicate that $f(t)/g(t) \to 1$ for t tending to some specified limit. By a Tauberian argument, the two conditions in (2.12) are equivalent under suitable regularity conditions. Recall that, under the assumptions of Papers B and D, the unique solution to (2.6) is a moving average driven by $(L_t)_{t\in\mathbb{R}}$ with a kernel g satisfying $\mathcal{F}[g](y) = (iy - \mathcal{F}[\eta](y))^{-1}$ for $y \in \mathbb{R}$. It is not too difficult to verify that $g \in L^1 \cap L^2$ (Lemma 2.2 of Paper B) and $f_X(y) = (2\pi)^{-1}|iy - \mathcal{F}[\eta](y)|^{-2}$ (Plancherel's theorem), and hence the solution does not possess any of the properties in (2.12).

The general equation considered in Paper E is

$$X_t - X_s = \int_{-\infty}^t \big(D_-^\beta \mathbb{1}_{(s,t]}\big)(u) \int_{[0,\infty)} X_{u-v} \, \eta(\mathrm{d}v) \, \mathrm{d}u + L_t - L_s, \qquad s < t, \tag{2.13}$$

where

$$\big(D_-^\beta \mathbb{1}_{(s,t]}\big)(u) = \frac{1}{\Gamma(1-\beta)}\big[(t-u)_+^{-\beta} - (s-u)_+^{-\beta}\big], \qquad u \in \mathbb{R},$$

is the right-sided Riemann–Liouville fractional derivative of the indicator function $\mathbb{1}_{(s,t]}$. While solutions to (2.13) may indeed be viewed as solutions to (2.6) in some

cases (see Example 4.5 of Paper E), (2.13) is generally better suited for studying long-memory processes. To motivate this statement, note that while both (2.6) and (2.13) can be written as

$$X_t - X_s = \int_0^\infty X_{t-u}\, \mu_{t-s}(\mathrm{d}u) + L_t - L_s, \qquad s < t, \tag{2.14}$$

for a suitable family of finite measures $(\mu_h)_{h>0}$, it can be checked that, as $y \to 0$ and for each $h > 0$, $\mathcal{F}[\mu_h](y) \sim h\eta([0,\infty))$ in the former case and $\mathcal{F}[\mu_h](y) \sim h\eta([0,\infty))(iy)^\beta$ in the latter case. When also keeping in mind that the autoregressive coefficients $(\pi_j)_{j\in\mathbb{N}_0}$ of discrete-time fractional (ARFIMA type) processes satisfy $\sum_{j=0}^\infty \pi_j e^{-ijy} \sim \alpha(iy)^\beta$ as $y \to 0$ for some $\alpha > 0$ (see, e.g., [11, Section 13.2]), this should indicate that (2.13) might be well-suited for the construction of long-memory processes.

In order to show existence and uniqueness of solutions to (2.13) it is assumed that $\int_{[0,\infty)} t|\eta|(\mathrm{d}t) < \infty$ and that the equation

$$h_{\eta,\beta}(z) := z^{1-\beta} - \int_{[0,\infty)} e^{-zt}\, \eta(\mathrm{d}t) = 0$$

has no solution $z \in \mathbb{C}$ with $\mathrm{Re}(z) \geq 0$. Here we define z^γ as $r^\gamma e^{i\gamma\theta}$, where $r > 0$ and $\theta \in (-\pi, \pi]$ correspond to the polar representation $z = re^{i\theta}$ of $z \in \mathbb{C} \setminus \{0\}$. Theorem 3.2 (Paper E) shows that these assumptions are sufficient to ensure that the unique solution to (2.13) is a backward moving average of the form (1.6) with $\mathcal{F}[g](y) = (iy)^{-\beta} h_{\eta,\beta}(iy)^{-1}$ for $y \in \mathbb{R}$. The notion of uniqueness is, however, weaker than in the non-fractional setting considered in Papers B and D; it is the only stationary process $(X_t)_{t\in\mathbb{R}}$ which satisfies (2.13), and which is *purely non-deterministic* in the sense that $\mathbb{E}[X_0] = 0$, $\mathbb{E}[X_0^2] < \infty$ and

$$\bigcap_{t\in\mathbb{R}} \overline{\mathrm{sp}}\{X_s : s \leq t\} = \{0\}.$$

Note that if $\mu_h((0,\infty)) = 0$ for all $h > 0$, (2.14) reveals immediately that translations of solutions remain solutions, and hence we cannot have the same strong type of uniqueness as in Papers B and D. Proposition 3.7 (Paper E) shows that the model generates exactly the type of long memory behavior that we asked for in (2.12):

$$\gamma_X(h) \sim \frac{\Gamma(1-2\beta)}{\Gamma(\beta)\Gamma(1-\beta)\eta([0,\infty))^2} h^{2\beta-1} \qquad \text{as } h \to \infty$$

and

$$f_X(y) \sim \frac{1}{\eta([0,\infty))^2} |y|^{-2\beta} \qquad \text{as } y \to 0.$$

An interesting feature of generating long memory processes in this way is that, in contrast to the long memory models in continuous time which are based on a fractional noise, the local path properties do not depend on β and $(X_t)_{t\geq 0}$ is a semimartingale (see Remarks 3.9 and 3.10 as well as the comment in relation to Proposition 3.6 of Paper E). Based on the close relation between CARMA processes and SDDEs with a certain type of delay (cf. (2.10)), this subclass is studied in detail and related to the fractionally integrated CARMA processes introduced in [13].

While the proofs of the paper do indeed make use of some of the same type of results as in Papers B and D, theory from fractional calculus as well as spectral representations of stationary processes also play a significant role.

2.4   Paper G

In Papers B and D it was argued that, under some additional assumptions, a unique stationary solution to (2.6) exists if $\eta([0,\infty)) \neq 0$, and Example 4.5 of Paper E shows that a stationary solution can sometimes exist even when $\mathcal{F}[\eta](y) \sim \alpha(iy)^\beta$ as $y \to 0$ for some $\alpha > 0$ and $\beta \in (0,1/2)$. But what happens if the convergence $\mathcal{F}[\eta](y) \to 0$ is fast? An extreme example is $\eta \equiv 0$, where a stationary solution to (2.6) cannot exist unless $(L_t)_{t\in\mathbb{R}}$ is identically zero. A more moderate example could be $\eta(\mathrm{d}t) = (Df)(t)\,\mathrm{d}t$ with $f, Df \in L^1$. To be able to find solutions in such situations it seems reasonable to allow that a solution is not stationary, but only has stationary increments. In the literature ([16]), a process with these characteristics is often referred to as being *integrated* (of order one).

The purpose of Paper G is to study solutions to SDDEs which are possibly integrated and, in the multivariate setting, *cointegrated*. Cointegration refers to the phenomenon that an n-dimensional process $(X_t)_{t\in\mathbb{R}}$ is integrated, but $(\beta^\top X_t)_{t\in\mathbb{R}}$ is stationary for some cointegration vector $\beta \in \mathbb{R}^n \setminus \{0\}$. In the paper we prove a Granger type representation theorem, which characterizes the class of integrated solutions to SDDEs under appropriate assumptions. This representation reveals in particular that increments of solutions are uniquely determined, but the possible translations as well as the number of linearly independent cointegration vectors are tied to the rank of $\eta([0,\infty))$. Such type of results should indeed indicate that the findings of Paper G are particularly interesting in the multidimensional setting—in fact, several parallels can be drawn to the celebrated cointegrated VAR model in this case. However, to avoid introducing too much notation and to agree with the level of details given in the above descriptions of Papers B–E, we only formulate the results in the univariate case. The reader is encouraged to consult Paper G (in particular, its introduction) for further details.

The interest will specifically be on stochastic processes $(X_t)_{t\in\mathbb{R}}$ with the following properties:

(i) $(X_t)_{t\in\mathbb{R}}$ is measurable and $\mathbb{E}[X_t^2] < \infty$ for all $t \in \mathbb{R}$.

(ii) $(X_t, Z_t)_{t\in\mathbb{R}}$ has stationary increments.

(iii) $(X_t)_{t\in\mathbb{R}}$ satisfies the SDDE

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

Properties (i) and (iii) implicitly impose the assumption that $\mathbb{E}[Z_t^2] < \infty$ for all $t \in \mathbb{R}$. The results obtained in Paper G are based on the assumptions that $\int_{[0,\infty)} e^{\delta t}\,|\eta|(\mathrm{d}t) < \infty$ for some $\delta > 0$ and that the equation

$$h_\eta(z) := z - \int_{[0,\infty)} e^{-zt}\,\eta(\mathrm{d}t) = 0$$

has no solutions $z \in \mathbb{C} \setminus \{0\}$ with $\mathrm{Re}(z) \geq 0$. Suppose also that

$$\eta([0,\infty)) = 0 \quad \text{and} \quad C_0 := 1 + \int_0^\infty \eta((t,\infty))\,\mathrm{d}t \neq 0. \tag{2.15}$$

Under these assumptions, one of the main results in the paper states that a process $(X_t)_{t\in\mathbb{R}}$ satisfies (i)–(iii) above if and only if

$$X_t = \xi + C_0 Z_t + \int_{-\infty}^t C(t-u)\,\mathrm{d}Z_u, \qquad t\in\mathbb{R}, \tag{2.16}$$

for some $\xi\in L^2(\mathbb{P})$ and with $C\colon\mathbb{R}\to\mathbb{R}$ characterized by $\mathcal{F}[C](y)=h_\eta(iy)^{-1}-C_0\,(iy)^{-1}$ for $y\in\mathbb{R}$ (cf. Theorem 1.2 and Corollary 3.7 of Paper G). This shows that solutions can always be decomposed into an initial value, a "random walk" and a moving average—and that the last two of them are uniquely determined by η. As in Papers B and D the result can also be formulated without assuming that $(Z_t)_{t\in\mathbb{R}}$ is a regular integrator (cf. Theorem 3.5 of Paper G). Based on this result and the relation between SDDEs and stationary invertible CARMA processes, we discuss how one can define (co)integrated CARMA processes as solutions to certain SDDEs. Although a detailed analysis of the multivariate setting will not be presented here, it should be mentioned that in this situation, the solution will still admit the representation (2.16) with C_0 being a deterministic $n\times n$ matrix, C a deterministic function with values in the space of $n\times n$ matrices and ξ a random vector which belongs to the column space of C_0. Consequently, if $\beta\in\mathbb{R}^n$ satisfies $\beta^\top C_0 = 0$, it follows from (2.16) that $(\beta^\top X_t)_{t\in\mathbb{R}}$ is a moving average and, thus, β is a cointegration vector. Theorem 1.2 of Paper G reveals that the space of such vectors coincides with the row space of $\eta([0,\infty))$.

The ideas in the proofs are again based on attacking problems in the spectral domain by relying on Hardy space theory and spectral representations of stationary processes. By heuristically applying the Fourier transform to the equation (2.6), one easily arrives at the conclusion that

$$\mathcal{F}[X](y) = h_\eta(iy)^{-1}\mathcal{F}[DZ](y), \qquad y\in\mathbb{R}. \tag{2.17}$$

If $h_\eta(iy)\neq 0$ for all $y\in\mathbb{R}$ as in Papers B and D, there exists $g\in L^2$ with $\mathcal{F}[g](y)=h_\eta(iy)$, and (2.17) indicates that $(X_t)_{t\in\mathbb{R}}$ should take the form (2.9). Now, when $\eta([0,\infty))=0$ the assumption that $|\eta|$ integrates $t\mapsto e^{\delta t}$ allows us to use the machinery of complex analysis to study the pole of $1/h_\eta$ at 0. Although it could be of any order $m\in\mathbb{N}$, the second assumption of (2.15) ensures that $m=1$ (the pole is simple). It is not too difficult to see that C_0 is the residue of $1/h_\eta$ at 0 and that, up to the discrepancy term ξ, (2.16) aligns with (2.17). Loosely speaking, m determines the order of integration of the solution, and hence the two assumptions of (2.15) result in solutions which are non-stationary, but have stationary increments.

It should be stressed that, since $(Z_t)_{t\in\mathbb{R}}$ is not necessarily a Lévy process and could in principle be stationary, $(X_t)_{t\in\mathbb{R}}$ given by (2.16) can also be stationary, and hence it might be misleading to call it "integrated". To make the definitions independent of the stationary properties of $(Z_t)_{t\in\mathbb{R}}$, one can rely on the particular framework and *define* an integrated process to be a stochastic process which satisfies

$$X_t - X_0 = \int_\mathbb{R}[g(t-u)-g(-u)]\,\mathrm{d}Z_u \quad\text{and}\quad \int_\mathbb{R}[g(t+u)-g(u)]\,\mathrm{d}u \neq 0$$

for all $t\neq 0$ and a suitable function $g\colon\mathbb{R}\to\mathbb{R}$ with $(u\mapsto g(t+u)-g(u))\in L^1\cap L^2$ for $t>0$. This strategy is well-known in the discrete-time literature (cf. [27, Definition 1]).

3   Limit theorems for quadratic forms and related quantities of Lévy-driven moving averages

Sections 1–2 introduced Lévy-driven moving averages of the form (1.6) as a convenient class to model continuous-time stationary processes and discussed important subclasses such as CARMA processes and solutions to certain SDDEs. The next task could naturally concern estimation within one of these particular subclasses. In the continuous-time framework one often distinguishes between three types of regimes, namely low, mixed and high frequency. Since Papers F and H consider only the low frequency setting, this will be our focus in the following section.

Consider a sample $Y(n) := [Y_1, \ldots, Y_n]^\top$ from a discrete-time stationary process $(Y_t)_{t \in \mathbb{N}}$, from which we want to infer a parameter θ_0 belonging to some set Θ. The attention will be restricted to *parametric* estimation as we will assume that Θ is a compact subset of $\mathbb{R}^d$ for some $d \in \mathbb{N}$. For instance, if $(Y_t)_{t \in \mathbb{N}}$ is the ARMA process satisfying (2.2) for some P and Q, one could be interested in estimating the coefficients of these polynomials as well the variance $\sigma^2 := \mathbb{E}[\varepsilon_0^2]$ of the innovations. It could also be the case that the observations stem from an underlying continuous-time process, e.g.,

$$Y_t = X_{t\Delta}, \qquad t \in \mathbb{N}, \tag{3.1}$$

for some $\Delta > 0$ where $(X_t)_{t \in \mathbb{R}}$ is a CARMA process, a solution to an SDDE or, more generally, a moving average. Many parametric estimators based on $Y(n)$ can be characterized as

$$\hat{\theta}_n \in \arg\min_{\theta \in \Theta} \ell_n(\theta) \tag{3.2}$$

for a sufficiently regular objective function $\ell_n = \ell_n(\,\cdot\,; Y(n))$. To study second order asymptotic properties of $\hat{\theta}_n$ as $n \to \infty$ it is important to establish a limit theorem for a suitably scaled version of the first order derivative $\ell_n'(\theta_0)$ of ℓ_n at θ_0. For a wide range of popular choices of ℓ_n, such as squared linear prediction errors, the negative Gaussian likelihood and Whittle's approximation, $\ell_n'(\theta_0)$ is closely related to the sample autocovariances of $Y(n)$ as well as quantities of the form

$$Q_n = \sum_{t,s=1}^{n} b(t-s) Y_t Y_s, \qquad n \in \mathbb{N}, \tag{3.3}$$

where $b \colon \mathbb{Z} \to \mathbb{R}$ is an even function (see the introduction of Paper F for details).

One general way to prove limit theorems for such quantities is to impose strict assumptions on the dependence structure of $(Y_t)_{t \in \mathbb{N}}$, e.g., rapidly decaying mixing coefficients. Besides that such conditions are often too restrictive, they are very difficult to verify in many situations (see the discussion in [2]). Instead, by assuming a certain structure of $(Y_t)_{t \in \mathbb{N}}$, it is possible to analyze the quantities directly and prove limit theorems, even in cases where the autocovariance function is slowly decaying. For instance, if $(Y_t)_{t \in \mathbb{N}}$ is a discrete-time moving average as in the first equation of (2.1), one can give precise conditions on b, the moving average coefficients $(\psi_t)_{t \in \mathbb{N}_0}$ and the noise $(\varepsilon_t)_{t \in \mathbb{Z}}$ to ensure that the sample autocovariances and the quadratic form Q_n are asymptotically Gaussian (see [11, Section 7] and [23]). The situation where $(Y_t)_{t \in \mathbb{N}}$ is given by (3.1) for some Lévy-driven moving average $(X_t)_{t \in \mathbb{R}}$ is only partly covered. Indeed, asymptotic results concerning the sample autocovariances

are established ([15, 35]), but results on the asymptotic behavior of Q_n have been missing.

3.1 Paper F

The main purpose of this paper has been to give general sufficient conditions on b, g and $(L_t)_{t\in\mathbb{R}}$ to ensure that

$$\frac{Q_n - \mathbb{E}[Q_n]}{\sqrt{n}} \xrightarrow{\mathscr{D}} \mathcal{N}(0, \eta^2), \qquad n \to \infty, \tag{3.4}$$

when $(Q_n)_{n\in\mathbb{N}}$ is given by (3.3) and $Y_t = \int_{\mathbb{R}} g(t\Delta - u)\,\mathrm{d}L_u$ for $t \in \mathbb{N}$. Here $\mathcal{N}(\xi, \eta^2)$ is the Gaussian distribution with mean $\xi \in \mathbb{R}$ and variance $\eta^2 > 0$, and $\xrightarrow{\mathscr{D}}$ denotes convergence in law. The tricky part of studying the limiting behavior of Q_n is that it involves a double sum. To succeed we followed a strategy similar to that of [23], which goes by first proving a result of the type (3.4) with Q_n replaced by

$$S_n = \sum_{t=1}^{n} \int_{\mathbb{R}} g_1(t\Delta - u)\,\mathrm{d}L_u \int_{\mathbb{R}} g_2(t\Delta - u)\,\mathrm{d}L_u, \qquad n \in \mathbb{N}, \tag{3.5}$$

and next approximating Q_n by S_n with a clever choice of g_1 and g_2. Note that a special case of (3.5) is the sample autocovariance of moving averages (assuming the mean is known to be zero), and therefore the limiting behavior of $\ell_n'(\theta_0)$ as $n \to \infty$ can sometimes also be determined by relying on results for quantities of the same form as S_n (see Examples 3.3 and 3.4 of Paper F for details). This means that results concerning $(S_n)_{n\in\mathbb{N}}$ may be of independent interest, and hence we will discuss central limit theorems for both $(S_n)_{n\in\mathbb{N}}$ and $(Q_n)_{n\in\mathbb{N}}$ here.

Throughout the paper it is assumed that $\mathbb{E}[L_1] = 0$ and $\mathbb{E}[L_1^4] < \infty$. The most general result (Theorem 3.1 of Paper F) obtained for $(S_n)_{n\in\mathbb{N}}$ is that

$$\frac{S_n - \mathbb{E}[S_n]}{\sqrt{n}} \xrightarrow{\mathscr{D}} \mathcal{N}(0, \eta^2), \qquad n \to \infty, \tag{3.6}$$

for some $\eta^2 > 0$, if g_1 and g_2 satisfy the following conditions:

(S1) $\int_{\mathbb{R}} |g_i(t)g_i(t + \cdot\Delta)|\,\mathrm{d}t \in \ell^{\alpha_i}$ for $i = 1, 2$ and $\alpha_1, \alpha_2 \in [1, \infty]$ with $1/\alpha_1 + 1/\alpha_2 = 1$.

(S2) $\int_{\mathbb{R}} |g_1(t)g_2(t + \cdot\Delta)|\,\mathrm{d}t \in \ell^2$.

(S3) $\left(t \longmapsto \|g_1(t + \cdot\Delta)g_2(t + \cdot\Delta)\|_{\ell^1}\right) \in L^2([0, \Delta])$.

As usual, ℓ^p denotes the space of sequences $(a_t)_{t\in\mathbb{Z}}$ satisfying $\sum_{t\in\mathbb{Z}} |a_t|^p < \infty$ when $p \in [1, \infty)$ and $\sup_{t\in\mathbb{Z}} |a_t| < \infty$ when $p = \infty$, and $\|\cdot\|_{\ell^p}$ is the corresponding norm. Condition (S3) is not needed if the fourth cumulant $\kappa_4 := \mathbb{E}[L_1^4] - 3\mathbb{E}[L_1^2]^2$ is zero or, equivalently, if $(L_t)_{t\in\mathbb{R}}$ is a Brownian motion. A simple sufficient condition for (S1)–(S3) to hold is that

(S*) $g_1, g_2 \in L^4$ and $\sup_{t\in\mathbb{R}} |t|^{\alpha_i} |g_i(t)| < \infty$ for $i = 1, 2$ and some $\alpha_1, \alpha_2 \in (1/2, 1)$ with $\alpha_1 + \alpha_2 > 3/2$.

The general result for $(Q_n)_{n\in\mathbb{N}}$ (Theorem 3.5 of Paper F) establishes that (3.4) holds true under the conditions below:

(Q1) There exist $\alpha, \beta \in [1, \infty]$ with $1/\alpha + 1/\beta = 1$, such that $\int_{\mathbb{R}} |g(t)g(t + \cdot \Delta)| \, dt \in \ell^\alpha$ and $\int_{\mathbb{R}} (|b| \star |g|)(t)(|b| \star |g|)(t + \cdot \Delta) \, dt \in \ell^\beta$.

(Q2) $\int_{\mathbb{R}} |g(t)| (|b| \star |g|)(t + \cdot \Delta) \, dt \in \ell^2$.

(Q3) $\left(t \longmapsto \|g(t + \cdot \Delta)(|b| \star |g|)(t + \cdot \Delta)\|_{\ell^1} \right) \in L^2([0, \Delta])$.

In the statements above we have used the notation $a \star f(t) = \sum_{s \in \mathbb{Z}} a(s) f(t - s\Delta)$ for functions $a \colon \mathbb{Z} \to \mathbb{R}$ and $f \colon \mathbb{R} \to \mathbb{R}$ (and $t \in \mathbb{R}$ such that the sum is meaningful). Again, condition (Q3) can be discarded if $\kappa_4 = 0$. In this setup an easy-to-check condition, which implies (Q1)–(Q3), is

(Q*) $g \in L^4$ and, for some $\alpha, \beta > 0$ with $\alpha + \beta < 1/2$,

$$\sup_{t \in \mathbb{R}} |t|^{1-\alpha/2} |g(t)| < \infty \quad \text{and} \quad \sup_{t \in \mathbb{Z}} |t|^{1-\beta} |b(t)| < \infty.$$

In Theorems 1.1 and 1.2 (Paper F), (Q*) and (S*) can be found together with other sufficient conditions.

It is not too difficult to see that conditions (Q1)–(Q3) are slightly stronger than (S1)–(S3) with $g_1 = g$ and $g_2 = b \star g$. In fact, the key step in showing (3.4) under assumptions (Q1)–(Q3) is to use (3.6) and argue that $\mathrm{Var}(Q_n - S_n)/n \to 0$ as $n \to \infty$ with this particular choice of g_1 and g_2. The proof concerning $(S_n)_{n \in \mathbb{N}}$ involves two steps, namely to show that (i) the result is true when adequately truncating g_1 and g_2 by using a central limit theorem for m-dependent sequences, and (ii) it remains true when passing to the limit. The conditions (S1)–(S3) and (Q1)–(Q3) should indicate the rather delicate interplay between the discrete-time sampling scheme and the continuous-time convolution structure of the moving average. Specifically, the assumptions concern either the integrability of certain sums or summability of convolutions. To obtain easy-to-check conditions as given in Theorems 1.1 and 1.2 of Paper F (in particular, (S*) and (Q*)) it was necessary to prove a suitable Young type inequality in this mixed framework (Lemma 4.3 of Paper F). Among other things, this inequality was used to prove that (S*) implies (S1)–(S3) and that (Q*) implies (Q1)–(Q3).

3.2 Paper H

This paper demonstrates how to use the results of Paper F to obtain asymptotic normality of a certain type of least squares estimator. Specifically, it is assumed that $(Y_t)_{t \in \mathbb{N}}$ is of the form

$$Y_t = \int_{\mathbb{R}} g(t\Delta - u) \, dL_u, \qquad t \in \mathbb{N}, \tag{3.7}$$

for some Lévy process $(L_t)_{t \in \mathbb{R}}$ satisfying $\mathbb{E}[L_1] = 0$ and $\mathbb{E}[L_1^4] < \infty$ and kernel g belonging to a suitable parametrized class of functions $\{g_\theta : \theta \in \Theta\} \subseteq L^2$. For simplicity, it is also assumed that $\mathbb{E}[L_1^2] = 1$ and, to avoid trivial cases, the set of t such that $g_\theta(t) \neq 0$ is not a Lebesgue null set. The aim is to estimate the vector $\theta_0 \in \Theta$ with the property $g = g_{\theta_0}$ from the sample $Y(n)$ using the estimator $\hat{\theta}_n$ in (3.2) with

$$\ell_n(\theta) = \sum_{t=k+1}^{n} (Y_t - \pi_k(Y_t; \theta)), \qquad \theta \in \Theta,$$

for some $k \in \mathbb{N}$ with $k < n$. Here $\pi_k(Y_t; \theta)$ is the $L^2(\mathbb{P})$ projection of Y_t onto the linear span of $Y_{t-1}, \ldots, Y_{t-k}$ computed under the model (3.7) with $g = g_\theta$.

It is assumed that the functions

$$\theta \longmapsto \int_{\mathbb{R}} g_\theta(j\Delta + t) g_\theta(t) \, \mathrm{d}t, \qquad j = 0, 1, \ldots, k,$$

are twice continuously differentiable on the interior of Θ and an identifiability condition as well as a full rank condition are also imposed (see Condition 2.1(a)–(b) of Paper H). As already mentioned, the typical condition to impose in addition to those mentioned above is that the sequence $(Y_t)_{t \in \mathbb{N}}$ exhibits a particular mixing behavior. Due to the form of ℓ_n and the moving average structure of $(Y_t)_{t \in \mathbb{N}}$ we rely instead on the findings of Paper F and impose a condition directly on the driving kernel g:

(∗) The function $t \longmapsto \sum_{s \in \mathbb{Z}} (|g(t + s\Delta)|^{4/3} + |g(t + s\Delta)|^2)$ belongs to $L^2([0, \Delta])$.

A sufficient condition for (∗) to be satisfied is that $g \in L^4$ and $\sup_{t \in \mathbb{R}} |t|^\beta |g(t)| < \infty$ for some $\beta \in (3/4, 1)$. Under these conditions, and provided that θ_0 belongs to the interior of Θ, strong consistency and asymptotic normality of $\hat{\theta}_n$ is established (Theorem 2.4 of Paper H) using standard arguments. The quality of the estimator is assessed through a simulation study in two concrete cases: (i) g belongs to the class of gamma kernels, and (ii) $Y_t = X_{t\Delta}$ where $(X_t)_{t \in \mathbb{R}}$ is the stationary solution to an SDDE with delay measure $\eta = \alpha \delta_0 + \beta \delta_1$. See Examples 3.1 and 3.2 of Paper H for further details.

References

[1] Andersen, T. G. (2000). Some reflections on analysis of high-frequency data. *J. Bus. Econom. Stat.* 18(2), 146–153.

[2] Ango Nze, P., P. Bühlmann and P. Doukhan (2002). Weak dependence beyond mixing and asymptotics for nonparametric regression. *Ann. Statist.* 30(2), 397–430. DOI: 10.1214/aos/1021379859.

[3] Barndorff-Nielsen, O. E. (2000). Superposition of Ornstein–Uhlenbeck type processes. *Teor. Veroyatnost. i Primenen.* 45(2), 289–311. DOI: 10.1137/S0040585X97978166.

[4] Barndorff-Nielsen, O. E. (2011). Stationary infinitely divisible processes. *Braz. J. Probab. Stat.* 25(3), 294–322. DOI: 10.1214/11-BJPS140.

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

[6] Barndorff-Nielsen, O. E., T. Mikosch and S. I. Resnick (2012). *Lévy processes: theory and applications*. Springer Science & Business Media.

[7] 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.

[8] Bergstrom, A. R. (1990). *Continuous time econometric modelling*. Oxford University Press.

[9]   Bertoin, J. (1996). *Lévy processes*. Vol. 121. Cambridge Tracts in Mathematics. Cambridge University Press.

[10] Bichteler, K. (1981). Stochastic integration and L^p-theory of semimartingales. *Ann. Probab.* 9(1), 49–89.

[11] 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.

[12] 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.

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

[14] Cheridito, P. (2004). Gaussian moving averages, semimartingales and option pricing. *Stochastic Process. Appl.* 109(1), 47–68.

[15] Cohen, S. and A. Lindner (2013). A central limit theorem for the sample autocorrelations of a Lévy driven continuous time moving average process. *J. Statist. Plann. Inference* 143(8), 1295–1306. DOI: 10.1016/j.jspi.2013.03.022.

[16] Comte, F. (1999). Discrete and continuous time cointegration. *J. Econometrics* 88(2), 207–226. DOI: 10.1016/S0304-4076(98)00025-6.

[17] Cont, R. and P. Tankov (2004). *Financial modelling with jump processes*. Chapman & Hall/CRC Financial Mathematics Series. Chapman & Hall/CRC, Boca Raton, FL.

[18] Delbaen, F. and W. Schachermayer (1994). A general version of the fundamental theorem of asset pricing. *Math. Ann.* 300(3), 463–520.

[19] Dellacherie, C. (1980). "Un survol de la théorie de l'intégrale stochastique". *Measure theory, Oberwolfach 1979 (Proc. Conf., Oberwolfach, 1979)*. Vol. 794. Lecture Notes in Math. Springer, Berlin, 365–395.

[20] Durrett, R. (1996). *Stochastic calculus*. Probability and Stochastics Series. A practical introduction. CRC Press, Boca Raton, FL.

[21] 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].

[22] Gandolfo, G. (2012). *Continuous-time econometrics: theory and applications*. Springer Science & Business Media.

[23] Giraitis, L. and D. Surgailis (1990). A central limit theorem for quadratic forms in strongly dependent linear variables and its application to asymptotical normality of Whittle's estimate. *Probab. Theory Related Fields* 86(1), 87–104. DOI: 10.1007/BF01207515.

[24] 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.

[25] Hamilton, J.D. (1994). *Time series analysis*. Princeton University Press, Princeton, NJ, xvi+799.

[26] Jacod, J. and A.N. Shiryaev (2003). *Limit Theorems for Stochastic Processes*. Second. Vol. 288. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]. Springer-Verlag, Berlin. DOI: 10.1007/978-3-662-05265-5.

[27] Johansen, S. (2009). "Cointegration: Overview and development". *Handbook of financial time series*. Springer, 671–693.

[28] Karhunen, K. (1950). Über die Struktur stationärer zufälliger Funktionen. *Ark. Mat.* 1, 141–160. DOI: 10.1007/BF02590624.

[29] Lépingle, D. and J. Mémin (1978). Sur l'intégrabilité uniforme des martingales exponentielles. *Z. Wahrsch. Verw. Gebiete* 42(3), 175–203. DOI: 10.1007/BF00641409.

[30] 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.

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

[32] Rosiński, J. (1989). On path properties of certain infinitely divisible processes. *Stochastic Process. Appl.* 33(1), 73–87. DOI: 10.1016/0304-4149(89)90067-7.

[33] Samorodnitsky, G. and M.S. Taqqu (1994). *Stable Non-Gaussian Random Processes*. Stochastic Modeling. Stochastic models with infinite variance. New York: Chapman & Hall.

[34] 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.

[35] Spangenberg, F. (2015). Limit theorems for the sample autocovariance of a continuous-time moving average process with long memory. arXiv: 1502.04851.

Equivalent Martingale Measures for Lévy-Driven Moving Averages and Related Processes

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

Abstract

In the present paper we obtain sufficient conditions for the existence of equivalent local martingale measures for Lévy-driven moving averages and other non-Markovian jump processes. The conditions that we obtain are, under mild assumptions, also necessary. For instance, this is the case for moving averages driven by an α-stable Lévy process with $\alpha \in (1, 2]$.

Our proofs rely on various techniques for showing the martingale property of stochastic exponentials.

1 Introduction and a main result

Absolutely continuous change of measure for stochastic processes is a classical problem in probability theory and there is a vast literature devoted to it. One motivation is the fundamental theorem of asset pricing, see Delbaen and Schachermayer [12], which relates existence of an equivalent local martingale measure to absence of arbitrage (or, more precisely, to the concept of no free lunch with vanishing risk) of a financial market. Several sharp and general conditions for absolutely continuous change of measure are given in [10, 19, 21, 24], and in case of Markov processes and solutions to stochastic differential equations, strong and explicit conditions are available, see e.g. [8, 11, 13, 17, 22] and references therein.

The main aim of the present paper is to obtain explicit conditions for the existence of an equivalent local martingale measure (ELMM) for Lévy-driven moving averages, and these are only Markovian in very special cases. Moving averages are important in various fields, e.g. because they are natural to use when modelling long-range dependence (for other applications, see [23]). Recalling that Hitsuda's representation theorem characterizes when a Gaussian process admits an ELMM, see [15, Theorem 6.3'], and Lévy-driven moving averages are infinitely divisible processes, our study can also be seen as a contribution to a similar representation theorem for this class.

We will now introduce our framework. Consider a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ equipped on which a two-sided Lévy process $L = (L_t)_{t \in \mathbb{R}}$, $L_0 = 0$, is defined. Fix a time horizon $T > 0$ and let

$$X_t = \int_{-\infty}^{t} \varphi(t-s)\, \mathrm{d}L_s, \qquad t \in [0, T], \tag{1.1}$$

for a given function $\varphi \colon \mathbb{R}_+ \to \mathbb{R}$ such that the integral in (1.1) is well-defined. We will refer to $(X_t)_{t \in [0,T]}$ as a (stationary) L-driven moving average. To avoid trivial cases, we assume that the set of $t \geq 0$ with $\varphi(t) \neq 0$ is not a Lebesgue null set. We will fix a filtration $(\mathscr{F}_t)_{t \in [0,T]}$ with the property that

$$\sigma(L_s : -\infty < s \leq t) \subseteq \mathscr{F}_t, \qquad t \in [0, T], \tag{1.2}$$

and which satisfies the usual conditions (see [16, Definition 1.3 (Ch. I)]). Furthermore, it will be assumed that $(L_t)_{t \in [0,T]}$ is an $(\mathscr{F}_t)_{t \in [0,T]}$-Lévy process in the sense that $L_t - L_s$ is independent of $\mathscr{F}_s$ for all $0 \leq s < t \leq T$. Our aim is to find explicit conditions that ensure the existence of a probability measure $\mathbb{Q}$ on $(\Omega, \mathscr{F})$, equivalent to $\mathbb{P}$, under which $(X_t)_{t \in [0,T]}$ is a local martingale with respect to $(\mathscr{F}_t)_{t \in [0,T]}$. Furthermore, we are interested in the structure of $(X_t)_{t \in [0,T]}$ under $\mathbb{Q}$.

A necessary condition for a process to admit an ELMM is that it is a semimartingale, and this property is (under mild assumptions on the Lévy measure) characterized for L-driven moving averages in Basse-O'Connor and Rosiński [4] and Knight [20]. Other relevant references in this direction include [3, 7]. In the case where L is Gaussian, and relying on Knight [20, Theorem 6.5], Cheridito [7, Theorem 4.5] gives a complete characterization of the L-driven moving averages that admit an ELMM:

Theorem 1.1 (Cheridito [7]). *Suppose that L is a Brownian motion. Then the moving average $(X_t)_{t \in [0,T]}$ defined in (1.1) admits an ELMM if and only if $\varphi(0) \neq 0$ and φ is absolutely continuous with a density φ' satisfying $\varphi' \in L^2(\mathbb{R}_+)$.*

Despite that, in general, the existence of an ELMM is a stronger condition than being a semimartingale, Theorem 1.1 shows (together with [20, Theorem 6.5]) that for Gaussian moving averages of the form (1.1), the two concepts are equivalent when $\varphi(0) \neq 0$. If L has a non-trivial Lévy measure, explicit conditions for the existence of an ELMM have, to the best of our knowledge, not been provided. It would be natural to try to obtain such conditions using the same techniques as in Theorem 1.1. However, these techniques are based on a local version of the Novikov condition, which will not be fulfilled as soon as the driving Lévy process is non-Gaussian. This is an implication of the fact that $\int_{\mathbb{R}} e^{\varepsilon x^2} \varpi(\mathrm{d}x) = \infty$ for any $\varepsilon > 0$ and any non-Gaussian

infinitely divisible distribution ϖ, see [26, Theorem 26.1]. Consequently, to prove the existence of an ELMM in a non-Gaussian setting, a completely different approach has to be used. An implication of our results is the non-Gaussian counterpart of Theorem 1.1 which is formulated in Theorem 1.2 below. In this formulation, $c \geq 0$ denotes the Gaussian component of L and F is its Lévy measure. Moreover, we will write $A^c = \mathbb{R} \setminus A$ for the complement of a given set $A \subseteq \mathbb{R}$.

Theorem 1.2. *Suppose that L has sample paths of locally unbounded variation and let $(X_t)_{t \in [0,T]}$ be a Lévy-driven moving average given by* (1.1).

(1) *Assume that either $x \mapsto F((-x,x)^c)$ is regularly varying at ∞ of index $\beta \in [-2,-1)$ or $\int_{|x|>1} x^2 F(\mathrm{d}x) < \infty$ and that the support of F is unbounded on both $(-\infty,0]$ and $[0,\infty)$. Then $(X_t)_{t \in [0,T]}$ admits an ELMM if and only if $\varphi(0) \neq 0$ and φ is absolutely continuous with a density φ' satisfying*

$$\int_0^\infty \left(c\varphi'(t)^2 + \int_{\mathbb{R}} \left(|x\varphi'(t)| \wedge |x\varphi'(t)|^2 \right) F(\mathrm{d}x) \right) \mathrm{d}t < \infty. \tag{1.3}$$

(2) *Assume that the support of F is contained in a compact set and*

$$F((-\infty,0)), F((0,\infty)) > 0.$$

Then $(X_t)_{t \in [0,T]}$ admits an ELMM if $\varphi(0) \neq 0$ and φ is absolutely continuous with a density φ', which is bounded and satisfies (1.3).

If L is a symmetric α-stable Lévy process with $\alpha \in (1,2)$, $x \mapsto F((-x,x)^c)$ is regularly varying of index $-\alpha$ and condition (1.3) is equivalent to $\varphi' \in L^\alpha(\mathbb{R}_+)$ (see [4, Example 4.9]). Thus, since we clearly have that the support of F is unbounded on both $(-\infty,0]$ and $[0,\infty)$, we can apply Theorem 1.2(1) to obtain the following natural extension of Theorem 1.1:

Corollary 1.3. *Suppose that L is a symmetric α-stable Lévy process with index $\alpha \in (1,2]$. Then the moving average $(X_t)_{t \in [0,T]}$ defined in* (1.1) *admits an ELMM if and only if $\varphi(0) \neq 0$ and φ is absolutely continuous with a density φ' satisfying $\varphi' \in L^\alpha(\mathbb{R}_+)$.*

A result similar to Corollary 1.3 can be formulated when L is a symmetric tempered stable Lévy process that is, when the Lévy measure takes the form $F(\mathrm{d}x) = \eta|x|^{-\alpha-1}e^{-\lambda|x|}\,\mathrm{d}x$ for $\eta, \lambda > 0$ and $\alpha \in [1,2)$. Indeed, since $\int_{\mathbb{R}} x^2 F(\mathrm{d}x) < \infty$ and F has unbounded support on both $(-\infty,0]$ and $[0,\infty)$ in this setup, there exists an ELMM $\mathbb{Q}$ for $(X_t)_{t \in [0,T]}$ if and only if $\varphi(0) \neq 0$ and $\int_0^\infty (|\varphi'(t)|^\alpha \wedge |\varphi'(t)|^2)\,\mathrm{d}t < \infty$ (as the latter condition is equivalent to (1.3) cf. [4, Example 4.9]).

It may be stressed that the Gaussian case considered in Theorem 1.1 and the non-Gaussian case considered in Theorem 1.2 are of fundamental different structure. Indeed, when L is a Brownian motion, one can apply the martingale representation theorem (when $(\mathscr{F}_t)_{t \in [0,T]}$ is the smallest filtration that meets (1.2) and satisfies the usual conditions) to show that the ELMM is unique, and by invariance of the quadratic variation under equivalent change of measure, $(X_t - X_0)_{t \in [0,T]}$ is a Brownian motion under the ELMM (one may need a semimartingale decomposition of $(X_t)_{t \in [0,T]}$, see e.g. (4.21)). If L is a general Lévy process, Theorem 2.1 and Remark 2.2 in Section 2

show that the ELMM will not be unique, and $(X_t - X_0)_{t\in[0,T]}$ and $(L_t)_{t\in[0,T]}$ will not be Lévy processes under any of our constructed ELMMs.

Besides the moving average framework we will also study ELMMs for semimartingales of the form

$$X_t = L_t + \int_0^t Y_s \, \mathrm{d}s, \qquad t \in [0,T], \tag{1.4}$$

for a given $(\mathscr{F}_t)_{t\in[0,T]}$-Lévy process $(L_t)_{t\in[0,T]}$ and a predictable process $(Y_t)_{t\in[0,T]}$ such that $t \mapsto Y_t$ is integrable on $[0,T]$ almost surely. This study turns out to be useful in order to deduce results for moving averages.

We will shortly present the outline of this paper. Section 2 presents Theorem 2.1, which concerns precise and tractable conditions on $(L_t)_{t\in[0,T]}$ and $(Y_t)_{t\in[0,T]}$ ensuring the existence of an ELMM for $(X_t)_{t\in[0,T]}$ in (1.4). An implication of this result is Theorem 1.2 and in turn Corollary 1.3. Theorem 2.1 is followed by a predictable criterion ensuring the martingale property of stochastic exponentials, Theorem 2.5, and this is based on a general approach of Lépingle and Mémin [21]. Due to the nature of this criterion, it can be used for other purposes than verifying the existence of ELMMs for $(X_t)_{t\in[0,T]}$ and thus, the result is of independent interest. Both Theorem 2.1 and Theorem 2.5 are accompanied by remarks and examples that illustrate their applications. Subsequently, Section 3 recalls the most fundamental and important concepts in relation to change of measure and integrals with respect to random measures. These concepts will be used throughout Section 4 which is devoted to prove the statements of Section 2. During Section 4 one will also find additional remarks and examples of a more technical nature.

2 Further main results

Let $L = (L_t)_{t\in[0,T]}$ be an $(\mathscr{F}_t)_{t\in[0,T]}$-Lévy process with triplet (c, F, b^h) relative to some truncation function $h\colon \mathbb{R} \to \mathbb{R}$. (Recall that a truncation function is a measurable function $h\colon \mathbb{R} \to \mathbb{R}$ which is bounded and satisfies $h(x) = x$ for x in a neighborhood of 0.) Here $b^h \in \mathbb{R}$ is the drift component, $c \geq 0$ is the Gaussian component, and F is the Lévy measure. Throughout the paper we will assume that L_t is integrable for every $t \in [0,T]$ which, according to [26, Corollary 25.8], is equivalent to $\int_{|x|>1} |x| F(\mathrm{d}x) < \infty$. Then, we may set $\xi = \int_{\mathbb{R}} (x - h(x)) F(\mathrm{d}x) + b^h$ so that $\mathbb{E}[L_t] = \xi t$. We denote by μ the jump measure of L and by $\nu(\mathrm{d}t, \mathrm{d}x) = F(\mathrm{d}x)\,\mathrm{d}t$ its compensator. It will be assumed that L has both positive and negative jumps such that we can choose $a, b > 0$ with

$$F((-b,-a)), F((a,b)) > 0. \tag{2.1}$$

In Theorem 2.1 we will give conditions for the existence of an ELMM $\mathbb{Q}$ for $(X_t)_{t\in[0,T]}$ given by

$$X_t = L_t + \int_0^t Y_s \, \mathrm{d}s, \qquad t \in [0,T], \tag{2.2}$$

where $(Y_t)_{t\in[0,T]}$ is a predictable process and $t \mapsto Y_t$ is Lebesgue integrable on $[0,T]$ almost surely. We will also provide the semimartingale (differential) characteristics of $(X_t)_{t\in[0,T]}$ under $\mathbb{Q}$ (these are defined in [16, Ch. II] and can be found in Section 3 as well). Recall that the notation A^c is used as the complement of a set $A \subseteq \mathbb{R}$.

Theorem 2.1. *Let $(X_t)_{t\in[0,T]}$ be given by (2.2). Consider the hypotheses:*

(h1) *The collection $(Y_t)_{t\in[0,T]}$ is tight and Y_t is infinitely divisible with a Lévy measure supported in $[-K,K]$ for all $t \in [0,T]$ and some $K > 0$.*

(h2) *The Lévy measure of L has unbounded support on both $(-\infty, 0]$ and $[0, \infty)$.*

If either (h1) *or* (h2) *holds, there exists an ELMM $\mathbb{Q}$ on $(\Omega, \mathscr{F})$ for $(X_t)_{t\in[0,T]}$ such that $d\mathbb{Q} = \mathscr{E}((\alpha - 1) * (\mu - \nu))_T \, d\mathbb{P}$ for some predictable function $\alpha \colon \Omega \times [0,T] \times \mathbb{R} \to (0, \infty)$, and the differential characteristics of $(X_t)_{t\in[0,T]}$ relative to h under $\mathbb{Q}$ are of the form*

$$\Big(c, \alpha(t,x)F(dx), b^h + Y_t + \int_{\mathbb{R}} (\alpha(t,x) - 1)h(x)F(dx)\Big), \qquad t \in [0,T]. \qquad (2.3)$$

For any $a, b > 0$ that meet (2.1), depending on the hypothesis, $\mathbb{Q}$ can be chosen such that:

(h1) *The function α is explicitly given by*

$$\alpha(t,x) = 1 + \frac{(Y_t + \xi)^- x}{\sigma_+^2}\mathbb{1}_{(a,b)}(x) - \frac{(Y_t + \xi)^+ x}{\sigma_-^2}\mathbb{1}_{(a,b)}(-x) \qquad (2.4)$$

where $\sigma_\pm^2 = \int_{\mathbb{R}} y^2 \mathbb{1}_{(a,b)}(\pm y)F(dy)$.

(h2) *With $\lambda = F([-a,a]^c)$, the relations*

$$\int_{[-a,a]^c} \alpha(t,x)F(dx) = \lambda \quad and \quad \int_{[-a,a]^c} x\alpha(t,x)F(dx) = -(Y_t + b^h) \qquad (2.5)$$

hold pointwise, and $\alpha(t,x) = 1$ whenever $|x| \le a$.

Remark 2.2. Suppose that Theorem 2.1 is applicable. Observe that, for instance by varying $a, b > 0$, an ELMM for $(X_t)_{t\in[0,T]}$ is not unique. In the following, fix an ELMM $\mathbb{Q}$ for $(X_t)_{t\in[0,T]}$, under which its characteristics have a differential form as in (2.3) relative to a truncation function h. As a first comment we see that, as long as $(Y_t)_{t\in[0,T]}$ is not deterministic, the characteristic triplet under $\mathbb{Q}$ of both $(L_t)_{t\in[0,T]}$ and $(X_t)_{t\in[0,T]}$ will not be deterministic. Consequently by [16, Theorem 4.15 (Ch. II)], none of them have independent increments, in particular they will never be Lévy processes, under $\mathbb{Q}$. Despite the fact that $(X_t)_{t\in[0,T]}$ does not have independent increments under $\mathbb{Q}$ we may still extract some useful information from the differential characteristics. Indeed, according to [16, Theorem 2.34 (Ch. II)], we may represent $(X_t)_{t\in[0,T]}$ through its canonical representation (under $\mathbb{P}$) as

$$X_t = X_t^c + h(x) * (\mu - \nu)_t + (x - h(x)) * \mu_t + \int_0^t (Y_s + b^h) \, ds, \qquad t \in [0,T], \qquad (2.6)$$

where $(X_t^c)_{t\in[0,T]}$ is the continuous martingale part of $(X_t)_{t\in[0,T]}$ under $\mathbb{P}$ and $*$ denotes integration, see Section 3 for more on the notation. Furthermore, recall that $b^h \in \mathbb{R}$ is the drift component of $(L_t)_{t\in[0,T]}$ relative to h and μ is the jump measure associated to $(X_t)_{t\in[0,T]}$ (or equivalently, $(L_t)_{t\in[0,T]}$). Consider the specific truncation function $h(x) = x\mathbb{1}_{(a,b)^c}(|x|)$ under (h1) and $h(x) = x\mathbb{1}_{[-a,a]}(x)$ under (h2). From (2.3) and (2.6) we deduce under $\mathbb{Q}$:

(i) The process X_t^c, $t \in [0,T]$, remains a Brownian motion with variance c.

(ii) It still holds that $h(x) * (\mu - \nu)_t$, $t \in [0, T]$, is a zero-mean Lévy process and its distribution is unchanged.

(iii) The process

$$(x - h(x)) * \mu_t + \int_0^t (Y_s + b^h) \, ds, \qquad t \in [0, T], \qquad (2.7)$$

is a local martingale, since $(X_t)_{t \in [0,T]}$ is a local martingale.

(iv) Except for the drift term involving $(Y_t)_{t \in [0,T]}$, it follows that the only component in (2.6) affected by the change of measure (under any of the hypotheses) is $(x - h(x)) * \mu_t$, $t \in [0, T]$, which goes from a compound Poisson process under $\mathbb{P}$ to a general càdlàg and piecewise constant process under $\mathbb{Q}$. Specifically, it will be affected in such a way that it is compensated according to (2.7). By exploiting the structure of the compensator of μ under $\mathbb{Q}$ it follows that the jumps of $(x - h(x)) * \mu_t$, $t \in [0, T]$, still arrive according to a Poisson process (with the same intensity as under $\mathbb{P}$) under (h2) while under (h1), they will arrive according to a counting process with a stochastic intensity. The (conditional) jump distribution is obtained from Lemma 4.5.

Note that although, strictly speaking, the function $h(x) = x \mathbb{1}_{(a,b)^c}(|x|)$ is not a genuine truncation function, we are allowed to use it as such, since $\int_{|x|>1} |x| F(dx) < \infty$ by assumption, which means the integrals in (2.6) will be well-defined.

Remark 2.3. As a first comment on the hypotheses presented in the statement of Theorem 2.1 we see that none of them is superior to the other one. Rather, there is a trade off between the restrictions on $(L_t)_{t \in [0,T]}$ and on $(Y_t)_{t \in [0,T]}$. In line with Remark 4.3, one may as well replace (h1) by

(h1') For any $t \in [0, T]$ and a suitable $\varepsilon > 0$, $Y_t \overset{\mathscr{D}}{=} Y_0$ and $\mathbb{E}[e^{\varepsilon |Y_0| \log(1 + |Y_0|)}] < \infty$.

The advantage of this hypothesis is that one is not restricted to the case where Y_t is infinitely divisible, however the price to pay is to require that $Y_t \overset{\mathscr{D}}{=} Y_0$ rather than the much weaker assumption of $(Y_t)_{t \in [0,T]}$ being tight.

Remark 2.4. Suppose that $(L_t)_{t \geq 0}$ and $(Y_t)_{t \geq 0}$ are defined on the probability space $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t \geq 0}, \mathbb{P})$ with $\mathscr{F} = \bigvee_{t \geq 0} \mathscr{F}_t$ and that Theorem 2.1 is applicable on the truncated space $(\Omega, \mathscr{F}_T, (\mathscr{F}_t)_{t \in [0,T]}, \mathbb{P}|_{\mathscr{F}_T})$ for any $T > 0$. Then one can sometimes extend it to a locally equivalent measure $\mathbb{Q}$ on $(\Omega, \mathscr{F})$. A probability space having this property is often referred to as being full. An example is the space of all càdlàg functions taking values in a Polish space when equipped with its standard filtration. For more details, see [5] and [10]. As is the case for Lévy processes, we believe that such $\mathbb{Q}$ will usually not be equivalent to $\mathbb{P}$, and we have chosen not to pursue this direction further.

Despite of a common structure in (2.3) under (h1) and (h2), the choices of α that we suggest under the different hypotheses in Theorem 2.1 differ by their very nature. This is a consequence of different ways of constructing the ELMM.

The proof of the existence of an ELMM for $(X_t)_{t \in [0,T]}$ consists of two steps. One step is to identify an appropriate possible probability density Z, that is, a positive random

variable which, given that $\mathbb{E}[Z] = 1$, defines an ELMM $\mathbb{Q}$ on $(\Omega, \mathscr{F})$ for $(X_t)_{t\in[0,T]}$ through $\mathrm{d}\mathbb{Q} = Z\,\mathrm{d}\mathbb{P}$. The candidate will always take the form $Z = \mathscr{E}((\alpha - 1) * (\mu - \nu))_T$ for some positive predictable function α. The remaining step is to check that $\mathbb{E}[Z] = 1$ or, equivalently, $\mathscr{E}((\alpha - 1) * (\mu - \nu))$ is a martingale. Although there exist several sharp results on when local martingales are true martingales, there has been a need for a tractable condition which is suited for the specific setup in question, and this was the motivation for Theorem 2.5. Specifically, it will be used to show Theorem 2.1 under hypothesis (h1). As mentioned, the proof of Theorem 2.5 is based on a very general approach presented by Lépingle and Mémin [21].

Theorem 2.5. *Let $W : \Omega \times [0,T] \times \mathbb{R} \to \mathbb{R}_+$ be a predictable function. Suppose that*

$$W(\omega, t, x) \leq |P_t(\omega)| g(x) \qquad \text{for all } (\omega, t, x) \in \Omega \times [0,T] \times \mathbb{R}, \qquad (2.8)$$

where the following hold:

(a) *The process $(P_t)_{t\in[0,T]}$ is predictable and satisfies that*

   (i) *for some fixed $K > 0$ and any $t \in [0,T]$, P_t is infinitely divisible with Lévy measure supported in $[-K, K]$, and*

   (ii) *the collection of random variables $(P_t)_{t\in[0,T]}$ is tight.*

(b) *The function $g : \mathbb{R} \to \mathbb{R}_+$ satisfies $g + g\log(1 + g) \in L^1(F)$.*

Then $W * (\mu - \nu)$ is well-defined and $\mathscr{E}(W * (\mu - \nu))$ is a martingale.

The following example shows how this result compares to other classical references for measure changes, when specializing to the case where μ is the jump measure of a Poisson process.

Example 2.6. Suppose that L is a (homogeneous) Poisson process with intensity $\lambda > 0$ and consider a density $Z = \mathscr{E}((\alpha - 1) * (\mu - \nu))_T$ for some positive predictable process $(\alpha_t)_{t\in[0,T]}$ which paths are integrable on $[0,T]$ almost surely. Within the literature of (marked) point processes, with this setup as a special case, one explicit and standard criterion ensuring that $\mathbb{E}[Z] = 1$ is the existence of constants $K_1, K_2 > 0$ and $\gamma > 1$ such that

$$\alpha_t^\gamma \leq K_1 + K_2(L_t + \lambda t) \qquad (2.9)$$

for all $t \in [0,T]$ almost surely, see [6, Theorem T11 (Ch. VIII)] or [14, Eq. (25)]. We observe that the inequality in (2.9) implies that (2.8) holds with $g \equiv 1$ and $P_t = 2 + K_1 + K_2(L_t + \lambda t)$, $t \in [0,T]$, where $(P_t)_{t\in[0,T]}$ meets (i)–(ii) in Theorem 2.5, and thus this criterion is implied by our result. Clearly, this also indicates that we cover other, less restrictive, choices of $(\alpha_t)_{t\in[0,T]}$. For instance, one could take $\gamma = 1$ and replace L by any Lévy process with a compactly supported Lévy measure in (2.9). Note that, although we might have $\alpha_t - 1 < 0$, Theorem 2.5 may still be applied according to Remark 4.2. For other improvements of (2.9), see also [27].

Section 4 contains proofs of the statements above accompanied by a minor supporting result and a discussion of the techniques. However, we start by recalling some fundamental concepts which will be (and already has been) used repeatedly.

3 Preliminaries

The following consists of a short recap of fundamental concepts. For a more formal and extensive treatment, see [16].

The stochastic exponential $\mathscr{E}(M) = (\mathscr{E}(M)_t)_{t \in [0,T]}$ of a semimartingale $(M_t)_{t \in [0,T]}$ is characterized as the unique càdlàg and adapted process with

$$\mathscr{E}(M)_t = 1 + \int_0^t \mathscr{E}(M)_{s-} \, \mathrm{d}M_s, \qquad t \in [0,T].$$

It is explicitly given as

$$\mathscr{E}(M)_t = e^{M_t - M_0 - \frac{1}{2}\langle M^c \rangle_t} \prod_{s \leq t} (1 + \Delta M_s) e^{-\Delta M_s}, \qquad t \in [0,T], \tag{3.1}$$

where $(M_t^c)_{t \in [0,T]}$ is the continuous martingale part of $(M_t)_{t \in [0,T]}$. If $(M_t)_{t \in [0,T]}$ is a local martingale, $\mathscr{E}(M)$ is a local martingale as well. Consequently whenever $\mathscr{E}(M)_t \geq 0$, equivalently $\Delta M_t \geq -1$, for all $t \in [0,T]$ almost surely, $\mathscr{E}(M)$ is a supermartingale. (Here, and in the following, we have adopted the definition of a semimartingale from [16], which in particular means that the process is càdlàg.)

A random measure on $[0,T] \times \mathbb{R}$ is a family of measures μ such that for each $\omega \in \Omega$, $\mu(\omega; \mathrm{d}t, \mathrm{d}x)$ is a measure on $([0,T] \times \mathbb{R}, \mathscr{B}([0,T]) \otimes \mathscr{B}(\mathbb{R}))$ satisfying $\mu(\omega; \{0\} \times \mathbb{R}) = 0$. For our purpose, μ will also satisfy that $\mu(\omega; [0,T] \times \{0\}) = 0$. Integration of a function $W: \Omega \times [0,T] \times \mathbb{R} \to \mathbb{R}$ with respect to μ over the set $(0,t] \times \mathbb{R}$ is denoted $W * \mu_t$ for $t \in [0,T]$. In this paper, μ will always be the jump measure of some adapted càdlàg process. To any such μ, one can associate a unique (up to a null set) predictable random measure ν, which is called its compensator. We will always be in the case where $\nu(\omega; \mathrm{d}t, \mathrm{d}x) = F_t(\omega; \mathrm{d}x) \, \mathrm{d}t$ with $(F_t(B))_{t \in [0,T]}$ being a predictable process for every $B \in \mathscr{B}(\mathbb{R})$. One can define the stochastic integral with respect to the compensated random measure $\mu - \nu$ for any predictable function $W: \Omega \times [0,T] \times \mathbb{R} \to \mathbb{R}$ satisfying that $(W^2 * \mu_t)^{1/2}$, $t \in [0,T]$, is locally integrable. The associated integral process is denoted $W * (\mu - \nu)$.

Let $h: \mathbb{R} \to \mathbb{R}$ be a bounded function with $h(x) = x$ in a neighbourhood of 0. The characteristics of a semimartingale $(M_t)_{t \in [0,T]}$, relative to the truncation function h, are then denoted (C, ν, B^h), which is unique up to a null set. Here C is the quadratic variation of the continuous martingale part of $(M_t)_{t \in [0,T]}$, ν is the predictable compensator of its jump measure, and B^h is the predictable finite variation part of the special semimartingale given by $M_t^h = M_t - \sum_{s \leq t}[\Delta M_s - h(\Delta M_s)]$ for $t \in [0,T]$. In the case where

$$C_t = \int_0^t c_s \, \mathrm{d}s, \quad \nu(\omega; \mathrm{d}t, \mathrm{d}x) = F_t(\omega; \mathrm{d}x) \, \mathrm{d}t \quad \text{and} \quad B_t^h = \int_0^t b_s^h \, \mathrm{d}s$$

for some predictable processes $(b_t^h)_{t \in [0,T]}$ and $(c_t)_{t \in [0,T]}$ and transition kernel $F_t(\omega; \mathrm{d}x)$, we call (c_t, F_t, b_t^h) the differential characteristics of $(M_t)_{t \in [0,T]}$.

Suppose that we have another probability measure $\mathbb{Q}$ on $(\Omega, \mathscr{F})$ such that $\mathrm{d}\mathbb{Q} = \mathscr{E}(W * (\mu - \nu))_T \, \mathrm{d}\mathbb{P}$, where μ is the jump measure of an $(\mathscr{F}_t)_{t \in [0,T]}$-Lévy process $(L_t)_{t \in [0,T]}$ with characteristic triplet (c, F, b^h) relative to a given truncation function h and ν is the compensator of μ. Then a version of Girsanov's theorem, see [2] or [18], implies

that under $\mathbb{Q}$, $(L_t)_{t\in[0,T]}$ is a semimartingale with differential characteristics (c, F_t, b_t^h), where

$$F_t(\mathrm{d}x) = (1 + W(t,x))F(\mathrm{d}x) \quad \text{and} \quad b_t^h = b^h + \int_{\mathbb{R}} W(t,x)h(x)F(\mathrm{d}x). \qquad (3.2)$$

4 Proofs

In the following, let $f \colon (-1, \infty) \to \mathbb{R}_+$ be defined by

$$f(x) = (1+x)\log(1+x) - x, \qquad x > -1. \qquad (4.1)$$

In order to show Theorem 2.5 we will state and prove a local version of [21, Theorem 1 (Section III)] below.

Lemma 4.1. *Let $(M_t)_{t\in[0,T]}$ be a purely discontinuous local martingale with $\Delta M_t > -1$ for all $t \in [0,T]$ almost surely. Suppose that the process*

$$\sum_{s\leq t} f(\Delta M_s), \qquad t \in [0,T],$$

has compensator $(\tilde{A}_t)_{t\in[0,T]}$ and that there exist stopping times $0 = \tau_0 < \tau_1 < \cdots < \tau_n = T$ such that

$$\mathbb{E}\big[\exp\big\{\tilde{A}_{\tau_k} - \tilde{A}_{\tau_{k-1}}\big\}\big] < \infty \qquad \text{for all } k = 1,\ldots,n. \qquad (4.2)$$

Then $\mathscr{E}(M)$ is a martingale.

Proof. The following technique of proving the result is similar to the one used in the proof of [24, Lemma 13]. For a given $k \in \{1,\ldots,n\}$ define the process

$$M_t^{(k)} = M_{t\wedge\tau_k} - M_{t\wedge\tau_{k-1}}, \qquad t \in [0,T].$$

Note that $(M_t^{(k)})_{t\in[0,T]}$ is a (purely discontinuous) local martingale and consequently, $\mathscr{E}(M^{(k)})$ is a local martingale. Due to the jump structure

$$\Delta M_t^{(k)} = \begin{cases} \Delta M_t & \text{if } t \in (\tau_{k-1}, \tau_k], \\ 0 & \text{otherwise}, \end{cases} \qquad (4.3)$$

it holds that

$$\sum_{s\leq t} f\big(\Delta M_s^{(k)}\big) = \sum_{s\leq t\wedge\tau_k} f(\Delta M_s) - \sum_{s\leq t\wedge\tau_{k-1}} f(\Delta M_s), \qquad t \in [0,T]. \qquad (4.4)$$

Consequently, the compensator of (4.4) is $(\tilde{A}_{t\wedge\tau_k} - \tilde{A}_{t\wedge\tau_{k-1}})_{t\in[0,T]}$, and due to the assumption in (4.2) it follows by [24, Theorem 8] that $\mathscr{E}(M^{(k)})$ is a martingale.

By [19, p. 404] we know that for $k \in \{1,\ldots,n-1\}$,

$$\mathscr{E}\big(M^{(k)}\big)\mathscr{E}\big(M^{(k+1)}\big) = \mathscr{E}\big(M^{(k)} + M^{(k+1)} + [M^{(k)}, M^{(k+1)}]\big).$$

Using (4.3) and that $(M_t^{(k)})_{t\in[0,T]}$ is purely discontinuous, one finds that $[M^{(k)}, M^{(k+1)}] = 0$, so for any $t \in [0, \tau_k]$,

$$\mathscr{E}(M)_t = \mathscr{E}\Big(\sum_{l=1}^k M^{(l)}\Big)_t = \prod_{l=1}^k \mathscr{E}\big(M^{(l)}\big)_t.$$

Since $\mathscr{E}(M^{(l)})_t = \mathscr{E}(M^{(l)})_{\tau_{k-1}}$ for all $t \geq \tau_{k-1}$ and $l < k$,

$$\mathbb{E}[\mathscr{E}(M)_{\tau_k}] = \mathbb{E}\Big[\mathbb{E}[\mathscr{E}(M^{(k)})_{\tau_k} \mid \mathscr{F}_{\tau_{k-1}}]\prod_{l=1}^{k-1}\mathscr{E}(M^{(l)})_{\tau_{k-1}}\Big] = \mathbb{E}[\mathscr{E}(M)_{\tau_{k-1}}].$$

As a consequence, we get inductively that $\mathbb{E}[\mathscr{E}(M)_T] = \mathbb{E}[\mathscr{E}(M)_0] = 1$. By using the fact that $\mathscr{E}(M)$ is a supermartingale, we have the result. □

Proof of Theorem 2.5. We divide the proof into two steps; the first step is to show that assumptions (i)–(ii) on $(P_t)_{t\in[0,T]}$ imply that for any $\varepsilon \in (0, 1/K)$,

$$\sup_{t\in[0,T]} \mathbb{E}\big[e^{\varepsilon|P_t|\log(1+|P_t|)}\big] < \infty, \tag{4.5}$$

and the second step will use this fact to prove that $W*(\mu-\nu)_t$, $t \in [0,T]$, is well-defined and that $\mathscr{E}(W*(\mu-\nu))$ is a martingale.

Step 1: The idea is to use a procedure similar to the one in [26, Lemma 26.5] and exploit the tightness property of $(P_t)_{t\in[0,T]}$ to get a uniform result across t. In the following we write

$$\Psi_t(u) := \log\mathbb{E}\big[e^{uP_t}\big] = \frac{1}{2}c_t u^2 + \int_{\mathbb{R}}\big(e^{ux} - 1 - ux\mathbb{1}_{[-1,1]}(x)\big)F_t(\mathrm{d}x) + b_t u, \qquad u \in \mathbb{R},$$

for the Laplace exponent of P_t with associated triplet (c_t, F_t, b_t), $t \in [0,T]$, relative to the truncation function $h(x) = x\mathbb{1}_{[-1,1]}(x)$. By the compact support of F_t, it follows from [26, Theorem 25.17] that $\Psi_t(u) \in \mathbb{R}$ is well-defined for all $u \in \mathbb{R}$ and $t \in [0,T]$. For fixed t, it holds that $\Psi_t \in C^\infty$,

$$\Psi_t'(u) = c_t u + \int_{\mathbb{R}}\big(xe^{ux} - x\mathbb{1}_{[-1,1]}(x)\big)F_t(\mathrm{d}x) + b_t, \qquad u \in \mathbb{R}, \tag{4.6}$$

and $\Psi_t'' > 0$, see [26, Lemma 26.4]. From (4.6) and the inequality $|e^{ux} - 1| \leq e^{uK}|x|$ for $x \in [-K, K]$ and $u \geq 0$, we get the bound

$$\Psi_t'(u) \leq c_t u + e^{uK}\int_{\mathbb{R}}x^2 F_t(\mathrm{d}x) + b_t + KF_t((1, K]). \tag{4.7}$$

Now suppose that $\sup_{t\in[0,T]}\int_{\mathbb{R}}x^2 F_t(\mathrm{d}x) = \infty$. Then, by the tightness of $(P_t)_{t\in[0,T]}$, we may according to Prokhorov's theorem choose a sequence $(t_n)_{n\geq 1} \subseteq [0,T]$ and a random variable P such that

$$P_{t_n} \xrightarrow{\mathscr{D}} P \quad \text{and} \quad \lim_{n\to\infty}\int_{\mathbb{R}}x^2 F_{t_n}(\mathrm{d}x) = \infty. \tag{4.8}$$

Since P is infinitely divisible, it has an associated characteristic triplet (c, ρ, b). By [16, Theorem 2.9 (Ch. VII)] it holds that

$$\lim_{n\to\infty}\int_{\mathbb{R}}g\,\mathrm{d}F_{t_n} = \int_{\mathbb{R}}g\,\mathrm{d}\rho$$

for all $g \colon \mathbb{R} \to \mathbb{R}$ which are continuous, bounded, and vanishing in a neighbourhood of 0. In particular, by the uniformly compact support of $(F_{t_n})_{n\geq 1}$, we get that ρ is

compactly supported. As a consequence, [16, Theorem 2.14 (Ch. VII)] and (4.8) imply that

$$c + \int_{\mathbb{R}} x^2 \rho(\mathrm{d}x) = \lim_{n \to \infty} \left(c_{t_n} + \int_{\mathbb{R}} x^2 F_{t_n}(\mathrm{d}x) \right) = \infty,$$

a contradiction, and we conclude that $\sup_{t \in [0,T]} \int_{\mathbb{R}} x^2 F_t(\mathrm{d}x) < \infty$. The same reasoning gives that both $\sup_{t \in [0,T]} c_t$ and $\sup_{t \in [0,T]} (b_t + F_t((1,K]))$ are finite as well. From these observations and (4.7) we deduce the existence of a constant $C > 0$ such that

$$\Psi_t'(u) \le C\left(1 + u + e^{uK}\right) \tag{4.9}$$

for $u \ge 0$. We may without loss of generality assume that

$$\lim_{u \to \pm\infty} \Psi_t'(u) = \infty \tag{4.10}$$

for all $t \in [0,T]$. To see this, let N^+ and N^- be standard Poisson random variables which are independent of each other and of $(P_t)_{t \in [0,T]}$, and consider the process

$$\tilde{P}_t = P_t + K(N^+ - N^-), \qquad t \in [0,T].$$

This process still satisfies assumptions (i)–(ii) stated in Theorem 2.5, and the derivative of the associated Laplace exponents will necessarily satisfy (4.10) by the structure in (4.6), since $\tilde{P}_t$ has a Lévy measure with mass on both $(-\infty, 0]$ and $[0, \infty)$. Moreover, the inequality

$$\mathbb{P}(N^+ = 0)^{-2} \sup_{t \in [0,T]} \mathbb{E}\left[e^{\varepsilon|\tilde{P}_t|\log(1+|\tilde{P}_t|)}\right] \ge \sup_{t \in [0,T]} \mathbb{E}\left[e^{\varepsilon|P_t|\log(1+|P_t|)}\right]$$

implies that it suffices to show (4.5) for $(\tilde{P}_t)_{t \in [0,T]}$. Thus, we will continue under the assumption that (4.10) holds. Now, by [26, Lemma 26.4] we may find a constant $\xi_0 > 0$ such that for any t, the inverse of Ψ_t', denoted by θ_t, exists on (ξ_0, ∞) and

$$\mathbb{P}(P_t \ge x) \le \exp\left\{ -\int_{\xi_0}^x \theta_t(\xi) \, \mathrm{d}\xi \right\} \qquad \text{for any } x > \xi_0. \tag{4.11}$$

Since $\lim_{\xi \to \infty} \theta_t(\xi) = \infty$ and $K - 1/\varepsilon' < 0$ for $\varepsilon' \in (\varepsilon, 1/K)$, it follows by (4.9) that $\lim_{\xi \to \infty} \xi e^{-\theta_t(\xi)/\varepsilon'} = 0$. In particular, by (4.9) once again, we can choose a $\xi_1 \ge \xi_0$ (independent of t) such that $-\theta_t(\xi) \le -\varepsilon' \log \xi$ for every $\xi \ge \xi_1$. Combining this fact with (4.11) gives that

$$\mathbb{P}(P_t \ge x) \le \exp\left\{ -\varepsilon' \int_{\xi_1}^x \log \xi \, \mathrm{d}\xi \right\} \le \tilde{C} e^{-\varepsilon' x(\log x - 1)} \qquad \text{for } x > \xi_1 \text{ and } t \in [0,T],$$

where $\tilde{C}$ is some constant independent of t. By estimating the probability $\mathbb{P}(P_t \le -x) = \mathbb{P}(-P_t \ge x)$ in a similar way it follows that ξ_1 and $\tilde{C}$ can be chosen large enough to ensure that

$$G(x) := \sup_{t \in [0,T]} \mathbb{P}(|P_t| \ge x) \le \tilde{C} e^{-\varepsilon' x \log x} \qquad \text{for all } t \in [0,T] \text{ and } x \ge \xi. \tag{4.12}$$

If we set $G_t(x) = \mathbb{P}(|P_t| \ge x)$ for $x \ge 0$, we have

$$\mathbb{E}\left[e^{\varepsilon|P_t|\log(1+|P_t|)}\right] = -\int_0^\infty e^{\varepsilon x \log(1+x)} \, G_t(\mathrm{d}x)$$

$$= 1 + \varepsilon \int_0^\infty e^{\varepsilon x \log(1+x)} \left(\log(1+x) + \frac{x}{1+x} \right) G_t(x) \, \mathrm{d}x$$

using integration by parts, and this implies in turn that

$$\sup_{t\in[0,T]} \mathbb{E}\big[e^{\varepsilon|P_t|\log(1+|P_t|)}\big] \leq 1 + \varepsilon \int_0^\infty e^{\varepsilon x \log(1+x)}(\log(1+x)+1)G(x)\,\mathrm{d}x < \infty$$

by (4.12). Consequently, we have shown that (4.5) does indeed hold.

Step 2: Arguing that $W * (\mu - \nu)_t$, $t \in [0,T]$, is well-defined amounts to showing that $\mathbb{E}[|W| * \nu_T] < \infty$. This is clearly the case since by (2.8),

$$\mathbb{E}[W * \nu_T] \leq T \sup_{t\in[0,T]} \mathbb{E}[|P_t|] \int_{\mathbb{R}} g(x)F(\mathrm{d}x),$$

and the right-hand side is finite by (4.5). By definition we have the equality

$$f(W) * \mu_t = \sum_{s\leq t} f(\Delta(W * (\mu - \nu))_s), \qquad t \in [0,T],$$

and the compensator of the process exists and is given as $\tilde{A} = f(W) * \nu$, since

$$\mathbb{E}[\tilde{A}_T] \leq T\bigg[\Big(1 + \sup_{t\in[0,T]} \mathbb{E}[|P_t|]\Big) \int_{\mathbb{R}} g(x)\log(1+g(x))F(\mathrm{d}x)$$

$$+ \sup_{t\in[0,T]} \mathbb{E}[|P_t|\log(1+|P_t|)] \int_{\mathbb{R}} g(x)F(\mathrm{d}x)\bigg],$$

which is finite by assumption (b) and (4.5). In the following we will argue that (4.2) in Lemma 4.1 is satisfied for $\tau_k \equiv t_k$, $k = 0, 1, \ldots, n$, for suitable numbers $0 = t_0 < t_1 < \cdots < t_n = T$, which subsequently allows us to conclude that $\mathscr{E}(W * (\mu - \nu))$ is a martingale. Fix $0 \leq s < t \leq T$ and note that (2.8) implies

$$\tilde{A}_t - \tilde{A}_s \leq \int_s^t \int_{\mathbb{R}} f(|P_u|g(x))F(\mathrm{d}x)\,\mathrm{d}u, \qquad (4.13)$$

since f is increasing on $\mathbb{R}_+$. We want to obtain a bound on $h(y) := \int_{\mathbb{R}} f(yg(x))F(\mathrm{d}x)$ for $y \geq 0$. First note that $f(x) \leq x\log(1+x)$ whenever $x \geq 0$, so

$$\frac{h(y)}{y\log y} \leq \int_{\mathbb{R}} g(x)\bigg[1 + \frac{\log(1+g(x))}{\log y}\bigg]F(\mathrm{d}x) \qquad y > 0.$$

Consequently, due to assumption (b), we can apply Lebesgue's theorem on dominated convergence to deduce that

$$\limsup_{y\to\infty} \frac{h(y)}{y\log(1+y)} < \gamma_1 \qquad \text{for some } \gamma_1 \in (0,\infty).$$

By monotonicity of h we may find $\gamma_2 \in (0,\infty)$ so that we obtain the bound $h(y) \leq \gamma_1 y\log(1+y) + \gamma_2$ for all $y \geq 0$. Thus for all $0 \leq s < t \leq T$, we have established the estimate

$$\int_s^t h(|P_u|)\,\mathrm{d}u \leq \gamma_1 \int_s^t |P_u|\log(1+|P_u|)\,\mathrm{d}u + \gamma_2(t-s). \qquad (4.14)$$

Now choose a partition $0 = t_0 < t_1 < \cdots < t_n = T$ with $t_k - t_{k-1} \leq \varepsilon/\gamma_1$ for some small number ε satisfying (4.5) holds. By (4.13) and (4.14) it follows by an application of Jensen's inequality and Tonelli's theorem that

$$
\begin{aligned}
e^{-\gamma_2(t_k - t_{k-1})} \mathbb{E}\big[e^{\tilde{A}_{t_k} - \tilde{A}_{t_{k-1}}}\big] &\leq \mathbb{E}\Big[\exp\Big\{\gamma_1 \int_{t_{k-1}}^{t_k} |P_t| \log(1 + |P_t|)\, \mathrm{d}t\Big\}\Big] \\
&\leq \frac{1}{t_k - t_{k-1}} \int_{t_{k-1}}^{t_k} \mathbb{E}\big[e^{\varepsilon |P_t| \log(1 + |P_t|)}\big]\, \mathrm{d}t \\
&\leq \sup_{t \in [0,T]} \mathbb{E}\big[e^{\varepsilon |P_t| \log(1 + |P_t|)}\big],
\end{aligned}
$$

which is finite and, thus, the proof is completed. □

Remark 4.2. It appears from (4.13) above that if $F(\mathbb{R}) < \infty$ or $W(u,x) = 0$ for $x \in (-\delta, \delta)$ with $\delta > 0$, one may allow that W takes values in $(-1, \infty)$ by assuming that $|W(t,x)| \leq |P_t| g(x)$ and replacing the inequality with

$$
\tilde{A}_t - \tilde{A}_s \leq M(t - s) + \int_s^t \int_{\mathbb{R}} f(|P_u| g(x)) F(\mathrm{d}x)\, \mathrm{d}u
$$

for a suitable $M > 0$. From this point, one can complete the proof in the same way as above and get that $\mathscr{E}(W * (\mu - \nu))$ is a martingale.

Remark 4.3. Note that there are other sets of assumptions that can be used to show Theorem 2.5, but they will not be superior to those suggested. Furthermore, the assumptions that we suggest are natural in order to formulate Theorem 2.1 in a way which in turn is suited for proving Theorem 1.2 in the introduction. However, by adjusting the set of assumptions in Theorem 2.5, one may obtain similar adjusted versions of Theorem 2.1 (see the discussion in Remark 2.3). In the bullet points below we shortly point out which properties the assumptions should imply and suggest other choices as well.

- The importance of (i)–(ii) is that they ensure (4.5) holds. Thus, it follows that one may replace these by $P_t \overset{\mathscr{D}}{=} P_0$ for $t \in [0,T]$ and $\mathbb{E}[e^{\varepsilon |P_0| \log(1 + |P_0|)}] < \infty$ for some $\varepsilon > 0$.

- Instead of assuming that $(P_t)_{t \in [0,T]}$ is a process satisfying (4.5) and $g + g \log(1 + g) \in L^1(F)$, one may do a similar proof under the assumptions that

$$
\sup_{t \in [0,T]} \mathbb{E}\big[e^{\varepsilon |P_t|^\gamma}\big] < \infty
$$

and $g \in L^\gamma(F)$ for some $\varepsilon > 0$ and $\gamma \in (1,2]$. In particular, one may allow for less integrability of F around zero for the cost of more integrability of $(P_t)_{t \in [0,T]}$.

Example 4.4 below shows that one cannot relax assumption (i) in Theorem 2.5 and still apply (a localized version of) the approach of Lépingle and Mémin [21]. Moreover, it appears that this approach cannot naturally be improved in the sense of obtaining a weaker condition than (4.5) in order to relax assumption (i).

Example 4.4. Consider the case where $W(t,x) = |Yx|$ for some $\mathscr{F}_0$-measurable infinitely divisible random variable Y with an associated Lévy measure which has unbounded support. Moreover, suppose that $\mathbb{E}[Y^2] < \infty$ and that F is given such that Theorem 2.5(b) holds with $g(x) = |x|$. Then $W * (\mu - \nu)$ is well-defined, and the compensator of $f(W) * \mu$ exists and is given by $f(W) * \nu$ (in the notation of (4.1)). Following the same arguments as in the proof of Theorem 2.5 we obtain that

$$f(W) * \nu_t \geq c_1 t Y \log(1 + Y) - c_2 t, \qquad t \in [0, T],$$

for suitable $c_1, c_2 > 0$. Consequently,

$$\mathbb{E}\big[e^{f(W)*\nu_t}\big] \geq \mathbb{E}\big[e^{c_1 t Y \log(1+Y)}\big]e^{-c_2 t} = \infty$$

for any $t > 0$ by [26, Theorem 26.1]. Thus, Lemma 4.1 cannot be applied if we remove assumption (i) in Theorem 2.5. Naturally, one can ask if it will be sufficient that $\mathbb{E}[e^{\tilde{f}(W)*\nu_t}] < \infty$ for another measurable function $\tilde{f} : (-1, \infty) \to \mathbb{R}_+$? The idea in the proof of [21, Theorem 1 (Section III)] is build on the assumption that $\tilde{f}$ is a function with $(1 - \lambda)\tilde{f}(x) \geq 1 + \lambda x - (1 + x)^\lambda$ for all $x > -1$ and $\lambda \in (0, 1)$. In particular, this requires that

$$\tilde{f}(x) \geq \lim_{\lambda \uparrow 1} \frac{1 + \lambda x - (1 + x)^\lambda}{1 - \lambda} = f(x)$$

for all $x > -1$, and thus any other candidate function will be (uniformly) worse than f.

Before proving Theorem 2.1 we will need a small result, which is stated and proven in Lemma 4.5 below. While the result may be well-known, we have not been able to find an appropriate reference. To a given adapted process $(M_t)_{t \in [0,T]}$ such that $t \mapsto M_t(\omega)$ is a càdlàg step function for each $\omega \in \Omega$, we define for $n \geq 1$ its nth jump time and size by

$$T_n = \inf\{t \in (T_{n-1}, T) : \Delta M_t \neq 0\} \in (0, T] \quad \text{and} \quad Z_n = \Delta M_{T_n}, \qquad (4.15)$$

respectively. Here we set $T_0 \equiv 0$ and $\inf \emptyset = T$.

Lemma 4.5. *Assume that the jump measure J of some càdlàg adapted process $(M_t)_{t \in [0,T]}$ has a predictable compensator ρ of the form $\rho(dt, dx) = G_t(dx)\,dt$, where $(G_t(B))_{t \in [0,T]}$ is a predictable process for every $B \in \mathscr{B}(\mathbb{R})$ and $\lambda_t := G_t(\mathbb{R}) \in (0, \infty)$ for $t \in [0, T]$. Then, in the notation of (4.15), it holds that*

$$\mathbb{P}(Z_n \in B \,|\, \mathscr{F}_{T_n-}) = \Phi_{T_n}(B) \quad \text{on} \quad \{T_n < T\} \qquad (4.16)$$

for any $n \geq 1$ and $B \in \mathscr{B}(\mathbb{R})$, where $\Phi_t := G_t / \lambda_t$.

Proof. To show (4.16), fix $n \geq 1$ and $B \in \mathscr{B}(\mathbb{R})$. Note that $\mathscr{F}_{T_n-}$ is generated by sets of the form $A \cap \{t < T_n\}$ for $t \in [0, T)$ and $A \in \mathscr{F}_t$. Consequently, it suffices to argue that

$$\mathbb{E}[\mathbb{1}_{A \cap \{t < T_n < T\}}\mathbb{1}_B(Z_n)] = \mathbb{E}[\mathbb{1}_{A \cap \{t < T_n < T\}}\Phi_{T_n}(B)]. \qquad (4.17)$$

Define the functions $\phi, \psi : \Omega \times [0, T] \times \mathbb{R} \to \mathbb{R}$ by

$$\phi(s,x) = \mathbb{1}_{A \cap \{t < T_n\}}\big[\mathbb{1}_{\{T_{n-1} \leq t\}}\mathbb{1}_{(t,T_n] \times B}(s,x) + \mathbb{1}_{\{T_{n-1} > t\}}\mathbb{1}_{(T_{n-1}, T_n] \times B}(s,x)\big]\mathbb{1}_{(0,T)}(s)$$

and

$$\psi(s,x) = \mathbb{1}_{A \cap \{t < T_n\}} \Phi_s(B) [\mathbb{1}_{\{T_{n-1} \le t\}} \mathbb{1}_{(t,T_n]}(s) + \mathbb{1}_{\{T_{n-1} > t\}} \mathbb{1}_{(T_{n-1},T_n]}(s)] \mathbb{1}_{(0,T)}(s),$$

and note that they are both predictable. Furthermore, we observe that the functions are defined such that

$$\phi * J_T = \mathbb{1}_{A \cap \{t < T_n < T\}} \mathbb{1}_B(Z_n), \quad \psi * J_T = \mathbb{1}_{A \cap \{t < T_n < T\}} \Phi_{T_n}(B),$$

and

$$\phi * \rho_T = \mathbb{1}_{A \cap \{t < T_n < T\}} \int_0^T G_s(B) [\mathbb{1}_{\{T_{n-1} \le t\}} \mathbb{1}_{(t,T_n]}(s) + \mathbb{1}_{\{T_{n-1} > t\}} \mathbb{1}_{(T_{n-1},T_n]}(s)] \, \mathrm{d}s = \psi * \rho_T.$$

Using these properties together with the dual relations $\mathbb{E}[\phi * J_T] = \mathbb{E}[\phi * \rho_T]$ and $\mathbb{E}[\psi * \rho_T] = \mathbb{E}[\psi * \rho_T]$ we obtain (4.17), and this gives the result. $\quad\square$

Using Lemma 4.5 it follows by a Monotone Class argument that on $\{T_n < T\}$,

$$\mathbb{E}[g(Z_n) \mid \mathscr{F}_{T_n-}] = \int_{\mathbb{R}} g(x) \Phi_{T_n}(\mathrm{d}x) \tag{4.18}$$

for any function $g \colon \Omega \times \mathbb{R} \to \mathbb{R}_+$ which is $\mathscr{F}_{T_n-} \otimes \mathscr{B}(\mathbb{R})$-measurable. With this fact and Theorem 2.5 in hand, we are ready to prove Theorem 2.1.

Proof of Theorem 2.1. We prove the result depending on the different hypotheses. In both cases the proof goes by arguing that $\mathscr{E}((\alpha-1)*(\mu-\nu))$ is a martingale and that the probability measure $\mathbb{Q}$ defined by $\mathrm{d}\mathbb{Q} = \mathscr{E}((\alpha-1)*(\mu-\nu))_T \, \mathrm{d}\mathbb{P}$ is an ELMM for $(X_t)_{t \in [0,T]}$. Since the differential characteristics of $(L_t)_{t \in [0,T]}$ under $\mathbb{P}$ coincide with its characteristic triplet (c, F, b^h), it follows directly from (3.2) that if $\mathbb{Q}$ is a probability measure, the differential characteristics of $(X_t)_{t \in [0,T]}$ under $\mathbb{Q}$ are given as in (2.3). In the following we have fixed $a, b > 0$ such that (2.1) holds.

Case (h1)*:* Consider the specific predictable function α given by (2.4). Then $\alpha(t,x) \ge 1$ and in particular, $\mathscr{E}((\alpha-1)*(\mu-\nu))_t > 0$. Moreover, $\alpha(t,x) - 1 \le |P_t| g(x)$ with $P_t = Y_t + \xi$ and $g(x) = C \mathbb{1}_{(a,b)}(|x|)|x|$ for some constant $C > 0$. Since ξ is just a constant, $(P_t)_{t \in [0,T]}$ inherits the properties in (h1) of $(Y_t)_{t \in [0,T]}$, thus (a) in Theorem 2.5 is satisfied. Likewise,

$$\int_{\mathbb{R}} g[1 + \log(1 + g)] \, \mathrm{d}F = \int_{|x| \in (a,b)} C|x| [1 + \log(1 + C|x|)] F(\mathrm{d}x) < \infty,$$

which shows that (b) is satisfied as well, and we conclude by Theorem 2.5 that $\mathscr{E}((\alpha-1)*(\mu-\nu))$ is a martingale. To argue that $(X_t)_{t \in [0,T]}$ is a local martingale under the associated probability measure $\mathbb{Q}$, note that it suffices to show that

$$\int_{|x| \in (a,b)} x \alpha(t,x) F(\mathrm{d}x) = -(Y_t + b^h)$$

by (2.3) and (2.7) in Remark 2.2, where $b^h \in \mathbb{R}$ is the drift component in the characteristic triplet of L with respect to the (pseudo) truncation function $h(x) = x \mathbb{1}_{(a,b)^c}(|x|)$.

Thus, we compute

$$\int_{|x|\in(a,b)} x\alpha(t,x)F(dx)$$

$$= \int_{|x|\in(a,b)} x F(dx) + \frac{(Y_t+\xi)^-}{\sigma_+^2} \int_{(a,b)} x^2 F(dx) - \frac{(Y_t+\xi)^+}{\sigma_-^2} \int_{(-b,-a)} x^2 F(dx)$$

$$= \int_{|x|\in(a,b)} x F(dx) - (Y_t+\xi)$$

$$= \int_{|x|\in(a,b)} x F(dx) - Y_t - \int_{|x|\in(a,b)} x F(dx) - b^h$$

$$= -(Y_t + b^h),$$

and the result is shown under hypothesis (h1).

Case (h2): Set $F^a = F(\,\cdot\,\cap[-a,a]^c)$. Note that $F^a((-\infty,\zeta)), F^a((\zeta,\infty)) > 0$ for any $\zeta \in \mathbb{R}$ by assumption, and this implies that we may find a strictly positive density $f_\zeta \colon \mathbb{R} \to (0,\infty)$ such that

$$\int_{[-a,a]^c} \frac{f_\zeta(x)}{F^a(\mathbb{R})} F(dx) = 1 \quad \text{and} \quad \int_{[-a,a]^c} x \frac{f_\zeta(x)}{F^a(\mathbb{R})} F(dx) = \zeta. \tag{4.19}$$

To see this, assume that X is a random variable on $(\Omega, \mathscr{F}, \mathbb{P})$ with $X \overset{\mathscr{D}}{=} F^a/F^a(\mathbb{R})$. Then, since $\mathbb{E}[X \mid X < \zeta] < \zeta < \mathbb{E}[X \mid X \geq \zeta]$, we may define

$$\lambda(\zeta) = \frac{\zeta - \mathbb{E}[X \mid X < \zeta]}{\mathbb{E}[X \mid X \geq \zeta] - \mathbb{E}[X \mid X < \zeta]} \in (0,1) \tag{4.20}$$

and

$$\omega_\zeta(B) = (1 - \lambda(\zeta))\mathbb{P}(X \in B \mid X < \zeta) + \lambda(\zeta)\mathbb{P}(X \in B \mid X \geq \zeta), \qquad B \in \mathscr{B}(\mathbb{R}).$$

Note that ω_ζ is a probability measure which is equivalent to $\mathbb{P}(X \in \,\cdot\,) = F^a/F^a(\mathbb{R})$ and has mean ζ. Thus, the density $d\omega_\zeta/d\mathbb{P}(X \in \,\cdot\,)$ is a function that satisfies (4.19). Moreover, such a density is explicitly given by

$$f_\zeta(x) = \frac{1-\lambda(\zeta)}{\mathbb{P}(X < \zeta)}\mathbb{1}_{(-\infty,\zeta)}(x) + \frac{\lambda(\zeta)}{\mathbb{P}(X \geq \zeta)}\mathbb{1}_{[\zeta,\infty)}(x), \qquad x \in \mathbb{R},$$

and we thus see that the map $(x,\zeta) \mapsto f_\zeta(x)$ is $\mathscr{B}(\mathbb{R}^2)$-measurable. By letting

$$\alpha(t,x) = f_{-(Y_t+b^h)/F^a(\mathbb{R})}(x)$$

for $|x| > a$ and $\alpha(t,x) = 1$ for $|x| \leq a$, we obtain a predictable function α, which is strictly positive and satisfies (2.5). Thus, it suffices to argue that an α with these properties defines an ELMM for $(X_t)_{t\in[0,T]}$ through $\mathscr{E}((\alpha-1)*(\mu-\nu))_T$. First observe that $(\alpha-1)*(\mu-\nu)$ is well-defined, since

$$|\alpha-1|*\nu_T = \int_0^T \int_{[-a,a]^c} |\alpha(s,x)-1|F(dx)\,ds \leq 2F^a(\mathbb{R})T.$$

The first property in (2.5) and the fact that $\alpha(t,x) = 1$ when $|x| \leq a$ imply that $(\alpha-1)*\nu_t = 0$ for $t \in [0,T]$. Consequently,

$$\mathscr{E}((\alpha-1)*(\mu-\nu))_t = \mathscr{E}((\alpha-1)*\mu)_t = e^{(\alpha-1)*\mu_t + (\log\alpha-(\alpha-1))*\mu_t} = \prod_{n=1}^{N_t} \alpha(T_n, Z_n),$$

where $(T_n, Z_n)_{n \geq 1}$ is defined as in (4.15) for the compound Poisson process $x\mathbb{1}_{[-a,a]^c}(x) * \mu_t$, $t \in [0, T]$, and $N_t = \mathbb{1}_{[-a,a]^c}(x) * \mu_t$, $t \in [0, T]$, is the underlying Poisson process that counts the number of jumps. In particular, for any given $n \geq 1$, we have

$$\mathbb{E}[\mathscr{E}((\alpha - 1) * \mu)_{T_n} \mid \mathscr{F}_{T_{n-1}}] = \mathscr{E}((\alpha - 1) * \mu)_{T_{n-1}} \mathbb{E}[\mathbb{E}[\alpha(T_n, Z_n) \mid \mathscr{F}_{T_n-}] \mid \mathscr{F}_{T_{n-1}}]$$
$$= \mathscr{E}((\alpha - 1) * \mu)_{T_{n-1}}$$

almost surely by the inclusion $\mathscr{F}_{T_{n-1}} \subseteq \mathscr{F}_{T_n-}$, if we can show that

$$\mathbb{E}[\alpha(T_n, Z_n) \mid \mathscr{F}_{T_n-}] = 1.$$

(Here we recall that $T_0 \equiv 0$.) However, this follows from the observations that $\alpha(T_n, Z_n) = 1$ almost surely on the set $\{T_n = T\}$ (since $Z_n = 0$) and

$$\mathbb{E}[\alpha(T_n, Z_n) \mid \mathscr{F}_{T_n-}] = F^a(\mathbb{R})^{-1} \int_{[-a,a]^c} \alpha(T_n, x) F(\mathrm{d}x) = 1$$

almost surely on $\{T_n < T\}$. The latter observation is implied by (2.5) and (4.18), since $(\omega, x) \mapsto \alpha(\omega, T_n(\omega), x)$ is $\mathscr{F}_{T_n-} \otimes \mathscr{B}(\mathbb{R})$-measurable. Consequently, $(\mathscr{E}((\alpha - 1) * \mu)_{T_n})_{n \geq 0}$ is a positive $\mathbb{P}$-martingale with respect to the filtration $(\mathscr{F}_{T_n})_{n \geq 0}$ and its mean is constantly equal to one, so we may define a probability measure $\mathbb{Q}^n$ on $\mathscr{F}_{T_n}$ by $\mathrm{d}\mathbb{Q}^n/\mathrm{d}\mathbb{P} = \mathscr{E}((\alpha - 1) * \mu)_{T_n}$ for each $n \geq 1$. By (3.2) it follows that the compensator of μ under $\mathbb{Q}^n$ is $[\alpha(t, x)\mathbb{1}_{\{t \leq T_n\}} + \mathbb{1}_{\{t > T_n\}}] F(\mathrm{d}x) \, \mathrm{d}t$. From this we get that the counting process $\mathbb{1}_{[-a,a]^c}(x) * \mu_t$, $t \in [0, T]$, is compensated by

$$\int_0^t \int_{\mathbb{R}} \mathbb{1}_{[-a,a]^c}(x)[\alpha(s, x)\mathbb{1}_{\{s \leq T_n\}} + \mathbb{1}_{\{s > T_n\}}] F(\mathrm{d}x) \, \mathrm{d}s = F^a(\mathbb{R})t, \qquad t \in [0, T],$$

under $\mathbb{Q}^n$ using (2.5). This shows that jumps continue to arrive according to a Poisson process with intensity $F^a(\mathbb{R})$ (see, e.g., [16, Theorem 4.5 (Ch. II)]), which in turn implies that

$$\mathbb{E}[\mathscr{E}((\alpha - 1) * \mu)_{T_n} \mathbb{1}_{\{T_n < T\}}] = \mathbb{Q}^n(T_n < T) = \mathbb{P}(T_n < T) \to 0, \qquad n \to \infty.$$

As a consequence,

$$1 = \lim_{n \to \infty} \mathbb{E}[\mathscr{E}((\alpha - 1) * \mu)_{T_n} \mathbb{1}_{\{T_n < T\}}] + \lim_{n \to \infty} \mathbb{E}[\mathscr{E}((\alpha - 1) * \mu)_T \mathbb{1}_{\{T_n = T\}}]$$
$$= \lim_{n \to \infty} \mathbb{E}[\mathscr{E}((\alpha - 1) * \mu)_T \mathbb{1}_{\{T_n = T\}}]$$
$$= \mathbb{E}[\mathscr{E}((\alpha - 1) * \mu)_T].$$

This shows that $\mathbb{Q}$ defined by $\mathrm{d}\mathbb{Q} = \mathscr{E}((\alpha - 1) * \mu)_T \, \mathrm{d}\mathbb{P}$ is a probability measure on $(\Omega, \mathscr{F})$. To show that $(X_t)_{t \in [0, T]}$ is a local martingale under $\mathbb{Q}$ we just observe that the compensator of $x\mathbb{1}_{[-a,a]^c}(x) * \mu_t$, $t \in [0, T]$, is given by

$$\int_0^t \int_{[-a,a]^c} x\alpha(s, x) F(\mathrm{d}x) \, \mathrm{d}s = -\int_0^t (Y_s + b^h) \, \mathrm{d}s, \qquad t \in [0, T],$$

according to (2.5). Thus, (2.7) holds and the proof is complete by Remark 2.2. □

Finally, we use Theorem 2.1 to prove Theorem 1.2.

Proof of Theorem 1.2. Without loss of generality we assume that $\mathbb{E}[L_1] = 0$. Suppose that $(X_t)_{t \in [0,T]}$ admits an ELMM. Then, it is a semimartingale and the assumptions imposed imply by [4, Theorem 4.1 and Corollary 4.8] that φ is absolutely continuous with a density φ' satisfying (1.3). Moreover, we find for any $t \in [0,T]$ that

$$
\begin{aligned}
X_t - X_0 &= \int_0^t \varphi(t-s)\,\mathrm{d}L_s + \int_{-\infty}^0 [\varphi(t-s) - \varphi(-s)]\,\mathrm{d}L_s \\
&= \varphi(0)L_t + \int_0^t \int_{-s}^{t-s} \varphi'(u)\,\mathrm{d}u\,\mathrm{d}L_s + \int_{-\infty}^0 \int_{-s}^{t-s} \varphi'(u)\,\mathrm{d}u\,\mathrm{d}L_s \\
&= \varphi(0)L_t + \int_0^t \int_0^u \varphi'(u-s)\,\mathrm{d}L_s\,\mathrm{d}u + \int_0^t \int_{-\infty}^0 \varphi'(u-s)\,\mathrm{d}L_s\,\mathrm{d}u \\
&= \varphi(0)L_t + \int_0^t Y_u\,\mathrm{d}u
\end{aligned}
\tag{4.21}
$$

where $Y_u = \int_{-\infty}^u \varphi'(u-s)\,\mathrm{d}L_s$. (Here we have applied a stochastic Fubini result, which may be found in [1, Theorem 3.1]. Moreover, we have extended the functions φ and φ' from $\mathbb{R}_+$ to $\mathbb{R}$ by setting $\varphi(t) = \varphi'(t) = 0$ for $t < 0$.) Note that, according to [9] and [16, Theorem 2.28 (Ch. I)], we may choose $(Y_t)_{t \in [0,T]}$ predictable. From this representation we find that $\varphi(0) \neq 0$, since otherwise an ELMM for $(X_t)_{t \in [0,T]}$ would imply $\varphi \equiv 0$.

Conversely, if φ has $\varphi(0) \neq 0$, is absolutely continuous, and the density φ' meets (1.3), we get from [4, Theorem 4.1 and Corollary 4.8] that $(X_t)_{t \in [0,T]}$ is a semimartingale that takes the form (4.21). Since the support of the Lévy measure of $(\varphi(0)L_t)_{t \in [0,T]}$ is unbounded on both $(-\infty, 0]$ and $[0, \infty)$, hypothesis (h2) of Theorem 2.1 is satisfied and we deduce the existence of an ELMM for $(X_t)_{t \in [0,T]}$. Suppose now instead that the density φ' is bounded, the support of F (the Lévy measure of L) is bounded, and $F((-\infty, 0)), F((0, \infty)) > 0$. Then we observe initially that, according to [25], $(Y_t)_{t \in [0,T]}$ is a stationary process, in particular tight, under $\mathbb{P}$ and the law of Y_0 is infinitely divisible with a Lévy measure given by

$$
F^Y(B) = (F \times \mathrm{Leb})\big(\{(x,s) \in \mathbb{R} \times (0,\infty) \,:\, x\varphi'(s) \in B \setminus \{0\}\}\big), \qquad B \in \mathscr{B}(\mathbb{R}).
$$

(Here Leb denotes the Lebesgue measure on $(0,\infty)$.) In particular, if $C > 0$ is a constant that bounds φ' we get the inequality

$$
F^Y([-M,M]^c) \leq (F \times \mathrm{Leb})\big([-\tfrac{M}{C}, \tfrac{M}{C}]^c \times (0,\infty)\big)
$$

for any $M > 0$, and this shows that the Lévy measure of Y_0 is compactly supported since the same holds for F. In this case, (h1) of Theorem 2.1 holds and we can conclude that an ELMM for $(X_t)_{t \in [0,T]}$ exists. $\qquad\square$

Remark 4.6. A natural comment is on the existence of $a, b > 0$ with the property (2.1). In light of the structure of the ELMM presented in Theorem 2.1, discussed in Remark 2.2, this assumption seems very natural. Indeed, assume that the triplet of L is given relative to the truncation function $h(x) = x\mathbb{1}_{[-a,a]}(x)$ and set $\tilde{Y}_t = Y_t + b^h$. Then, according to (2.7), we try to find $\mathbb{Q}$ under which

$$
\begin{aligned}
&x\mathbb{1}_{\{|x|>a\}} * \mu_t + \int_0^t \tilde{Y}_s\,\mathrm{d}s \\
&= \Big[x\mathbb{1}_{\{x>a\}} * \mu_t - \int_0^t \tilde{Y}_s^-\,\mathrm{d}s\Big] - \Big[|x|\mathbb{1}_{\{x<-a\}} * \mu_t - \int_0^t \tilde{Y}_s^+\,\mathrm{d}s\Big], \qquad t \in [0,T], \tag{4.22}
\end{aligned}
$$

is a local martingale. Intuitively, $\mathbb{Q}$ should ensure that positive jumps are compensated by the negative drift part and vice versa. Clearly, this construction is not possible if (2.1) does not hold for any $a, b > 0$ and all jumps are of same sign. In case all jumps are of the same sign, it may sometimes be possible to construct $\mathbb{Q}$, although the recipe becomes rather case specific. For instance, if all jumps of L are positive, one may still make the desired change of measure under (h1) or under the hypothesis that F has unbounded support on $(0, \infty)$, provided that the second term in (4.22) is not present. Even in the case where the term $\int_0^t (Y_s + b^h)^+ \, ds$, $t \in [0, T]$, is non-zero, it might possibly be absorbed by a change of drift of the Gaussian component in L if such exists.

Acknowledgments

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]   Barndorff-Nielsen, O. E. and A. Shiryaev (2015). *Change of Time and Change of Measure*. Second. Advanced Series on Statistical Science & Applied Probability, 21. World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ, xviii+326. DOI: 10.1142/9609.

[3]   Basse-O'Connor, A. and J. Pedersen (2009). Lévy driven moving averages and semimartingales. *Stochastic Process. Appl.* 119(9), 2970–2991. DOI: 10.1016/j.spa.2009.03.007.

[4]   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.

[5]   Bichteler, K. (2002). *Stochastic integration with jumps*. Vol. 89. Encyclopedia of Mathematics and its Applications. Cambridge University Press, xiv+501. DOI: 10.1017/CBO9780511549878.

[6]   Brémaud, P. (1981). *Point Processes and Queues*. Martingale dynamics, Springer Series in Statistics. Springer-Verlag, New York-Berlin, xviii+354.

[7]   Cheridito, P. (2004). Gaussian moving averages, semimartingales and option pricing. *Stochastic Process. Appl.* 109(1), 47–68.

[8]   Cheridito, P., D. Filipović and M. Yor (2005). Equivalent and absolutely continuous measure changes for jump-diffusion processes. *Ann. Appl. Probab.* 15(3), 1713–1732. DOI: 10.1214/105051605000000197.

[9]   Cohn, D. L. (1972). Measurable choice of limit points and the existence of separable and measurable processes. *Z. Wahrsch. Verw. Gebiete* 22, 161–165.

[10] Criens, D. (2016). Structure Preserving Equivalent Martingale Measures for SCII Models. arXiv: 1606.02593.

[11] Dawson, D. A. (1968). Equivalence of Markov processes. *Trans. Amer. Math. Soc.* 131, 1–31.

[12] Delbaen, F. and W. Schachermayer (1994). A general version of the fundamental theorem of asset pricing. *Math. Ann.* 300(3), 463–520.

[13] Eberlein, E. and J. Jacod (1997). On the range of options prices. English. *Finance Stoch.* 1(2), 131–140.

[14] Gjessing, H. K., K. Røysland, E. A. Pena and O. O. Aalen (2010). Recurrent events and the exploding Cox model. *Lifetime Data Anal.* 16(4), 525–546. DOI: 10.1007/s10985-010-9180-y.

[15] Hida, T. and M. Hitsuda (1993). *Gaussian Processes*. Vol. 120. Translations of Mathematical Monographs. Translated from the 1976 Japanese original by the authors. Providence, RI: American Mathematical Society, xvi+183.

[16] Jacod, J. and A. N. Shiryaev (2003). *Limit Theorems for Stochastic Processes*. Second. Vol. 288. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]. Springer-Verlag, Berlin. DOI: 10.1007/978-3-662-05265-5.

[17] Kabanov, Y. M, R. S. Liptser and A. N. Shiryayev (1980). "On absolute continuity of probability measures for Markov-Itô processes". *Stochastic differential systems (Proc. IFIP-WG 7/1 Working Conf., Vilnius, 1978)*. Vol. 25. Lecture Notes in Control and Information Sci. Springer, Berlin-New York, 114–128.

[18] Kallsen, J. (2006). "A didactic note on affine stochastic volatility models". *From stochastic calculus to mathematical finance*. Springer, Berlin, 343–368. DOI: 10.1007/978-3-540-30788-4_18.

[19] Kallsen, J. and A. N. Shiryaev (2002). The cumulant process and Esscher's change of measure. *Finance Stoch.* 6(4), 397–428. DOI: 10.1007/s007800200069.

[20] Knight, F. B. (1992). *Foundations of the Prediction Process*. Vol. 1. Oxford Studies in Probability. Oxford Science Publications. New York: The Clarendon Press Oxford University Press, xii+248.

[21] Lépingle, D. and J. Mémin (1978). Sur l'intégrabilité uniforme des martingales exponentielles. *Z. Wahrsch. Verw. Gebiete* 42(3), 175–203. DOI: 10.1007/BF00641409.

[22] Mijatović, A. and M. Urusov (2012). On the martingale property of certain local martingales. *Probab. Theory Related Fields* 152(1-2), 1–30. DOI: 10.1007/s00440-010-0314-7.

[23] Podolskij, M. (2015). "Ambit fields: survey and new challenges". *XI Symposium on Probability and Stochastic Processes*. Springer, 241–279.

[24] Protter, P. and K. Shimbo (2008). "No arbitrage and general semimartingales". *Markov processes and related topics: a Festschrift for Thomas G. Kurtz*. Vol. 4. Inst. Math. Stat. Collect. Inst. Math. Statist., Beachwood, OH, 267–283. DOI: 10.1214/074921708000000426.

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

[26] 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.

[27] Sokol, A. and N.R. Hansen (2015). Exponential martingales and changes of measure for counting processes. *Stoch. Anal. Appl.* 33(5), 823–843. DOI: 10.1080/07362994.2015.1040890.

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: 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 $\psi \colon \mathbb{R} \to \mathbb{R}$ some measurable function which is integrable with respect to $(L_t)_{t \in \mathbb{R}}$ (in the sense of [23]). Processes of the form

$$X_t = \int_{\mathbb{R}} \psi(t - u) \, \mathrm{d}L_u, \qquad t \in \mathbb{R}, \tag{1.1}$$

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 $\{z \in \mathbb{C} : \text{Re}(z) = 0\}$, choosing $\psi \colon \mathbb{R} \to \mathbb{R}$ to be the function characterized by

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

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

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

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

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, eventually, 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 $\theta \colon \mathbb{R} \to \mathbb{R}$ 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

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

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

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 $\phi(dt) = \alpha e^{-\beta t} \mathbb{1}_{[0,\infty)}(t)\, dt$ and $\phi = \sum_{j=1}^{p} \phi_j \delta_j$ for $\alpha, \phi_j \in \mathbb{R}$ and $\beta > 0$. 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

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

for suitable kernels $\theta, \theta_0 \colon \mathbb{R} \to \mathbb{R}$, 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 $\mu \colon \mathscr{B}(\mathbb{R}) \to \mathbb{R}$ of the form $\mu = \mu^+ - \mu^-$, where μ^+ and μ^- are two finite measures which are mutually singular. Integration of a function $f \colon \mathbb{R} \to \mathbb{R}$ 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

$$\int_{\mathbb{R}} e^{-\operatorname{Re}(z)t} |\mu|(dt) < \infty,$$

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

$$\mathcal{L}[\mu](z) = \int_{\mathbb{R}} e^{-zt} \,\mu(\mathrm{d}t).$$

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 $\mathcal{L}[f] := \mathcal{L}[f(t)\,\mathrm{d}t]$. 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

$$\mu * \nu(B) = \int_{\mathbb{R}} \int_{\mathbb{R}} \mathbb{1}_B(t+u) \,\mu(\mathrm{d}t)\,\nu(\mathrm{d}u)$$

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

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

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

$$\int_{\mathbb{R}} \left(f(u)^2 \sigma^2 + \int_{\mathbb{R}} \left(|xf(u)| \wedge |xf(u)|^2 \right) \nu(\mathrm{d}x) \right) \mathrm{d}u < \infty, \tag{1.6}$$

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

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

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 $x_0\colon \mathbb{R} \to \mathbb{R}$ 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 nth 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 $\mathrm{Re}(z) \geq -\delta$ 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 $x_0\colon \mathbb{R} \to \mathbb{R}$, 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*

$$x_0(t) = \mathbb{1}_{[0,\infty)}(t) + \int_{-\infty}^{t} \int_{[0,\infty)} x_0(u - v)\,\eta(\mathrm{d}v)\,\mathrm{d}u, \qquad t \in \mathbb{R}. \tag{2.3}$$

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

   (i) *If η has nth 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 $\mathrm{Re}(z) \geq 0$, 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*

$$\mu_{x_0}(\mathrm{d}u) = \delta_0(\mathrm{d}u) + \left(\int_{[0,\infty)} x_0(u - v)\,\eta(\mathrm{d}v) \right) \mathrm{d}u.$$

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

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

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 nth 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

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

The last term on the right-hand side of (2.5) is finite, since $x_0 \in L^1$ according to 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*

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

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 $\mathrm{Re}(z) \geq 0$, $(X_t)_{t \in \mathbb{R}}$ admits the following causal representation:

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

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

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

for some integrable Lévy process $(L_t)_{t\in\mathbb{R}}$ with $\mathbb{E}[L_1] = 0$ and measurable function $\theta\colon \mathbb{R} \to \mathbb{R}$ 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*

$$X_t = \int_{\mathbb{R}} \theta * \mu_{x_0}(t-u)\, \mathrm{d}L_u, \qquad t \in \mathbb{R}. \tag{2.9}$$

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

$$X_t = \int_{\mathbb{R}} x_0(t-u)\, \mathrm{d}L_u, \qquad t \in \mathbb{R}.$$

Remark 2.7. Let the situation be as in Corollary 2.6 with $h(z) \neq 0$ whenever $\operatorname{Re}(z) \geq 0$. 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 $(L_t)_{t\in\mathbb{R}}$:

$$X_t = \int_{-\infty}^{t} \theta * \mu_{x_0}(t-u)\, \mathrm{d}L_u, \qquad t \in \mathbb{R}.$$

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 $\operatorname{Re}(z) \geq 0$ 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 further 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

$$X_t - X_s = -\lambda \int_{s}^{t} X_u\, \mathrm{d}u + Z_t - Z_s, \qquad s < t. \tag{2.10}$$

In the setup of (2.1) this corresponds to $\eta = -\lambda\delta_0$. With h given by (2.2), we have $h(z) = z + \lambda \neq 0$ for every $z \in \mathbb{C}$ with $\operatorname{Re}(z) \neq -\lambda$, 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 $\{z \in \mathbb{C} : 0 < \mathrm{Re}(z) < a\}$ for a suitable $a > 0$ as

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

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

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

where the kernel g is characterized by

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

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, $iy + b_0 \neq 0$ for all $y \in \mathbb{R}$ and, thus, we may write

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

By choosing $\eta(dt) = (b_0 - a_1)\delta_0(dt) - (a_2 - b_0(a_1 - b_0))e^{-b_0 t}\mathbb{1}_{[0,\infty]}(t) \, dt$ (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

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

where $\beta, \gamma > 0$ and Γ is the gamma function. In this case, $h(z) = z - \alpha_1 - \alpha_2(z + \gamma)^{-\beta}$, and hence h is of the form $P_1(\cdot + \gamma)/P_2(\cdot + \gamma)$, where $P_i(z) = z^{a_i} + b_i z^{c_i} + d_i$ for suitable

constants $a_i, c_i > 0$ 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

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

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

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

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

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

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,

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

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

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

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

$$X_t = \int_{\mathbb{R}} \theta * \mu_{x_0}(t-u) \, \mathrm{d}L_u, \qquad t \in \mathbb{R}.$$

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 $X_t = \int_{\mathbb{R}} x_0(t-u) \, \mathrm{d}Z_u$, 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 ∞:

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

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. Let $(B_t)_{t \in \mathbb{R}}$ be a Brownian motion with respect to a filtration $(\mathscr{F}_t)_{t \in \mathbb{R}}$. Moreover, let $(\sigma_t)_{t \in \mathbb{R}}$ be a predictable process with $\sigma_0 \in L^2(\mathbb{P})$, and 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

$$Z_t = \int_0^t \sigma_u \, \mathrm{d}B_u, \qquad t \in \mathbb{R},$$

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 $\mathrm{Re}(z) \geq 0$ 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

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

for $t \in \mathbb{R}$, where we have used Corollary 2.3, (4.9) 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

$$
X_t = \int_0^\infty X_{t-u}\,\phi(\mathrm{d}u) + \int_{-\infty}^t \theta(t-u)\,\mathrm{d}L_u, \qquad t \in \mathbb{R}, \tag{3.1}
$$

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 $\theta \colon \mathbb{R} \to \mathbb{R}$ 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 stationary 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*

$$
X_t = \int_{-\infty}^t \psi(t-u)\,\mathrm{d}L_u, \qquad t \in \mathbb{R},
$$

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 $\mathrm{Re}(z) \in (0, a)$ and

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

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 $c \in (0, a)$ and

$$\mathcal{L}[\psi](z) = \frac{\mathcal{L}[\theta](z)}{1 - \mathcal{L}[\phi](z)}, \qquad 0 < \mathrm{Re}(z) < a. \tag{3.3}$$

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

$$\mathcal{L}[\psi(t - \cdot)](-z) = e^{zt}(\mathcal{L}[\psi](z)\mathcal{L}[\phi](z) + \mathcal{L}[\theta](z))$$
$$= \mathcal{L}\left[\int_0^{\infty} \psi(t - u - \cdot)\phi(\mathrm{d}u) + \theta(t - \cdot) \right](-z)$$

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

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

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 $\mathrm{d}L_r$ and using a stochastic Fubini result (e.g., [2, Theorem 3.1]) it follows that the moving average $X_t = \int_{\mathbb{R}} \psi(t - r)\,\mathrm{d}L_r$, $t \in \mathbb{R}$, 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 $\mathrm{Re}(z) \in (0, a)$ by the inequality $|\mathcal{L}[\phi](z)| \leq |\phi|((0, \infty))$, and

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

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

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

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 does 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 $\phi(dt) = \alpha e^{-\beta t} \mathbb{1}_{[0,\infty)}(t) \, dt$ and define the measure $\xi(dt) = e^{\alpha t} \phi(dt) = \alpha e^{-(\beta - \alpha)t} \mathbb{1}_{[0,\infty)}(t) \, dt$.

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 $\mathcal{L}[\phi](z) = \alpha/(z + \beta) \neq 1$ and

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

when $\text{Re}(z) > 0$. 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

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

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

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

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 $f, f_0 \colon \mathbb{R} \to \mathbb{R}$, 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

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

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

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

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 is possible to 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 $\Phi, \Theta \colon \mathbb{C} \to \mathbb{C}$ 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

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

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 $\text{Re}(z) \in (0, a)$. Thus, since $\theta \in L^2$, Remark 3.3 implies that there exists a

stationary solution to (3.1), and it is given by $X_t = \int_{\mathbb{R}} \psi(t-u)\,\mathrm{d}L_u$, $t \in \mathbb{R}$, 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

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

for all $z \in \mathbb{C}$ with a positive 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 $-\infty \leq a < b \leq \infty$. Suppose that $F \colon \mathbb{C} \to \mathbb{C}$ is a function which is analytic on the strip $\mathcal{S}_{a,b}$ and satisfies*

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

Then there exists a function $f \colon \mathbb{R} \to \mathbb{C}$ 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 $\int_{\mathbb{R}} e^{zu} f(u)\,\mathrm{d}u = F(z)$ 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 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 $c_1 + iy$:

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

Since

$$\int_{\mathbb{R}}\Big|\int_{c_1}^{c_2} e^{-(x+iy)u}F(x + iy)\,\mathrm{d}x\Big|^2 \,\mathrm{d}y$$
$$\leq e^{-2(c_1 u \wedge c_2 u)}(c_2 - c_1)^2 \sup_{a < x < b}\int_{\mathbb{R}} |F(x + iy)|^2 \,\mathrm{d}y < \infty,$$

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

$$\int_{c_1}^{c_2} e^{-(x \pm i y_n) u} F(x \pm i y_n) \, dx \longrightarrow 0.$$

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

$$\left(u \longmapsto \int_{-y}^{y} e^{-i x u} F(c_k + i x) \, dx \right) \longrightarrow \left(u \longmapsto \mathcal{F}[F(c_k + i \cdot)](u) \right)$$

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 also assume that

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

for Lebesgue almost all $u \in \mathbb{R}$. Combining this with (4.2) yields $e^{-c_1 u} \mathcal{F}[F(c_1 + i \cdot)](u) = e^{-c_2 u} \mathcal{F}[F(c_2 + i \cdot)](u)$ for almost all $u \in \mathbb{R}$. Consequently, there exists a function $f \colon \mathbb{R} \to \mathbb{C}$ with the property that $f(u) = (2\pi)^{-1} e^{-c u} \mathcal{F}[F(c + i \cdot)](u)$ for almost all $u \in \mathbb{R}$ for any given $c \in (a, b)$. For such c we compute

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

Consequently, $(u \mapsto f(u) e^{c u}) \in L^2$ 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

$$\left(\int_{\mathbb{R}} |f(u)| e^{c u} \, du \right)^2$$
$$\leq \left(\int_0^{\infty} |f(u) e^{(c+\varepsilon) u}|^2 \, du + \int_{-\infty}^0 |f(u) e^{(c-\varepsilon) u}|^2 \, du \right) \int_0^{\infty} e^{-2 \varepsilon u} \, du < \infty$$

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

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

and this completes the proof. $\qquad \square$

Proof of Lemma 2.2. Observe that, generally, $h(z) \neq 0$ if $\mathrm{Re}(z) \geq 0$ and $|z| > |\eta|([0, \infty))$ and thus, under the assumption that $h(i y) \neq 0$ 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 $z \in \mathcal{S}_{0,a}$. The fact that $|h(z)| \sim |z|$ as $|z| \to \infty$ when $\mathrm{Re}(z) \geq 0$ and, once again, the continuity of h imply that (4.1) is satisfied for $1/h(- \cdot)$ (over the interval $(-a, 0)$), and thus we get the existence of a function $\tilde{x}_0 \colon \mathbb{R} \to \mathbb{R}$ such that $\mathcal{L}[\tilde{x}_0] = 1/h$ on $\mathcal{S}_{0,a}$ and $t \mapsto e^{c t} \tilde{x}_0(t) \in L^2$ 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 $x_0 \colon \mathbb{R} \to \mathbb{R}$ given by

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

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 $\mathcal{S}_{0,a}$. However, this follows from the following calculation, which holds for an arbitrary $z \in \mathcal{S}_{0,a}$:

$$\mathcal{L}\left[\mathbb{1}_{[0,\infty)} + \int_{-\infty}^{\cdot} \int_{[0,\infty)} \tilde{x}_0(u-v)\,\eta(\mathrm{d}v)\,\mathrm{d}u\right](z)$$

$$= z^{-1}\left[1 + \mathcal{L}[\tilde{x}_0](z)\mathcal{L}[\eta](z)\right] = z^{-1}\frac{z}{z - \mathcal{L}[\eta](z)} = \frac{1}{h(z)}.$$

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

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

for $k \in \{1,\dots,n\}$ (D^k denoting the kth 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

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

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

$$\int_{\mathbb{R}} |x_0(u)|^q\,\mathrm{d}u$$

$$\le \left(\int_{\mathbb{R}} \left(x_0(u)(1+|u|^n)\right)^2\,\mathrm{d}u\right)^{q/2}\left(\int_{\mathbb{R}} (1+|u|^n)^{-2q/(2-q)}\,\mathrm{d}u\right)^{1-q/2} < \infty,$$

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

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

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 $\varepsilon \in (0, \delta)$ such that $1/h(-\cdot)$ satisfies (4.1) over the interval $(-a, \varepsilon)$ and therefore, we have that $u \mapsto x_0(u)e^{cu} \in L^2$ for $c \in [-a, \varepsilon]$. If $h(z) \neq 0$ for all $z \in \mathbb{C}$ with $\mathrm{Re}(z) \ge 0$ we can argue that (4.1) holds for $1/h(-\cdot)$ with $a = -\infty$ and $b = 0$ in the same way as above and, thus, Lemma 4.1 implies $x_0(u) = 0$ for $u < 0$. $\square$

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 $(Y_t)_{t \le s}$, a process $(X_t)_{t \in \mathbb{R}}$ satisfies*

$$X_t = \begin{cases} X_s + \int_s^t \int_{[0,\infty)} X_{u-v}\,\eta(\mathrm{d}v)\,\mathrm{d}u & \text{if} \quad t \ge s, \\ Y_t & \text{if} \quad t < s. \end{cases} \tag{4.3}$$

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

$$X_t = X_s x_0(t-s) + \int_s^\infty \int_{(u-s,\infty)} Y_{u-v}\, \eta(\mathrm{d}v)\, x_0(t-u)\, \mathrm{d}u$$

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 $z \in \mathcal{S}_{0,a}$ (this is possible due to the assumption $h(iy) \neq 0$ for all $y \in \mathbb{R}$). Note that

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

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

$$\mathcal{L}[X\mathbb{1}_{[s,\infty)}](z) = \mathcal{L}\left[\mathbb{1}_{[s,\infty)}\left\{X_s + \int_s^\cdot \int_{[0,\infty)} X_{u-v}\, \eta(\mathrm{d}v)\, \mathrm{d}u\right\}\right](z)$$

$$= \frac{X_s e^{-zs}}{z} + \int_s^\infty e^{-zt} \int_s^t \int_{[0,\infty)} X_{u-v}\, \eta(\mathrm{d}v)\, \mathrm{d}u\, \mathrm{d}t$$

$$= \frac{X_s e^{-zs}}{z} + \int_{[0,\infty)} \int_s^\infty X_{u-v} \int_u^\infty e^{-zt}\, \mathrm{d}t\, \mathrm{d}u\, \eta(\mathrm{d}v)$$

$$= \frac{1}{z}\left(X_s e^{-zs} + \int_{[0,\infty)} \int_{s-v}^\infty X_u e^{-z(u+v)}\, \mathrm{d}u\, \eta(\mathrm{d}v)\right)$$

$$= \frac{1}{z}\left(X_s e^{-zs} + \mathcal{L}[\eta](z)\mathcal{L}[X\mathbb{1}_{[s,\infty)}](z)\right.$$

$$\left. + \mathcal{L}\left[\mathbb{1}_{[s,\infty)} \int_{(\cdot-s,\infty)} Y_{\cdot-v}\, \eta(\mathrm{d}v)\right](z)\right).$$

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 $\mathcal{S}_{0,a}$ almost surely. For instance, Fubini's theorem is applicable in the third equality above for any $z \in \mathcal{S}_{0,a}$ almost surely, since

$$\mathbb{E}\left[\int_s^\infty \int_s^t \int_{[0,\infty)} e^{-n^{-1}t}|X_{u-v}|\, |\eta|(\mathrm{d}v)\, \mathrm{d}u\, \mathrm{d}t\right]$$

$$= |\eta|([0,\infty)) \int_s^\infty (t-s)e^{-n^{-1}t}\, \mathrm{d}t \sup_{t \in \mathbb{R}} \mathbb{E}[|X_t|] < \infty$$

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

$$\mathcal{L}[X\mathbb{1}_{[s,\infty)}](z) = \frac{X_s e^{-zs}}{h(z)} + \frac{\mathcal{L}\left[\mathbb{1}_{[s,\infty)} \int_{(\cdot-s,\infty)} Y_{\cdot-v}\, \eta(\mathrm{d}v)\right](z)}{h(z)}. \tag{4.4}$$

By applying the expectation operator, we note that

$$\int_s^\infty \int_{(u-s,\infty)} |Y_{u-v}|\, |\eta|(\mathrm{d}v)\, |x_0(t-u)|\, \mathrm{d}u < \infty \tag{4.5}$$

almost surely for each $t \in \mathbb{R}$ if $\int_s^\infty |\eta|((u-s,\infty))|x_0(t-u)|\,\mathrm{d}u < \infty$. 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 (4.5) 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

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

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

$$\mathcal{L}\!\left[\int_s^\infty \int_{(u-s,\infty)} Y_{u-v}\,\eta(\mathrm{d}v)\,x_0(\,\cdot\,-u)\,\mathrm{d}u\right]\!(z) = \frac{\mathcal{L}\!\left[\mathbb{1}_{[s,\infty)} \int_{(\,\cdot\,-s,\infty)} Y_{\cdot\,-v}\,\eta(\mathrm{d}v)\right]\!(z)}{h(z)}$$

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

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 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 $s \in \mathbb{R}$,

$$U_t = U_s + \int_s^t \int_{[0,\infty)} U_{u-v}\,\eta(\mathrm{d}v)\,\mathrm{d}u \tag{4.6}$$

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,

$$U_r = U_s x_0(r-s) + \int_s^\infty \int_{(u-s,\infty)} U_{u-v}\,\eta(\mathrm{d}v)\,x_0(r-u)\,\mathrm{d}u \tag{4.7}$$

for Lebesgue almost all $r > t-1$ and all $s \in \mathbb{Q}$ with $s \leq t-1$. For any such r we observe that the right-hand side of (4.7) 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 $\mathbb{E}[|Z_u|] \leq a + b|u|$ for any u with suitably chosen $a, b > 0$ (see Corollary 4.4), and this implies that

$$\mathbb{E}\!\left[\int_\mathbb{R} |Z_u| \int_{[0,\infty)} |x_0(t-u-v)|\,|\eta|(\mathrm{d}v)\,\mathrm{d}u\right]$$

$$\leq a|\eta|([0,\infty)) \int_\mathbb{R} |x_0(u)|\,\mathrm{d}u + b \int_\mathbb{R} |u| \int_{[0,\infty)} |x_0(t-u-v)|\,|\eta|(\mathrm{d}v)\,\mathrm{d}u$$

$$\leq \left(a|\eta|([0,\infty)) + b \int_{[0,\infty)} v\,|\eta|(\mathrm{d}v)\right) \int_\mathbb{R} |x_0(u)|\,\mathrm{d}u$$

$$+ b|\eta|([0,\infty)) \int_\mathbb{R} (|t| + |u|)|x_0(u)|\,\mathrm{d}u.$$

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

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

$$
\int_s^t \int_{[0,\infty)} X_{u-v}\,\eta(\mathrm{d}v)\,\mathrm{d}u - \int_s^t \int_{[0,\infty)} Z_{u-v}\,\eta(\mathrm{d}v)\,\mathrm{d}u
$$

$$
= \int_s^t \int_{[0,\infty)} \int_{\mathbb{R}} Z_r \int_{[0,\infty)} x_0(u-v-r-w)\,\eta(\mathrm{d}w)\,\mathrm{d}r\,\eta(\mathrm{d}v)\,\mathrm{d}u
$$

$$
= \int_{\mathbb{R}} Z_r \int_{[0,\infty)} \int_{s-r-w}^{t-r-w} \int_{[0,\infty)} x_0(u-v)\,\eta(\mathrm{d}v)\,\mathrm{d}u\,\eta(\mathrm{d}w)\,\mathrm{d}r
$$

$$
= \int_{\mathbb{R}} Z_r \int_{[0,\infty)} [x_0(t-r-w) - x_0(s-r-w)]\,\eta(\mathrm{d}w)\,\mathrm{d}r
$$

$$
- \int_{\mathbb{R}} \int_{[0,\infty)} Z_r[\mathbb{1}_{[0,\infty)}(t-r-w) - \mathbb{1}_{[0,\infty)}(s-r-w)]\,\eta(\mathrm{d}w)\,\mathrm{d}r
$$

$$
= \int_{\mathbb{R}} Z_r \int_{[0,\infty)} [x_0(t-r-w) - x_0(s-r-w)]\,\eta(\mathrm{d}w)\,\mathrm{d}r
$$

$$
- \int_s^t \int_{[0,\infty)} Z_{r-w}\,\eta(\mathrm{d}w)\,\mathrm{d}r.
$$

Next, we write

$$
X_t = \int_{\mathbb{R}} (Z_t - Z_{t-u}) \int_{[0,\infty)} x_0(u-v)\,\eta(\mathrm{d}v)\,\mathrm{d}u, \qquad t \in \mathbb{R}, \tag{4.8}
$$

using that

$$
\int_{\mathbb{R}} \int_{[0,\infty)} x_0(u-v)\,\eta(\mathrm{d}v)\,\mathrm{d}u = \int_{\mathbb{R}} x_0(u)\,\mathrm{d}u\,\eta([0,\infty)) = h(0)\eta([0,\infty)) = -1. \tag{4.9}
$$

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

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

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 $\mathrm{Re}(z) \geq 0$, the causal representation (2.7) of $(X_t)_{t \in \mathbb{R}}$ follows from (4.8) and the fact that $x_0(t) = 0$ for $t < 0$ by Lemma 2.2(iii). This completes the proof. □

Proof of Corollary 2.6. It follows from (4.9) and Corollary 2.3 that

$$Z_t + \int_{\mathbb{R}} Z_{t-u} \int_{[0,\infty)} x_0(u-v)\,\eta(\mathrm{d}v)\,\mathrm{d}u$$

$$= \int_{\mathbb{R}} [Z_{t-u} - Z_t] \int_{[0,\infty)} x_0(u-v)\,\eta(\mathrm{d}v)\,\mathrm{d}u$$

$$= \int_{\mathbb{R}} \int_{\mathbb{R}} [\theta(t-u-r) - \theta(t-r)][\mu_{x_0}(\mathrm{d}u) - \delta_0(\mathrm{d}u)]\,\mathrm{d}L_r$$

$$= \int_{\mathbb{R}} \int_{\mathbb{R}} \theta(t-u-r)\,\mu_{x_0}(\mathrm{d}u)\,\mathrm{d}L_r$$

$$= \int_{\mathbb{R}} \theta * x_0(t-r)\,\mathrm{d}L_r,$$

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). $\qquad\square$

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

$$\int_{\mathbb{R}} (\psi_n(t) - \psi_m(t))^2\,\mathrm{d}t = \frac{1}{2\pi} \int_{\mathbb{R}} \left| \mathcal{F}\left[\sum_{k=m+1}^{n} \theta * \phi^{*k} \right](y) \right|^2 \mathrm{d}y \tag{4.10}$$

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

$$\left| \mathcal{F}\left[\sum_{k=m+1}^{n} \theta * \phi^{*k} \right](y) \right| \le |\mathcal{F}[\theta](y)| \sum_{k=m+1}^{n} |\phi|((0,\infty))^k \le \frac{|\mathcal{F}[\theta](y)|}{1 - |\phi|((0,\infty))}. \tag{4.11}$$

The first inequality in (4.11) 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 (4.11) and dominated convergence together with the relation (4.10) 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,

$$X_t = X * \phi^{*n}(t) + \sum_{k=0}^{n-1} \left(\int_{\mathbb{R}} \theta(\cdot - u)\,\mathrm{d}L_u \right) * \phi^{*k}(t)$$

$$= X * \phi^{*n}(t) + \int_{\mathbb{R}} \psi_{n-1}(t-u)\,\mathrm{d}L_u \tag{4.12}$$

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

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

Since $\mathbb{E}[X_0^2] < \infty$ and $|\phi^{*n}|((0,\infty)) = |\phi|((0,\infty))^n \to 0$ as $n \to \infty$, we establish that $X * \phi^{*n}(t) \to 0$ in $L^2(\mathbb{P})$ as $n \to \infty$. Consequently, (4.12) shows that $\int_{\mathbb{R}} \psi_n(t-u)\,\mathrm{d}L_u \to X_t$ 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

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

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

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

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

Acknowledgments

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.

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 of the noise.

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., Barndorff-Nielsen and Shephard [1] use it as the stochastic volatility component in option pricing modeling and Schwartz [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, Brockwell et al. [8] model turbulent wind speed data as a CAR(2) process, García et al. [11] and

Benth et al. [3] fit a CARMA$(2,1)$ process to electricity spot prices, and Benth et al. [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 Basse-O'Connor et al. [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 on and supports the related work of Brockwell et al. [7].

Section 2 recalls a few central definitions and gives a dynamic interpretation of CARMA processes by relating them to solutions of stochastic delay 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 càdlàg 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

$$\log \mathbb{E}[e^{iyL_t}] = t \log \mathbb{E}[e^{iyL_1}]$$

for $y \in \mathbb{R}$ and $t \geq 0$. The definition is extended to a two-sided Lévy process $(L_t)_{t\in\mathbb{R}}$, $L_0 = 0$, 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 $L_t = -L_{(-t)_-}^2$ 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

$$P(z) = z^p + a_1 z^{p-1} + \cdots + a_p \quad \text{and} \quad Q(z) = b_0 + b_1 z + \cdots + b_{p-1} z^{p-1} \qquad (2.1)$$

for $z \in \mathbb{C}$ and $a_1, \ldots, a_p, b_0, \ldots, b_{p-1} \in \mathbb{R}$. Define

$$
\tilde{A}_p = \begin{bmatrix}
0 & 1 & 0 & \cdots & 0 \\
0 & 0 & 1 & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \cdots & 1 \\
-a_p & -a_{p-1} & -a_{p-2} & \cdots & -a_1
\end{bmatrix},
$$

$e_p = [0, 0, \ldots, 0, 1]^\top \in \mathbb{R}^p$, and $b = [b_0, b_1, \ldots, b_{p-2}, b_{p-1}]^\top$. 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

$$
\mathrm{d}X_t = \tilde{A}_p X_t \, \mathrm{d}t + e_p \, \mathrm{d}L_t, \tag{2.2}
$$

and it is explicitly given by $X_t = \int_{-\infty}^{t} e^{\tilde{A}_p(t-u)} e_p \, \mathrm{d}L_u$ 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

$$
Y_t = b^\top X_t, \qquad t \in \mathbb{R}. \tag{2.3}
$$

This is the state-space representation of the formal stochastic differential equation

$$
P(D)Y_t = Q(D)DL_t, \qquad t \in \mathbb{R}, \tag{2.4}
$$

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.
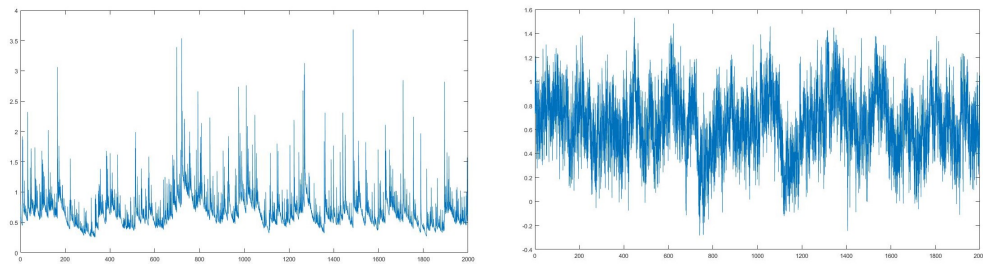
Figure 1: A simulation of a CARMA$(2, 1)$ process with parameters $a_1 = 1.3619$, $a_2 = 0.0443$, and $b_0 = 0.2061$. It is driven by a gamma (Lévy) process with parameters $\lambda = 0.2488$ and $\xi = 0.5792$ on the left and a Brownian motion with mean $\mu = 0.1441$ and standard deviation $\sigma = 0.2889$ 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

$$
\mathrm{d}Y_t = \int_{[0, \infty)} Y_{t-u} \, \eta(\mathrm{d}u) \, \mathrm{d}t + \mathrm{d}L_t, \qquad t \in \mathbb{R}. \tag{2.5}
$$

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 $\eta_\beta(dt) = \mathbb{1}_{[0,\infty)}(t)e^{\beta t}\, dt$ for $\beta \in \mathbb{C}$ with $\text{Re}(\beta) < 0$, 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, Corollary 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,

$$\eta(dt) = \gamma_0\,\delta_0(dt) + \Big(\mathbb{1}_{[0,\infty)}(t)\sum_{i=1}^{p-1}\gamma_i e^{\beta_i t}\Big)\, dt \tag{2.8}$$

where

$$\gamma_0 = -\Big(a_1 + \sum_{j=1}^{p-1}\beta_j\Big) \quad and \quad \gamma_i = -\frac{P(\beta_i)}{Q'(\beta_i)} \qquad for\ i = 1,\ldots,p-1.$$

Proof. It follows immediately from [2, Corollary 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

$$\eta_j(dt) = \mathbb{1}_{[0,\infty)}(t)\sum_{i=1}^{j}e^{\beta_i t}\prod_{k=1,k\neq i}^{j}(\beta_i - \beta_k)^{-1}\, dt,$$

from which it follows that

$$\eta(dt) - \alpha_0\delta_0(dt) = \sum_{j=1}^{p-1}\alpha_j\Big(\mathbb{1}_{[0,\infty)}(t)\sum_{i=1}^{j}e^{\beta_i t}\prod_{k=1,k\neq i}^{j}(\beta_i - \beta_k)^{-1}\, dt\Big)$$

$$= \mathbb{1}_{[0,\infty)}(t)\sum_{i=1}^{p-1}e^{\beta_i t}\sum_{j=i}^{p-1}\alpha_j\prod_{k=1,k\neq i}^{j}(\beta_i - \beta_k)^{-1}\, dt.$$

Finally, observe that the definition of $\alpha_0, \alpha_1,\ldots,\alpha_{p-1}$ implies that

$$\gamma_i = \frac{\sum_{j=i}^{p-1}\alpha_j\prod_{k=j+1}^{p-1}(\beta_i - \beta_k)}{\prod_{k=1,k\neq i}^{p-1}(\beta_i - \beta_k)} = \sum_{j=i}^{p-1}\alpha_j\prod_{k=1,k\neq i}^{j}(\beta_i - \beta_k)^{-1},\ i = 1,\ldots,p-1,$$

which concludes the proof. □

Remark 2.2. In Brockwell et al. [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 $\mathcal{F}[f](y) = \int_{\mathbb{R}} e^{-ixy} f(x) \, \mathrm{d}x$ for $f \in L^1$. If we then heuristically take this Fourier transform on both sides of (2.4) we get

$$P(iy)\mathcal{F}[Y](y) = Q(iy)\mathcal{F}[DL](y), \qquad y \in \mathbb{R}.$$

For $\gamma_0 \in \mathbb{R}$, this can be rewritten as

$$\mathcal{F}[DL](y) = \left(\frac{P(iy) - (iy - \gamma_0)Q(iy)}{Q(iy)} - \gamma_0 \right) \mathcal{F}[Y](y) + \mathcal{F}[DY](y), \qquad y \in \mathbb{R}.$$

If we let $\gamma_0 = -\left(a_1 + \sum_{j=1}^{p-1} \beta_j \right)$ then

$$y \longmapsto \frac{P(iy) - (iy - \gamma_0)Q(iy)}{Q(iy)} \in L^2,$$

and consequently, there exists $f \in L^2$ such that

$$\left(\frac{P(iy) - (iy - \gamma_0)Q(iy)}{Q(iy)} - \gamma_0 \right) \mathcal{F}[Y](y) = \mathcal{F}[-f * Y - \gamma_0 Y](y), \qquad y \in \mathbb{R}.$$

We conclude that $(Y_t)_{t \in \mathbb{R}}$ satisfies the formal equation

$$DY_t = f * Y_t + \gamma_0 Y_t + DL_t, \qquad y \in \mathbb{R}.$$

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

$$f(t) = -\mathbb{1}_{[0,\infty)}(t) \sum_{i=1}^{p-1} \frac{P(\beta_i)}{Q'(\beta_i)} e^{\beta_i t}$$

in line with Theorem 2.1.

The simplest example beyond the 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

$$\mathrm{d}Y_t = \alpha_0 Y_t \, \mathrm{d}t + \alpha_1 \int_0^\infty e^{\beta_1 u} Y_{t-u} \, \mathrm{d}u \, \mathrm{d}t + \mathrm{d}L_t, \qquad t \in \mathbb{R},$$

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*

$$\mathbb{E}[Y_t] = -\frac{\mu}{\eta([0, \infty))} \quad and \quad \gamma(h) := \mathrm{Cov}(Y_h, Y_0) = \sigma^2 e_p^\top e^{A_p |h|} \Sigma e_p \qquad for\ h \in \mathbb{R},$$

where

$$\mu = \mathbb{E}[L_1], \quad \sigma^2 = \mathrm{Var}(L_1), \quad and \quad \Sigma = \int_0^\infty e^{A_p y} e_p e_p^\top e^{A_p^\top y}\,\mathrm{d}y.$$

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

$$\theta_0 = [\alpha_0, \alpha_1, \beta_1, \alpha_2, \beta_2, \ldots, \alpha_{p-1}, \beta_{p-1}]^\top$$

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 $\pi_{k-1}(Y_{k\Delta}; \theta)$ be the $L^2(\mathbb{P}_\theta)$ projection of $Y_{k\Delta}$ onto the linear span of $Y_0, Y_\Delta, Y_{2\Delta}, \ldots, Y_{(k-1)\Delta}$ and set $\epsilon_k(\theta) = Y_{k\Delta} - \pi_{k-1}(Y_{k\Delta}; \theta)$. Then the least squares estimator $\hat{\theta}_n$ of θ_0 is the point that minimizes

$$\theta \longmapsto \sum_{k=1}^n \epsilon_k(\theta)^2.$$

In practice, the projections $\pi_{k-1}(Y_{k\Delta}; \theta)$, $k = 1, \ldots, n$, 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 $\alpha_0, \alpha_1, \beta_1, \ldots, \alpha_{p-1}, \beta_{p-1}$ (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 $Y_{k\Delta} = e_p^\top Z_k$, $k \in \mathbb{N}_0$, with $(Z_k)_{k\in\mathbb{N}_0}$ satisfying the state-equation*

$$Z_k = e^{A_p\Delta}Z_{k-1} + U_k, \qquad k \in \mathbb{N},$$

where $(U_k)_{k\in\mathbb{N}}$ is a sequence of i.i.d. random vectors with mean 0 and covariance matrix $\int_0^\Delta e^{A_p u} e_p e_p^\top e^{A_p^\top u}\, du$.

Proof. From [2, Proposition 3.13] we have that $Y_t = e_p^\top \tilde{Z}_t$, $t \in \mathbb{R}$, where $(\tilde{Z}_t)_{t\in\mathbb{R}}$ is the $\mathbb{R}^p$-valued Ornstein–Uhlenbeck process given by

$$\tilde{Z}_t = \int_{-\infty}^t e^{A_p(t-u)} e_p\, dL_u, \qquad t \in \mathbb{R}.$$

Thus, by defining $Z_k = \tilde{Z}_{k\Delta}$ so that $Y_{k\Delta} = e_p^\top Z_k$ for $k \in \mathbb{N}_0$, 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 follows immediately. $\qquad\square$

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 $n + 1 = 2000$, 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

$$f(x) = \frac{1}{\Gamma(\lambda)\xi^\lambda} x^{\lambda-1} e^{-\frac{x}{\xi}}, \qquad x > 0,$$

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 a 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.

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 $\beta_1 = -0.2061$.

Noise	Spacing	Parameter	Mean	Bias	Std
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

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 $\int_0^\infty e^{\beta_1 s}(Y_{t-s} - \mathbb{E}_{\theta_0}[Y_0]) \, ds$ 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 1950 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 1950 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

$$\mathrm{d}L_t = \frac{P(D)}{Q(D)} Y_t \, \mathrm{d}t, \qquad t \in \mathbb{R}. \tag{5.1}$$

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 $\mathrm{d}Y_t - \int_{[0,\infty)} Y_{t-s}\,\eta(\mathrm{d}s)\,\mathrm{d}t$. It seems that this statement continues to hold when $p > q + 1$ as well when $\mathrm{d}Y_t$ is replaced by a suitable linear combination of $\mathrm{d}Y_t, \mathrm{d}(DY)_t, \ldots, \mathrm{d}(D^{p-q-1}Y)_t$.
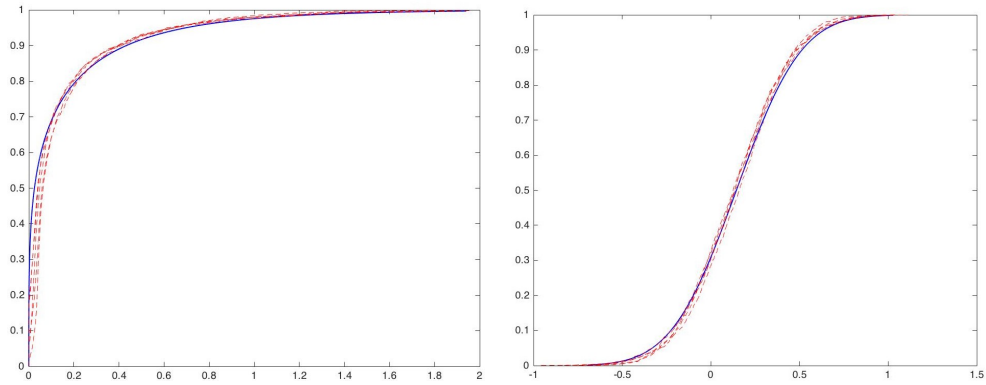
Figure 2: Five estimations of the distribution function of L_1, based on estimates of α_0, α_1, and β_1, plotted against the true distribution function for a sampling frequency of $\Delta = 1$. The left corresponds to gamma noise and the right to Gaussian noise.
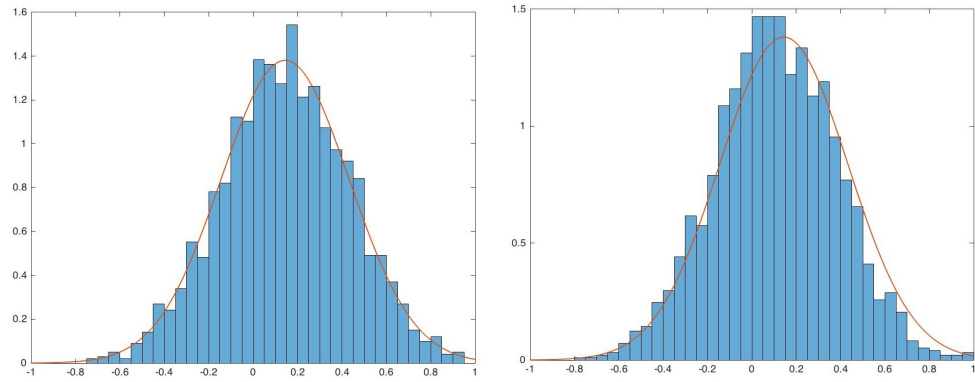
Figure 3: Histograms of the true increments on the left and estimated increments, based on estimates of α_0, α_1, and β_1 for a sampling frequency of $\Delta = 1$, on the right plotted against the theoretical (Gaussian) probability density function.

Acknowledgments

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 (2017). A continuous-time framework for ARMA processes. arXiv: 1704.08574v1.

[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.

Multivariate Stochastic Delay Differential Equations and CAR Representations of 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 may be thought of as a CAR(∞) representation. Furthermore, we show how the CAR(∞) 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 which will have particular focus on existence, uniqueness and representations.

MSC: 60G05; 60G22; 60G51; 60H05; 60H10

Keywords: CARMA processes; FICARMA processes; Long memory; MCARMA processes; Multivariate Ornstein–Uhlenbeck processes; Multivariate stochastic delay differential equations; Noise recovery; Prediction

1   Introduction and main ideas

The class of autoregressive moving average (ARMA) processes 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 Box and Jenkins [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

$$P(D)X_t = Q(D)DZ_t, \qquad t \in \mathbb{R}, \tag{1.1}$$

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 $\mathrm{Re}(z) \geq 0$. Under this assumption, $(X_t)_{t \in \mathbb{R}}$ can be rigorously defined 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, 27]. Moreover, there exists a vast amount of literature on theoretical results for CARMA processes (and variations of these), and a few references are [6, 7, 8, 10, 18, 19, 26].

It is well-known that any causal CARMA process has a continuous time moving average representation of CMA(∞) type

$$X_t = \int_{-\infty}^{t} g(t-u) \, \mathrm{d}Z_u, \qquad t \in \mathbb{R},$$

see the references above or 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 a CAR(∞) representation of CARMA processes and, by the arguments below, it will take the form

$$R(D)X_t = \int_{0}^{\infty} X_{t-u} f(u) \, \mathrm{d}u + DZ_t, \qquad t \in \mathbb{R}, \tag{1.2}$$

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 for 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 (1.2) holds with $R = P$ and $f = 0$. If $q \geq 1$, it is convenient to rephrase (1.1) in the frequency domain (that is, apply the Fourier transform $\mathcal{F}$ on both sides of the equation and rearrange terms):

$$\frac{P(iy)}{Q(iy)} \mathcal{F}[X](y) = \mathcal{F}[DZ](y), \qquad y \in \mathbb{R}. \tag{1.3}$$

Using 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

$$\begin{aligned}
\frac{P(iy)}{Q(iy)} \mathcal{F}[X](y) &= \left(R(iy) - \frac{S(iy)}{Q(iy)} \right) \mathcal{F}[X](y) \\
&= \mathcal{F}[R(D)X](y) - \mathcal{F}[f](y)\mathcal{F}[X](y),
\end{aligned}$$

where $f : \mathbb{R} \to \mathbb{R}$ is the L^2 function characterized by the relation $\mathcal{F}[f](y) = S(iy)/Q(iy)$ 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 formulas for MCARMA processes have, to the best of our knowledge, not been studied in the literature.

Autoregressive 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 autoregressive 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 $Z_t = [Z_t^1, \ldots, Z_t^n]^\top$, $t \in \mathbb{R}$, with stationary increments and a finite signed $n \times n$ matrix-valued measure η, concentrated on $[0, \infty)$, a stationary process $X_t = [X_t^1, \ldots, X_t^n]^\top$, $t \in \mathbb{R}$, is a solution to the associated MSDDE if it satisfies

$$\mathrm{d}X_t = \eta * X(t)\,\mathrm{d}t + \mathrm{d}Z_t, \qquad t \in \mathbb{R}. \tag{1.4}$$

By equation (1.4) we mean that

$$X_t^j - X_s^j = \sum_{k=1}^n \int_s^t \int_{[0,\infty)} X_{u-v}^k \, \eta_{jk}(\mathrm{d}v)\,\mathrm{d}u + Z_t^j - Z_s^j, \qquad j = 1, \ldots, n, \tag{1.5}$$

almost surely for each $s < t$. This system of equations is an extension of the stochastic delay differential equation (SDDE) in [3, Section 2] 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

$$\int_{\mathbb{R}} |f_{il}(t)|^p \mu_{lj}(\mathrm{d}t) < \infty$$

for $l = 1, \ldots, k$, $i = 1, \ldots, m$ and $j = 1, \ldots, n$. Provided that $f \in L^1(\mu)$, we set

$$\int_{\mathbb{R}} f(t)\,\mu(\mathrm{d}t) = \sum_{l=1}^k \begin{bmatrix} \int_{\mathbb{R}} f_{1l}(t)\,\mu_{l1}(\mathrm{d}t) & \cdots & \int_{\mathbb{R}} f_{1l}(t)\,\mu_{ln}(\mathrm{d}t) \\ \vdots & \ddots & \vdots \\ \int_{\mathbb{R}} f_{ml}(t)\,\mu_{l1}(\mathrm{d}t) & \cdots & \int_{\mathbb{R}} f_{ml}(t)\,\mu_{ln}(\mathrm{d}t) \end{bmatrix}. \tag{2.1}$$

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 is 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

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

provided that $f(t - \cdot) \in L^1(\mu)$. In case that μ is the Lebesgue–Stieltjes measure of a function $g \colon \mathbb{R} \to \mathbb{R}^{k \times n}$ 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

$$D(\mu) = \left\{ z \in \mathbb{C} : \int_{\mathbb{R}} e^{-\mathrm{Re}(z)t} |\mu_{ij}|(\mathrm{d}t) < \infty \quad \text{for } i = 1, \ldots, k \text{ and } j = 1, \ldots, n \right\}$$

and define its Laplace transform $\mathcal{L}[\mu]$ as

$$\mathcal{L}[\mu]_{ij}(z) = \int_{\mathbb{R}} e^{-zt} \mu_{ij}(\mathrm{d}t) \qquad \text{for } z \in D(\mu), i = 1, \ldots, k \text{ and } j = 1, \ldots, n.$$

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 $\mu(\mathrm{d}t) = f(t)\,\mathrm{d}t$ 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

$$\int_{\mathbb{R}} |t|^p |\mu_{jk}|(\mathrm{d}t) < \infty \qquad \text{for all } j, k = 1, \ldots, n.$$

Finally, for two functions $f, g \colon \mathbb{R} \to \mathbb{C}$ 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

$$\lim_{t \to a} \frac{f(t)}{g(t)} = 0, \quad \lim_{t \to a} \frac{f(t)}{g(t)} = 1 \quad \text{and} \quad \limsup_{t \to a} \left| \frac{f(t)}{g(t)} \right| < \infty,$$

respectively.

3 Stochastic delay differential equations

Consider the general MSDDE in (1.4), where the noise $(Z_t)_{t\in\mathbb{R}}$ is a measurable process, which is integrable and has stationary increments, and the delay measure η is a finite (signed) $n \times n$ matrix-valued measure concentrated on $[0,\infty)$. 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 $h\colon D(\eta) \to \mathbb{C}^{n\times n}$ given by

$$h(z) = I_n z - \mathcal{L}[\eta](z), \qquad z \in D(\mu), \tag{3.1}$$

where I_n is the $n \times n$ identity matrix.

Theorem 3.1. *Let h be given as 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 $g\colon \mathbb{R} \to \mathbb{R}^{n\times n}$ in L^2 characterized by*

$$\mathcal{F}[g](y) = h(iy)^{-1}, \qquad y \in \mathbb{R}, \tag{3.2}$$

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 $\mathrm{Re}(z) \geq 0$ then the solution in (3.3) is casual in the sense that $(X_t)_{t\in\mathbb{R}}$ is adapted to the filtration $\{\sigma(Z_t - Z_s : s < t)\}_{t\in\mathbb{R}}$.

As discussed in Section 4.1, 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,

$$X_t = \int_{\mathbb{R}} g(t-u) \, dZ_u, \qquad t \in \mathbb{R}. \tag{3.4}$$

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. [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 $\mathrm{Re}(z) \geq 0$ 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*

$$\hat{Z}_u = \mathbb{E}[Z_u - Z_s \mid Z_s - Z_r, r < s], \qquad u > s, \tag{3.5}$$

it holds that

$$\mathbb{E}[X_t \mid X_u, u \leq s]$$
$$= g(t-s)X_s + \int_s^t g(t-u)\eta * \{\mathbb{1}_{(-\infty,s]}X\}(u) \, du + g * \{\mathbb{1}_{(s,\infty)}\hat{Z}\}(t),$$

using the notation

$$\big[\eta * \{\mathbb{1}_{(-\infty,s]}X\}(u)\big]_j := \sum_{k=1}^{n} \int_{[u-s,\infty)} X_{u-v}^k \, \eta_{jk}(\mathrm{d}v)$$

$$\text{and} \qquad \big[g * \{\mathbb{1}_{(s,\infty)}\hat{Z}\}(u)\big]_j := \sum_{k=1}^{n} \int_{[0,u-s)} \hat{Z}_{u-v}^k \, g_{jk}(\mathrm{d}v)$$

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

$$\mathbb{E}[X_t \mid X_u, u \le s]$$
$$= g(t-s)X_s + \int_s^t g(t-u)\eta * \{\mathbb{1}_{(-\infty,s]}X\}(u) \, \mathrm{d}u + \int_s^t g(t-u) \, \mathrm{d}u \, \mathbb{E}[Z_1],$$

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 $(\hat{Z}_u)_{s<u\le t}$ and then use the formula given in Theorem 3.2 to compute the prediction $\mathbb{E}[X_t \mid X_u, u \le s]$.

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

$$X_t^j = \bigg[\int_{\mathbb{R}} g(t-u) \, \mathrm{d}Z_u\bigg]_j = \sum_{k=1}^{n} \int_{\mathbb{R}} g_{jk}(t-u) \, \mathrm{d}Z_u^k, \qquad t \in \mathbb{R}, \tag{4.1}$$

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 $y \in \mathbb{R}$, $\det(h(iy)) \ne 0$. 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 $\mathcal{I}_k : L^1 \cap L^2 \to L^1(\mathbb{P})$ which has the following properties:*

(i) *For all $s < t$, $\mathcal{I}_k(\mathbb{1}_{(s,t]}) = Z_t^k - Z_s^k$.*

(ii) *If μ is a finite Borel measure on $\mathbb{R}$ having first moment then*

$$\mathcal{I}_k\left(\int_{\mathbb{R}} f_r(t - \cdot)\,\mu(\mathrm{d}r)\right) = \int_{\mathbb{R}} \mathcal{I}_k(f_r(t - \cdot))\,\mu(\mathrm{d}r) \tag{4.2}$$

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

$$X_t^j = \sum_{k=1}^n \mathcal{I}_k(g_{jk}(t - \cdot)), \qquad j = 1,\ldots,n, \tag{4.3}$$

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 $\int \cdot \; \mathrm{d}Z^k = \mathcal{I}_k$.

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 if $f \in L^1 \cap L^2$ the integral $\int_{\mathbb{R}} f(u)\,\mathrm{d}Z_u^j$ is well-defined in the sense of [22] and belongs to $L^1(\mathbb{P})$. (The latter fact is easily derived from [22, Theorem 3.3].) 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 [23, 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*

$$I_-^\beta f : t \longmapsto \frac{1}{\Gamma(\beta)} \int_t^\infty f(u)(u - t)^{\beta-1}\,\mathrm{d}u \tag{4.4}$$

is well-defined and belongs to L^α for any $\beta \in (0, 1 - 1/\alpha)$.

Example 4.4. Fix $\alpha_1,\ldots,\alpha_n \in (1,2]$ and consider an n-dimensional Lévy process $(L_t)_{t\in\mathbb{R}}$, where its kth coordinate $(L_t^k)_{t\in\mathbb{R}}$ is symmetric α_k-stable if $\alpha_k \in (1,2)$ and mean zero and square integrable if $\alpha_k = 2$. Then, for a given vector $\beta = [\beta_1,\ldots,\beta_n]$ with $\beta_k \in (0, 1 - 1/\alpha_k)$ for $k = 1,\ldots,n$ the corresponding fractional Lévy process $(Z_t)_{t\in\mathbb{R}}$ with parameter β is defined entrywise as

$$\begin{aligned} Z_t^k &= \int_{\mathbb{R}} \left(I_-^{\beta_k}[\mathbb{1}_{(-\infty,t]} - \mathbb{1}_{(-\infty,0]}]\right)(u)\,\mathrm{d}L_u^k \\ &= \frac{1}{\Gamma(1 + \beta_k)} \int_{\mathbb{R}} \left[(t - u)_+^{\beta_k} - (-u)_+^{\beta_k}\right]\mathrm{d}L_u^k \end{aligned} \tag{4.5}$$

for $t \in \mathbb{R}$ and $k = 1,\ldots,n$, where $x_+ = \max\{x,0\}$. Proposition 4.3 shows that $I_-^{\beta_k} f \in L^{\alpha_k}$ for any $f \in L^1 \cap L^2$, and hence we can define integration of f with respect to $(Z_t^k)_{t \in \mathbb{R}}$ through $(L_t^k)_{t \in \mathbb{R}}$ as

$$\int_{\mathbb{R}} f(t) \, dZ_t^k = \int_{\mathbb{R}} \big(I_-^{\beta_k} f\big)(t) \, dL_t^k.$$

Note that the integral belongs to $L^2(\mathbb{P})$ if $\alpha_k = 2$ and to $L^\gamma(\mathbb{P})$ for any $\gamma < \alpha_k$ if $\alpha_k \in (1,2)$. While the integral clearly satisfies assumption (i) of Proposition 4.1 in light of (4.5), one can rely on the stochastic Fubini result for $(L_t^k)_{t \in \mathbb{R}}$ given in [2, Theorem 3.1] to verify that assumption (ii) is satisfied as well. Consequently, $(Z_t)_{t \in \mathbb{R}}$ 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 $\varpi_0, \varpi_1, \ldots, \varpi_{m-1}$ 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 mth order MSDDE if it is $m-1$ times differentiable and

$$dX_t^{(m-1)} = \sum_{j=0}^{m-1} \varpi_j * X^{(j)}(t) \, dt + dZ_t, \qquad t \in \mathbb{R}, \tag{4.6}$$

where $(X_t^{(j)})_{t \in \mathbb{R}}$ denotes the entrywise jth derivative of $(X_t)_{t \in \mathbb{R}}$ with respect to t. By (4.6) we mean that

$$\big(X_t^{(m-1)}\big)^k - \big(X_s^{(m-1)}\big)^k = \sum_{j=0}^{m-1} \sum_{l=1}^{n} \int_s^t \int_{[0,\infty)} \big(X_{u-v}^{(j)}\big)^l (\varpi_j)_{kl}(dv) \, du + Z_t^k - Z_s^k$$

for $k = 1,\ldots,n$ and each $s < t$ almost surely. Equation (4.6) corresponds to the mn-dimensional MSDDE in (1.4) with noise $[0,\ldots,0,Z_t^\top]^\top \in \mathbb{R}^{mn}$ and

$$\eta = \begin{bmatrix} 0 & I_n \delta_0 & 0 & \cdots & 0 \\ 0 & 0 & I_n \delta_0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & I_n \delta_0 \\ \varpi_0 & \varpi_1 & \varpi_2 & \cdots & \varpi_{m-1} \end{bmatrix}. \tag{4.7}$$

(If $n = 1$ then $\eta = \varpi_0$.) With η given by (4.7) it follows that

$$D(\eta) = \bigcap_{j=0}^{m-1} D(\varpi_j)$$

and

$$h(z) = \begin{bmatrix} zI_n & -I_n & 0 & \cdots & 0 \\ 0 & zI_n & -I_n & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & zI_n & -I_n \\ -\mathcal{L}[\varpi_0](z) & -\mathcal{L}[\varpi_1](z) & \cdots & -\mathcal{L}[\varpi_{m-2}](z) & zI_n - \mathcal{L}[\varpi_{m-1}](z) \end{bmatrix}$$

for $z \in D(\eta)$. In general, we know from Theorem 3.1 that a solution to (4.6) exists if $\det h(iy) \neq 0$ for all $y \in \mathbb{R}$, and in this case the unique solution is given by

$$X_t = \int_{\mathbb{R}} g_{1m}(t - u) \, \mathrm{d}Z_u, \qquad t \in \mathbb{R}, \tag{4.8}$$

where $\mathcal{F}[g_{1m}]$ corresponds to entry $(1, m)$ in the $n \times n$ block representation of $h(i \cdot)^{-1}$. In other words, if e_j denotes the jth canonical basisvector of $\mathbb{R}^m$ and $\otimes$ the Kronecker product then

$$\mathcal{F}[g_{1m}](y) = (e_1 \otimes I_n)^\top h(iy)^{-1} (e_m \otimes I_n), \qquad y \in \mathbb{R}.$$

However, due to the particular structure of η in (4.7) we can simplify these expressions:

Theorem 4.5. *Let the setup be as above. Then it holds that*

$$\det h(z) = \det\Big(I_n z^m - \sum_{j=0}^{m-1} \mathcal{L}[\varpi_j](z)z^j\Big), \qquad z \in D(\eta), \tag{4.9}$$

and if $\det h(iy) \neq 0$ for all $y \in \mathbb{R}$, there exists a unique solution to (4.6) and it is given as (4.8) where $g_{1m} \colon \mathbb{R} \to \mathbb{R}^{n \times n}$ is characterized by

$$\mathcal{F}[g_{1m}](y) = \Big(I_n(iy)^m - \sum_{j=0}^{m-1} \mathcal{F}[\varpi_j](y)(iy)^j\Big)^{-1}, \qquad y \in \mathbb{R}. \tag{4.10}$$

The solution is causal if $\det h(z) \neq 0$ whenever $\mathrm{Re}(z) \geq 0$.

Observe that, as should be the case, we are back to the first order MSDDE when $m = 1$ and (4.9)–(4.10) agree with Theorem 3.1. As we will see in Section 4.3 below, one motivation for introducing higher order MSDDEs of the form (4.6) 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

$$\mathrm{d}X_t^{(p-1)} = -\sum_{j=0}^{p-1} A_{p-j} X_t^{(j)} \, \mathrm{d}t + \mathrm{d}Z_t, \qquad t \in \mathbb{R}, \tag{4.11}$$

which is of the form (4.6) with $m = p$ and $\omega_j = -A_{p-j}\delta_0$ 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 $p = 1$, (4.11) collapses to the multivariate Ornstein–Uhlenbeck equation

$$\mathrm{d}X_t = -A_1 X_t \,\mathrm{d}t + \mathrm{d}Z_t, \qquad t \in \mathbb{R},$$

and if the real part of all the eigenvalues of A_1 are 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

$$X_t = \int_{-\infty}^{t} e^{-A_1(t-u)} \,\mathrm{d}Z_u, \qquad t \in \mathbb{R}. \tag{4.12}$$

Lévy-driven multivariate Ornstein–Uhlenbeck processes have been studied extensively in the literature, and the moving average structure (4.12) of the solution is well-known when $(Z_t)_{t \in \mathbb{R}}$ is a Lévy process. We refer to [1, 24, 25] 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

$$P(z) = I_n z^p + A_1 z^{p-1} + \cdots + A_p$$
$$\text{and} \qquad Q(z) = B_0 + B_1 z + \cdots + B_{p-1} z^{p-1} \tag{4.13}$$

for $z \in \mathbb{C}$ and suitable $A_1, \ldots, A_p, B_0, \ldots, B_{p-1} \in \mathbb{R}^{n \times n}$. 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 satisfies

$$\mathcal{F}[\tilde{g}](y) = P(iy)^{-1} Q(iy), \qquad y \in \mathbb{R}. \tag{4.14}$$

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

$$X_t = \int_{\mathbb{R}} \tilde{g}(t-u) \,\mathrm{d}Z_u, \qquad t \in \mathbb{R}, \tag{4.15}$$

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

$$\tilde{g}(t) = (e_1^p \otimes I_n)^\top e^{At} E, \qquad t \geq 0, \tag{4.16}$$

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

$$z \longmapsto P(z)E(z) - Q(z)z^p$$

is at most of degree $p - 1$. (Above, and henceforth, we use the notation e_j^k for the jth 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 square integrable Lévy process with mean zero, and

$$Z_t^j = \frac{1}{\Gamma(1 + \beta_j)} \int_{\mathbb{R}} \left[(t - u)_+^{\beta_j} - (-u)_+^{\beta_j} \right] dL_u^j, \qquad t \in \mathbb{R},$$

for $\beta_j \in (0, 1/2)$ and $j = 1, \ldots, n$, then $(X_t)_{t \in \mathbb{R}}$ is an MFICARMA(p, β, q) process, $\beta = [\beta_1, \ldots, \beta_n]$, 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 (4.15) 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 (4.15) is the unique solution to a certain (possibly higher order) MSDDE of the form (4.6). Before formulating the main result of this section we introduce some notation. To P and Q defined in (4.13) 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

$$z \longmapsto Q(z)R(z) - P(z) \tag{4.17}$$

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 (4.13), 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 (4.6) with*

$$m = p - q, \quad \omega_0(du) = -C_0 \delta_0(du) + f(u)\,du \quad and \quad \omega_j = -C_j \delta_0,$$

for $1 \leq j \leq m - 1$ or, written out,

$$dX_t^{(m-1)} = -\sum_{j=0}^{m-1} C_j X_t^{(j)} \,dt + \left(\int_0^\infty f(u) X_{t-u} \,du \right) dt + dZ_t, \qquad t \in \mathbb{R}, \tag{4.18}$$

where $C_0, \ldots, C_{m-1} \in \mathbb{R}^{n \times n}$ are defined as in (4.17) above, $(X_t^{(j)})_{t \in \mathbb{R}}$ is the jth derivative of $(X_t)_{t \in \mathbb{R}}$, and where $f \colon \mathbb{R} \to \mathbb{R}^{n \times n}$ is characterized by

$$\mathcal{F}[f](y) = R(iy) - Q(iy)^{-1} P(iy), \qquad y \in \mathbb{R}. \tag{4.19}$$

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$(2, 1)$ 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 [7, 8, 9, 21] and for (Lévy-driven) MCARMA processes in [11]. In both cases, different approaches that do not rely on MSDDEs are 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

$$f(t) = (e_1^q \otimes I_n)^\top e^{Bt} F, \qquad t \geq 0, \tag{4.20}$$

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

$$z \longmapsto Q(z)F(z) - [Q(z)R(z) - P(z)]z^q$$

is at most of degree $q - 1$ (see (4.14) and (4.16)).

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

$$\hat{Z}_u = \mathbb{E}[Z_u - Z_s \mid Z_s - Z_r, r < s], \qquad u > s,$$

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*

$$\tilde{g}_j(t) = (e_1^p \otimes I_n)^\top e^{At} \sum_{k=j}^{p-q} A^{k-j} E C_k, \qquad t \geq 0,$$

for $j = 1, \ldots, p - q$, *where* $C_0, \ldots, C_{p-q-1}$ *are given in (4.17) 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

$$\mathbb{E}[X_t \mid X_u, u \le s] = \sum_{j=1}^{p-q} \tilde{g}_j(t-s) X_s^{(j-1)}$$

$$+ \int_{-\infty}^{s} \int_{s}^{t} \tilde{g}(t-u) f(u-v) \, du \, X_v \, dv + \tilde{g} * \{\hat{Z} \mathbb{1}_{(s,\infty)}\}(t),$$

where $\tilde{g}$ and f are given in (4.16) and (4.20), respectively, and

$$\tilde{g} * \{\hat{Z} \mathbb{1}_{(s,\infty)}\}(t) = \mathbb{1}_{\{p=q+1\}} \hat{Z}_u + (e_1^p \otimes I_n)^\top A e^{At} \int_{s}^{t} e^{-Av} E \hat{Z}_v \, dv.$$

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

$$P(z) = I_n z^3 + A_1 z^2 + A_2 z + A_3 \quad \text{and} \quad Q(z) = B_0 + I_n z$$

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 $\operatorname{Re}(z) \ge 0$. According to (4.16), $(X_t)_{t \in \mathbb{R}}$ may be written as

$$X_t = \int_{-\infty}^{t} (e_1^3 \otimes I_n)^\top e^{A(t-u)} E \, dZ_u, \qquad t \in \mathbb{R},$$

where $E_1 = 0$, $E_2 = I_n$ and $E_3 = B_0 - A_1$. With $C_1 = A_1 - B_0$, $C_0 = A_2 + B_0(B_0 - A_1)$ and $F = B_0(A_2 - B_0(A_1 - B_0)) - A_3$, Theorem 4.8 and Remark 4.10 imply that

$$dX_t^{(1)} = -C_1 X_t^{(1)} \, dt - C_0 X_t \, dt + \left(\int_{0}^{\infty} e^{-B_0 u} F X_{t-u} \, du \right) dt + dZ_t, \qquad t \in \mathbb{R}.$$

Moreover, by Corollary 4.11, we have the prediction formula

$$\mathbb{E}[X_t \mid X_u, u \le s] = (e_1^3 \otimes I_n)^\top e^{At} \Big[(EC_1 + AE) X_s + E X_s^{(1)}$$

$$+ \int_{s}^{t} e^{-Au} E \Big(e^{B_0 u} \int_{-\infty}^{s} e^{-B_0 v} F X_v \, dv + \hat{Z}_u \Big) du \Big].$$

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 definition of $h: D(\eta) \to \mathbb{C}^{n \times n}$ in (3.1). Note that we always have $\{z \in \mathbb{C} : \operatorname{Re}(z) \ge 0\} \subseteq D(\eta)$ and $h(iy) = I_n iy - \mathcal{F}[\eta](y)$ 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*

$$\mathcal{F}[g](y) = h(iy)^{-1}, \qquad y \in \mathbb{R}. \tag{5.1}$$

Moreover, the following statements hold:

(i) *The function g satisfies*

$$g(t-r) - g(s-r) = \mathbb{1}_{(s,t]}(r)I_n + \int_s^t g * \eta(u-r)\, \mathrm{d}u$$

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*

$$g(t) = \mathbb{1}_{[0,\infty)}(t)I_n + \int_{-\infty}^t g * \eta(u)\, \mathrm{d}u \tag{5.2}$$

for almost all $t \in \mathbb{R}$. In particular,

$$\int_{\mathbb{R}} g * \eta(u)\, \mathrm{d}u = -I_n. \tag{5.3}$$

(iii) *If $\int_{[0,\infty)} e^{\delta u} |\eta_{jk}|(\mathrm{d}u) < \infty$ for all $j,k = 1,\ldots,n$ and some $\delta > 0$, then there exists $\varepsilon > 0$ such that*

$$\sup_{t \in \mathbb{R}} \max_{j,k=1,\ldots,n} |g_{jk}(t)| e^{\varepsilon |t|} \leq C$$

for a suitable constant $C > 0$.

(iv) *If $\det h(z) \neq 0$ for all $z \in \mathbb{C}$ with $\operatorname{Re}(z) \geq 0$, then g is vanishing on $(-\infty, 0)$ almost everywhere.*

Proof. In order to show the existence of g it suffices to argue that

$$\left(y \longmapsto [h(iy)^{-1}]_{jk}\right) \in L^2 \qquad \text{for } j,k = 1,\ldots,n, \tag{5.4}$$

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 $H(y)$ denote the $n \times n$ matrix which has the same rows as $h(iy)$, but where the jth column is replaced by the kth canonical basis vector (that is, the vector with all entries equal to zero except of the kth entry which equals one). Then it follows by Cramer's rule that

$$[h(iy)^{-1}]_{jk} = \frac{\det H(y)}{\det h(iy)}, \qquad y \in \mathbb{R}.$$

Recalling that $h(iy) = I_n iy - \mathcal{F}[\eta](y)$ and that $\mathcal{F}[\eta](y)$ is bounded in y we get by the Leibniz formula that $|\det H(iy)| \sim |y|^n$ and $|\det A(y)| = O(|y|^{n-1})$ as $|y| \to \infty$. This shows in particular that

$$\left|[h(iy)^{-1}]_{jk}\right| = O(|y|^{-1}), \qquad |y| \to \infty. \tag{5.5}$$

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

$$\mathcal{F}\left[g(t-\,\cdot\,) - g(s-\,\cdot\,) - \int_s^t g * \eta(u-\,\cdot\,)\, \mathrm{d}u\right](y)$$

$$= (e^{-ity} - e^{-isy})\mathcal{F}[g](-y) - \mathcal{F}[\mathbb{1}_{(s,t]}](y)\mathcal{F}[g](-y)\mathcal{F}[\eta](-y)$$

$$= \mathcal{F}[\mathbb{1}_{(s,t]}](y)h(-iy)^{-1}(-I_n iy - \mathcal{F}[\eta](-y))$$

$$= \mathcal{F}[\mathbb{1}_{(s,t]}](y)I_n,$$

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 $\tilde{h}\colon y \mapsto h(iy)$ is (entrywise) p times differentiable with the mth derivative given by

$$\left[i\delta_0(\{m-1\} \cap \{j-k\}) - (-i)^m \int_{[0,\infty)} e^{iuy} u^m \eta_{jk}(\mathrm{d}u) \right], \qquad m = 1,\dots,p,$$

and in particular all the the entries of $(D^m \tilde{h})(y)$ are bounded in y. Observe that, clearly, if a function $A\colon \mathbb{R} \to \mathbb{C}^{n \times n}$ takes the form

$$A(t) = B(t)C(t)D(t), \qquad t \in \mathbb{R}, \tag{5.6}$$

where all the entries of $B, D\colon \mathbb{R} \to \mathbb{C}^{n \times n}$ decay at least as $|y|^{-1}$ as $|y| \to \infty$ and all the entries of $C\colon \mathbb{R} \to \mathbb{C}^{n \times n}$ 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

$$(D\tilde{h}^{-1})(y) = -\tilde{h}(y)^{-1}(D\tilde{h})(y)\tilde{h}(y)^{-1}, \qquad y \in \mathbb{R},$$

it follows recursively that $D^m \tilde{h}^{-1}$ is a sum of functions of the form (5.6), and thus all its entries decay at least as $|y|^{-1}$ as $|y| \to \infty$ for $m = 1,\dots,p$. 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 $\mathcal{F}^{-1}[D^p \tilde{h}] = \left(t \mapsto (-it)^p g(t) \right)$ is an L^2 function. This implies in turn that $t \mapsto g_{jk}(t)(1 + |t|)^p \in L^2$ and, thus,

$$\int_{\mathbb{R}} |g_{jk}(t)|^q \, \mathrm{d}t \le \left(\int_{\mathbb{R}} \left(g_{jk}(t)(1+|t|)^p \right)^2 \mathrm{d}t \right)^{\frac{q}{2}} \left(\int_{\mathbb{R}} (1+|t|)^{-\frac{2pq}{2-q}} \, \mathrm{d}t \right)^{1-\frac{q}{2}} < \infty$$

for any $q \in [1/p, 2)$ and $j,k = 1,\dots,n$. By using the particular observation that $g \in L^1$ and (i) we obtain that

$$g(t) = \mathbb{1}_{[0,\infty)}(t)I_n + \int_{-\infty}^{t} g * \eta(u) \, \mathrm{d}u \tag{5.7}$$

for (almost) all $t \in \mathbb{R}$. This shows that

$$|g_{jk}(t)| \le 1 + \int_{\mathbb{R}} |[g * \eta(u)]_{jk}| \, \mathrm{d}u \le 1 + \sum_{l=1}^{n} \int_{\mathbb{R}} |g_{jl}(u)| \, \mathrm{d}u \, |\eta_{lk}|([0,\infty))$$

for all $t \in \mathbb{R}$ and for every $j,k = 1,\dots,n$ 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

$$\int_{\mathbb{R}} g * \eta(u) \, \mathrm{d}u = -I_n,$$

which concludes the proof of (ii).

Now suppose that $\int_{[0,\infty)} e^{\delta u} |\eta_{jk}|(du) < \infty$ for all $j,k = 1,\dots,n$ and some $\delta > 0$. In this case, $\mathcal{S}_\delta := \{z \in \mathbb{C} : \mathrm{Re}(z) \in [-\delta,\delta]\} \subseteq D(\eta)$ and

$$z \longmapsto \det h(z) = \det\left(I_n z - \int_{[0,\infty)} e^{zu} \eta(du) \right)$$

is strictly separated from 0 when $z \in \mathcal{S}_\delta$ and $|z|$ is sufficiently large. Indeed, the dominating term in $\det h(z)$ is z^n when $|z|$ is large, since

$$\left| \left[\int_{[0,\infty)} e^{zu} \, \eta(\mathrm{d}u) \right]_{jk} \right| \le \max_{l,m=1,\ldots,n} \int_{[0,\infty)} e^{\delta u} |\eta_{lm}|(\mathrm{d}u) \qquad \text{for } j,k = 1,\ldots,n.$$

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 $\mathcal{S}_{\tilde{\delta}} := \{z \in \mathbb{C} : \mathrm{Re}(z) \in [-\tilde{\delta}, \tilde{\delta}]\}$. 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,

$$\sup_{x \in [-\tilde{\delta}, \tilde{\delta}]} \int_{\mathbb{R}} \left| [h(x+iy)^{-1}]_{jk} \right|^2 \mathrm{d}y < \infty,$$

and this implies that $t \mapsto g_{jk}(t) e^{\varepsilon t} \in L^1$ for all $\varepsilon \in (-\tilde{\delta}, \tilde{\delta})$. This implication is a slight extension of the characterization of Hardy functions given in [13, Theorem 1 (Section 3.4)]; a general statement and the corresponding proof can be found in [3, Lemma 4.1].

Now fix any $\varepsilon \in (0, \tilde{\delta})$ and $j, k \in \{1, \ldots, n\}$, 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

$$|g_{jk}(t)| e^{\varepsilon t} \le |g_{jk}(0)| + \int_{\mathbb{R}} |[g * \eta(u)]_{jk}| e^{\varepsilon u} \, \mathrm{d}u + \varepsilon \int_{\mathbb{R}} |g_{jk}(u)| e^{\varepsilon u} \, \mathrm{d}u. \tag{5.8}$$

Since

$$\int_{\mathbb{R}} |[g * \eta(u)]_{jk}| e^{\varepsilon u} \, \mathrm{d}u \le \sum_{l=1}^{n} \int_{\mathbb{R}} |g_{jl}(u)| e^{\varepsilon u} \, \mathrm{d}u \int_{[0,\infty)} e^{\varepsilon u} |\eta_{lk}|(\mathrm{d}u)$$

it follows from (5.8) that

$$\max_{j,k=1,\ldots,n} |g_{jk}(t)| \le C e^{-\varepsilon t}$$

for all $t > 0$ with

$$C := 1 + \max_{j,k=1,\ldots,n} \left(\sum_{l=1}^{n} \int_{\mathbb{R}} |g_{jl}(u)| e^{\varepsilon |u|} \, \mathrm{d}u \int_{[0,\infty)} e^{\varepsilon u} |\eta_{lk}|(\mathrm{d}u) + \varepsilon \int_{\mathbb{R}} |g_{jk}(u)| e^{\varepsilon |u|} \, \mathrm{d}u \right).$$

By considering $-\varepsilon$ rather than ε in the above calculations one reaches the conclusion that

$$\max_{j,k=1,\ldots,n} |g_{jk}(t)| \le C e^{\varepsilon t}, \qquad t < 0,$$

and this verifies (iii).

Finally, suppose that $\det h(z) \ne 0$ for all $z \in \mathbb{C}$ with $\mathrm{Re}(z) \ge 0$. Then it holds that h and, thus, $z \mapsto h(z)^{-1}$ is continuous on $\{z \in \mathbb{C} : \mathrm{Re}(z) \ge 0\}$ and analytic on $\{z \in \mathbb{C} : \mathrm{Re}(z) > 0\}$. 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

$$\sup_{x>0} \int_{\mathbb{R}} |(h(x+iy)^{-1})_{jk}| \, \mathrm{d}y < \infty.$$

From the theory on Hardy spaces, see [12] or [13, Section 3.4], this implies that g is vanishing on $(-\infty, 0)$ almost everywhere, which verifies (iv) and ends the proof. $\qquad\square$

From Proposition 5.1 it becomes evident that we may (and, hence, do) choose the kernel g to satisfy (5.2) pointwise, so that the function induces a finite Lebesgue–Stieltjes measure $g(\mathrm{d}t)$. 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 $g\colon \mathbb{R} \to \mathbb{R}^{n \times n}$ characterized in* (5.1) *induces an $n \times n$ finite Lebesgue–Stieltjes measure, which is given by*

$$g(\mathrm{d}t) = I_n \delta_0(\mathrm{d}t) + g * \eta(t)\, \mathrm{d}t. \tag{5.9}$$

A function $f = [f_{jk}]\colon \mathbb{R} \to \mathbb{C}^{m \times n}$ is in $L^1(g(\mathrm{d}t))$ if

$$\int_{\mathbb{R}} |f_{jl}(t)[g * \eta]_{lk}(t)|\, \mathrm{d}t < \infty, \qquad l = 1, \ldots, n,$$

for $j = 1, \ldots, m$ and $k = 1, \ldots, n$. Moreover, the measure $g(\mathrm{d}t)$ has $(p-1)$th moment whenever η has pth 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 $f = [f_{jk}]\colon \mathbb{R} \to \mathbb{C}^{m \times n}$ to be integrable with respect to $g(\mathrm{d}t) = [g_{jk}(\mathrm{d}t)]$ we require that $f_{jl} \in L^1(|g_{lk}|(\mathrm{d}t))$, $l = 1, \ldots, n$, for each choice of $j = 1, \ldots, m$ and $k = 1, \ldots, n$. Since the variation measure $|g_{lk}|(\mathrm{d}t)$ of $g_{lk}(\mathrm{d}t)$ is given by

$$|g_{lk}|(\mathrm{d}t) = \delta_0(\{l-k\})\delta_0(\mathrm{d}t) + |[g * \eta(t)]_{lk}|\, \mathrm{d}t,$$

we see that this condition is equivalent to the statement in the result. Finally, suppose that η has pth moment for some $p \in \mathbb{N}$. Then, for any $j, k \in \{1, \ldots, n\}$, we get that

$$\int_{\mathbb{R}} |t|^{p-1} |g_{jk}|(\mathrm{d}t) \leq \sum_{l=1}^{n} \Big(|\eta_{lk}|([0, \infty)) \int_{\mathbb{R}} |t^{p-1} g_{jl}(t)|\, \mathrm{d}t$$
$$+ \int_{[0,\infty)} |t|^{p-1} |\eta_{lk}|(\mathrm{d}t) \int_{\mathbb{R}} |g_{jl}(t)|\, \mathrm{d}t \Big).$$

From the assumptions on η and Proposition 5.1(ii) we get immediately that $|\eta_{lk}|([0, \infty))$, $\int_{[0,\infty)} |t|^{p-1} |\eta_{lk}|(\mathrm{d}t)$ and $\int_{\mathbb{R}} |g_{jl}(t)|\, \mathrm{d}t$ are finite for all $l = 1, \ldots, n$. Moreover, for any such l we compute that

$$\int_{\mathbb{R}} |t^{p-1} g_{jl}(t)|\, \mathrm{d}t$$
$$\leq \int_{|t| \leq 1} |t^{p-1} g_{jl}(t)|\, \mathrm{d}t + \Big(\int_{|t|>1} t^{-2}\, \mathrm{d}t \Big)^{\frac{1}{2}} \Big(\int_{|t|>1} (t^p g_{jl}(t))^2\, \mathrm{d}t \Big)^{\frac{1}{2}}$$

which is finite, since $(t \mapsto t^p g_{jl}(t)) \in L^2$ 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,*

$$X_t = g(t-s)X_s + \int_s^\infty g(t-u)\,\eta * \{\mathbb{1}_{(-\infty,s]}X\}(u)\,\mathrm{d}u + g * \{\mathbb{1}_{(s,\infty)}(Z - Z_s)\}(t) \qquad (5.10)$$

for Lebesgue almost all $t > s$, *using the notation*

$$\big[\eta * \{\mathbb{1}_A X\}\big]_j(t) := \sum_{k=1}^n \int_{[0,\infty)} \mathbb{1}_A(t-u)X_{t-u}^k\,\eta_{jk}(\mathrm{d}u)$$

$$and \qquad \big[g * \{\mathbb{1}_{(s,\infty)}(Z - Z_s)\}\big]_j(t) := \sum_{k=1}^n \int_{\mathbb{R}} \mathbb{1}_{(s,\infty)}(t-u)(Z_{t-u}^k - Z_s^k)\,g_{jk}(\mathrm{d}u)$$

for $j = 1,\ldots,n$ *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 $0 \le \mathrm{Re}(z) < \delta$ and

$$\sup_{0 < x < \delta} \int_{\mathbb{R}} \big|\big[h(x+iy)^{-1}\big]_{jk}\big|^2\,\mathrm{d}y < \infty \qquad \text{for all } j,k = 1,\ldots,n.$$

Thus, a slight extension of [13, Theorem 1 (Section 3.4)] (which can be found in [3, Lemma 4.1]) ensures that $\mathcal{L}[g](z) = h(z)^{-1}$ when $\mathrm{Re}(z) \in (0,\delta)$. 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),

$$\mathbb{1}_{(s,\infty)}(t)X_t = \mathbb{1}_{(s,\infty)}(t)X_s + \int_{-\infty}^t \mathbb{1}_{(s,\infty)}(u)\,\eta * X(u)\,\mathrm{d}u + \mathbb{1}_{(s,\infty)}(t)(Z_t - Z_s)$$

for Lebesgue almost all $t \in \mathbb{R}$ outside a $\mathbb{P}$-null set (this is a consequence of Tonelli's theorem). In particular, this shows that

$$z\mathcal{L}[\mathbb{1}_{(s,\infty)}X](z)$$
$$= z\Big\{X_s\mathcal{L}[\mathbb{1}_{(s,\infty)}](z) + \mathcal{L}\Big[\int_{-\infty}^{\cdot} \mathbb{1}_{(s,\infty)}(u)\,\eta * X(u)\,\mathrm{d}u\Big](z)$$
$$\quad + \mathcal{L}[\mathbb{1}_{(s,\infty)}(Z - Z_s)](z)\Big\}$$
$$= \mathcal{L}[X_s\delta_0(\,\cdot\, - s)](z) + \mathcal{L}[\mathbb{1}_{(s,\infty)}\,\eta * X](z) + z\mathcal{L}[\mathbb{1}_{(s,\infty)}(Z - Z_s)](z).$$

By noticing that

$$\mathcal{L}[\mathbb{1}_{(s,\infty)}\,\eta * X](z) = \mathcal{L}\big[\mathbb{1}_{(s,\infty)}\,\eta * \{\mathbb{1}_{(-\infty,s]}X\}\big](z) + \mathcal{L}\big[\eta * \{\mathbb{1}_{(s,\infty)}X\}\big](z)$$
$$= \mathcal{L}\big[\mathbb{1}_{(s,\infty)}\,\eta * \{\mathbb{1}_{(-\infty,s]}X\}\big](z) + \mathcal{L}[\eta](z)\mathcal{L}[\mathbb{1}_{(s,\infty)}X](z),$$

it follows that

$$h(z)\mathcal{L}[\mathbb{1}_{(s,\infty)}X](z)$$
$$= \mathcal{L}\big[X_s\delta_0(\,\cdot\, - s) + \mathbb{1}_{(s,\infty)}\,\eta * \{\mathbb{1}_{(-\infty,s]}X\}\big](z) + z\mathcal{L}[\mathbb{1}_{(s,\infty)}(Z - Z_s)](z).$$

(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, by using that $\mathcal{L}[g](z) = h(z)^{-1}$, we find

$$zh(z)^{-1}\mathcal{L}[\mathbb{1}_{(s,\infty)}(Z - Z_s)](z) = \mathcal{L}[g(\mathrm{d}t)](z)\mathcal{L}[\mathbb{1}_{(s,\infty)}(Z - Z_s)](z)$$
$$= \mathcal{L}\big[g * \{\mathbb{1}_{(s,\infty)}(Z - Z_s)\}\big](z)$$

and, thus,

$$X_t = g(t-s)X_s + \int_s^\infty g(t-u)\eta * \{\mathbb{1}_{(-\infty,s]}X\}(u)\,\mathrm{d}u + g * \{\mathbb{1}_{(s,\infty)}(Z - Z_s)\}$$

for Lebesgue almost all $t > s$ with probability one. $\qquad\square$

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 of Theorem 3.1. Fix $t \in \mathbb{R}$. The convolution in (3.3) is well-defined if $u \mapsto Z_{t-u}^\top$ is $g^\top$-integrable (by Corollary 5.2) which means that $u \mapsto Z_{t-u}^k$ belongs to $L^1(|g_{jk}|(\mathrm{d}u))$ for all $j,k = 1,\ldots,n$. 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 $\mathbb{E}[|Z_u^k|] \leq \alpha + \beta|u|$ for all $u \in \mathbb{R}$. Consequently,

$$\mathbb{E}\bigg[\int_\mathbb{R} |Z_{t-u}^k|\,\mu(\mathrm{d}u)\bigg] \leq (\alpha + \beta|t|)\mu(\mathbb{R}) + \beta\int_\mathbb{R}|u|\,\mu(\mathrm{d}u) < \infty$$

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 $|g_{jk}|(\mathrm{d}u)$ for $j = 1,\ldots,n$ (according to Corollary 5.2).

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

$$\int_s^t X^\top * \eta^\top(u)\,\mathrm{d}u = \int_s^t Z^\top * \eta^\top(u)\,\mathrm{d}u + \int_s^t \bigg(\int_\mathbb{R} g * \eta(r)Z_{\cdot-r}\,\mathrm{d}u\bigg)^\top * \eta^\top(u)\,\mathrm{d}u$$

it is clear by the definition of $(X_t)_{t\in\mathbb{R}}$ that it suffices to argue that

$$\int_s^t \bigg(\int_\mathbb{R} g * \eta(r)Z_{\cdot-r}\,\mathrm{d}u\bigg)^\top * \eta^\top(u)\,\mathrm{d}u$$
$$= \int_\mathbb{R} Z_r^\top[g * \eta(t-r) - g * \eta(s-r)]^\top\,\mathrm{d}r - \int_s^t Z^\top * \eta^\top(r)\,\mathrm{d}r.$$

We do this componentwise, so we fix $i \in \{1,\dots,n\}$ and compute that

$$\left[\int_s^t \left(\int_{\mathbb{R}} g * \eta(r) Z_{\cdot - r} \, dr \right)^{\top} * \eta^{\top}(u) \, du \right]_i$$

$$= \sum_{j=1}^{n} \sum_{k=1}^{n} \sum_{l=1}^{n} \int_s^t \left(\int_{\mathbb{R}} g_{jl} * \eta_{lk}(v) Z_{\cdot - r}^k \, dr \right) * \eta_{ij}(u) \, du$$

$$= \sum_{j=1}^{n} \sum_{k=1}^{n} \sum_{l=1}^{n} \int_{\mathbb{R}} Z_r^k \int_{[0,\infty)} \int_s^t \int_{[0,\infty)} g_{jl}(u - v - r - w) \eta_{ij}(dv) \, du \, \eta_{lk}(dw) \, dr$$

$$= \sum_{k=1}^{n} \sum_{l=1}^{n} \int_{\mathbb{R}} Z_r^k \int_{[0,\infty)} \int_s^t (g * \eta)_{il}(u - r - w) \, du \, \eta_{lk}(dw) \, dr$$

$$= \sum_{k=1}^{n} \sum_{l=1}^{n} \left(\int_{\mathbb{R}} Z_r^k \int_{[0,\infty)} [g_{il}(t - r - w) - g_{il}(s - r - w)] \eta_{lk}(dw) \, dr \right.$$

$$\left. - \int_{\mathbb{R}} Z_r^k \int_{[0,\infty)} \delta_0(\{i - l\}) \mathbb{1}_{(s,t]}(r + w) \eta_{lk}(dw) \, dr \right)$$

$$= \sum_{k=1}^{n} \left(\int_{\mathbb{R}} Z_r^k [(g * \eta)_{ik}(t - r) - (g * \eta)_{ik}(s - r)] \, dr - \int_s^t Z^k * \eta_{ik}(r) \, dr \right)$$

$$= \left[\int_{\mathbb{R}} Z_r^{\top} [g * \eta(t - r) - g * \eta(s - r)]^{\top} \, dr - \int_s^t Z^{\top} * \eta^{\top}(r) \, dr \right]_i$$

where we have used (i) in Proposition 5.1 and the fact that g and η commute in a convolution sense, $g * \eta = (g^{\top} * \eta^{\top})^{\top}$ (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 $-m = t_0^k < t_1^k < \cdots < t_k^k = m$ be a partition of $[-m, m]$ with $\max_{j=1,\dots,k}(t_j^k - t_{j-1}^k) \to 0$, $k \to \infty$, and define the Riemann sum

$$X_t^{m,k} = \sum_{j=1}^{k} g * \eta(t_{j-1}^k) [Z_{t-t_{j-1}^k} - Z_t](t_j^k - t_{j-1}^k).$$

Observe that $(X_t^{m,k})_{t \in \mathbb{R}}$ is stationary. Moreover, the ith component of $X_t^{m,k}$ converges to the ith component of

$$X_t^m = \int_{-m}^{m} g * \eta(u) [Z_{t-u} - Z_t] \, du$$

in $L^1(\mathbb{P})$ as $k \to \infty$. To see this, we start by noting that

$$\mathbb{E}\left[|[X_t^m]_i - [X_t^{m,k}]_i| \right] \leq \sum_{j=1}^{n} \int_{\mathbb{R}} \sum_{l=1}^{k} \mathbb{1}_{(t_{l-1}^k, t_l^k]}(u) \mathbb{E}\left[|[g * \eta]_{ij}(u)[Z_{t-u}^j - Z_t^j] \right.$$

$$\left. - [g * \eta]_{ij}(t_{l-1}^k)[Z_{t-t_{l-1}^k}^j - Z_t^j]| \right] du.$$

Then, for each $j \in \{1, \ldots, n\}$,

$$\max_{l=1,\ldots,k} \mathbb{1}_{(t^k_{l-1}, t^k_l]}(u) \, \mathbb{E}\big[\big|[g*\eta]_{ij}(u)[Z^j_{t-u} - Z^j_t] - [g*\eta]_{ij}(t^k_{l-1})[Z^j_{t-t^k_{l-1}} - Z^j_t]\big|\big]$$

$$\leq \max_{l=1,\ldots,k} \mathbb{1}_{(t^k_{l-1}, t^k_l]}(u)\Big(|(g*\eta)_{ij}(u)|\, \mathbb{E}\big[|Z^j_{t-u} - Z^j_{t-t^k_{l-1}}|\big]$$

$$+ \mathbb{E}\big[|Z^j_{t-t^k_{l-1}} - Z^j_t|\big]|[g*\eta]_{ij}(u) - [g*\eta]_{ij}(t^k_{l-1})|\Big) \to 0$$

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

$$X_t = g(t-s)X_s + \int_s^\infty g(t-u)\eta * \{\mathbb{1}_{(-\infty,s]}X\}(u)\,\mathrm{d}u$$

holds for each $s \in \mathbb{R}$ and Lebesgue almost all $t > s$. For a given j we thus find that

$$\mathbb{E}\big[|X^j_t|\big] \leq C \sum_{k=1}^n \Big(|g_{jk}(t-s)| + \sum_{l=1}^n \int_s^\infty |g_{jk}(t-u)||\eta_{kl}|([u-s,\infty))\,\mathrm{d}u\Big)$$

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 $\mathrm{Re}(z) \geq 0$ 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

$$X_t = Z_t + \int_0^\infty g*\eta(u)Z_{t-u}\,\mathrm{d}u = -\int_0^\infty g*\eta(u)[Z_t - Z_{t-u}]\,\mathrm{d}u, \qquad t \in \mathbb{R},$$

by (5.3). $\qquad\square$

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. $\qquad\square$

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 $\mathcal{I}_k(f_r(t - \cdot)) = Z^k_t - Z^k_{t-r}$ and thus, since $(Z^k_t)_{t\in\mathbb{R}}$ is integrable and has stationary increments, $\mathbb{E}[|\mathcal{I}_k(f_r(t - \cdot))|] \leq \alpha + \beta|r|$ for all $r \in \mathbb{R}$ and suitably chosen $\alpha, \beta > 0$ (see, e.g., [2, Corollary A.3]). In particular,

$$\mathbb{E}\Big[\int_\mathbb{R} |\mathcal{I}_k(f_r(t - \cdot))|\,|\mu|(\mathrm{d}r)\Big] \leq \alpha|\mu|(\mathbb{R}) + \beta \int_\mathbb{R} |r|\,|\mu|(\mathrm{d}r) < \infty,$$

and thus $\mathcal{I}_k(f_r(t - \cdot))$ is integrable with respect to μ and 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), fix $t \in \mathbb{R}$ and $j, k \in \{1, \dots, n\}$, and note 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 $\mathcal{I}_k$ to get

$$
\begin{aligned}
\int_{\mathbb{R}} [g * \eta]_{jk}(r)\big[Z_{t-r}^k - Z_t^k\big]\,dr &= \int_{\mathbb{R}} \mathcal{I}_k(\mathbb{1}_{(t, t-r]})[g * \eta]_{jk}(r)\,dr \\
&= \mathcal{I}_k\Big(\int_{\mathbb{R}} \mathbb{1}_{(t, t-r]}[g * \eta]_{jk}(r)\,dr\Big) \\
&= \mathcal{I}_k\Big(\delta_0(\{j - k\})\mathbb{1}_{[0, \infty)}(t - \cdot) + \int_{-\infty}^{t - \cdot} [g * \eta]_{jk}(u)\,du\Big) \\
&= \mathcal{I}_k(g_{jk}(t - \cdot))
\end{aligned}
$$

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

$$
X_t^j = \sum_{k=1}^n \int_{\mathbb{R}} [g * \eta]_{jk}(r)\big[Z_{t-r}^k - Z_t^k\big]\,dr = \sum_{k=1}^n \mathcal{I}_k(g_{jk}(t - \cdot)),
$$

which was to be shown. $\qquad\square$

Proof of Proposition 4.3. Let $\alpha \in (1, 2]$ and $\beta \in (0, 1 - 1/\alpha)$, and consider a function $f \colon \mathbb{R} \to \mathbb{R}$ in $L^1 \cap L^\alpha$. We start by writing

$$
\int_t^\infty |f(u)|(u - t)^{\beta - 1}\,du = \int_0^1 |f(t + u)|u^{\beta - 1}\,du + \int_1^\infty |f(t + u)|u^{\beta - 1}\,du.
$$

For the left term we find that

$$
\begin{aligned}
\int_{\mathbb{R}} \Big(\int_0^1 |f(t + u)|u^{\beta - 1}\,du\Big)^\alpha\,dt &\le \Big(\int_0^1 u^{\beta - 1}\,du\Big)^{\alpha - 1} \int_{\mathbb{R}} \int_0^1 |f(t + u)|^\alpha u^{\beta - 1}\,du\,dt \\
&= \Big(\int_0^1 u^{\beta - 1}\,du\Big)^\alpha \int_{\mathbb{R}} |f(t)|^\alpha\,dt < \infty.
\end{aligned}
$$

For the right term we find

$$
\begin{aligned}
\int_{\mathbb{R}} \Big(\int_1^\infty |f(t + u)|u^{\beta - 1}\,du\Big)^\alpha\,dt &\le \Big(\int_{\mathbb{R}} f(u)\,du\Big)^{\alpha - 1} \int_{\mathbb{R}} \int_1^\infty |f(t + u)|u^{\alpha(\beta - 1)}\,du\,dt \\
&= \Big(\int_{\mathbb{R}} f(u)\,du\Big)^\alpha \int_1^\infty u^{\alpha(\beta - 1)}\,du < \infty.
\end{aligned}
$$

We conclude that $(I_-^\beta f)(u) \in L^\alpha$. $\qquad\square$

Proof of Theorem 4.5. The identity (4.9) 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 $A = I_n z$, $B = -(e_1 \otimes I_n)^\top \in \mathbb{R}^{n \times (m-1)n}$, $C = -e_{m-1} \otimes \mathcal{L}[\omega_0](z) \in \mathbb{R}^{(m-1)n \times n}$ and

$$
D = \begin{bmatrix}
I_n z & -I_n & 0 & \cdots & 0 \\
0 & I_n z & -I_n & \cdots & 0 \\
\vdots & \vdots & \ddots & \ddots & \vdots \\
0 & 0 & \cdots & I_n z & -I_n \\
-\mathcal{L}[\omega_1](z) & -\mathcal{L}[\omega_2](z) & \cdots & -\mathcal{L}[\omega_{m-2}](z) & I_n z - \mathcal{L}[\omega_{m-1}](z)
\end{bmatrix}.
$$

Here e_1 and e_{m-1} refer to the first 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 $[h(iy)^{-1}]_{jk}$ is an $n \times n$ matrix. Then we know from Theorem 3.1 that the unique solution to (4.6) is a $(Z_t)_{t \in \mathbb{R}}$-driven moving average of the form (4.8) with $\mathcal{F}[g_{1m}](y) = [h(iy)^{-1}]_{1m}$. 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

$$
\left(I_n z^m - \sum_{j=0}^{m-1} \mathcal{L}[\omega_j](z) z^j \right)^{-1}
$$

using the representation (5.11) and standard rules for inverting block matrices. This means in particular that (4.10) is true. $\qquad\square$

Proof of Theorem 4.8. We start by arguing that there exists an integrable function f, which is vanishing on $(-\infty, 0)$ and has Fourier transform given by (4.19). 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 $z \in \mathcal{S}_\delta := \{x + iy : x \geq -\delta, y \in \mathbb{R}\}$ for a suitably chosen $\delta > 0$. By a slight modification of [13, Theorem 1 (Section 3.4)], or by [3, Lemma 4.1], it suffices to argue that there exists $\varepsilon \in (0, \delta]$ such that

$$
\sup_{x > -\varepsilon} \int_{\mathbb{R}} |H(x + iy)_{jk}|^2 \, \mathrm{d}y < \infty \qquad \text{for all } j, k = 1, \ldots, n. \tag{5.12}
$$

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 $\omega_0, \omega_1, \ldots, \omega_{p-q-1}$ 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 (4.7). Then it follows from (4.9) 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 $\det h(iy)$ is non-zero for all $y \in \mathbb{R}$. In light of Proposition 4.5, in particular (4.10), we may therefore conclude that the unique solution to (4.6) is a $(Z_t)_{t\in\mathbb{R}}$-driven moving average, where the driving kernel has Fourier transform

$$\left(I_n(iy)^{p-q} + \sum_{j=0}^{p-q-1} R_j(iy)^j - \mathcal{F}[f](y)\right)^{-1} = P(iy)^{-1}Q(iy), \qquad y \in \mathbb{R}.$$

In other words, the unique solution is the $(Z_t)_{t\in\mathbb{R}}$-driven MCARMA 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* $C_0,\ldots,C_{p-q-1}$ *be given in (4.17) and* $C_{p-q} = I_n$*. Define*

$$R_j(z) = \sum_{k=j}^{p-q} C_k z^{k-j}, \qquad j = 1,\ldots,p-q-1.$$

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*

$$(e_1^{p-q} \otimes I_n)^\top g = [\tilde{g}R_1(D),\ldots,\tilde{g}R_{p-q-1}(D),\tilde{g}] \tag{5.13}$$

where

$$\tilde{g}R_j(D)(t) = \sum_{k=j}^{p-q} D^{k-j}\tilde{g}(t)C_k$$
$$= \mathbb{1}_{[0,\infty)}(t)(e_1^p \otimes I_n)^\top e^{At} \sum_{k=j}^{p-q} A^{k-j}EC_k \tag{5.14}$$

for $j = 1,\ldots,p-q-1$ *and* $g\colon \mathbb{R} \to \mathbb{R}^{n\times n}$ *is characterized by* $\mathcal{F}[g](y) = h(iy)^{-1}$ *with* $h\colon \mathbb{C} \to \mathbb{C}^{n(p-q)\times n(p-q)}$ *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. The fact 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 $\mathcal{F}[\tilde{g}](y) = P(iy)^{-1}Q(iy)$. Consequently, (5.13) follows since

$$[P(iy)^{-1}Q(iy)R_1(iy),$$
$$\ldots,P(iy)^{-1}Q(iy)R_{p-q-1}(iy), P(iy)^{-1}Q(iy)]h(z) = (e_1^{p-q} \otimes I_n)^\top.$$

The relation in (5.14) is due to the representation of $\tilde{g}$ given in (4.16). □

Proof of Corollary 4.11. The prediction formula is a consequence of Lemma 5.4 combined with Theorems 3.2 and 4.8. Furthermore, to get the expression for $\tilde{g} *$ $\{\hat{Z}\mathbb{1}_{(s,\infty)}\}$, note that

$$\tilde{g}(\mathrm{d}v) = \mathbb{1}_{\{p=q+1\}}\delta_0(\mathrm{d}v) + (e_1^p \otimes I_n)^\top e^{Av}AE \, \mathrm{d}v,$$

which follows from the representation of $\tilde{g}$ in (4.16). □

Acknowledgments

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

[23]   Samorodnitsky, G. and M. S. Taqqu (1994). *Stable Non-Gaussian Random Processes*. Stochastic Modeling. Stochastic models with infinite variance. New York: Chapman & Hall.

[24]   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.

[25]   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.

[26]   Stelzer, R. (2011). CARMA Processes driven by Non-Gaussian Noise. arXiv: 1201.0155.

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

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 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.

MSC: 60G22; 60H10; 60H20; 60G17; 60H05

Keywords: Long-range dependence; Moving average processes; Semimartingales; Stochastic differential equations

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, 25, 28, 29] 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 Granger and Joyeux [14] and Hosking [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 $\pi_j = \prod_{0 < k \le j} \frac{k-1-\gamma}{k}$. 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

$$P(B)(1 - B)^\beta X_t = Q(B)\varepsilon_t, \qquad t \in \mathbb{Z}, \tag{1.1}$$

where P and Q are real polynomials with no zeroes on $\{z \in \mathbb{C} : |z| \le 1\}$, $(\varepsilon_t)_{t \in \mathbb{Z}}$ is an i.i.d. sequence with $\mathbb{E}[\varepsilon_0] = 0$, $\mathbb{E}[\varepsilon_0^2] \in (0, \infty)$ 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,

$$P(B)X_t = Q(B)(1 - B)^{-\beta}\varepsilon_t, \qquad t \in \mathbb{Z}. \tag{1.2}$$

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

$$\gamma_X(t) \sim ct^{2\beta-1}, \qquad t \to \infty, \tag{1.3}$$

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

$$\mathrm{d}X_t = -\kappa X_t \, \mathrm{d}t + \mathrm{d}I^\beta L_t, \qquad t \in \mathbb{R}, \tag{1.4}$$

where $\kappa > 0$ and

$$I^\beta L_t := \frac{1}{\Gamma(1 + \beta)} \int_{-\infty}^t \left[(t - u)^\beta - (-u)_+^\beta\right] \mathrm{d}L_u, \qquad t \in \mathbb{R}, \tag{1.5}$$

with $(L_t)_{t \in \mathbb{R}}$ being a Lévy process which satisfies $\mathbb{E}[L_1] = 0$ and $\mathbb{E}[L_1^2] \in (0, \infty)$. 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.,

$$X_t - X_s = -\frac{\kappa}{\Gamma(1 - \beta)} \int_{-\infty}^t \left[(t - u)^{-\beta} - (s - u)_+^{-\beta}\right] X_u \, \mathrm{d}u + L_t - L_s, \qquad s < t. \tag{1.6}$$

One can write (1.6) compactly as

$$dX_t = -\kappa D^\beta X_t \, dt + dL_t, \qquad t \in \mathbb{R}, \tag{1.7}$$

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) are 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 [21, 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, and Jusselin and Rosenbaum [19] show that the no-arbitrage assumption implies that the volatility of the macroscopic price process is rough. These findings suggest 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 Brockwell and Marquardt [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

$$X_t = \int_{-\infty}^{t} g(t-u) \, dI^\beta L_u, \qquad t \in \mathbb{R}, \tag{1.8}$$

with $g\colon \mathbb{R} \to \mathbb{R}$ being the L^1 function characterized by

$$\mathcal{F}[g](y) := \int_{\mathbb{R}} e^{-iyu} g(u)\, \mathrm{d}u = \frac{Q(iy)}{P(iy)}, \qquad y \in \mathbb{R}.$$

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

$$P(D)X_t = Q(D)DI^{\beta}L_t, \qquad t \in \mathbb{R}.$$

(See, e.g., [21].) 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

$$\mathrm{d}X_t = \int_{[0,\infty)} X_{t-u}\, \eta(\mathrm{d}u)\, \mathrm{d}t + \mathrm{d}I^{\beta}L_t, \qquad t \in \mathbb{R}, \tag{1.9}$$

for a suitable finite signed measure η. See [2, 22] 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

$$X_t - X_s = \int_{-\infty}^{t} \big(D_-^{\beta}\mathbb{1}_{(s,t]}\big)(u) \int_{[0,\infty)} X_{u-v}\, \eta(\mathrm{d}v)\, \mathrm{d}u + L_t - L_s \tag{1.10}$$

almost surely for each $s < t$, where η is a given finite signed measure and

$$\big(D_-^{\beta}\mathbb{1}_{(s,t]}\big)(u) = \frac{1}{\Gamma(1-\beta)}\big[(t-u)_+^{-\beta} - (s-u)_+^{-\beta}\big], \qquad u \in \mathbb{R}.$$

We will refer to (1.10) as a *stochastic fractional delay differential equation* (SFDDE). Equation (1.10) can be compactly written as

$$\mathrm{d}X_t = \int_{[0,\infty)} D^{\beta}X_{t-u}\, \eta(\mathrm{d}u)\, \mathrm{d}t + \mathrm{d}L_t, \qquad t \in \mathbb{R}, \tag{1.11}$$

with $(D^{\beta}X_t)_{t\in\mathbb{R}}$ defined in Proposition 3.6. The 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,

$$\eta(\mathrm{d}t) = -\kappa\delta_0(\mathrm{d}t) + f(t)\, \mathrm{d}t, \tag{1.12}$$

where $f(t) = \mathbb{1}_{[0,\infty)}(t)b^{\top}e^{At}e_1$ with $e_1 = [1,0,\dots,0]^{\top} \in \mathbb{R}^n$, $b \in \mathbb{R}^n$ 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

$$dX_t = \int_0^\infty D^\beta X_{t-u} f(u) \, du \, dt + dL_t, \qquad t \in \mathbb{R}. \tag{1.13}$$

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 [3], 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. Section 6 contains the proofs of all the results presented in Sections 3 and 4. 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 $\mu \colon \mathscr{B}(\mathbb{R}) \to \mathbb{R}$ 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 $f, g \colon \mathbb{R} \to \mathbb{C}$ 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 $\mathcal{L}[\mu] \colon D(\mu) \to \mathbb{C}$ of μ by

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

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 càdlàg 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 $L_t = -L_{(-t)-}^2$ if $t < 0$. If $\mathbb{E}[L_1^2] < \infty$, $\mathbb{E}[L_1] = 0$ and $f \in L^2$, the integral $\int_{\mathbb{R}} f(u)\, \mathrm{d}L_u$ is well-defined as an L^2 limit of integrals of step functions, and the following isometry property holds:

$$\mathbb{E}\left[\left(\int_{\mathbb{R}} f(u)\, \mathrm{d}L_u\right)^2\right] = \mathbb{E}[L_1^2] \int_{\mathbb{R}} f(u)^2\, \mathrm{d}u.$$

For more on Lévy processes and integrals with respect to these, see [26, 31]. Finally, for two functions $f, g \colon \mathbb{R} \to \mathbb{C}$ 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

$$\lim_{t \to a} \frac{f(t)}{g(t)} = 0, \quad \limsup_{t \to a}\left|\frac{f(t)}{g(t)}\right| < \infty \quad \text{and} \quad \lim_{t \to a} \frac{f(t)}{g(t)} = 1,$$

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 $\mathbb{E}[L_1^2] = 1$. Moreover, denote by η a finite (possibly signed) measure on $[0, \infty)$ with

$$\int_{[0,\infty)} t\, |\eta|(\mathrm{d}t) < \infty \tag{3.1}$$

and set

$$\left(D_-^\beta \mathbb{1}_{(s,t]}\right)(u) = \frac{1}{\Gamma(1-\beta)}\left[(t-u)_+^{-\beta} - (s-u)_+^{-\beta}\right], \qquad u \in \mathbb{R}. \tag{3.2}$$

(In line with [12] we write $D_-^\beta \mathbb{1}_{(s,t]}$ rather than $D^\beta \mathbb{1}_{(s,t]}$ in (3.2) to emphasize that it is the right-sided version of the Riemann–Liouville fractional derivative of $\mathbb{1}_{(s,t]}$.) 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

$$X_t - X_s = \int_{-\infty}^t \left(D_-^\beta \mathbb{1}_{(s,t]}\right)(u) \int_{[0,\infty)} X_{u-v}\, \eta(\mathrm{d}v)\, \mathrm{d}u + L_t - L_s \tag{3.3}$$

almost surely for each $s < t$. Note that equation (3.3) is indeed well-defined, since η is finite, $(X_t)_{t \in \mathbb{R}}$ is bounded in $L^1(\mathbb{P})$ and $D_-^\beta \mathbb{1}_{(s,t]} \in L^1$. As noted in the introduction, we will often write (3.3) shortly as

$$\mathrm{d}X_t = \int_{[0,\infty)} D^\beta X_{t-u}\, \eta(\mathrm{d}u)\, \mathrm{d}t + \mathrm{d}L_t, \qquad t \in \mathbb{R}, \tag{3.4}$$

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 lead to a stationary solution to (3.3) we introduce the function $h = h_{\beta,\eta} \colon \{z \in \mathbb{C} : \mathrm{Re}(z) \geq 0\} \to \mathbb{C}$ given by

$$h(z) = z^{1-\beta} - \int_{[0,\infty)} e^{-zu}\, \eta(\mathrm{d}u), \qquad \mathrm{Re}(z) \geq 0. \tag{3.5}$$

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 $\mathbb{C} \setminus \{z \in \mathbb{R} : z \leq 0\}$. 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 $\mathrm{Re}(z) \geq 0$. Then there exists a unique $g \colon \mathbb{R} \to \mathbb{R}$, which belongs to L^γ for $(1-\beta)^{-1} < \gamma \leq 2$ and is vanishing on $(-\infty, 0)$, such that*

$$\mathcal{F}[g](y) = \frac{(iy)^{-\beta}}{h(iy)}, \qquad y \in \mathbb{R}. \tag{3.6}$$

Moreover, the following statements hold:

(i) *For $t > 0$ the Marchaud fractional derivative $D^\beta g(t)$ at t of g given by*

$$D^\beta g(t) = \frac{\beta}{\Gamma(1-\beta)} \lim_{\delta \downarrow 0} \int_\delta^\infty \frac{g(t) - g(t-u)}{u^{1+\beta}}\, \mathrm{d}u \tag{3.7}$$

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,*

$$g(t) = \frac{1}{\Gamma(\beta)} \int_0^t D^\beta g(u)(t-u)^{\beta-1}\, \mathrm{d}u, \qquad t > 0. $$

(iii) *The function g satisfies*

$$g(t) = 1 + \int_0^t \left(D^\beta g\right) * \eta(u)\, \mathrm{d}u, \qquad t \geq 0, \tag{3.8}$$

and for $v \in \mathbb{R}$ and with $D_-^\beta \mathbb{1}_{(s,t]}$ given in (3.2),

$$g(t-v) - g(s-v) = \int_{-\infty}^t \left(D_-^\beta \mathbb{1}_{(s,t]}\right)(u)\, g * \eta(u-v)\, \mathrm{d}u + \mathbb{1}_{(s,t]}(v). \tag{3.9}$$

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

$$\bigcap_{t \in \mathbb{R}} \overline{\mathrm{sp}}\{X_s : s \leq t\} = \{0\}, \tag{3.10}$$

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 $\mathrm{Re}(z) \geq 0$ and let g be the function introduced in Proposition 3.1. Then the process*

$$X_t = \int_{-\infty}^t g(t-u)\, \mathrm{d}L_u, \qquad t \in \mathbb{R}, \tag{3.11}$$

is well-defined, centered and square integrable, and it is the unique purely non-deterministic solution to the SFDDE (3.3).

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 (3.3) 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 (3.3) 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

$$\mathbb{E}\left[e^{iyL_1} \right] = e^{-\sigma^\alpha |y|^\alpha}, \qquad y \in \mathbb{R},$$

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 (3.3) 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 [30].

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, (3.3) with $D_-^\beta \mathbb{1}_{(s,t]}$ replaced by $\mathbb{1}_{(s,t]}$) is

$$z \longmapsto z - \mathcal{L}[\eta](z), \tag{3.12}$$

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 [2]). 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 $L_t - L_s$, $s \leq t$. 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 $\text{Re}(z) = 0$ rather than when $\text{Re}(z) \geq 0$ in terms of causality is similar to that of non-fractional SDDEs (see, e.g., [2]).

The next result shows why one may view (3.3) 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 (3.3) 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 $\mathrm{Re}(z) \geq 0$ and let $(X_t)_{t\in\mathbb{R}}$ be the solution to (3.3) given in Theorem 3.2. Then, for $t \in \mathbb{R}$, the limit*

$$D^\beta X_t := \frac{\beta}{\Gamma(1-\beta)} \lim_{\delta\downarrow 0} \int_\delta^\infty \frac{X_t - X_{t-u}}{u^{1+\beta}} \, \mathrm{d}u \tag{3.13}$$

exists in $L^2(\mathbb{P})$, $D^\beta X_t = \int_{-\infty}^t D^\beta g(t-u) \, \mathrm{d}L_u$, and it holds that

$$
\frac{1}{\Gamma(1-\beta)} \int_{-\infty}^t \Big[(t-u)^{-\beta} - (s-u)_+^{-\beta}\Big] \int_{[0,\infty)} X_{u-v}\, \eta(\mathrm{d}v) \, \mathrm{d}u
$$
$$
= \int_s^t \int_{[0,\infty)} D^\beta X_{u-v}\, \eta(\mathrm{d}v) \, \mathrm{d}u
\tag{3.14}
$$

almost surely for each $s < t$.

We will now provide some properties of the solution $(X_t)_{t\in\mathbb{R}}$ to (3.3) given in (3.11). Since the autocovariance function γ_X takes the form

$$\gamma_X(t) = \int_\mathbb{R} g(t+u)g(u) \, \mathrm{d}u, \qquad t \in \mathbb{R}, \tag{3.15}$$

it follows by Plancherel's theorem that $(X_t)_{t\in\mathbb{R}}$ admits a spectral density f_X which is given by

$$f_X(y) = |\mathcal{F}[g](y)|^2 = \frac{1}{|h(iy)|^2} |y|^{-2\beta}, \qquad y \in \mathbb{R}. \tag{3.16}$$

(See the appendix for a brief recap of the spectral theory.) The following result concerning γ_X and f_X shows that solutions to (3.3) 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 $\mathrm{Re}(z) \geq 0$ and let γ_X and f_X be the functions introduced in (3.15)–(3.16). Then it holds that*

$$\gamma_X(t) \sim \frac{\Gamma(1-2\beta)}{\Gamma(\beta)\Gamma(1-\beta)\eta([0,\infty))^2} t^{2\beta-1} \qquad \text{as } t \to \infty$$

$$\text{and} \qquad f_X(y) \sim \frac{1}{\eta([0,\infty))^2} |y|^{-2\beta} \qquad \text{as } y \to 0.$$

In particular, $\int_\mathbb{R} |\gamma_X(t)| \, \mathrm{d}t = \infty$.

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 $\mathrm{Re}(z) \geq 0$, let $(X_t)_{t\in\mathbb{R}}$ be the solution to (3.3) and denote by ρ_X its ACF. Then it holds that $1 - \rho_X(h) \sim h$ as $h \downarrow 0$.*

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

$$\frac{\mathbb{E}[(X_t - X_0)^2]}{t^{2\gamma}} \leq C$$

for all sufficiently small $t > 0$. By defining the semi-variogram

$$\gamma_V(t) := \frac{1}{2}\mathbb{E}[(X_t - X_0)^2], \qquad t \in \mathbb{R},$$

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 (3.3) 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 (3.3), 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 [21, 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 [3, 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 $\mathrm{Re}(z) \geq 0$ and let $(X_t)_{t\in\mathbb{R}}$ denote the solution to (3.3). Then for any $s < t$, it holds that*

$$\mathbb{E}[X_t \mid X_u, u \leq s] = g(t - s)X_s$$
$$+ \int_{[0,t-s)} \int_{-\infty}^s X_w \int_{[0,\infty)} \left(D_-^\beta \mathbb{1}_{(s,t-u]}\right)(v + w)\,\eta(dv)\,dw\,g(du),$$

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

$$\eta(dt) = -\kappa\delta_0(dt) + f(t)\,dt \quad \text{with} \quad f(t) = \mathbb{1}_{[0,\infty)}(t)b^\top e^{At}e_1, \tag{4.1}$$

where $e_1 := [1, 0, \ldots, 0]^\top \in \mathbb{R}^n$. 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

$$\mathcal{F}[f](y) = b^\top (I_n iy - A)^{-1} e_1, \qquad y \in \mathbb{R},$$

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 $\kappa + b^\top A^{-1} e_1 \neq 0$ and*

$$Q(z)[z + \kappa z^\beta] + R(z)z^\beta \neq 0 \tag{4.3}$$

for all $z \in \mathbb{C} \setminus \{0\}$ with $\operatorname{Re}(z) \geq 0$. Then there exists a unique purely non-deterministic solution $(X_t)_{t \in \mathbb{R}}$ to (3.3) with η given by (4.1) and it is given by (3.11) with $g : \mathbb{R} \to \mathbb{R}$ characterized through the relation

$$\mathcal{F}[g](y) = \frac{Q(iy)}{Q(iy)[iy + \kappa(iy)^\beta] + R(iy)(iy)^\beta}, \qquad y \in \mathbb{R}. \tag{4.4}$$

Before giving examples we state Proposition 4.2, which shows that the general SFDDE (3.3) can be written as

$$dX_t = -\kappa D^\beta X_t \, dt + \int_0^\infty X_{t-u} D^\beta f(u) \, du \, dt + dL_t, \qquad t \in \mathbb{R}, \tag{4.5}$$

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 [2] 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*

$$D^\beta f(t) = \frac{1}{\Gamma(1-\beta)} b^\top \left(A e^{At} \int_0^t e^{-Au} u^{-\beta} \, du + t^{-\beta} I_n \right) e_1$$

for $t > 0$ belongs to $L^1 \cap L^2$. If in addition (4.3) holds, $\kappa + b^\top A^{-1} e_1 \neq 0$ 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

$$\bigl(Q(D)[D + \kappa D^\beta] + R(D)D^\beta\bigr)X_t = Q(D)DL_t, \qquad t \in \mathbb{R}, \tag{4.6}$$

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 $\{z \in \mathbb{C} : \operatorname{Re}(z) < 0\}$, 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

$$\bigl(Q(D)[D^{1+\beta} + \kappa D^\beta] + R(D)D^\beta\bigr)X_t = Q(D)DL_t, \qquad t \in \mathbb{R}. \tag{4.7}$$

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 [3], 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 (3.3) becomes

$$X_t - X_s = -\frac{\kappa}{\Gamma(1-\beta)} \int_{-\infty}^t \bigl[(t-u)^{-\beta} - (s-u)_+^{-\beta}\bigr]X_u \, du + L_t - L_s, \qquad s < t, \tag{4.8}$$

or, in short,

$$dX_t = -\kappa D^\beta X_t \, dt + dL_t, \qquad t \in \mathbb{R}. \tag{4.9}$$

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 $\operatorname{Re}(z) \geq 0$. 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

$$\bigl(r\cos(\theta) + \kappa r^\beta\cos(\beta\theta)\bigr) + i\bigl(r\sin(\theta) + \kappa r^\beta\sin(\beta\theta)\bigr) \neq 0. \tag{4.10}$$

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

$$\gamma_X(t) = \int_{\mathbb{R}} \frac{e^{ity}}{y^2 + 2\kappa\sin(\frac{\beta\pi}{2})|y|^{1+\beta} + \kappa^2|y|^{2\beta}} \, dy, \qquad t \in \mathbb{R}. \tag{4.11}$$

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

$$\gamma_\beta(t) = \int_{\mathbb{R}} \frac{e^{ity}}{|y|^{2(1+\beta)} + \kappa^2 |y|^{2\beta}} \, \mathrm{d}y, \qquad t \in \mathbb{R}. \tag{4.12}$$

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).
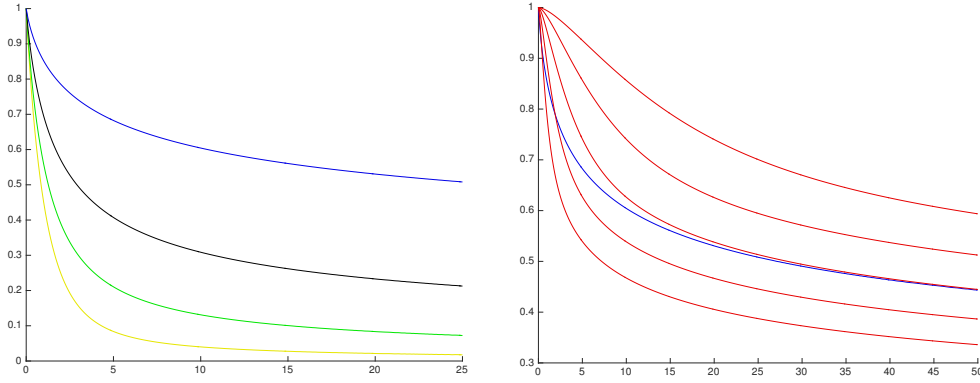
Figure 1: The left plot is the ACF based on (4.11) with $\beta = 0.1$ (yellow), $\beta = 0.2$ (green), $\beta = 0.3$ (black) and $\beta = 0.4$ (blue). With $\beta = 0.4$ fixed, the plot on the right compares the ACF based on (4.11) with $\kappa = 1$ (blue) to the ACF based on (4.12) for $\kappa = 0.125, 0.25, 0.5, 1, 2$ (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 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

$$\mathrm{d}X_t = \frac{\kappa_2}{\Gamma(1-\beta)} \int_0^\infty X_{t-u}\Big(\kappa_1 e^{-\kappa_1 u} \int_0^u e^{\kappa_1 v} v^{-\beta} \, \mathrm{d}v - u^{-\beta}\Big) \mathrm{d}u \, \mathrm{d}t + \mathrm{d}L_t, \quad t \in \mathbb{R}, \tag{4.13}$$

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). With $\gamma_1 = \cos(\beta\pi/2)$ and $\gamma_2 = \sin(\beta\pi/2)$ the autocovariance function γ_X is given by

$$\gamma_X(t) = \int_{\mathbb{R}} e^{ity} \frac{y^2 + \kappa_1^2}{y^4 + 2\kappa_2(\kappa_1\gamma_2|y|^{1+\beta} - \gamma_1|y|^{2+\beta}) + \kappa_1^2 y^2 + \kappa_2^2|y|^{2\beta}} \, dy, \quad t \in \mathbb{R}. \quad (4.15)$$

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

$$\gamma_\beta(t) = \int_{\mathbb{R}} e^{ity} \frac{y^2 + \kappa_1^2}{|y|^{4+2\beta} + (\kappa_1^2 - 2\kappa_2)|y|^{2+2\beta} + \kappa_2^2|y|^{2\beta}} \, dy, \qquad t \in \mathbb{R}. \quad (4.16)$$

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. From these plots we see that both the ACF corresponding to (4.15) and (4.16) are decreasing in κ_2, which is similar to the role of κ in Example 4.4. It appears as well that a larger κ_2 causes more curvature, although this effect is less pronounced for (4.15) than for (4.16).
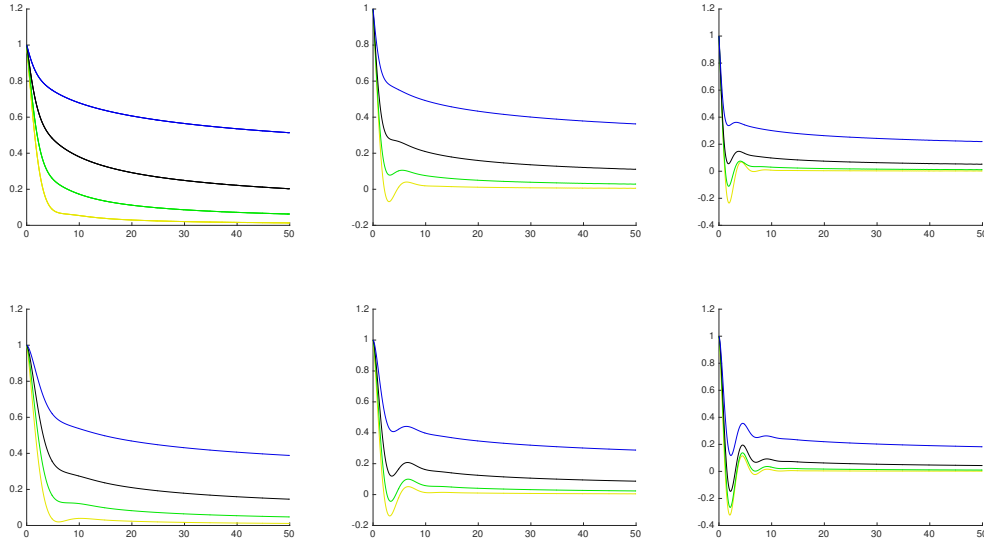
Figure 2: First row is ACF based on (4.15), second row is ACF based on (4.16), and the columns correspond to $\kappa_2 = 0.5$, $\kappa_2 = 1$ and $\kappa_2 = 2$, respectively. Within each plot, the lines correspond to $\beta = 0.1$ (yellow), $\beta = 0.2$ (green), $\beta = 0.3$ (black) and $\beta = 0.4$ (blue). In all plots, $\kappa_1 = 1$.

5 Simulation from the SFDDE

In the following we will focus on simulating from (3.3). 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

$$\int_{\mathbb{R}} \left[(n\Delta - u)_+^{-\beta} - ((n-1)\Delta - u)_+^{-\beta} \right] X_u \, du$$

$$\simeq \frac{1}{1-\beta} \Delta^{1-\beta} X_{(n-1)\Delta}$$

$$+ \sum_{k=n-c_1}^{n-1} \frac{X_{k\Delta} + X_{(k-1)\Delta}}{2} \int_{(k-1)\Delta}^{k\Delta} \left[(n\Delta - u)_+^{-\beta} - ((n-1)\Delta - u)_+^{-\beta} \right] du$$

$$= \frac{1}{1-\beta} \Delta^{1-\beta} X_{(n-1)\Delta} + \frac{1}{1-\beta} \sum_{k=n-c_1}^{n-1} \frac{X_{k\Delta} + X_{(k-1)\Delta}}{2}$$

$$\cdot \left(2((n-k-1)\Delta)^{1-\beta} - ((n-k)\Delta)^{1-\beta} - ((n-k-2)\Delta)^{1-\beta} \right)$$

for $n = -c_2 + 1, \ldots, 3700/\Delta - c_2 - c_1$. 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 $X_1, \ldots, X_{1600}$. This is repeated 200 times, and in every repetition the sample ACF based on $X_1, \ldots, X_L$ is computed for $t = 1, \ldots, 25$ 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., [23]). Due to this bias, it may be difficult to see if we succeed in simulating from (3.3), 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

$$\left[\bar{\rho}(k) - 1.96 \frac{\hat{\sigma}(k)}{\sqrt{200}}, \bar{\rho}(k) + 1.96 \frac{\hat{\sigma}(k)}{\sqrt{200}} \right],$$

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 results in 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 [23], and hence we may expect this to apply to solutions to (3.3) 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.
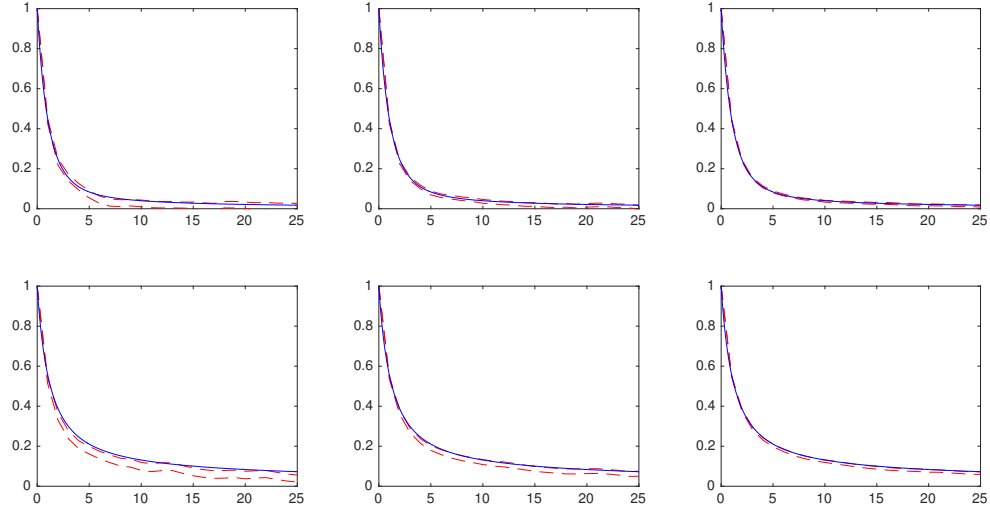
Figure 3: Theoretical ACF and 95 % confidence intervals of the mean of the sample ACF based on 200 replications of $X_1, \dots, X_L$. Columns correspond to $L = 100$, $L = 400$ and $L = 1600$, respectively, and rows correspond to $\beta = 0.1$ and $\beta = 0.2$, respectively. The model is (4.8).
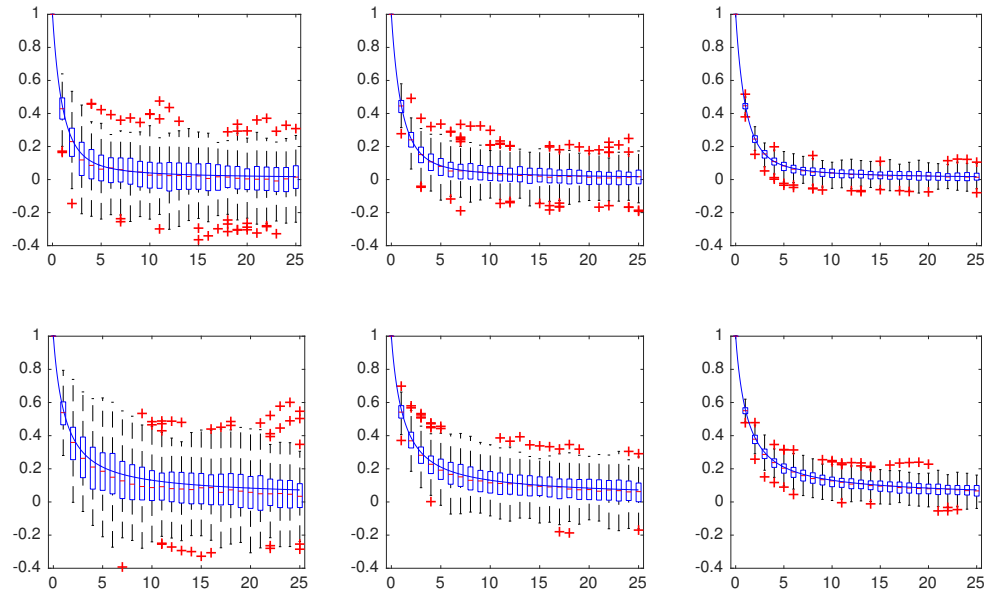
Figure 4: Box plots for the sample ACF based on 200 replications of $X_1, \dots, X_L$ together with the theoretical ACF. Columns correspond to $L = 100$, $L = 400$ and $L = 1600$, respectively, and rows correspond to $\beta = 0.1$ and $\beta = 0.2$, respectively. The model is (4.8).

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 (3.3), restricted to the case where $\eta(\mathrm{d}t) = -e^{-t}\,\mathrm{d}t$ and $(L_t)_{t\in\mathbb{R}}$ is a standard Brownian motion. In this case, we use the approximation

$$\int_{\mathbb{R}} \left[(n\Delta - u)_+^{-\beta} - ((n-1)\Delta - u)_+^{-\beta} \right] \int_0^\infty X_{u-v}\, e^{-v}\,\mathrm{d}v\,\mathrm{d}u$$

$$= \int_0^\infty X_{n\Delta - v} \int_0^v \left[(u-\Delta)_+^{-\beta} - u_+^{-\beta} \right] e^{u-v}\,\mathrm{d}u\,\mathrm{d}v$$

$$\simeq \frac{1}{2}\Delta X_{(n-1)\Delta} f(\Delta)$$

$$+ \sum_{k=2}^{c_1} \frac{1}{4}\Delta \big(X_{(n-k)\Delta} + X_{(n-k+1)\Delta} \big) \big(\varphi(k\Delta) + \varphi((k-1)\Delta) \big)$$

where $\varphi\colon \mathbb{R} \to \mathbb{R}$ is given by

$$\varphi(v) = \int_0^v \left[(u-\Delta)_+^{-\beta} - u^{-\beta} \right] e^{u-v}\,\mathrm{d}u.$$

We approximate φ recursively by noting that

$$\varphi(k\Delta) = \int_0^{k\Delta} \left[(u-\Delta)_+^{-\beta} - u^{-\beta} \right] e^{u-k\Delta}\,\mathrm{d}v$$

$$\simeq \frac{1+e^{-\Delta}}{2} \int_{(k-1)\Delta}^{k\Delta} \left[(u-\Delta)_+^{-\beta} - u_+^{-\beta} \right] \mathrm{d}v + e^{-\Delta}\varphi((k-1)\Delta)$$

$$= \frac{1}{1-\beta} \frac{1+e^{-\Delta}}{2} \left[((k-1)\Delta)^{1-\beta} - (k\Delta)^{1-\beta} \right] + e^{-\Delta}\varphi((k-1)\Delta)$$

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 the first example that we considered in the sense of convergence of the sample ACF and the effect of memory (the value of β).

6 Proofs

Proof of Proposition 3.1. For $\gamma > 0$ define $h_\gamma(z) = z^\gamma/h(z)$ for each $z \in \mathbb{C} \setminus \{0\}$ with $\mathrm{Re}(z) \geq 0$. 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

$$\sup_{x>0} \int_{\mathbb{R}} |h_\gamma(x+iy)|^2\,\mathrm{d}y < \infty \tag{6.1}$$

for $\gamma \in (-1/2, 1/2)$. In other words, h_γ is a certain Hardy function, and thus there exists a function $f_\gamma\colon \mathbb{R} \to \mathbb{R}$ 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 [2, 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 [24, 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,

$$g(t) = \frac{1}{\Gamma(\beta)} \int_0^t f_0(u)(t-u)^{\beta-1}\,\mathrm{d}u, \qquad t > 0.$$
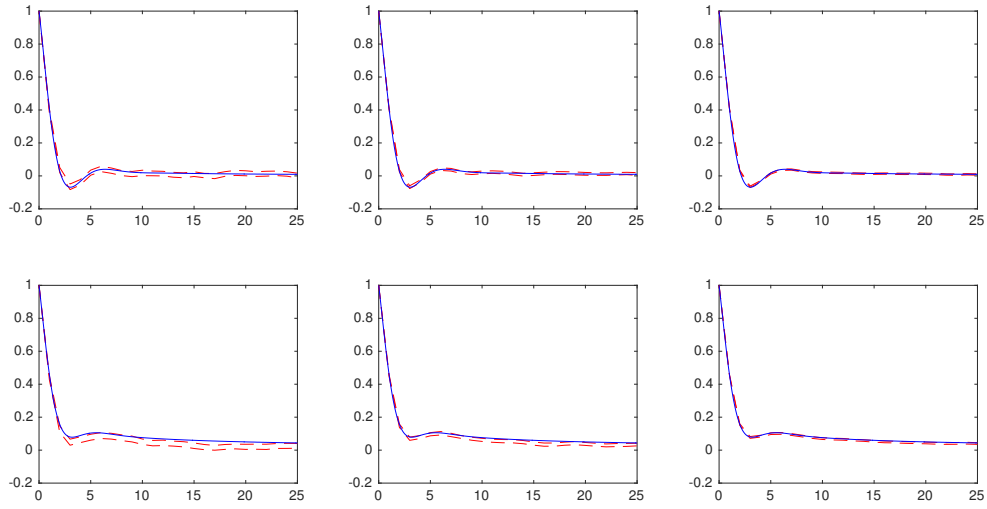
Figure 5: Theoretical ACF and 95 % confidence intervals of the mean of the sample ACF sample based on 200 replications of $X_1,\dots,X_L$. Columns correspond to $L = 100$, $L = 400$ and $L = 1600$, respectively, and rows correspond to $\beta = 0.1$ and $\beta = 0.2$, respectively. The model is (4.13).
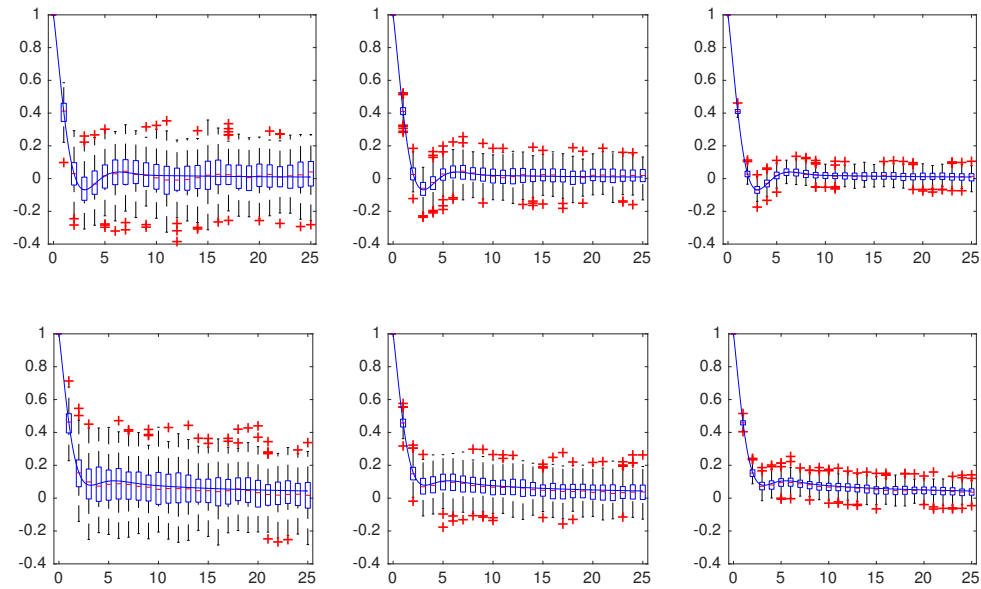
Figure 6: Box plots for the sample ACF based on 200 replications of $X_1,\dots,X_L$ together with the theoretical ACF. Columns correspond to $L = 100$, $L = 400$ and $L = 1600$, respectively, and rows correspond to $\beta = 0.1$ and $\beta = 0.2$, respectively. The model is (4.13).

Conversely, [24, 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 $x \geq 0$ (except at 0 when $x = 0$) and

$$
\begin{aligned}
\mathcal{L}[u \mapsto u f_0(u)](x + iy) &= i \frac{\mathrm{d}}{\mathrm{d}y} \mathcal{L}[f_0](x + iy) \\
&= \frac{\mathcal{L}[u \, \eta(\mathrm{d}u)](x + iy) + (1 - \beta)(x + iy)^{-\beta}}{h(x + iy)^2}.
\end{aligned}
\tag{6.2}
$$

The function $\mathcal{L}[u \mapsto u f_0(u)]$ is analytic on $\{z \in \mathbb{C} : \mathrm{Re}(z) < 0\}$ and from the identity (6.2) it is not too difficult to see that it also satisfies the Hardy condition (6.1). 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 $f_0 \in L^1 \cap L^2$, [3, 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

$$
\int_s^t \left(D^\beta g\right) * \eta(u - \, \cdot \,) \, \mathrm{d}u = \int_\mathbb{R} \left(D^\beta_- \mathbb{1}_{(s,t]}\right)(u) \, g * \eta(u - \, \cdot \,) \, \mathrm{d}u
$$

almost everywhere by comparing their Fourier transforms. This establishes the relation

$$
g(t - v) - g(s - v) = \int_s^t \left(D^\beta g\right) * \eta(u - v) \, \mathrm{d}u + \mathbb{1}_{(s,t]}(v).
$$

By letting $s \to -\infty$, and using that $D^\beta g$ and g are both vanishing on $(-\infty, 0)$, we deduce that

$$
g(t) = \mathbb{1}_{[0,\infty)}(t)\left(1 + \int_0^t \left(D^\beta g\right) * \eta(u) \, \mathrm{d}u\right),
$$

for almost all $t \in \mathbb{R}$ which shows (3.8) and, thus, finishes the proof. □

Proof of Theorem 3.2. Since $g \in L^2$, according to Proposition 3.1, and $\mathbb{E}[L_1^2] < \infty$ and $\mathbb{E}[L_1] = 0$,

$$
X_t = \int_{-\infty}^t g(t - u) \, \mathrm{d}L_u, \qquad t \in \mathbb{R},
$$

is a well-defined process (e.g., in the sense of [26]) 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

$$
X_t - X_s = \int_\mathbb{R} \left(\int_\mathbb{R} \left(D^\beta_- \mathbb{1}_{(s,t]}\right)(u) \, g * \eta(u - r) \, \mathrm{d}u\right) \mathrm{d}L_r + L_t - L_s.
$$

By a stochastic Fubini result (e.g., [1, Theorem 3.1]) we can change the order of integration (twice) and obtain

$$
\int_\mathbb{R} \left(\int_\mathbb{R} \left(D^\beta_- \mathbb{1}_{(s,t]}\right)(u) \, g * \eta(u - r) \, \mathrm{d}u\right) \mathrm{d}L_r = \int_\mathbb{R} \left(D^\beta_- \mathbb{1}_{(s,t]}\right)(u) \, X * \eta(u) \, \mathrm{d}u.
$$

This shows that $(X_t)_{t \in \mathbb{R}}$ is a solution to (3.3). To show uniqueness, note that the spectral process Λ_X (with spectral distribution, say, F_X) of any purely non-deterministic solution $(X_t)_{t \in \mathbb{R}}$ satisfies

$$
\int_\mathbb{R} \mathcal{F}[\mathbb{1}_{(s,t]}](-y)(iy)^\beta h(iy) \Lambda_X(\mathrm{d}y) = L_t - L_s
\tag{6.3}
$$

almost surely for all choices of $s < t$. This follows from the results in the supplementary material on spectral representations (see Section 7). 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 (6.3) 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

$$\int_{\mathbb{R}} \mathcal{F}[f](-y)(iy)^\beta h(iy) \Lambda_X(\mathrm{d}y) = \int_{\mathbb{R}} f(u)\, \mathrm{d}L_u \tag{6.4}$$

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 $\mathcal{F}[f](y) = e^{-ity}(-iy)^{-\beta}/h(-iy)$ and (6.4) thus implies that $X_t = \int_{-\infty}^{t} g(t - u)\, \mathrm{d}L_u$, which ends the proof. $\qquad\square$

Proof of Proposition 3.6. We start by arguing that the limit in (3.13) exists and is equal to $\int_{-\infty}^{t} D^\beta g(t - u)\, \mathrm{d}L_u$. For a given $\delta > 0$ it follows by a stochastic Fubini result that

$$\frac{\beta}{\Gamma(1 - \beta)} \int_{\delta}^{\infty} \frac{X_t - X_{t-u}}{u^{1+\beta}}\, \mathrm{d}u = \int_{\mathbb{R}} D_\delta^\beta g(t - r)\, \mathrm{d}L_r, \tag{6.5}$$

where

$$D_\delta^\beta g(t) = \frac{\beta}{\Gamma(1 - \beta)} \int_{\delta}^{\infty} \frac{g(t) - g(t - u)}{u^{1+\beta}}\, \mathrm{d}u, \qquad t > 0,$$

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 (3.3). Then, almost surely, $u \mapsto (X_t - X_{t-u})/u^{1+\beta}$ is in L^1 and the relation (6.5) thus shows that

$$\int_{\mathbb{R}} \left[D_\delta^\beta g(t - r) - D_{\delta'}^\beta g(t - r) \right] \mathrm{d}L_r \xrightarrow{\mathbb{P}} 0 \qquad \text{as } \delta, \delta' \to 0,$$

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 (6.5) and the isometry property of the integral map $\int_{\mathbb{R}} \cdot\, \mathrm{d}L$ that the limit in (3.13) exists and that $D^\beta X_t = \int_{-\infty}^{t} D^\beta g(t - u)\, \mathrm{d}L_u$. To show (3.14) we start by recalling the definition of $D_-^\beta \mathbb{1}_{(s,t]}$ in (3.2) 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 $\int_0^\infty e^{\pm iv} v^{\gamma-1}\, \mathrm{d}v$ is equal to $\Gamma(\gamma) e^{\pm i\pi\gamma/2}$ for any $\gamma \in (0, 1)$. Now observe that

$$\mathcal{F}\left[\int_{\mathbb{R}} (D_-^\beta \mathbb{1}_{(s,t]})(u)\, g * \eta(u - \cdot)\, \mathrm{d}u \right](y) = (-iy)^\beta \mathcal{F}[\mathbb{1}_{(s,t]}](y) \mathcal{F}[g](-y) \mathcal{F}[\eta](-y)$$

$$= \mathcal{F}[\mathbb{1}_{(s,t]}](y) \mathcal{F}\left[(D^\beta g) * \eta \right](-y)$$

$$= \mathcal{F}\left[\int_s^t (D^\beta g) * \eta(u - \cdot)\, \mathrm{d}u \right](y),$$

and hence $\int_{\mathbb{R}} (D_-^\beta \mathbb{1}_{(s,t]})(u)\, g * \eta(u - \cdot)\, \mathrm{d}u = \int_s^t (D^\beta g) * \eta(u - \cdot)\, \mathrm{d}u$ almost everywhere. Consequently, using that $D^\beta X_t = \int_{-\infty}^{t} D^\beta g(t - u)\, \mathrm{d}L_u$ and applying a stochastic Fubini

result twice,

$$
\begin{aligned}
\int_s^t (D^\beta X) * \eta(u) \, \mathrm{d}u &= \int_\mathbb{R} \int_s^t (D^\beta g) * \eta(u - r) \, \mathrm{d}u \, \mathrm{d}L_r \\
&= \int_\mathbb{R} \int_\mathbb{R} (D_-^\beta \mathbb{1}_{(s,t]})(u) \, g * \eta(u - r) \, \mathrm{d}u \, \mathrm{d}L_r \\
&= \frac{1}{\Gamma(1 - \beta)} \int_\mathbb{R} \left[(t - u)_+^{-\beta} - (s - u)_+^{-\beta} \right] X * \eta(u) \, \mathrm{d}u.
\end{aligned}
$$

The semimartingale property of $(X_t)_{t \in \mathbb{R}}$ is now an immediate consequence of (3.3). $\square$

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 $u, v \in \mathbb{R}$,

$$
\int_{u \vee v}^\infty (s - u)^{\beta - 1} (s - v)^{\beta - 1} \, \mathrm{d}s = \frac{\Gamma(\beta) \Gamma(1 - 2\beta)}{\Gamma(1 - \beta)} |u - v|^{2\beta - 1}
$$

by [16, p. 404]. Having this relation in mind we use Proposition 3.1(ii) and (3.15) to do the computations

$$
\begin{aligned}
\gamma_X(t) &= \frac{1}{\Gamma(\beta)^2} \int_\mathbb{R} \int_\mathbb{R} \int_\mathbb{R} D^\beta g(u) \, D^\beta g(v) (s + t - u)_+^{\beta - 1} (s - v)_+^{\beta - 1} \, \mathrm{d}v \, \mathrm{d}u \, \mathrm{d}s \\
&= \frac{1}{\Gamma(\beta)^2} \int_\mathbb{R} \int_\mathbb{R} D^\beta g(u) \, D^\beta g(v) \\
&\quad \cdot \int_{(u-t) \vee v}^\infty (s - (u - t))^{\beta - 1} (s - v)^{\beta - 1} \, \mathrm{d}s \, \mathrm{d}v \, \mathrm{d}u \\
&= \frac{\Gamma(1 - 2\beta)}{\Gamma(\beta) \Gamma(1 - \beta)} \int_\mathbb{R} \int_\mathbb{R} D^\beta g(u) \, D^\beta g(v) |u - v - t|^{2\beta - 1} \, \mathrm{d}v \, \mathrm{d}u \\
&= \frac{\Gamma(1 - 2\beta)}{\Gamma(\beta) \Gamma(1 - \beta)} \int_\mathbb{R} \gamma(u) |u - t|^{2\beta - 1} \, \mathrm{d}u,
\end{aligned}
\tag{6.6}
$$

where $\gamma(u) = \int_\mathbb{R} D^\beta g(u + v) \, D^\beta g(v) \, \mathrm{d}v$. Note that $\gamma \in L^1$ since $D^\beta g \in L^1$ by Proposition 3.1 and, using Plancherel's theorem,

$$
\gamma(u) = \int_\mathbb{R} e^{-iuy} |\mathcal{F}[D^\beta g](y)|^2 \, \mathrm{d}y = \mathcal{F}[|h(i \cdot)|^{-2}](u).
$$

In particular $\int_\mathbb{R} \gamma(u) \, \mathrm{d}u = |h(0)|^{-2} = \eta([0, \infty))^{-2}$, and hence it follows from (6.6) that we have shown the result if we can argue that

$$
\frac{\int_\mathbb{R} \gamma(u) |u - t|^{2\beta - 1} \, \mathrm{d}u}{t^{2\beta - 1}} = \int_\mathbb{R} \frac{\gamma(u)}{|\frac{u}{t} - 1|^{1 - 2\beta}} \, \mathrm{d}u \to \int_\mathbb{R} \gamma(u) \, \mathrm{d}u \qquad \text{as } t \to \infty.
\tag{6.7}
$$

It is clear by Lebesgue's theorem on dominated convergence that

$$
\int_{-\infty}^0 \frac{\gamma(u)}{|\frac{u}{t} - 1|^{1 - 2\beta}} \, \mathrm{d}u \to \int_{-\infty}^0 \gamma(u) \, \mathrm{d}u \qquad \text{as } t \to \infty.
$$

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, $\gamma(u) = \mathcal{F}[\phi](u)/(iu)$ and, thus,

$$\int_{t/2}^{\infty} \frac{\gamma(u)}{|\frac{u}{t} - 1|^{1-2\beta}} \, du = \int_{1/2}^{\infty} \frac{t\gamma(tu)}{|u - 1|^{1-2\beta}} \, du = -i \int_{1/2}^{\infty} \frac{\mathcal{F}[\phi](tu)}{u|u - 1|^{1-2\beta}} \, du. \qquad (6.8)$$

By the Riemann–Lebesgue lemma and Lebesgue's theorem on dominated convergence it follows that the right-hand side of expression in (6.8) tends to zero as t tends to infinity. Finally, integration by parts and the symmetry of γ yields

$$\int_0^{t/2} \gamma(u)\Big(1 - \frac{1}{|\frac{u}{t} - 1|^{1-2\beta}}\Big) \, du = \int_0^{1/2} t\gamma(tu)\Big(1 - \frac{1}{(1-u)^{1-2\beta}}\Big) \, du$$

$$= \big(2^{1-2\beta} - 1\big) \int_{-\infty}^{-t/2} \gamma(u) \, du$$

$$- \int_0^{1/2} \frac{1 - 2\beta}{(1-u)^{2-2\beta}} \int_{-\infty}^{-tu} \gamma(v) \, dv \, du,$$

where both terms on the right-hand side converge to zero as t tends to infinity. Thus, we have shown (6.7), and this completes the 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 $\int_{\mathbb{R}} \cdot \, d\Lambda_X \colon L^2(F_X) \to L^2(\mathbb{P})$, see [15, p. 389], we have that

$$\frac{\mathbb{E}[(X_t - X_0)^2]}{t} = t^{-2} \int_{\mathbb{R}} |1 - e^{iy}|^2 f_X(y/t) \, dy$$

$$= \int_{\mathbb{R}} \frac{|1 - e^{iy}|^2}{|y|^{2\beta} |(iy)^{1-\beta} - t^{1-\beta}\mathcal{F}[\eta](y/t)|^2} \, dy. \qquad (6.9)$$

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

$$\frac{|1 - e^{iy}|^2}{|y|^{2\beta} |(iy)^{1-\beta} - t^{1-\beta}\mathcal{F}[\eta](y/t)|^2} \leq 2\frac{|1 - e^{iy}|^2}{y^2}.$$

If $|y| < C_1 t$, we note that the assumption on the function in (3.5) implies that

$$C_2 := \inf_{|x| \leq C_1} \big|(ix)^{1-\beta} - \mathcal{F}[\eta](x)\big| > 0,$$

which shows that

$$\big|(iy)^{1-\beta} - t^{1-\beta}\mathcal{F}[\eta](y/t)\big| \geq t^{1-\beta}C_2 \geq \frac{C_2}{C_1^{1-\beta}}|y|^{1-\beta}.$$

This establishes that

$$\frac{|1 - e^{iy}|^2}{|y|^{2\beta}\big|(iy)^{1-\beta} - t^{1-\beta}\mathcal{F}[\eta](y/t)\big|^2} \leq \frac{C_1^{2(1-\beta)}}{C_2^2} \frac{|1 - e^{iy}|^2}{y^2}.$$

Consequently, it follows from (6.9) and Lebesgue's theorem on dominated convergence that

$$\frac{\mathbb{E}[(X_t - X_0)^2]}{t} \to \int_{\mathbb{R}} \frac{|1 - e^{iy}|^2}{y^2} \, dy = \int_{\mathbb{R}} |\mathcal{F}[\mathbb{1}_{(0,1]}](y)|^2 \, dy = 1 \qquad \text{as } t \downarrow 0,$$

which was to be shown. □

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

$$
\mathbb{E}\Big[\int_0^{t-s}\int_{-\infty}^s |X_w|\int_{[0,\infty)} \big|(D_-^\beta \mathbb{1}_{(s,t-u]})(v+w)\big|\,|\eta|(\mathrm{d}v)\,\mathrm{d}w\,|g|(\mathrm{d}u)\Big]
$$
$$
\leq \mathbb{E}[|X_0|]\int_0^{t-s}\int_{[0,\infty)}\int_{-\infty}^s \big|(D_-^\beta \mathbb{1}_{(s,t-u]})(v+w)\big|\,\mathrm{d}w\,|\eta|(\mathrm{d}v)\,|g|(\mathrm{d}u)
\tag{6.10}
$$

is finite. This is implied by the facts that

$$
\Gamma(1-\beta)\int_{-\infty}^s \big|(D_-^\beta \mathbb{1}_{(s,t-u]})(v+w)\big|\,\mathrm{d}w
$$
$$
\leq \int_{u+s-t}^0 (t-s-u+w)^{-\beta}\,\mathrm{d}w + \int_0^1 \big[w^{-\beta}-(t-s-u+w)^{-\beta}\big]\,\mathrm{d}w
$$
$$
\quad + (1+\beta)\int_1^\infty w^{-1-\beta}(t-s-u)\,\mathrm{d}w
$$
$$
= \frac{1}{1-\beta}\big(2(t-s-u)^{1-\beta}+1-(t-s-u+1)^{1-\beta}\big)+\frac{(1+\beta)}{\beta}(t-s-u)
$$
$$
\leq \frac{2}{1-\beta}(t-s)^{1-\beta}+\frac{(1+\beta)}{\beta}(t-s)
$$

for $u \in [0, t-s]$ and $g(\mathrm{d}u)$ 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 (3.3) that

$$
\mathcal{L}[X\mathbb{1}_{(s,\infty)}](z) = X_s\mathcal{L}[\mathbb{1}_{(s,\infty)}](z) + \mathcal{L}[\mathbb{1}_{(s,\infty)}(L_{\cdot}-L_s)](z)
$$
$$
\quad + \mathcal{L}\Big[\mathbb{1}_{(s,\infty)}\int_{\mathbb{R}} X_u\int_{[0,\infty)}\big(D_-^\beta \mathbb{1}_{(s,\cdot]}\big)(u+v)\,\eta(\mathrm{d}v)\,\mathrm{d}u\Big](z).
\tag{6.11}
$$

By noting that $(D_-^\beta \mathbb{1}_{(s,t]})(u) = 0$ when $t \leq s < u$ we obtain

$$
\mathcal{L}\Big[\mathbb{1}_{(s,\infty)}\int_s^\infty X_u\int_{[0,\infty)}\big(D_-^\beta \mathbb{1}_{(s,\cdot]}\big)(u+v)\,\eta(\mathrm{d}v)\,\mathrm{d}u\Big](z)
$$
$$
= \frac{1}{\Gamma(1-\beta)}\mathcal{L}\Big[\int_s^\infty X_u\int_{[0,\infty)}(\cdot-u-v)_+^{-\beta}\,\eta(\mathrm{d}v)\,\mathrm{d}u\Big](z)
$$
$$
= \mathcal{L}[\mathbb{1}_{(s,\infty)}X](z)\mathcal{L}[\eta](z)z^{\beta-1}.
$$

Combining this observation with (6.11) we get the relation

$$
\big(z-z^\beta\mathcal{L}[\eta](z)\big)\mathcal{L}[\mathbb{1}_{(s,\infty)}X](z)
$$
$$
= zX_s\mathcal{L}[\mathbb{1}_{(s,\infty)}](z) + z\mathcal{L}[\mathbb{1}_{(s,\infty)}(L-L_s)](z)
$$
$$
\quad + z\mathcal{L}\Big[\mathbb{1}_{(s,\infty)}\int_{-\infty}^s X_u\int_{[0,\infty)}\big(D_-^\beta \mathbb{1}_{(s,\cdot]}\big)(u+v)\,\eta(\mathrm{d}v)\,\mathrm{d}u\Big](z),
$$

which implies

$$
\begin{aligned}
\mathcal{L}[&\mathbb{1}_{(s,\infty)}X](z) \\
&= \mathcal{L}[g](z)\mathcal{L}[X_s\delta_0(s-\,\cdot\,)](z) + z\mathcal{L}[g](z)\mathcal{L}[\mathbb{1}_{(s,\infty)}(L-L_s)](z) \\
&\quad + z\mathcal{L}[g](z)\mathcal{L}\Big[\mathbb{1}_{(s,\infty)}\int_{-\infty}^{s} X_u \int_{[0,\infty)} \big(D_-^\beta\mathbb{1}_{(s,\,\cdot\,]}\big)(u+v)\,\eta(\mathrm{d}v)\,\mathrm{d}u\Big](z) \\
&= \mathcal{L}[g(\,\cdot\,-s)X_s](z) + \mathcal{L}\Big[\int_s^{\,\cdot} g(\,\cdot\,-u)\,\mathrm{d}L_u\Big](z) \\
&\quad + \mathcal{L}\Big[\int_0^{\,\cdot\,-s}\int_{-\infty}^{s} X_w \int_{[0,\infty)} \big(D_-^\beta\mathbb{1}_{(s,\,\cdot\,-u]}\big)(v+w)\,\eta(\mathrm{d}v)\,\mathrm{d}w\,g(\mathrm{d}u)\Big](z).
\end{aligned}
$$

This establishes the identity

$$
\begin{aligned}
X_t = {}& g(t-s)X_s + \int_s^t g(t-u)\,\mathrm{d}L_u \\
& + \int_0^{t-s}\int_{-\infty}^{s} X_w \int_{[0,\infty)} \big(D_-^\beta\mathbb{1}_{(s,t-u]}\big)(v+w)\,\eta(\mathrm{d}v)\,\mathrm{d}w\,g(\mathrm{d}u)
\end{aligned}
\tag{6.12}
$$

almost surely for Lebesgue almost all $t > s$. Since both sides of (6.12) 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 $\mathbb{E}[\,\cdot\mid X_u, u \leq s]$ on both sides of (6.12) we obtain the result. $\square$

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 $\{z \in \mathbb{C} : \mathrm{Re}(z) \geq 0\}$) if and only if

$$
Q(z)[z^{1-\beta} + \kappa] + R(z) \neq 0 \qquad \text{for all } z \in \mathbb{C} \text{ with } \mathrm{Re}(z) \geq 0.
\tag{6.13}
$$

Condition (6.13) 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 $h(0) = \kappa + b^\top A^{-1}e_1 \neq 0$, 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

$$
\mathcal{F}[g](y) = \frac{(iy)^{-\beta}}{(iy)^{1-\beta} + \kappa + \frac{R(iy)}{Q(iy)}} = \frac{Q(iy)}{Q(iy)[iy + \kappa(iy)^\beta] + R(iy)(iy)^\beta}, \qquad y \in \mathbb{R}.
$$

This finishes the proof. $\square$

Proof of Proposition 4.2. We will first show that $D^\beta f \in L^1$. By using that $\int_0^\infty e^{Au}\,\mathrm{d}u = -A^{-1}$ we can rewrite $D^\beta f$ as

$$
D^\beta f(t) = \frac{1}{\Gamma(1-\beta)}b^\top A\Big(\int_0^t e^{Au}\big[(t-u)^{-\beta} - t^{-\beta}\big]\,\mathrm{d}u - \int_t^\infty e^{Au}t^{-\beta}\,\mathrm{d}u\Big)e_1, \quad t > 0,
$$

from which we see that it suffices to argue that (each entry of)

$$
t \longmapsto \int_0^t e^{Au}\big[(t-u)^{-\beta} - t^{-\beta}\big]\,\mathrm{d}u
$$

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 $\gamma > 0$,

$$\int_0^\infty \int_0^t e^{-\gamma u} \big|(t-u)^{-\beta} - t^{-\beta}\big| \, \mathrm{d}u \, \mathrm{d}t$$
$$\leq \int_0^\infty e^{-\gamma u} \Big(\int_u^{u+1} \big[(t-u)^{-\beta} + t^{-\beta}\big] \, \mathrm{d}t + \beta u \int_1^\infty t^{-\beta-1} \, \mathrm{d}t \Big) \, \mathrm{d}u < \infty.$$

Here we have used the mean value theorem to establish the inequality

$$\big|(t-u)^{-\beta} - t^{-\beta}\big| \leq \beta u (t-u)^{-\beta-1}$$

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,

$$D^\beta f(t) = \frac{1}{\Gamma(1-\beta)} \frac{\mathrm{d}}{\mathrm{d}t} \int_0^t f(t-u) u^{-\beta} \, \mathrm{d}u, \qquad t > 0.$$

Consequently, it follows by [27, Theorem 7.1] that the Fourier transform $\mathcal{F}[D^\beta f]$ of f is given by

$$\mathcal{F}[D^\beta f](y) = (iy)^\beta \mathcal{F}[f](y) = (iy)^\beta b^\top (iy - A)^{-1} e_1, \qquad y \in \mathbb{R},$$

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

$$\int_0^\infty D^\beta X_{t-u} f(u) \, \mathrm{d}u = \int_{\mathbb{R}} (D^\beta g) * f(t-r) \, \mathrm{d}L_r = \int_0^\infty X_{t-u} D^\beta f(u) \, \mathrm{d}u$$

using Proposition 3.6 and a stochastic Fubini result. This finishes the proof. $\qquad\square$

7 Supplement to "Stochastic differential equations with a fractionally filtered delay: a semimartingale model for long-range dependent processes"

This supplement provides an exposition of the spectral representation and related results for continuous-time stationary, measurable, centered and square integrable processes. The content of the results should be well-known and is mainly provided for reference.

7.1 Spectral representations of continuous-time stationary processes

In the following we present and prove a few results related to the spectral theory for continuous-time stationary, measurable, centered and square integrable processes. Although the results should be well-known, we have not been able to find an appropriate reference to earlier literature. However, the results presented here rely heavily on [15, Section 9.4] and [20, Appendix A2.1], in which an extensive treatment of the spectral theory is given.

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 $u \leq v \leq s \leq t$,

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

$$G((s,t]) = \mathbb{E}[|S(t) - S(s)|^2], \qquad s < t.$$

We have the following stochastic Fubini result for this type of integral:

Proposition 7.1. *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*

$$\int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(x,y)\,\mu(\mathrm{d}x) \right) S(\mathrm{d}y) = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(x,y)\,S(\mathrm{d}y) \right) \mu(\mathrm{d}x) \qquad (7.1)$$

almost surely.

Suppose that $(X_t)_{t \in \mathbb{R}}$ is a measurable and stationary process with $\mathbb{E}[X_0^2] < \infty$ and $\mathbb{E}[X_0] = 0$, and denote by γ_X its autocovariance function. Since $(X_t)_{t \in \mathbb{R}}$ is continuous in $L^2(\mathbb{P})$ (cf. [1, Corollary A.3]), it follows by Bochner's theorem that there exists a finite Borel measure F_X on $\mathbb{R}$ such that

$$\gamma_X(t) = \int_{\mathbb{R}} e^{ity}\,F_X(\mathrm{d}y), \qquad t \in \mathbb{R}.$$

The measure F_X is referred to as the spectral distribution of $(X_t)_{t \in \mathbb{R}}$.

Theorem 7.2. *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*

$$X_t = \int_{\mathbb{R}} e^{ity}\,\Lambda_X(\mathrm{d}y) \qquad (7.2)$$

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 (7.2) is referred to as its spectral representation.

Remark 7.3. Let the situation be as in Theorem 7.2 and note that if there exists another process $\tilde{\Lambda}_X = \{\tilde{\Lambda}_X(y) : y \in \mathbb{R}\}$ such that

$$X_t = \int_{\mathbb{R}} e^{ity}\,\tilde{\Lambda}_X(\mathrm{d}y), \qquad t \in \mathbb{R},$$

then its control measure is necessarily given by F_X and

$$\int_{\mathbb{R}} f(y)\,\Lambda_X(\mathrm{d}y) = \int_{\mathbb{R}} f(y)\,\tilde{\Lambda}_X(\mathrm{d}y)$$

almost surely for all $f \in L^2(F_X)$.

Proof of Proposition 7.1. First, note that (7.1) is trivially true when f is of the form

$$f(x,y) = \sum_{j=1}^{n} \alpha_j \mathbb{1}_{A_j}(x)\mathbb{1}_{B_j}(y) \tag{7.3}$$

for $\alpha_1,\ldots,\alpha_n \in \mathbb{C}$ and Borel sets $A_1,B_1,\ldots,A_n,B_n \subseteq \mathbb{R}$. 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 (7.3) such that $f_n \to f$ in $L^2(\mu \times G)$ as $n \to \infty$. Set

$$X_n = \int_{\mathbb{R}}\left(\int_{\mathbb{R}} f_n(x,y)\,\mu(\mathrm{d}x)\right)S(\mathrm{d}y), \quad X = \int_{\mathbb{R}}\left(\int_{\mathbb{R}} f(x,y)\,\mu(\mathrm{d}x)\right)S(\mathrm{d}y)$$

$$\text{and} \quad Y = \int_{\mathbb{R}}\left(\int_{\mathbb{R}} f(x,y)\,S(\mathrm{d}y)\right)\mu(\mathrm{d}x)$$

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,

$$\int_{\mathbb{R}}\left|\int_{\mathbb{R}} f(x,y)\,\mu(\mathrm{d}x)\right|^2 G(\mathrm{d}y) \le \mu(\mathbb{R})\int_{\mathbb{R}^2} |f(x,y)|^2\,(\mu \times G)(\mathrm{d}x,\mathrm{d}y) < \infty$$

$$\text{and} \quad \mathbb{E}\left[\int_{\mathbb{R}}\left|\int_{\mathbb{R}} f(x,y)\,S(\mathrm{d}y)\right|^2 \mu(\mathrm{d}x)\right] = \int_{\mathbb{R}^2} |f(x,y)|^2\,(\mu \times G)(\mathrm{d}x,\mathrm{d}y) < \infty.$$

Next, we find that

$$\mathbb{E}[|X - X_n|^2] = \int_{\mathbb{R}}\left|\int_{\mathbb{R}} (f(x,y) - f_n(x,y))\,\mu(\mathrm{d}x)\right|^2 G(\mathrm{d}y)$$

$$\le \mu(\mathbb{R})\int_{\mathbb{R}^2} |f(x,y) - f_n(x,y)|^2\,(\mu \times G)(\mathrm{d}x,\mathrm{d}y)$$

which tends to zero by the choice of $(f_n)_{n\in\mathbb{N}}$. Since $X_n = \int_{\mathbb{R}}\left(\int_{\mathbb{R}} f_n(x,y)\,S(\mathrm{d}y)\right)\mu(\mathrm{d}x)$, one shows in a similar way that $X_n \to Y$ in $L^2(\mathbb{P})$, and hence we conclude that $X = Y$ almost surely. $\qquad\square$

Proof of Theorem 7.2. 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

$$\langle X_s, X_t\rangle_{L^2(\mathbb{P})} = \mathbb{E}[X_s X_t] = \int_{\mathbb{R}} e^{i(t-s)x} F_X(\mathrm{d}y) = \langle f_s, f_t\rangle_{L^2(F_X)}, \qquad s, t \in \mathbb{R},$$

we can define a linear isometric isomorphism $\mu\colon \overline{H}_F \to \overline{H}_X$ as the one satisfying

$$\mu\left(\sum_{j=1}^{n} \alpha_j f_{t_j}\right) = \sum_{j=1}^{n} \alpha_j X_{t_j}$$

for any given $n \in \mathbb{N}$, $\alpha_1,\ldots,\alpha_n \in \mathbb{C}$ and $t_1 < \cdots < t_n$. Since $\mathbb{1}_{(-\infty,y]} \in \overline{H}_F$ for each $y \in \mathbb{R}$ (cf. [32, 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

$$\Lambda_X(y) = \mu(\mathbb{1}_{(-\infty,y]}), \qquad y \in \mathbb{R}.$$

133

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

$$\mathbb{E}[|\Lambda_X(y_2) - \Lambda_X(y_1)|^2] = F_X((y_1, y_2]), \qquad y_1 < y_2.$$

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 $n \in \mathbb{N}$, $\alpha_1, \ldots, \alpha_n \in \mathbb{C}$ and $t_0 < t_1 < \cdots < t_n$, we have

$$\int_{\mathbb{R}} \left(\sum_{j=1}^{n} \alpha_j \mathbb{1}_{(t_{j-1}, t_j]}(y) \right) \Lambda_X(\mathrm{d}y) = \sum_{j=1}^{n} \alpha_j \mu(\mathbb{1}_{(t_{j-1}, t_j]}) = \mu\left(\sum_{j=1}^{n} \alpha_j \mathbb{1}_{(t_{j-1}, t_j]} \right).$$

Since $f \mapsto \int_{\mathbb{R}} f(y) \Lambda_X(\mathrm{d}y)$ 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

$$\int_{\mathbb{R}} f(y) \Lambda_X(\mathrm{d}y) = \mu(f)$$

almost surely for any $f \in \overline{H}_F$. In particular, it shows that

$$X_t = \mu(f_t) = \int_{\mathbb{R}} e^{ity} \Lambda_X(\mathrm{d}y), \qquad t \in \mathbb{R},$$

which is the spectral representation of $(X_t)_{t \in \mathbb{R}}$. □

Acknowledgments

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 (2019). Stochastic delay differential equations and related autoregressive models. *Stochastics.* Forthcoming. DOI: 10.1080/17442508.2019.1635601.

[3] 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.

[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] Jusselin, P. and M. Rosenbaum (2018). No-arbitrage implies power-law market impact and rough volatility. arXiv: 1805.07134.

[20] Koopmans, L. H. (1995). *The spectral analysis of time series*. Academic Press.

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

[22] 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.

[23] Newbold, P. and C. Agiakloglou (1993). Bias in the sample autocorrelations of fractional noise. *Biometrika* 80(3), 698–702.

[24] 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.

[25] Pipiras, V. and M.S. Taqqu (2017). *Long-range dependence and self-similarity*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.

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

[27] Samko, S.G., A.A. Kilbas, O.I. Marichev, et al. (1993). Fractional integrals and derivatives. *Theory and Applications*, *Gordon and Breach*, *Yverdon* 1993.

[28] Samorodnitsky, G. (2016). *Stochastic processes and long range dependence*. Vol. 26. Springer.

[29] Samorodnitsky, G. et al. (2007). Long range dependence. *Foundations and Trends® in Stochastic Systems* 1(3), 163–257.

[30] Samorodnitsky, G. and M.S. Taqqu (1994). *Stable Non-Gaussian Random Processes*. Stochastic Modeling. Stochastic models with infinite variance. New York: Chapman & Hall.

[31] 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.

[32] Yaglom, A.M (1987). *Correlation theory of stationary and related random functions. Vol. I*. Springer Series in Statistics. Basic results. New York: Springer-Verlag.

Limit Theorems for Quadratic Forms and Related Quantities of Discretely Sampled Continuous-Time Moving Averages

Mikkel Slot Nielsen and Jan Pedersen

Abstract

The limiting behavior of Toeplitz type quadratic forms of stationary processes has received much attention through decades, particularly due to its importance in statistical estimation of the spectrum. In the present paper we study such quantities in the case where the stationary process is a discretely sampled continuous-time moving average driven by a Lévy process. We obtain sufficient conditions, in terms of the kernel of the moving average and the coefficients of the quadratic form, ensuring that the centered and adequately normalized version of the quadratic form converges weakly to a Gaussian limit.

1 Introduction

Let $(Y_t)_{t\in\mathbb{Z}}$ be a stationary sequence of random variables with $\mathbb{E}[Y_0] = 0$ and $\mathbb{E}[Y_0^2] < \infty$, and suppose that $(Y_t)_{t\in\mathbb{Z}}$ is characterized by a parameter θ which we, for simplicity, assume to be an element of $\mathbb{R}$. If one wants to infer the true value θ_0 of θ from a sample $Y(n) = [Y_1,\dots,Y_n]^\top$, a typical estimator is obtained as

$$\hat{\theta}_n = \arg\min_\theta \ell_n(\theta),$$

where $\ell_n = \ell_n(\,\cdot\,; Y(n))$ is a suitable objective function. On an informal level, the usual strategy for showing asymptotic normality of the estimator $\hat{\theta}_n$ is to use a Taylor series

expansion to write

$$\frac{\ell'_n(\theta_0)}{\sqrt{n}} = -\frac{\ell''_n(\theta^*_n)}{n}\sqrt{n}(\hat{\theta}_n - \theta_0),$$

and then show that $\ell''_n(\theta^*_n)/n$ converges in probability to a non-zero constant and $\ell'_n(\theta_0)/\sqrt{n}$ converges in distribution to a centered Gaussian random variable. Here ℓ'_n and ℓ''_n refer to the first and second order derivative of ℓ_n with respect to θ, respectively, and θ^*_n is a point in the interval formed by $\hat{\theta}_n$ and θ_0. While the convergence of $\ell''_n(\theta^*_n)/n$ usually can be shown by an ergodic theorem under the assumptions of consistency of $\hat{\theta}_n$ and ergodicity of $(Y_t)_{t\in\mathbb{Z}}$, showing the desired convergence of $\ell'_n(\theta_0)/\sqrt{n}$ may be much more challenging. In particular, if the quantity $\ell'_n(\theta_0)$ corresponds to a rather complicated function of $Y(n)$, one often needs to impose restrictive assumptions on the dependence structure of $(Y_t)_{t\in\mathbb{Z}}$, e.g., rapidly decaying mixing coefficients. In addition to the concern that such type of mixing conditions do not hold in the presence of long memory, they may generally be difficult to verify.

When ℓ_n has an explicit form, one can sometimes exploit the particular structure to prove asymptotic normality of $\ell'_n(\theta_0)/\sqrt{n}$. To be concrete, let $\gamma_Y(\,\cdot\,;\theta)$ denote the autocovariance function of $(Y_t)_{t\in\mathbb{Z}}$ and $\Sigma_n(\theta) = [\gamma_Y(j-k;\theta)]_{j,k=1,\ldots,n}$ the covariance matrix of $Y(n)$. A very popular choice of ℓ_n is the (scaled) negative Gaussian log-likelihood,

$$\ell_n(\theta) = \log\det(\Sigma_n(\theta)) + Y(n)^\top \Sigma_n(\theta)^{-1} Y(n). \tag{1.1}$$

In order to avoid the cumbersome and, in the presence of long memory, unstable computations related to the inversion of $\Sigma_n(\theta)$, one sometimes instead uses Whittle's approximation of (1.1), which is given by

$$\begin{aligned}
\ell_{n,Whittle}(\theta) &= \frac{n}{2\pi}\int_{-\pi}^{\pi} \log(2\pi f_Y(y;\theta))\,\mathrm{d}y + Y(n)^\top A_n(\theta) Y(n) \\
&= \frac{n}{2\pi}\int_{-\pi}^{\pi}\left[\log(2\pi f_Y(y;\theta))\,\mathrm{d}y + \frac{I_Y(y)}{2\pi f_Y(y;\theta)}\right]\mathrm{d}y,
\end{aligned} \tag{1.2}$$

where $f_Y(\,\cdot\,;\theta)$ is the spectral density of Y, I_Y is the periodogram of Y and

$$A_n(\theta) = \left[\frac{1}{(2\pi)^2}\int_{-\pi}^{\pi} e^{i(j-k)y}\frac{1}{f_Y(y;\theta)}\,\mathrm{d}y\right]_{j,k=1,\ldots,n}.$$

(For details about the relation between the Gaussian likelihood and Whittle's approximation, and for some justification for their use, see [4, 16, 22].) An important feature of both (1.1) and (1.2) is that, under suitable assumptions on $\gamma_Y(\,\cdot\,;\theta)$ and $f_Y(\,\cdot\,;\theta)$, the quantities $\ell'_n(\theta_0)/\sqrt{n}$ and $\ell'_{n,Whittle}(\theta_0)/\sqrt{n}$ are of the form $(Q_n - \mathbb{E}[Q_n])/\sqrt{n}$, where

$$Q_n = \sum_{t,s=1}^{n} b(t-s)Y_t Y_s \tag{1.3}$$

and $b: \mathbb{Z} \to \mathbb{R}$ is an even function. Consequently, proving asymptotic normality of $\ell'_n(\theta_0)/\sqrt{n}$ and $\ell'_{n,Whittle}(\theta_0)/\sqrt{n}$ reduces to determining for which processes $(Y_t)_{t\in\mathbb{Z}}$ and functions b, $(Q_n - \mathbb{E}[Q_n])/\sqrt{n}$ converges in distribution to a centered Gaussian random variable. In the case where $(Y_t)_{t\in\mathbb{Z}}$ is Gaussian and $b(t) = \int_{-\pi}^{\pi} e^{ity}\hat{b}(y)\,\mathrm{d}y$, the papers [1, 14] give conditions on $\hat{b}$ and the spectral density of $(Y_t)_{t\in\mathbb{Z}}$ ensuring that such weak convergence holds. Moreover, Fox and Taqqu [13] proved non-central limit

theorems for (an adequately normalized version of) (1.3) in case $Y_t = H(X_t)$ where H is a Hermite polynomial and $(X_t)_{t\in\mathbb{Z}}$ is a normalized Gaussian sequence with a slowly decaying autocovariance function. In particular, they showed that the limit can be both Gaussian and non-Gaussian depending on the decay rate of the autocovariances. Later, Giraitis and Surgailis [15] left the Gaussian framework and considered instead general linear processes of the form

$$Y_t = \sum_{s\in\mathbb{Z}} \varphi_{t-s}\varepsilon_s, \qquad t \in \mathbb{Z}, \tag{1.4}$$

where $(\varepsilon_t)_{t\in\mathbb{Z}}$ is an i.i.d. sequence with $\mathbb{E}[\varepsilon_0] = 0$ and $\mathbb{E}[\varepsilon_0^4] < \infty$, and $\sum_{t\in\mathbb{Z}} \varphi_t^2 < \infty$. They provided sufficient conditions (in terms of b and the autocovariance function of $(Y_t)_{t\in\mathbb{Z}}$) ensuring that $(Q_n - \mathbb{E}[Q_n])/\sqrt{n}$ tends to a Gaussian limit. Many interesting processes are given by (1.4), the short-memory ARMA processes and the long-memory ARFIMA processes being the main examples, and their properties have been studied extensively. The literature on these processes is overwhelming, and the following references form only a small sample: [7, 11, 16, 18].

The continuous-time analogue of (1.4) is the moving average process $(X_t)_{t\in\mathbb{R}}$ given by

$$X_t = \int_{\mathbb{R}} \varphi(t-s)\,\mathrm{d}L_s, \qquad t \in \mathbb{R}, \tag{1.5}$$

where $(L_t)_{t\in\mathbb{R}}$ is a two-sided Lévy process with $\mathbb{E}[L_1] = 0$ and $\mathbb{E}[L_1^4] < \infty$, and where $\varphi\colon \mathbb{R} \to \mathbb{R}$ is a function in L^2. Among popular and well-studied continuous-time moving averages are the CARMA processes, particularly the Ornstein–Uhlenbeck process, and solutions to linear stochastic delay differential equations (see [6, 9, 17, 19] for more on these processes). Bai et al. [2] considered a continuous-time version of (1.3), where sums are replaced by integrals and $(Y_t)_{t\in\mathbb{Z}}$ by $(X_t)_{t\in\mathbb{R}}$ defined in (1.5), and they obtained conditions on b and φ ensuring both a Gaussian and non-Gaussian limit for (a suitably normalized version of) the quadratic form.

Our main contribution is Theorem 1.1, which gives sufficient conditions on φ and b ensuring that $(Q_n - \mathbb{E}[Q_n])/\sqrt{n}$ converges in distribution to a centered Gaussian random variable when $Y_t = X_{t\Delta}$, $t \in \mathbb{Z}$, for some fixed $\Delta > 0$. In the formulation we denote by κ_4 the fourth cumulant of L_1 and by γ_X the autocovariance function of $(X_t)_{t\in\mathbb{R}}$ (see the formula in (3.3)).

Theorem 1.1. *Let $(X_t)_{t\in\mathbb{R}}$ be given by (1.5) and define Q_n as in (1.3) with $Y_t = X_{t\Delta}$ for some $\Delta > 0$. Suppose that one of the following statements holds:*

(i) *There exist $\alpha,\beta \in [1,2]$ with $2/\alpha + 1/\beta \geq 5/2$, such that $\sum_{t\in\mathbb{Z}} |b(t)|^\beta < \infty$ and*

$$\left(t \longmapsto \sum_{s\in\mathbb{Z}} |\varphi(t+s\Delta)|^\kappa\right) \in L^{4/\kappa}([0,\Delta]) \qquad \text{for } \kappa = \alpha, 2.$$

(ii) *The function φ belongs to L^4 and there exist $\alpha,\beta > 0$ with $\alpha + \beta < 1/2$, such that*

$$\sup_{t\in\mathbb{R}} |t|^{1-\alpha/2}|\varphi(t)| < \infty \quad \text{and} \quad \sup_{t\in\mathbb{Z}} |t|^{1-\beta}|b(t)| < \infty.$$

Then, as $n \to \infty$, $(Q_n - \mathbb{E}[Q_n])/\sqrt{n}$ tends to a Gaussian random variable with mean zero and variance

$$\eta^2 = \kappa_4 \int_0^\Delta \left(\sum_{s \in \mathbb{Z}} \varphi(t + s\Delta) \sum_{u \in \mathbb{Z}} b(u)\varphi(t + (s + u)\Delta) \right)^2 dt$$

$$+ 2 \sum_{s \in \mathbb{Z}} \left(\sum_{u \in \mathbb{Z}} b(u)\gamma_X((s + u)\Delta) \right)^2.$$

While the statement in (i) is more general than the statement in (ii) of Theorem 1.1, the latter provides an easy-to-check condition in terms of the decay of φ and b at infinity. This decay condition is mild enough to apply to many interesting choices of $(X_t)_{t \in \mathbb{R}}$, including some situations where long memory is present (see, e.g., Example 3.11). Theorem 1.1 relies on an approximation of Q_n by a quantity of the type

$$S_n = \sum_{t=1}^n X_{t\Delta}^1 X_{t\Delta}^2, \tag{1.6}$$

where $(X_t^1)_{t \in \mathbb{R}}$ and $(X_t^2)_{t \in \mathbb{R}}$ are moving averages of the form (1.5), and a limit theorem for $(S_n - \mathbb{E}[S_n])/\sqrt{n}$. This idea is borrowed from [15]. Although we can use the same overall idea, $(X_{t\Delta})_{t \in \mathbb{Z}}$ is generally not of the form (1.4) and, due to the interplay between the continuous-time specification (1.5) and the discrete-time (low frequency) sampling scheme, the spectral density and related quantities become less tractable. The conditions of Theorem 1.1 are similar to the rather general results of [2], which concerned the continuous-time version of (1.3). A reason that we obtain conditions of the same type as [2] is that our proofs, too, rely on (various modifications of) Young's inequality for convolutions. Since the setup of that paper requires a continuum of observations of $(X_t)_{t \in \mathbb{R}}$, those results cannot be applied in our case.

In addition to its purpose as a tool in the proof of Theorem 1.1, a limit theorem for $(S_n - \mathbb{E}[S_n])/\sqrt{n}$ is of independent interest, e.g., since it is of the same form as the (scaled) sample autocovariance of (1.5) and of $\ell_n'(\theta_0)/\sqrt{n}$ when ℓ_n is a suitable least squares objective function (see Examples 3.3 and 3.4 for details). For this reason, we present our limit theorem for $(S_n - \mathbb{E}[S_n])/\sqrt{n}$ here:

Theorem 1.2. *Let $(X_t^1)_{t \in \mathbb{R}}$ and $(X_t^2)_{t \in \mathbb{R}}$ be as in (1.5) with corresponding kernels $\varphi_1, \varphi_2 \in L^2$ and define S_n by (1.6). Suppose that one of the following statements holds:*

(i) *There exist $\alpha_1, \alpha_2 \in [1, 2]$ with $1/\alpha_1 + 1/\alpha_2 \geq 3/2$, such that*

$$\left(t \longmapsto \sum_{s \in \mathbb{Z}} \left(|\varphi_i(t + s\Delta)|^{\alpha_i} + \varphi_i(t + s\Delta)^2 \right) \right) \in L^2([0, \Delta]) \qquad \text{for } i = 1, 2.$$

(ii) *The functions φ_1 and φ_2 belong to L^4 and there exist $\alpha_1, \alpha_2 \in (1/2, 1)$ with $\alpha_1 + \alpha_2 > 3/2$, such that*

$$\sup_{t \in \mathbb{R}} |t|^{\alpha_i} |\varphi_i(t)| < \infty \qquad \text{for } i = 1, 2.$$

Then, as $n \to \infty$, $(S_n - \mathbb{E}[S_n])/\sqrt{n}$ tends to a Gaussian random variable with mean zero and variance

$$\eta^2 = \kappa_4 \int_0^\Delta \left(\sum_{s \in \mathbb{Z}} \varphi_1(t + s\Delta)\varphi_2(t + s\Delta) \right)^2 \mathrm{d}t + \mathbb{E}[L_1^2]^2 \sum_{s \in \mathbb{Z}} \left(\int_{\mathbb{R}} \varphi_1(t)\varphi_1(t + s\Delta) \, \mathrm{d}t \right.$$

$$\left. \cdot \int_{\mathbb{R}} \varphi_2(t)\varphi_2(t + s\Delta) \, \mathrm{d}t + \int_{\mathbb{R}} \varphi_1(t)\varphi_2(t + s\Delta) \, \mathrm{d}t \int_{\mathbb{R}} \varphi_2(t)\varphi_1(t + s\Delta) \, \mathrm{d}t \right).$$

As was the case in Theorem 1.1, statement (i) is more general than statement (ii) of Theorem 1.2, but the latter may be convenient as it gives conditions on the decay rate of φ_1 and φ_2 at infinity. In relation to Theorem 1.2, it should be mentioned that limit theorems for the sample autocovariances of moving average processes (1.5) have been studied in [5, 10, 25].

The paper is organized as follows: Section 2 recalls the most relevant concepts in relation to Lévy processes and the corresponding integration theory. Section 3 presents Theorems 3.1 and 3.5, which are our most general central limit theorems for S_n and Q_n, and from which we will deduce Theorems 1.1 and 1.2 as special cases. Moreover, Section 3 provides examples demonstrating that the imposed conditions on φ (or φ_1 and φ_2) are satisfied for CARMA processes, solutions to stochastic delay equations and certain fractional (Lévy) noise processes. Finally, Section 4 contains proofs of all the statements of the paper together with a few supporting results.

2 Preliminaries

In this section we introduce some notation that will be used repeatedly and we recall a few concepts related to Lévy processes and integration of deterministic functions with respect to them. For a detailed exposition of Lévy processes and the corresponding integration theory, see [23, 24].

For a given measurable function $f \colon \mathbb{R} \to \mathbb{R}$ and $p \geq 1$ we write $f \in L^p$ if $|f|^p$ is integrable with respect to the Lebesgue measure and $f \in L^\infty$ if f is bounded almost everywhere. For a given function $a \colon \mathbb{Z} \to \mathbb{R}$ (or sequence $(a(t))_{t \in \mathbb{Z}}$) we write $a \in \ell^p$ if $\|a\|_{\ell^p} := (\sum_{t \in \mathbb{Z}} |a(t)|^p)^{1/p} < \infty$ and $a \in \ell^\infty$ if $\|a\|_{\ell^\infty} := \sup_{t \in \mathbb{Z}} |a(t)| < \infty$.

A stochastic process $(L_t)_{t \geq 0}$, $L_0 = 0$, is called a one-sided Lévy process if it is càdlàg and has stationary and independent increments. The distribution of $(L_t)_{t \geq 0}$ is characterized by L_1 as a consequence of the relation $\log \mathbb{E}[\exp\{iyL_t\}] = t \log \mathbb{E}[\exp\{iyL_1\}]$. By the Lévy–Khintchine representation it holds that

$$\log \mathbb{E}\big[e^{iyL_1}\big] = iy\gamma - \frac{1}{2}\rho^2 y^2 + \int_{\mathbb{R}} (e^{iyx} - 1 - iyx \mathbb{1}_{\{|x| \leq 1\}}) \, \nu(\mathrm{d}x), \qquad y \in \mathbb{R},$$

for some $\gamma \in \mathbb{R}$, $\rho^2 \geq 0$ and Lévy measure ν, and hence (the distribution of) $(L_t)_{t \geq 0}$ may be summarized as a triplet (γ, ρ^2, ν). The same holds for a (two-sided) Lévy process $(L_t)_{t \in \mathbb{R}}$ which is constructed as $L_t = L_t^1 \mathbb{1}_{t \geq 0} - L_{(-t)-}^2 \mathbb{1}_{t < 0}$, where $(L_t^1)_{t \geq 0}$ and $(L_t^2)_{t \geq 0}$ are one-sided Lévy processes which are independent copies.

Let $(L_t)_{t \in \mathbb{R}}$ be a Lévy process with $\mathbb{E}[|L_1|] < \infty$ and $\mathbb{E}[L_1] = 0$. Then, for a given measurable function $f \colon \mathbb{R} \to \mathbb{R}$, the integral $\int_{\mathbb{R}} f(t) \, \mathrm{d}L_t$ is well-defined (as a limit in probability of integrals of simple functions) and belongs to $L^p(\mathbb{P})$, $p \geq 1$, if

$$\int_{\mathbb{R}} \int_{\mathbb{R}} \big(|f(t)x|^p \wedge |f(t)x|^2\big) \, \nu(\mathrm{d}x) \, \mathrm{d}t < \infty. \tag{2.1}$$

In particular, (2.1) is satisfied if $f \in L^2 \cap L^p$ and $\int_{|x|>1} |x|^p \, \nu(dx) < \infty$, the latter condition being equivalent to $\mathbb{E}[|L_1|^p] < \infty$. Finally, when (2.1) holds for $p = 2$ we will often make use of the isometry property of the integral map:

$$\mathbb{E}\left[\left(\int_{\mathbb{R}} f(t) \, dL_t\right)^2\right] = \mathbb{E}[L_1^2] \int_{\mathbb{R}} f(t)^2 \, dt.$$

3 Further results and examples

As in the introduction, it will be assumed throughout that $(L_t)_{t \in \mathbb{R}}$ is a two-sided Lévy process with $\mathbb{E}[L_1] = 0$ and $\mathbb{E}[L_1^4] < \infty$. Set $\sigma^2 = \mathbb{E}[L_1^2]$ and $\kappa_4 = \mathbb{E}[L_1^4] - 3\sigma^4$. Moreover, for functions $\varphi, \varphi_1, \varphi_2 : \mathbb{R} \to \mathbb{R}$ in L^2 define

$$X_t = \int_{\mathbb{R}} \varphi(t-s) \, dL_s \quad \text{and} \quad X_t^i = \int_{\mathbb{R}} \varphi_i(t-s) \, dL_s \tag{3.1}$$

for $t \in \mathbb{R}$ and $i = 1, 2$. We will be interested in the quantities

$$S_n = \sum_{t=1}^{n} X_{t\Delta}^1 X_{t\Delta}^2 \quad \text{and} \quad Q_n = \sum_{t,s=1}^{n} b(t-s) X_{t\Delta} X_{s\Delta} \tag{3.2}$$

for a given $\Delta > 0$ and an even function $b \colon \mathbb{Z} \to \mathbb{R}$. Our main results, Theorems 3.1 and 3.5, provide a central limit theorem for the quantities in (3.2) and are more general than Theorems 1.1 and 1.2 which were presented in Section 1. Before the formulations we define the autocovariance function of $(X_t)_{t \in \mathbb{R}}$,

$$\gamma_X(h) = \mathbb{E}[X_0 X_h] = \sigma^2 \int_{\mathbb{R}} \varphi(t)\varphi(t+h) \, dt, \qquad h \in \mathbb{R}, \tag{3.3}$$

as well as the autocovariance (crosscovariance) functions of $(X_t^1)_{t \in \mathbb{R}}$ and $(X_t^2)_{t \in \mathbb{R}}$,

$$\gamma_{ij}(h) = \mathbb{E}[X_0^i X_h^j] = \sigma^2 \int_{\mathbb{R}} \varphi_i(t)\varphi_j(t+h) \, dt, \qquad h \in \mathbb{R}. \tag{3.4}$$

Theorem 3.1. *Suppose that the following conditions hold:*

(i) $\int_{\mathbb{R}} |\varphi_i(t)\varphi_i(t + \cdot \Delta)| \, dt \in \ell^{\alpha_i}$ *for* $i = 1, 2$ *and* $\alpha_1, \alpha_2 \in [1, \infty]$ *with* $1/\alpha_1 + 1/\alpha_2 = 1$.

(ii) $\int_{\mathbb{R}} |\varphi_1(t)\varphi_2(t + \cdot \Delta)| \, dt \in \ell^2$.

(iii) $\left(t \longmapsto \kappa_4 \|\varphi_1(t + \cdot \Delta)\varphi_2(t + \cdot \Delta)\|_{\ell^1}\right) \in L^2([0, \Delta])$.

Then, as $n \to \infty$, $(S_n - \mathbb{E}[S_n])/\sqrt{n}$ *tends to a Gaussian random variable with mean zero and variance*

$$\eta^2 = \kappa_4 \int_0^{\Delta} \left(\sum_{s \in \mathbb{Z}} \varphi_1(t + s\Delta)\varphi_2(t + s\Delta)\right)^2 dt + \sum_{s \in \mathbb{Z}} \gamma_{11}(s\Delta)\gamma_{22}(s\Delta)$$
$$+ \sum_{s \in \mathbb{Z}} \gamma_{12}(s\Delta)\gamma_{21}(s\Delta). \tag{3.5}$$

Remark 3.2. If $\kappa_4 = 0$, equivalently $(L_t)_{t\in\mathbb{R}}$ is a Brownian motion, assumption (iii) of Theorem 3.1 is trivially satisfied and the first term in the variance formula (3.5) vanishes.

Loosely speaking, assumptions (i)–(ii) of Theorem 3.1 concern summability of continuous-time convolutions. Hence, by relying on a modification of Young's convolution inequality, Theorem 1.2 can be shown to be a special case of Theorem 3.1 (see Lemma 4.3 and the following proof of Theorem 1.2 in Section 4). Examples 3.3 and 3.4 are possible applications of Theorem 3.1.

Example 3.3. Let $n, m \in \mathbb{N}$ with $m < n-1$, define the sample autocovariance of $(X_t)_{t\in\mathbb{R}}$ based on $X_\Delta, X_{2\Delta}, \ldots, X_{n\Delta}$ up to lag m as

$$\hat{\gamma}_n(j) = n^{-1} \sum_{t=1}^{n-j} X_{t\Delta} X_{(t+j)\Delta}, \qquad j = 1,\ldots,m, \tag{3.6}$$

and set $\hat{\gamma}_n = [\hat{\gamma}_n(1),\ldots,\hat{\gamma}_n(m)]^\top$. Moreover, let $\tilde{\varphi}(t) = [\varphi(t+\Delta),\ldots,\varphi(t+m\Delta)]^\top$ and $\gamma_s = [\gamma_X((s+1)\Delta),\ldots,\gamma_X((s+m)\Delta)]^\top$ using the notation as in (3.1) and (3.3). Then, for a given $\alpha = [\alpha_1,\ldots,\alpha_m]^\top \in \mathbb{R}^m$, it holds that

$$\alpha^\top \hat{\gamma}_n - \alpha^\top \gamma_0 = n^{-1} \sum_{t=1}^{n} \left(X_{t\Delta}^1 X_{t\Delta}^2 - \mathbb{E}[X_0^1 X_0^2] \right) + O_p(n^{-1}), \tag{3.7}$$

where $(X_t^1)_{t\in\mathbb{R}}$ and $(X_t^2)_{t\in\mathbb{R}}$ are given by (3.1) with $\varphi_1 = \varphi$ and $\varphi_2(t) = \alpha^\top \tilde{\varphi}(t)$. Here $O_p(n^{-1})$ in (3.7) means that the equality holds up to a term ε_n which is stochastically bounded by n^{-1} (that is, $(n\varepsilon_n)_{n\in\mathbb{N}}$ is tight). Then if

$$\int_{\mathbb{R}} |\varphi(t)\varphi(t + \cdot \Delta)| \, dt \in \ell^2 \quad \text{and} \quad \left(t \longmapsto \|\varphi(t + \cdot \Delta)\|_{\ell^2}^2 \right) \in L^2([0,\Delta]), \tag{3.8}$$

assumptions (i)–(iii) of Theorem 3.1 hold and we deduce that $\alpha^\top \sqrt{n}(\hat{\gamma}_n - \gamma_0)$ converges in distribution to a centered Gaussian random variable with variance $\alpha^\top \Sigma \alpha$, where

$$\Sigma = \kappa_4 \int_0^\Delta K(t)K(t)^\top \, dt + \sum_{s\in\mathbb{Z}} (\gamma_s + \gamma_{-s})\gamma_s^\top, \quad K(t) := \sum_{s\in\mathbb{Z}} \varphi(t + s\Delta)\tilde{\varphi}(t + s\Delta).$$

By the Cramér–Wold theorem we conclude that $\sqrt{n}(\hat{\gamma}_n - \gamma_0)$ converges in distribution to a centered Gaussian vector with covariance matrix Σ. This type of central limit theorem for the sample autocovariances of continuous-time moving averages was established in [10] under the same assumptions on φ as imposed above.

Example 3.4. Motivated by the discussion in the introduction, this example will illustrate how Theorem 3.1 can be applied to show asymptotic normality of the (adequately normalized) derivative of a least squares objective function. Fix $k \in \mathbb{N}$, let $v \colon \mathbb{R} \to \mathbb{R}^k$ be a differentiable function with derivative v' and consider

$$\ell_n(\theta) = \sum_{t=k+1}^{n} \left(X_{t\Delta} - v(\theta)^\top X(t) \right)^2, \qquad \theta \in \mathbb{R}, \tag{3.9}$$

where $X(t) = [X_{(t-1)\Delta}, \ldots, X_{(t-k)\Delta}]^\top$. In this case

$$\ell_n'(\theta) = -2 \sum_{t=k+1}^{n} \left(X_{t\Delta} - v(\theta)^\top X(t) \right) v'(\theta)^\top X(t), \qquad \theta \in \mathbb{R},$$

and hence it is of the same form as S_n in (3.2) with $\varphi_1(t) = [-1, v(\theta)^\top] \tilde{\varphi}(t)$ and $\varphi_2(t) = [0, 2v'(\theta)^\top] \tilde{\varphi}(t)$, where $\tilde{\varphi}(t) = [\varphi(t), \varphi(t-\Delta), \ldots, \varphi(t-k\Delta)]^\top$. Suppose that $v(\theta_0)$ coincides with the vector of coefficients of the $L^2(\mathbb{P})$ projection of $X_{(k+1)\Delta}$ onto the linear span of $X_{k\Delta}, \ldots, X_\Delta$ for some $\theta_0 \in \mathbb{R}$. In this case $\mathbb{E}[\ell_n'(\theta_0)] = 0$, and if (3.8) holds it thus follows from Theorem 3.1 that $\ell_n'(\theta_0)/\sqrt{n}$ converges in distribution to a centered Gaussian random variable.

Theorem 3.5 is our most general result concerning the limiting behavior of $(Q_n - \mathbb{E}[Q_n])/\sqrt{n}$ as $n \to \infty$. For notational convenience we will, for given $a \colon \mathbb{Z} \to \mathbb{R}$ and $f \colon \mathbb{R} \to \mathbb{R}$, set

$$(a \star f)(t) := \sum_{s \in \mathbb{Z}} a(s) f(t - s\Delta) \tag{3.10}$$

for any $t \in \mathbb{R}$, such that $\sum_{s \in \mathbb{Z}} |a(s) f(t - s\Delta)| < \infty$. If a and f are non-negative, the definition in (3.10) is used for all $t \in \mathbb{R}$. Moreover, we write $|a|(t) = |a(t)|$ and $|f|(t) = |f(t)|$.

Theorem 3.5. *Suppose that the following statements hold:*

(i) *There exist $\alpha, \beta \in [1, \infty]$ with $1/\alpha + 1/\beta = 1$, such that $\int_{\mathbb{R}} |\varphi(t) \varphi(t + \cdot \Delta)| \, dt \in \ell^\alpha$ and $\int_{\mathbb{R}} (|b| \star |\varphi|)(t)(|b| \star |\varphi|)(t + \cdot \Delta) \, dt \in \ell^\beta$.*

(ii) *$\int_{\mathbb{R}} |\varphi(t)| (|b| \star |\varphi|)(t + \cdot \Delta) \, dt \in \ell^2$.*

(iii) *$\left(t \longmapsto \kappa_4 \|\varphi(t + \cdot \Delta)(|b| \star |\varphi|)(t + \cdot \Delta)\|_{\ell^1} \right) \in L^2([0, \Delta])$.*

Then, as $n \to \infty$, $(Q_n - \mathbb{E}[Q_n])/\sqrt{n}$ converges in distribution to a Gaussian random variable with mean zero and variance

$$\eta^2 = \kappa_4 \int_0^\Delta \left(\sum_{s \in \mathbb{Z}} \varphi(t + s\Delta)(b \star \varphi)(t + s\Delta) \right)^2 dt + 2\|(b \star \gamma_X)(\cdot \Delta)\|_{\ell^2}^2. \tag{3.11}$$

Remark 3.6. The idea in the proof of Theorem 3.5 is to approximate Q_n by S_n with $\varphi_1 = \varphi$ and $\varphi_2 = b \star \varphi$. The conditions imposed in Theorem 3.5 correspond to assuming that φ and $|b| \star |\varphi|$ satisfy (i)–(iii) of Theorem 3.1. In particular, these conditions ensure that S_n is well-defined and that Theorem 3.1 applies to this choice of φ_1 and φ_2. The only lacking part in order to deduce Theorem 3.5 from Theorem 3.1 is to show that S_n is in fact a proper approximation of Q_n in the sense that $\text{Var}(Q_n - S_n)/n \to 0$ as $n \to \infty$, but this is verified in Section 4 where the proofs of the stated results can be found.

Remark 3.7. Note that for any $s \in \mathbb{Z}$ with $b(s) \neq 0$, it holds that

$$|\varphi(t)| \leq |b(s)|^{-1} (|b| \star |\varphi|)(t + s\Delta) \qquad \text{for all } t \in \mathbb{R}. \tag{3.12}$$

This fact ensures that assumptions (i)–(ii) of Theorem 3.5 hold if there exists $\beta \in [1, 2]$ such that

$$\int_{\mathbb{R}} (|b| \star |\varphi|)(t)(|b| \star |\varphi|)(t + \cdot \Delta)\, dt \in \ell^\beta. \tag{3.13}$$

(Here we exclude the trivial case $b \equiv 0$.) Indeed, if (3.13) is satisfied we can choose $\alpha \geq \beta$ such that $1/\alpha + 1/\beta = 1$ and then assumptions (i)–(ii) are met due to the inequality in (3.12) and the fact that $\ell^\beta \subseteq \ell^\alpha \cap \ell^2$.

Remark 3.8. We will now briefly comment on the conditions of Theorems 1.1 and 1.2, particularly on sufficient conditions for applying Theorems 3.1 and 3.5. We will restrict our attention to assumptions of the type

$$\left(t \longmapsto \|\psi(t + \cdot \Delta)\|_{\ell^\kappa}^\kappa\right) \in L^2([0, \Delta]), \tag{3.14}$$

where $\psi \colon \mathbb{R} \to \mathbb{R}$ is a measurable function and $\kappa \geq 1$. First of all, note that the weaker condition $(t \mapsto \|\psi(t + \cdot \Delta)\|_{\ell^\kappa}^\kappa) \in L^1([0, \Delta])$ is satisfied if and only if $\psi \in L^\kappa$, and condition (3.14) implies $\psi \in L^{2\kappa}$. In particular, a necessary condition for (3.14) to hold is that $\psi \in L^\kappa \cap L^{2\kappa}$. On the other hand, one may decompose $\|\psi(t + \cdot \Delta)\|_{\ell^\kappa}^\kappa$ as

$$\|\psi(t + \cdot \Delta)\|_{\ell^\kappa}^\kappa = \sum_{s=-M}^{M} |\psi(t + s\Delta)|^\kappa + \sum_{s=M+1}^{\infty} \left(|\psi(t + s\Delta)|^\kappa + |\psi(t - s\Delta)|^\kappa\right) \tag{3.15}$$

for any $M \in \mathbb{N}$. The first term on right-hand side of (3.15) belongs to $L^2([0, \Delta])$ (viewed as a function of t) if $\psi \in L^{2\kappa}$. If in addition $\psi \in L^\kappa$, the second term on the right-hand tends to zero as $M \to \infty$ for (Lebesgue almost) all $t \in [0, \Delta]$. If this could be assumed to hold uniformly across all t, that is, if the second term belongs to $L^\infty([0, \Delta])$ for a sufficiently large M, then (3.14) would be satisfied. Therefore, loosely speaking, the difference between $L^\kappa \cap L^{2\kappa}$ and the space of functions satisfying (3.14) consists of functions ψ where the second term in (3.15) tends to zero pointwise, but not uniformly, in t as $M \to \infty$. Ultimately, this is a condition on the behavior of the tail of the function between grid points. For instance, if there exists a sequence $(\psi_s)_{s \in \mathbb{Z}}$ in ℓ^κ such that $\sup_{t \in [0, \Delta]} |\psi(t \pm s\Delta)| \leq \psi_s$ for all sufficiently large s, then (3.14) holds. An assumption such as (3.14) seems to be necessary and is the cost of considering a continuous-time process only on a discrete-time grid. In [10], where they prove a central limit theorem for the sample autocovariance of a continuous-time moving average in a low frequency setting, a similar condition is imposed.

In the following examples we will put some attention on concrete specifications of moving average processes, where the behavior of the corresponding kernel is known, and hence Theorems 1.1 and 1.2 may be applicable.

Example 3.9. Fix $p \in \mathbb{N}$ and let $P(z) = z^p + a_1 z^{p-1} + \cdots + a_p$ and $Q(z) = b_0 + b_1 z + \cdots + b^{p-1} z^{p-1}$, $z \in \mathbb{C}$, be two real polynomials where all the zeroes of P are contained in $\{z \in \mathbb{C} : \mathrm{Re}(z) < 0\}$. Moreover, let $q \in \mathbb{N}_0$ with $q < p$ and suppose that $b_q = 1$ and $b_k = 0$ for $q < k \leq p - 1$. Finally, define

$$A = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ -a_p & -a_{p-1} & -a_{p-2} & \cdots & -a_1 \end{bmatrix}, \quad b = \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_{p-2} \\ b_{p-1} \end{bmatrix} \quad \text{and} \quad e_p = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}.$$

Then the corresponding (causal) CARMA(p,q) process $(X_t)_{t\in\mathbb{R}}$ is given by

$$X_t = \int_{-\infty}^{t} b^\top e^{A(t-u)} e_p \, dL_u, \quad t \in \mathbb{R}. \tag{3.16}$$

(See [21, Remark 3.2].) The definition in (3.16) is based on a state-space representation of the more intuitive formal differential equation

$$P(D)X_t = Q(D)DL_t, \qquad t \in \mathbb{R}, \tag{3.17}$$

where D denotes differentiation with respect to time. Equation (3.17) should be compared to the corresponding representation of an ARMA process in terms of the backward-shift operator. Since it can be shown that the eigenvalues of A correspond to the roots of P, the kernel $\varphi\colon t \mapsto \mathbb{1}_{[0,\infty)}(t)b^\top e^{At}e_p$ is exponentially decaying at infinity. Combining this with the (absolute) continuity of φ on $[0,\infty)$ ensures that the kernel belongs to L^∞ as well. In particular, this shows that Theorem 1.1(i) holds as long as $b \in \ell^2$. For more on CARMA processes, we refer to [6, 8, 9].

Example 3.10. Let η be a finite signed measure on $[0,\infty)$ and suppose that

$$z - \int_{[0,\infty)} e^{-zt}\eta(dt) \neq 0$$

for every $z \in \mathbb{C}$ with $\mathrm{Re}(z) \geq 0$. Then it follows from [3, Theorem 3.4] that the unique stationary solution $(X_t)_{t\in\mathbb{R}}$ to the corresponding stochastic delay differential equation

$$dX_t = \int_{[0,\infty)} X_{t-s}\,\eta(ds)\,dt + dL_t, \qquad t \in \mathbb{R},$$

takes the form $X_t = \int_{-\infty}^{t} \varphi(t-s)\,dL_s$, where $\varphi\colon \mathbb{R}\to\mathbb{R}$ is characterized as the unique L^2 function satisfying $\varphi(t) = 0$ for $t < 0$ and

$$\varphi(t) = 1 + \int_0^t \int_{[0,\infty)} \varphi(s-u)\,\eta(du)\,ds, \qquad t \geq 0.$$

Consequently, it follows form the integration by parts formula that

$$\begin{aligned}
\sup_{t\geq 0} t^p|\varphi(t)| \leq {} & p\int_0^\infty t^{p-1}|\varphi(t)|\,dt + 2^p|\eta|([0,\infty))\int_0^\infty t^p|\varphi(t)|\,dt \\
& + 2^p\int_{[0,\infty)} t^p\,|\eta|(dt)\int_0^\infty |\varphi(t)|\,dt
\end{aligned} \tag{3.18}$$

for a given $p \geq 1$. Here $|\eta|$ is the variation measure of η. If one assumes that $|\eta|$ has moments up to order $p+1$, that is,

$$\int_{[0,\infty)} t^{p+1}|\eta|(dt) < \infty,$$

it follows by [3, Lemma 3.2] that the measure $|\varphi(t)|\,dt$ is finite and has moments up to order p. Consequently, under this assumption we have that $\sup_{t\geq 0} t^p\,|\varphi(t)| < \infty$ by (3.18) and Theorem 1.1(ii) holds as long as $\sup_{t\in\mathbb{Z}} |t|^{1/2+\delta}|b(t)| < \infty$ for some $\delta > 0$.

Example 3.11. Suppose that $(X_t)_{t \in \mathbb{R}}$ is given by (3.1) with

$$\varphi(t) = \frac{1}{\Gamma(1+d)} \big[t_+^d - (t-1)_+^d \big], \qquad t \in \mathbb{R},$$

and $d \in (0, 1/4)$. (Here $\Gamma(1+d) = \int_0^\infty u^d e^{-u} \, \mathrm{d}u$ is the Gamma function at $1+d$.) In other words, we assume that $(X_t)_{t \in \mathbb{R}}$ is a fractional Lévy noise with parameter d. Recall that $\gamma_X(h) \sim ch^{2d-1}$ as $h \to \infty$ for a suitable constant $c > 0$ (see, e.g., [20, Theorem 6.3]), and hence we are in a setup where

$$\sum_{s \in \mathbb{Z}} |\gamma_X(s\Delta)| = \infty, \quad \text{but} \quad \sum_{s \in \mathbb{Z}} \gamma_X(s\Delta)^2 < \infty.$$

Moreover, it is shown in [10, Theorem A.1] that $(X_{t\Delta})_{t \in \mathbb{Z}}$ is not strongly mixing. However, Theorems 1.1 and 1.2 may still be applied in this setup, since φ is vanishing on $(-\infty, 0)$, continuous on $\mathbb{R}$, and $\varphi(t) \sim d \, t^{d-1}/\Gamma(1+d)$ as $t \to \infty$.

4 Proofs

The first observation will be used in the proof of Theorem 3.1.

Lemma 4.1. Let $g_1, g_2, g_3, g_4 \colon \mathbb{R} \to \mathbb{R}$ be functions in $L^2 \cap L^4$. Then it holds that

$$
\begin{aligned}
\mathbb{E}\bigg[\prod_{j=1}^4 \int_\mathbb{R} g_j(u) \, \mathrm{d}L_u \bigg] = {} & \kappa_4 \int_\mathbb{R} \prod_{j=1}^4 g_j(u) \, \mathrm{d}u \\
& + \sigma^4 \bigg(\int_\mathbb{R} g_1(u) g_2(u) \, \mathrm{d}u \int_\mathbb{R} g_3(u) g_4(u) \, \mathrm{d}u \\
& + \int_\mathbb{R} g_1(u) g_3(u) \, \mathrm{d}u \int_\mathbb{R} g_2(u) g_4(u) \, \mathrm{d}u \\
& + \int_\mathbb{R} g_1(u) g_4(u) \, \mathrm{d}u \int_\mathbb{R} g_2(u) g_3(u) \, \mathrm{d}u \bigg).
\end{aligned}
\tag{4.1}
$$

Proof. Set $Y_i = \int_\mathbb{R} g_i(u) \, \mathrm{d}L_u$. Then, using [16, Proposition 4.2.2], we obtain that

$$
\begin{aligned}
\mathbb{E}[Y_1 Y_2 Y_3 Y_4] = {} & \mathrm{Cum}(Y_1, Y_2, Y_3, Y_4) + \mathbb{E}[Y_1 Y_2]\mathbb{E}[Y_3 Y_4] \\
& + \mathbb{E}[Y_1 Y_3]\mathbb{E}[Y_2 Y_4] + \mathbb{E}[Y_1 Y_4]\mathbb{E}[Y_2 Y_3],
\end{aligned}
\tag{4.2}
$$

where

$$\mathrm{Cum}(Y_1, Y_2, Y_3, Y_4) = \frac{\partial^4}{\partial u_1 \cdots \partial u_4} \log \mathbb{E}\big[e^{i(u_1 Y_1 + \cdots + u_4 Y_4)} \big] \Big|_{u_1 = \cdots = u_4 = 0}.$$

Set $\psi_L(u) = \log \mathbb{E}[e^{iuL_1}]$ for $u \in \mathbb{R}$. It follows from the Lévy–Khintchine representation that we can find a constant $C > 0$ such that $|\psi_L^{(1)}(u)| \le C|u|$ and $|\psi_L^{(m)}(u)| \le C$ for $m = 2, 3, 4$. (Here $\psi_L^{(m)}$ is the mth derivative of ψ_L.) Using this together with the representation

$$\log \mathbb{E}\big[e^{i(u_1 Y_1 + \cdots + u_4 Y_4)} \big] = \int_\mathbb{R} \psi_L(u_1 g_1(t) + \cdots + u_4 g_4(t)) \, \mathrm{d}t,$$

see [23], we can interchange differentiation and integration to obtain

$$\mathrm{Cum}(Y_1, Y_2, Y_3, Y_4)$$

$$= \int_{\mathbb{R}} \psi_L^{(4)}(u_1 g_1(t) + \cdots + u_4 g_4(t)) \prod_{j=1}^{4} g_j(t) \, \mathrm{d}t \Big|_{u_1 = \cdots = u_4 = 0} = \kappa_4 \int_{\mathbb{R}} \prod_{j=1}^{4} g_j(t) \, \mathrm{d}t.$$

By combining this observation with the fact that $\mathbb{E}[Y_j Y_k] = \sigma^2 \int_{\mathbb{R}} g_j(u) g_k(u) \, \mathrm{d}u$ (using the isometry property), the result is an immediate consequence of (4.2). □

Remark 4.2. In case $g_0 = g_1 = g_2 = g_3$, Lemma 4.1 collapses to [10, Lemma 3.2], and if $\kappa_4 = 0$ then $(L_t)_{t \in \mathbb{R}}$ is a Brownian motion and the result is a special case of Isserlis' theorem.

We are now ready to prove Theorem 3.1.

Proof of Theorem 3.1. The proof goes by approximating $(X_{t\Delta}^1 X_{t\Delta}^2)_{t \in \mathbb{Z}}$ by a k-dependent sequence (cf. [7, Definition 6.4.3]), to which we can apply a classical central limit theorem. Fix $m > 0$, and set $\varphi_i^m = [(-m) \vee \varphi_i \wedge m] \mathbb{1}_{[-m,m]}$ and

$$X_t^{i,m} = \int_{\mathbb{R}} \varphi_i^m(t - s) \, \mathrm{d}L_s = \int_{t-m}^{t+m} \varphi_i^m(t - s) \, \mathrm{d}L_s, \qquad t \in \mathbb{R},$$

for $i = 1, 2$. Furthermore, set

$$S_n^m = \sum_{t=1}^{n} X_{t\Delta}^{1,m} X_{t\Delta}^{2,m}, \qquad n \in \mathbb{N}.$$

Note that since $\varphi_i^m \in L^2 \cap L^4$ and $\varphi_i^m(t) = 0$ when $|t| > m$, $(X_{t\Delta}^{1,m} X_{t\Delta}^{2,m})_{t \in \mathbb{Z}}$ is a $k(m)$-dependent sequence of square integrable random variables, where $k(m) = \inf\{n \in \mathbb{N} : n \geq 2m/\Delta\}$. Hence, we can apply [7, Theorem 6.4.2] to deduce that

$$\frac{S_n^m - \mathbb{E}[S_n^m]}{\sqrt{n}} \xrightarrow{\mathscr{D}} Y_m, \qquad n \to \infty,$$

where Y_m is a Gaussian random variable with mean zero and variance

$$\eta_m^2 = \sum_{s=-k(m)}^{k(m)} \gamma_{X^{1,m} X^{2,m}}(s\Delta). \tag{4.3}$$

Here $\gamma_{X^{1,m} X^{2,m}}$ denotes the autocovariance function of $(X_t^{1,m} X_t^{2,m})_{t \in \mathbb{R}}$. Next, we need to argue that $\eta_m^2 \to \eta^2$ with η^2 given by (3.5). Since $\varphi_i^m \in L^2 \cap L^4$ we can use Lemma 4.1 to compute $\gamma_{X^{1,m} X^{2,m}}(s\Delta)$ for each $s \in \mathbb{Z}$:

$$\gamma_{X^{1,m} X^{2,m}}(s\Delta)$$

$$= \kappa_4 \int_{\mathbb{R}} \varphi_1^m(t) \varphi_2^m(t) \varphi_1^m(t + s\Delta) \varphi_2^m(t + s\Delta) \, \mathrm{d}t + \sigma^4 \int_{\mathbb{R}} \varphi_1^m(t) \varphi_1^m(t + s\Delta) \, \mathrm{d}t$$

$$\cdot \int_{\mathbb{R}} \varphi_2^m(t) \varphi_2^m(t + s\Delta) \, \mathrm{d}t + \sigma^4 \int_{\mathbb{R}} \varphi_1^m(t) \varphi_2^m(t + s\Delta) \, \mathrm{d}t \cdot \int_{\mathbb{R}} \varphi_2^m(t) \varphi_1^m(t + s\Delta) \, \mathrm{d}t. \tag{4.4}$$

Note that $\sigma^2 \int_{\mathbb{R}} \varphi_i^m(t) \varphi_j^m(t+s\Delta)\,dt \to \gamma_{ij}(s\Delta)$, since $\varphi_i^m \to \varphi_i$ in L^2. By using assumption (iii) and that $F: t \mapsto \sum_{s\in\mathbb{Z}} |\varphi_1(t+s\Delta)\varphi_2(t+s\Delta)|$ is a periodic function with period Δ we establish as well that

$$
\sum_{s\in\mathbb{Z}} \int_{\mathbb{R}} |\kappa_4 \varphi_1(t)\varphi_2(t)\varphi_1(t+s\Delta)\varphi_2(t+s\Delta)|\,dt
$$
$$
= \kappa_4 \sum_{s\in\mathbb{Z}} \int_{s\Delta}^{(s+1)\Delta} |\varphi_1(t)\varphi_2(t)|F(t)\,dt = \kappa_4 \int_0^\Delta F(t)^2\,dt < \infty. \tag{4.5}
$$

In particular, Lebesgue's theorem on dominated convergence implies

$$
\kappa_4 \int_{\mathbb{R}} \varphi_1^m(t)\varphi_2^m(t)\varphi_1^m(t+s\Delta)\varphi_2^m(t+s\Delta)\,dt \to \kappa_4 \int_{\mathbb{R}} \varphi_1(t)\varphi_2(t)\varphi_1(t+s\Delta)\varphi_2(t+s\Delta)\,dt.
$$

Combining these observations with (4.4) shows that $\gamma_{X^{1,m}X^{2,m}}(s\Delta) \to \gamma_s$ for each $s \in \mathbb{Z}$, where

$$
\gamma_s = \kappa_4 \int_{\mathbb{R}} \varphi_1(t)\varphi_2(t)\varphi_1(t+s\Delta)\varphi_2(t+s\Delta)\,dt + \gamma_{11}(s\Delta)\gamma_{22}(s\Delta) + \gamma_{12}(s\Delta)\gamma_{21}(s\Delta)
$$

It follows as well from (4.4) that

$$
|\gamma_{X^{1,m}X^{2,m}}(s\Delta)|
$$
$$
\leq \kappa_4 \int_{\mathbb{R}} |\varphi_1(t)\varphi_2(t)\varphi_1(t+s\Delta)\varphi_2(t+s\Delta)|\,dt + \sigma^4 \int_{\mathbb{R}} |\varphi_1(t)\varphi_1(t+s\Delta)|\,dt
$$
$$
\cdot \int_{\mathbb{R}} |\varphi_2(t)\varphi_2(t+s\Delta)|\,dt + \sigma^4 \int_{\mathbb{R}} |\varphi_1(t)\varphi_2(t+s\Delta)|\,dt \cdot \int_{\mathbb{R}} |\varphi_2(t)\varphi_1(t+s\Delta)|\,dt.
$$
$$\tag{4.6}$$

Thus, if we can argue that the three terms on the right-hand side of (4.6) are summable over $s \in \mathbb{Z}$, we conclude from (4.3) that $\eta_m^2 \to \sum_{s\in\mathbb{Z}} \gamma_s = \eta^2$ by dominated convergence. In (4.5) it was shown that the first term is summable. For the second term we apply Hölder's inequality to obtain

$$
\left\| \int_{\mathbb{R}} |\varphi_1(t)\varphi_1(t+\,\cdot\,\Delta)|\,dt \int_{\mathbb{R}} |\varphi_2(t)\varphi_2(t+\,\cdot\,\Delta)|\,dt \right\|_{\ell^1} \leq \prod_{i=1}^2 \left\| \int_{\mathbb{R}} |\varphi_i(t)\varphi_i(t+\,\cdot\,\Delta)|\,dt \right\|_{\ell^{\alpha_i}},
$$

which is finite by assumption (i). The last term is handled in the same way using the Cauchy–Schwarz inequality and assumption (ii):

$$
\left\| \int_{\mathbb{R}} |\varphi_1(t)\varphi_2(t+\,\cdot\,\Delta)|\,dt \int_{\mathbb{R}} |\varphi_2(t)\varphi_1(t+\,\cdot\,\Delta)|\,dt \right\|_{\ell^1}
$$
$$
\leq \left\| \int_{\mathbb{R}} |\varphi_1(t)\varphi_2(t+\,\cdot\,\Delta)|\,dt \right\|_{\ell^2}^2 < \infty.
$$

Consequently, Y_m converges in distribution to a Gaussian random variable with mean zero and variance η^2. In light of this, the result is implied by [7, Proposition 6.3.10] if the following condition holds:

$$
\forall \varepsilon > 0: \lim_{m\to\infty} \limsup_{n\to\infty} \mathbb{P}\big(|n^{-1/2}(S_n - \mathbb{E}[S_n]) - n^{-1/2}(S_n^m - \mathbb{E}[S_n^m])| > \varepsilon\big) = 0. \tag{4.7}
$$

In order to show (4.7) we find for fixed m, using [7, Theorem 7.1.1],

$$\limsup_{n\to\infty} \mathbb{E}\big[\big(n^{-1/2}(S_n - \mathbb{E}[S_n]) - n^{-1/2}(S_n^m - \mathbb{E}[S_n^m])\big)^2\big]$$

$$= \limsup_{n\to\infty} n\mathbb{E}\Big[\Big(n^{-1}\sum_{s=1}^{n}\big(X_{s\Delta}^1 X_{s\Delta}^2 - X_{s\Delta}^{1,m}X_{s\Delta}^{2,m}\big) - \mathbb{E}\big[X_0^1 X_0^2 - X_0^{1,m}X_0^{2,m}\big]\Big)^2\Big]$$

$$= \sum_{s\in\mathbb{Z}} \gamma_{X^1 X^2 - X^{1,m}X^{2,m}}(s\Delta)$$

where $\gamma_{X^1 X^2 - X^{1,m}X^{2,m}}$ is the autocovariance function for $(X_t^1 X_t^2 - X_t^{1,m}X_t^{2,m})_{t\in\mathbb{R}}$. First, we will establish that $X_0^{1,m}X_0^{2,m} \to X_0^1 X_0^2$ in $L^2(\mathbb{P})$. To this end, recall that if a measurable function $f: \mathbb{R}^2 \to \mathbb{R}$ is square integrable (with respect to the Lebesgue measure on $\mathbb{R}^2$), and $t \mapsto f(t,t)$ and $t \mapsto \kappa_4 f(t,t)$ belong to L^1 and L^2, respectively, then the two-dimensional with-diagonal (Stratonovich type) integral $I^S(f)$ of f with respect to $(L_t)_{t\in\mathbb{R}}$ is well-defined and by the Hu–Meyer formula,

$$\mathbb{E}[I^S(f)^2] \le C\Big[\iint_{\mathbb{R}^2} f(s,t)^2 \, \mathrm{d}(s,t) + \kappa_4 \int_{\mathbb{R}} f(t,t)^2 \, \mathrm{d}t + \Big(\int_{\mathbb{R}} f(t,t) \, \mathrm{d}t\Big)^2\Big] \qquad (4.8)$$

for a suitable constant $C > 0$. A fundamental property of the Stratonovich integral is that it satisfies the relation

$$I^S(f) = \int_{\mathbb{R}} g(t) \, \mathrm{d}L_t \int_{\mathbb{R}} h(t) \, \mathrm{d}L_t,$$

when $f(s,t) = g \otimes h(s,t) := g(s)h(t)$ for given measurable functions $g, h: \mathbb{R} \to \mathbb{R}$ such that $g, h, gh \in L^2$. (See [2, 12] for details.) Since $\kappa_4 \varphi_1 \varphi_2 \in L^2$ according to (4.5), we can write $I^S(\varphi_1 \otimes \varphi_2(-\cdot) - \varphi_1^m \otimes \varphi_2^m(-\cdot)) = X_0^1 X_0^2 - X_0^{1,m}X_0^{2,m}$, and hence (4.8) shows that

$$\mathbb{E}\big[\big(X_0^1 X_0^2 - X_0^{1,m}X_0^{2,m}\big)^2\big]$$

$$\le C\Big[\iint_{\mathbb{R}^2} \big(\varphi_1(s)\varphi_2(t) - \varphi_1^m(s)\varphi_2^m(t)\big)^2 \, \mathrm{d}(s,t) + \kappa_4 \int_{\mathbb{R}} \big(\varphi_1(t)\varphi_2(t) - \varphi_1^m(t)\varphi_2^m(t)\big)^2 \, \mathrm{d}t$$

$$+ \Big(\int_{\mathbb{R}} \varphi_1(t)\varphi_2(t) \, \mathrm{d}t - \int_{\mathbb{R}} \varphi_1^m(t)\varphi_2^m(t) \, \mathrm{d}t\Big)^2\Big]$$

$$(4.9)$$

for a suitable constant $C > 0$. It is clear that the three terms on the right-hand side of (4.9) tend to zero as m tends to infinity by dominated convergence, and thus we have that $X_t^{1,m}X_t^{2,m} \to X_t^1 X_t^2$ in $L^2(\mathbb{P})$. In particular, this shows that $\gamma_{X^1 X^2 - X^{1,m}X^{2,m}}(s\Delta) \to 0$ as $m \to \infty$ for each $s \in \mathbb{Z}$. By using the same type of bound as in (4.6), we establish the existence of a function $h: \mathbb{Z} \to [0,\infty)$ in ℓ^1 with $|\gamma_{X^1 X^2 - X^{1,m}X^{2,m}}(s\Delta)| \le h(s)$ for all $s \in \mathbb{Z}$ and, consequently,

$$\lim_{m\to\infty} \limsup_{n\to\infty} \mathbb{E}\big[\big(n^{-1/2}(S_n - \mathbb{E}[S_n]) - n^{-1/2}(S_n^m - \mathbb{E}[S_n^m])\big)^2\big]$$

$$= \lim_{m\to\infty} \sum_{s\in\mathbb{Z}} \gamma_{X^1 X^2 - X^{1,m}X^{2,m}}(s\Delta) = 0$$

according to Lebesgue's theorem. In light of (4.7), we have finished the proof. □

Relying on the ideas of Young's convolution inequality, we obtain the following lemma:

Lemma 4.3. *Let* $\alpha, \beta, \gamma \in [1, \infty]$ *satisfy* $1/\alpha + 1/\beta - 1 = 1/\gamma$. *Suppose that*

$$\big(t \longmapsto \|f(t + \, \cdot \, \Delta)\|_{\ell^\alpha}\big) \in L^{2\alpha}([0, \Delta]) \quad and \quad \big(t \longmapsto \|g(t + \, \cdot \, \Delta)\|_{\ell^\beta}\big) \in L^{2\beta}([0, \Delta]).$$

Then it holds that $\int_{\mathbb{R}} |f(t) g(t + \, \cdot \, \Delta)| \, dt \in \ell^\gamma$.

Proof. First observe that, for any measurable function $h \colon \mathbb{R} \to \mathbb{R}$ and $p \in [1, \infty]$, $h \in L^p$ if and only if $t \mapsto \|h(t + \, \cdot \, \Delta)\|_{\ell^p}$ belongs to $L^p([0, \Delta])$. In particular, this ensures that $f \in L^\alpha$ and $g \in L^\beta$. If $\gamma = \infty$ then $1/\alpha + 1/\beta = 1$, and the result follows immediately from Hölder's inequality. Hence, we will restrict the attention to $\gamma < \infty$, in which case we necessarily also have that $\alpha, \beta < \infty$. First, consider the case where $\alpha, \beta \neq \gamma$, or equivalently $\alpha, \beta, \gamma > 1$, and set $\alpha' = \alpha/(\alpha - 1)$ and $\beta' = \beta/(\beta - 1)$. Note that these definitions ensure that $\alpha'(1 - \beta/\gamma) = \beta$, $\beta'(1 - \alpha/\gamma) = \alpha$ and $1/\alpha' + 1/\beta' + 1/\gamma = 1$. Hence, using the Hölder inequality and the facts that $f \in L^\alpha$ and $g \in L^\beta$,

$$\int_{\mathbb{R}} |f(t) g(t + s\Delta)| \, dt \leq \Big(\int_{\mathbb{R}} |f(t)|^\alpha |g(t + s\Delta)|^\beta \, dt\Big)^{1/\gamma} \cdot \Big(\int_{\mathbb{R}} |f(t)|^{\beta'(1 - \alpha/\gamma)} \, dt\Big)^{1/\beta'}$$
$$\cdot \Big(\int_{\mathbb{R}} |g(t + s\Delta)|^{\alpha'(1 - \beta/\gamma)} \, dt\Big)^{1/\alpha'}$$
$$= M^{1/\gamma} \Big(\int_{\mathbb{R}} |f(t)|^\alpha |g(t + s\Delta)|^\beta \, dt\Big)^{1/\gamma}$$

for a suitable constant $M < \infty$. By raising both sides to the γth power, summing over $s \in \mathbb{Z}$ and applying the Cauchy–Schwarz inequality we obtain that

$$\Big\| \int_{\mathbb{R}} |f(t) g(t + \, \cdot \, \Delta)| \, dt \Big\|_{\ell^\gamma}^\gamma$$
$$\leq M \int_{\mathbb{R}} |f(t)|^\alpha \|g(t + \, \cdot \, \Delta)\|_{\ell^\beta}^\beta \, dt \tag{4.10}$$
$$\leq M \Big(\int_0^\Delta \|f(t + \, \cdot \, \Delta)\|_{\ell^\alpha}^{2\alpha} \, dt\Big)^{1/2} \Big(\int_0^\Delta \|g(t + \, \cdot \, \Delta)\|_{\ell^\beta}^{2\beta} \, dt\Big)^{1/2},$$

which is finite, and thus we have finished the proof in case $\alpha, \beta \neq \gamma$. If, e.g., $\gamma = \alpha \neq \beta$ then $\alpha > 1$. Again, set $\alpha' = \alpha/(\alpha - 1)$ and note that $1/\alpha' + 1/\gamma = 1$, so the Hölder inequality ensures that

$$\int_{\mathbb{R}} |f(t) g(t + s\Delta)| \, dt \leq \Big(\int_{\mathbb{R}} |f(t)|^\alpha |g(t + s\Delta)|^\beta \, dt\Big)^{1/\gamma} \Big(\int_{\mathbb{R}} |g(t)|^\beta \, dt\Big)^{1/\alpha'},$$

and hence the inequalities in (4.10) hold in this case as well for a suitable constant $M > 0$. Finally if $\alpha = \beta = \gamma = 1$, we compute that

$$\Big\| \int_{\mathbb{R}} |f(t) g(t + \, \cdot \, \Delta)| \, dt \Big\|_{\ell^1} = \int_0^\Delta \|f(t + \, \cdot \, \Delta)\|_{\ell^1} \|g(t + \, \cdot \, \Delta)\|_{\ell^1} \, dt$$
$$\leq \Big(\int_0^\Delta \|f(t + \, \cdot \, \Delta)\|_{\ell^1}^2 \, dt\Big)^{1/2} \Big(\int_0^\Delta \|g(t + \, \cdot \, \Delta)\|_{\ell^1}^2 \, dt\Big)^{1/2} < \infty,$$

which finishes the proof. $\qquad \square$

Proof of Theorem 1.2. To show that statement (i) implies the stated weak convergence of $(S_n - \mathbb{E}[S_n])/\sqrt{n}$, it suffices to check that assumptions (i)–(iii) of Theorem 3.1 are satisfied. Initially note that, in view of the observation in the beginning of the proof of Lemma 4.3, the imposed assumptions imply that

$$\varphi_i \in L^\beta \quad \text{and} \quad (t \longmapsto \|\varphi_i(t + \cdot \Delta)\|_{\ell^\beta}) \in L^{2\beta}([0, \Delta]) \qquad \text{for all } \beta \in [\alpha_i, 2].$$

Since

$$\tfrac{1}{2} \in \left\{ \tfrac{1}{\beta_1} + \tfrac{1}{\beta_2} - 1 : \alpha_i \le \beta_i \le 2 \right\},$$

we can thus assume that $\alpha_1, \alpha_2 \in [1, 2]$ are given such that $1/\alpha_1 + 1/\alpha_2 - 1 = 1/2$. Next, define γ_i by the relation $1/\gamma_i = 2/\alpha_i - 1$ if $\alpha_i < 2$ and $\gamma_i = \infty$ if $\alpha_i = 2$. In this case, $1/\gamma_1 + 1/\gamma_2 = 1$. By applying Lemma 4.3 with $f = g = \varphi_i$, $\alpha = \beta = \alpha_i$ and $\gamma = \gamma_i$, we deduce that (i) of Theorem 3.1 holds. Assumption (ii) of Theorem 3.1 holds as well by Lemma 4.3 with $f = \varphi_1$, $g = \varphi_2$, $\alpha = \alpha_1$, $\beta = \alpha_2$ and $\gamma = 2$. Finally, we have that assumption (iii) of Theorem 3.1 is satisfied, since

$$\int_0^\Delta \|\varphi_1(t + \cdot \Delta)\varphi_2(t + \cdot \Delta)\|_{\ell^1}^2 \, dt$$

$$\le \left(\int_0^\Delta \|\varphi_1(t + \cdot \Delta)\|_{\ell^2}^4 \, dt \right)^{1/2} \left(\int_0^\Delta \|\varphi_2(t + \cdot \Delta)\|_{\ell^2}^4 \, dt \right)^{1/2} < \infty,$$

where we have applied the Cauchy–Schwarz inequality both for sums and integrals.

The last part of the proof (concerning statement (ii) of the theorem) amounts to showing that if $\varphi_1, \varphi_2 \in L^4$ and $\alpha_1, \alpha_2 \in (1/2, 1)$ are given such that $\alpha_1 + \alpha_2 > 3/2$ and

$$c_i := \sup_{t \in \mathbb{R}} |t|^{\alpha_i} |\varphi_i(t)| < \infty, \qquad i = 1, 2, \tag{4.11}$$

then $t \mapsto \|\varphi_i(t + \cdot \Delta)\|_{\ell^\kappa}^\kappa$ belongs to $L^2([0, \Delta])$ for $\kappa \in \{\beta_i, 2\}$ where $\beta_i \in (1/\alpha_i, 2]$, $i = 1, 2$, satisfy $1/\beta_1 + 1/\beta_2 \ge 3/2$. To show this, consider $\kappa \in \{\beta_i, 2\}$ and write

$$\|\varphi_i(t + \cdot \Delta)\|_{\ell^\kappa}^\kappa = |\varphi_i(t + \Delta)|^\kappa + |\varphi_i(t)|^\kappa + |\varphi_i(t - \Delta)|^\kappa$$

$$+ \sum_{s=2}^\infty |\varphi_i(t + s\Delta)|^\kappa + \sum_{s=2}^\infty |\varphi_i(t - s\Delta)|^\kappa \tag{4.12}$$

for $t \in [0, \Delta]$. Since $\varphi_i \in L^4$, the first three terms on the right-hand side of (4.12) belong to $L^2([0, \Delta])$. The last two terms belong to $L^\infty([0, \Delta])$, since

$$\sup_{t \in [0, \Delta]} \sum_{s=2}^\infty |\varphi_i(t \pm s\Delta)|^\kappa \le c_i^\kappa \Delta^{-\kappa\alpha_i} \sum_{s=1}^\infty s^{-\kappa\alpha_i} < \infty$$

by (4.11), and hence $(t \mapsto \|\varphi_i(t + \cdot \Delta)\|_{\ell^\kappa}^\kappa) \in L^2([0, \Delta])$. $\qquad \square$

Proof of Theorem 3.5. Initially, we note that

$$Q_n = \sum_{t=1}^n X_{t\Delta} \int_{\mathbb{R}} \sum_{s=t-n}^{t-1} b(s)\varphi((t-s)\Delta - u) \, dL_u = S_n - \varepsilon_n - \delta_n, \tag{4.13}$$

where

$$S_n = \sum_{t=1}^{n} X_{t\Delta} \int_{\mathbb{R}} b \star \varphi(t\Delta - u)\, \mathrm{d}L_u,$$

$$\varepsilon_n = \sum_{t=1}^{n} X_{t\Delta} \int_{\mathbb{R}} \sum_{s=t}^{\infty} b(s)\varphi((t-s)\Delta - u)\, \mathrm{d}L_u$$

$$\text{and} \qquad \delta_n = \sum_{t=1}^{n} X_{t\Delta} \int_{\mathbb{R}} \sum_{s=-\infty}^{t-n-1} b(s)\varphi((t-s)\Delta - u)\, \mathrm{d}L_u.$$

As pointed out in Remark 3.6, the imposed assumptions ensure that Theorem 3.1 is applicable with $\varphi_1 = \varphi$ and $\varphi_2 = |b| \star |\varphi|$ (in particular, when $\varphi_2 = b \star \varphi$), and thus $(S_n - \mathbb{E}[S_n])/\sqrt{n} \xrightarrow{\mathscr{D}} \mathscr{N}(0, \eta^2)$ where η^2 is given by (3.5). By using that b is even we compute

$$\sigma^2 \sum_{s \in \mathbb{Z}} \gamma_X(s\Delta) \int_{\mathbb{R}} (b \star \varphi)(t)(b \star \varphi)(t + s\Delta)\, \mathrm{d}t$$
$$= \sum_{s \in \mathbb{Z}} \sum_{u,v \in \mathbb{Z}} b(u)b(v)\gamma_X((s+u)\Delta)\gamma_X((s+v)\Delta) = \|(b \star \gamma_X)(\,\cdot\,\Delta)\|_{\ell^2}^2$$

and

$$\sigma^4 \sum_{s \in \mathbb{Z}} \int_{\mathbb{R}} \varphi(t)(b \star \varphi)(t + s\Delta)\, \mathrm{d}t \cdot \int_{\mathbb{R}} (b \star \varphi)(t)\varphi(t + s\Delta)\, \mathrm{d}t$$
$$= \sum_{s \in \mathbb{Z}} \sum_{u,v \in \mathbb{Z}} b(u)b(v)\gamma_X((s-u)\Delta)\gamma_X((s+v)\Delta) = \|(b \star \gamma_X)(\,\cdot\,\Delta)\|_{\ell^2}^2,$$

and it follows that η^2 coincides with (3.11). In light of the decomposition (4.13) and Slutsky's theorem, we have shown the result if we can argue that $\mathrm{Var}(\varepsilon_n)/n \to 0$ and $\mathrm{Var}(\delta_n)/n \to 0$ as $n \to \infty$. We will only show that $\mathrm{Var}(\varepsilon_n)/n \to 0$, since arguments verifying that $\mathrm{Var}(\delta_n)/n \to 0$ are similar. Define $a(t) = \int_{\mathbb{R}} \varphi(s)\varphi(t\Delta + s)\, \mathrm{d}s$ and note that we have the identities

$$\mathbb{E}[\varepsilon_n] = \sigma^2 \sum_{t=1}^{n} \sum_{s=-\infty}^{0} a(t-s)b(t-s)$$

$$\text{and} \qquad \mathbb{E}[\varepsilon_n^2] = \sum_{t,s=1}^{n} \sum_{u=t}^{\infty} \sum_{v=s}^{\infty} b(u)b(v)\mathbb{E}[X_{t\Delta}X_{s\Delta}X_{(t-u)\Delta}X_{(s-v)\Delta}]$$
$$= \sum_{t,s,u,v \in \mathbb{Z}} b(t-u)b(s-v)\mathbb{E}[X_{t\Delta}X_{s\Delta}X_{u\Delta}X_{v\Delta}]\mathbb{1}_{\{1 \le t,s \le n\}}\mathbb{1}_{\{u,v \le 0\}}.$$

Moreover, with

$$c(t,s,u) = \int_{\mathbb{R}} \varphi(t\Delta + v)\varphi(s\Delta + v)\varphi(u\Delta + v)\varphi(v)\, \mathrm{d}v,$$

it follows by Lemma 4.1 that

$$\mathbb{E}[X_{t\Delta}X_{s\Delta}X_{u\Delta}X_{v\Delta}] = \kappa_4 c(t-v, s-v, u-v) + \sigma^4 a(t-s)a(u-v)$$
$$+ \sigma^4 a(t-u)a(s-v) + \sigma^4 a(t-v)a(s-u)$$

for any $t, s, u, v \in \mathbb{Z}$. Thus, we establish the identity

$$
\begin{aligned}
n^{-1} \mathrm{Var}(\varepsilon_n) &= \kappa_4 n^{-1} \sum_{t,s,u,v \in \mathbb{Z}} b(t-u) b(s-v) c(t-v, s-v, u-v) \mathbb{1}_{\{1 \leq t, s \leq n\}} \mathbb{1}_{\{u, v \leq 0\}} \\
&\quad + \sigma^4 n^{-1} \sum_{t,s,u,v \in \mathbb{Z}} a(t-s) a(u-v) b(t-u) b(s-v) \mathbb{1}_{\{1 \leq t, s \leq n\}} \mathbb{1}_{\{u, v \leq 0\}} \\
&\quad + \sigma^4 n^{-1} \sum_{t,s,u,v \in \mathbb{Z}} a(t-v) a(s-u) b(t-u) b(s-v) \mathbb{1}_{\{1 \leq t, s \leq n\}} \mathbb{1}_{\{u, v \leq 0\}}.
\end{aligned}
\tag{4.14}
$$

It suffices to argue that each of the three terms on the right-hand side of (4.14) tends to zero as n tends to infinity. Regarding the first term, by a change of variables from (t, s, u, v) to $(t - v, s - u, u - v, v)$, we have

$$
\begin{aligned}
&n^{-1} \kappa_4 \sum_{t,s,u,v \in \mathbb{Z}} b(t-u) b(s-v) c(t-v, s-v, u-v) \mathbb{1}_{\{1 \leq t, s \leq n\}} \mathbb{1}_{\{u, v \leq 0\}} \\
&= \kappa_4 \sum_{t,s,u \in \mathbb{Z}} b(t-u) b(s+u) c(t, s+u, u) n^{-1} \sum_{v \in \mathbb{Z}} \mathbb{1}_{\{1 \leq t+v, s+u+v \leq n\}} \mathbb{1}_{\{u+v, v \leq 0\}}.
\end{aligned}
\tag{4.15}
$$

Since for fixed $t, s, u \in \mathbb{Z}$,

$$
\sum_{v \in \mathbb{Z}} \mathbb{1}_{\{1 \leq t+v, s+u+v \leq n\}} \mathbb{1}_{\{u+v, v \leq 0\}} \leq \min\{|t|, n\},
$$

it will follow that the expression in (4.15) tends to zero as n tends to infinity by Lebesgue's theorem on dominated convergence if

$$
\kappa_4 \sum_{t,s,u \in \mathbb{Z}} |b(t) b(s) c(t+u, s, u)| < \infty.
\tag{4.16}
$$

To show (4.16) we use that the function $t \mapsto \kappa_4 \| \varphi(t + \cdot \Delta)(|b| \star |\varphi|)(t + \cdot \Delta) \|_{\ell^1}$ belongs to $L^2([0, \Delta])$ (by assumption (iii)) and is periodic with period Δ:

$$
\begin{aligned}
&\kappa_4 \sum_{t,s,u \in \mathbb{Z}} |b(t) b(s) c(t+u, s, u)| \\
&\leq \kappa_4 \sum_{u \in \mathbb{Z}} \int_{\mathbb{R}} |\varphi(v)| (|b| \star |\varphi|)(v) |\varphi(v + u\Delta)| (|b| \star |\varphi|)(v + u\Delta) \, dv \\
&= \kappa_4 \int_0^\Delta \| \varphi(v + \cdot \Delta)(|b| \star |\varphi|)(v + \cdot \Delta) \|_{\ell^1}^2 \, dv < \infty.
\end{aligned}
$$

Hence, (4.15) tends to zero. We will handle the second term on the right-hand side of (4.14) in a similar way. In particular, by a change of variables from (t, s, u, v) to $(t, t - s, s - u, t - v)$,

$$
\begin{aligned}
&n^{-1} \sum_{t,s,u,v \in \mathbb{Z}} a(t-s) a(u-v) b(t-u) b(s-v) \mathbb{1}_{\{1 \leq t, s \leq n\}} \mathbb{1}_{\{u, v \leq 0\}} \\
&= \sum_{s,u,v \in \mathbb{Z}} a(s) a(v-u-s) b(s+u) b(v-s) n^{-1} \sum_{t \in \mathbb{Z}} \mathbb{1}_{\{1 \leq t, t-s \leq n\}} \mathbb{1}_{\{t-s-u, t-v \leq 0\}}.
\end{aligned}
\tag{4.17}
$$

For fixed $s, u, v \in \mathbb{Z}$,

$$
\sum_{t \in \mathbb{Z}} \mathbb{1}_{\{1 \leq t, t-s \leq n\}} \mathbb{1}_{\{t-s-u, t-v \leq 0\}} \leq \min\{|v|, n\},
$$

and since

$$\sum_{s,u,v\in\mathbb{Z}}|a(s)a(v-u-s)b(s+u)b(v-s)|$$

$$\leq \|a\|_{\ell^\alpha}\Big\|\sum_{u,v\in\mathbb{Z}}|a(v-u-\,\cdot\,)b(\,\cdot\,+u)b(v-\,\cdot\,)|\Big\|_{\ell^\beta} \tag{4.18}$$

$$\leq \Big\|\int_{\mathbb{R}}|\varphi(u)\varphi(u+\,\cdot\,\Delta)|\,\mathrm{d}u\Big\|_{\ell^\alpha}\Big\|\int_{\mathbb{R}}(|b|\star|\varphi|)(u)\,(|b|\star|\varphi|)(u+\,\cdot\,\Delta)\,\mathrm{d}u\Big\|_{\ell^\beta}$$

where the right-hand side is finite by assumption (i), it follows again by dominated convergence that (4.17) tends to zero as n tends to infinity. Finally, for the third term on the right-hand side of (4.14), we make a change of variables from (t,s,u,v) to $(t-u,s-t,u-v,v)$ and establish the inequality

$$n^{-1}\sum_{t,s,u,v\in\mathbb{Z}}a(t-v)a(s-u)b(t-u)b(s-v)\mathbb{1}_{\{1\leq t,s\leq n\}}\mathbb{1}_{\{u,v\leq 0\}}$$

$$\leq \sum_{t,s,u\in\mathbb{Z}}a(t+u)a(t+s)b(t)b(t+s+u)n^{-1}\min\{|t+u|,n\}. \tag{4.19}$$

The right-hand side of (4.19) tends to zero as n tends to infinity by dominated convergence using (4.18) and that a is even. Consequently, (4.14) shows that $\mathrm{Var}(\varepsilon_n)/n\to 0$ as $n\to\infty$, which ends the proof. $\qquad\square$

Proof of Theorem 1.1. To show (i), define $\gamma\in[1,2]$ by the relation $1/\gamma=1/\alpha+1/\beta-1$ and note that $1/\alpha+1/\gamma\geq 3/2$. According to Remark 3.6 it suffices to check that the assumptions of Theorem 3.1 are satisfied for the functions φ and $|b|\star|\varphi|$, which in turn follows from the same arguments as in the proof of Theorem 1.2 if

$$\big(t\longmapsto\|\varphi(t+\,\cdot\,\Delta)\|_{\ell^\alpha}^\alpha+\|\varphi(t+\,\cdot\,\Delta)\|_{\ell^2}^2\big)\in L^2([0,\Delta]) \tag{4.20}$$

$$\text{and}\qquad \big(t\longmapsto\|(|b|\star|\varphi|)(t+\,\cdot\,\Delta)\|_{\ell^\gamma}^\gamma+\|(|b|\star|\varphi|)(t+\,\cdot\,\Delta)\|_{\ell^2}^2\big)\in L^2([0,\Delta]). \tag{4.21}$$

Condition (4.20) holds by assumption (since $\alpha\leq 2$), so we only need to prove (4.21). If $\beta=1$ so that b is summable, it follows from Jensen's inequality that

$$(|b|\star|\varphi|)(t)^\kappa\leq\|b\|_{\ell^1}^{\kappa-1}\sum_{s\in\mathbb{Z}}|b(s)||\varphi(t+s\Delta)|^\kappa,$$

and thus $\|(|b|\star|\varphi|)(t+\,\cdot\,\Delta)\|_{\ell^\kappa}^\kappa\leq\|b\|_{\ell^1}^\kappa\|\varphi(t+\,\cdot\,\Delta)\|_{\ell^\kappa}^\kappa$ for any $\kappa\geq 1$. Since $\alpha=\gamma$ when $\beta=1$, this shows that (4.20) implies (4.21). Next if $\beta>1$, set $\beta'=\beta/(\beta-1)$. As in the proof of Lemma 4.3 (replacing integrals by sums), we can use the Hölder inequality to obtain the estimate

$$(|b|\star|\varphi|)(t)\leq M^{1/\gamma}\Big(\sum_{s\in\mathbb{Z}}|\varphi(t+s\Delta)|^\alpha\Big)^{1/\beta'}\Big(\sum_{s\in\mathbb{Z}}|b(s)|^\beta|\varphi(t+s\Delta)|^\alpha\Big)^{1/\gamma}$$

for some constant $M>0$. By raising both sides to the γth power and exploiting the periodicity of $t\mapsto\|\varphi(t+\,\cdot\,\Delta)\|_{\ell^\alpha}^\alpha$, it follows that

$$\|(|b|\star|\varphi|)(t+\,\cdot\,\Delta)\|_{\ell^\gamma}^\gamma\leq M\Big(\sum_{s\in\mathbb{Z}}|\varphi(t+s\Delta)|^\alpha\Big)^{\gamma/\beta'}\sum_{s\in\mathbb{Z}}|b(s)|^\beta\sum_{u\in\mathbb{Z}}|\varphi(t+(s+u)\Delta)|^\alpha$$

$$=M\|b\|_{\ell^\beta}^\beta\|\varphi(t+\,\cdot\,\Delta)\|_{\ell^\alpha}^\gamma \tag{4.22}$$

for a sufficiently large constant $M > 0$. Since $\gamma \leq 2$, (4.22) and the assumption $(t \mapsto \|\varphi(t + \cdot \Delta)\|_{\ell^\alpha}^\alpha) \in L^{4/\alpha}([0,\Delta])$ show that $(t \mapsto \|(|b| \star |\varphi|)(t + \cdot \Delta)\|_{\ell^\gamma}^\gamma) \in L^2([0,\Delta])$. To show $t \mapsto \|(|b| \star |\varphi|)(t + \cdot \Delta)\|_{\ell^2}^2 \in L^2([0,\Delta])$, we note that the assumption $2/\alpha + 1/\beta \geq 5/2$ ensures that we may choose $\beta^* \in [\beta, 2]$ such that $1/\alpha + 1/\beta^* = 3/2$. Using the same type of arguments as above, now with α, β^* and $\gamma^* = 2$ instead of α, β and γ, we obtain the inequality

$$\|(|b| \star |\varphi|)(t + \cdot \Delta)\|_{\ell^2}^2 \leq M \|b\|_{\ell^{\beta^*}}^{\beta^*} \|\varphi(t + \cdot \Delta)\|_{\ell^\alpha}^2.$$

Due to the fact that $(t \mapsto \|\varphi(t + \cdot \Delta)\|_{\ell^\alpha}^\alpha) \in L^{4/\alpha}([0,\Delta])$, this shows that $(t \mapsto \|(|b| \star |\varphi|)(t + \cdot \Delta)\|_{\ell^2}^2) \in L^2([0,\Delta])$ and, thus, ends the proof under statement (i).

In view of the above, to show the last part of the theorem (concerning statement (ii)), it suffices to argue that if $\varphi \in L^4$, then

$$c_1 := \sup_{t \in \mathbb{R}} |t|^{1-\alpha/2} |\varphi(t)| < \infty \quad \text{and} \quad c_2 := \sup_{t \in \mathbb{Z}} |t|^{1-\beta} |b(t)| < \infty$$

for some $\alpha, \beta > 0$ with $\alpha + \beta < 1/2$, then there exist $p, q \in [1,2]$ such that $2/p + 1/q \geq 5/2$, $b \in \ell^q$ and $(t \mapsto \|\varphi(t + \cdot \Delta)\|_{\ell^\kappa}) \in L^4([0,\Delta])$ for $\kappa \in \{p, 2\}$. To do so observe that

$$\tfrac{5}{2} \in \left\{ \tfrac{2}{p} + \tfrac{1}{q} : \tfrac{2}{2-\alpha} < p \leq 2, \tfrac{1}{1-\beta} < q \leq 2 \right\},$$

and hence we may (and do) fix $p, q \in [1,2]$ such that $2/p + 1/q \geq 5/2$, $p(\alpha/2 - 1) < -1$ and $q(\beta - 1) < -1$. With this choice it holds that $b \in \ell^q$, since

$$\|b\|_{\ell^q}^q \leq |b(0)|^q + 2c_2^q \sum_{s=1}^{\infty} s^{q(\beta-1)} < \infty.$$

We can use the same type of arguments as in the last part of the proof of Theorem 1.2 to conclude that $(t \mapsto \|\varphi(t + \cdot \Delta)\|_{\ell^\kappa}^\kappa) \in L^{4/\kappa}([0,\Delta])$ for $\kappa \in \{p, 2\}$. Indeed, in view of the decomposition (4.12) (with φ playing the role of φ_i) and the fact that $\varphi \in L^4$, it suffices to argue that $\sup_{t \in [0,\Delta]} \sum_{s=2}^{\infty} |\varphi(t \pm s\Delta)|^\kappa < \infty$. However, this is clearly the case as $\kappa(\alpha/2 - 1) \leq p(\alpha/2 - 1) < -1$ and, thus,

$$\sup_{t \in [0,\Delta]} \sum_{s=2}^{\infty} |\varphi(t + s\Delta)|^\kappa \leq c_1^\kappa \Delta^{\kappa(\alpha/2-1)} \sum_{s=1}^{\infty} s^{\kappa(\alpha/2-1)} < \infty.$$

This ends the proof of the result. □

Acknowledgments

References

[1] Avram, F. (1988). On bilinear forms in Gaussian random variables and Toeplitz matrices. *Probab. Theory Related Fields* 79(1), 37–45. DOI: 10.1007/BF00319101.

[2] Bai, S., M.S. Ginovyan and M.S. Taqqu (2016). Limit theorems for quadratic forms of Lévy-driven continuous-time linear processes. *Stochastic Process. Appl.* 126(4), 1036–1065. DOI: 10.1016/j.spa.2015.10.010.

[3] Basse-O'Connor, A., M.S. Nielsen, J. Pedersen and V. Rohde (2017). A continuous-time framework for ARMA processes. arXiv: 1704.08574v1.

[4] Beran, J., Y. Feng, S. Ghosh and R. Kulik (2016). *Long-Memory Processes*. Springer.

[5] Brandes, D.-P. and I.V. Curato (2018). On the sample autocovariance of a Lévy driven moving average process when sampled at a renewal sequence. arXiv: 1804.02254.

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

[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 (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] Cohen, S. and A. Lindner (2013). A central limit theorem for the sample autocorrelations of a Lévy driven continuous time moving average process. *J. Statist. Plann. Inference* 143(8), 1295–1306. DOI: 10.1016/j.jspi.2013.03.022.

[11] Doukhan, P., G. Oppenheim and M.S. Taqqu, eds. (2003). *Theory and applications of long-range dependence*. Boston, MA: Birkhäuser Boston Inc.

[12] Farré, M., M. Jolis and F. Utzet (2010). Multiple Stratonovich integral and Hu-Meyer formula for Lévy processes. *Ann. Probab.* 38(6), 2136–2169. DOI: 10.1214/10-AOP528.

[13] Fox, R. and M.S. Taqqu (1985). Noncentral limit theorems for quadratic forms in random variables having long-range dependence. *Ann. Probab.* 13(2), 428–446.

[14] Fox, R. and M.S. Taqqu (1987). Central limit theorems for quadratic forms in random variables having long-range dependence. *Probab. Theory Related Fields* 74(2), 213–240. DOI: 10.1007/BF00569990.

[15] Giraitis, L. and D. Surgailis (1990). A central limit theorem for quadratic forms in strongly dependent linear variables and its application to asymptotical normality of Whittle's estimate. *Probab. Theory Related Fields* 86(1), 87–104. DOI: 10.1007/BF01207515.

[16] Giraitis, L., H.L. Koul and D. Surgailis (2012). *Large sample inference for long memory processes*. Imperial College Press, London, xvi+577. DOI: 10.1142/p591.

[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] Hamilton, J.D. (1994). *Time series analysis*. Princeton University Press, Princeton, NJ.

[19] 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.

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

[21] Marquardt, T. and R. Stelzer (2007). Multivariate CARMA processes. *Stochastic Process. Appl.* 117(1), 96–120. DOI: 10.1016/j.spa.2006.05.014.

[22] Pipiras, V. and M.S. Taqqu (2017). *Long-range dependence and self-similarity*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.

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

[24] 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.

[25] Spangenberg, F. (2015). Limit theorems for the sample autocovariance of a continuous-time moving average process with long memory. arXiv: 1502.04851.

On Non-Stationary Solutions to MSDDEs: Representations and the Cointegration Space

Mikkel Slot Nielsen

Abstract

In this paper we study solutions to multivariate stochastic delay differential equations (MSDDEs) and their relation to the discrete-time cointegrated VAR model. In particular, we observe that an MSDDE can always be written in an error correction form and, under suitable conditions, we argue that a process with stationary increments is a solution to the MSDDE if and only if it admits a certain Granger type representation. A direct implication of these results is a complete characterization of the cointegration space. Finally, the relation between MSDDEs and invertible multivariate CARMA equations is used to introduce the cointegrated MCARMA processes.

MSC: 60G10; 60G12; 60H05; 60H10; 91G70

Keywords: Cointegration; Error correction form; Granger representation theorem; Multivariate CARMA processes; Multivariate SDDEs; Non-stationary processes

1 Introduction and main results

Cointegration refers to the phenomenon that some linear combinations of non-stationary time series are stationary. This concept goes at least back to Engle and Granger [9] who used the notion of cointegration to formalize the idea of a long run equilibrium between two or more non-stationary time series. Several models have been shown to be able to embed this idea, and one of the most popular among them is the VAR model:

$$X_t = \Gamma_1 X_{t-1} + \Gamma_2 X_{t-2} + \cdots + \Gamma_p X_{t-p} + \varepsilon_t, \qquad t \in \mathbb{Z}. \tag{1.1}$$

Here $(\varepsilon_t)_{t\in\mathbb{Z}}$ is an n-dimensional, say, i.i.d. sequence with $\mathbb{E}[\varepsilon_0] = 0$ and $\mathbb{E}[\varepsilon_0\varepsilon_0^\top]$ invertible, and $\Gamma_1,\ldots,\Gamma_p \in \mathbb{R}^{n\times n}$ are $n \times n$ matrices. If one is searching for a solution $(X_t)_{t\in\mathbb{Z}}$ which is only stationary in its differences, $\Delta X_t := X_t - X_{t-1}$, one often rephrases (1.1) in error correction form

$$\Delta X_t = \Pi_0 X_{t-1} + \sum_{j=1}^{p-1} \Pi_j \Delta X_{t-j} + \varepsilon_t, \qquad t \in \mathbb{Z}, \tag{1.2}$$

where $\Pi_0 = -I_n + \sum_{j=1}^{p}\Gamma_j$ and $\Pi_j = -\sum_{k=j+1}^{p}\Gamma_k$. (Here I_n denotes the $n \times n$ identity matrix.) Properties of solutions to (1.1) concerning existence, uniqueness and stationarity are determined by the characteristic polynomial $\Gamma(z) := I_n - \sum_{j=1}^{p}\Gamma_j z^j$. Let r be the rank of $\Pi_0 = -\Gamma(1)$ and, if $r < n$, let $\alpha^\perp, \beta^\perp \in \mathbb{R}^{n\times(n-r)}$ be matrices of rank $n-r$ satisfying $\Pi_0^\top \alpha^\perp = \Pi_0\beta^\perp = 0$. Standard existence and uniqueness results for VAR models and the Granger representation theorem yield the following:

Theorem 1.1. *Suppose that $\det\Gamma(z) = 0$ implies $|z| > 1$ or $z = 1$. Moreover, suppose either that $r = n$, or $r < n$ and $(\alpha^\perp)^\top(I_n - \sum_{j=1}^{p-1}\Pi_j)\beta^\perp$ is invertible. Then a process $(X_t)_{t\in\mathbb{Z}}$ with $\mathbb{E}[\|X_t\|^2] < \infty$ and stationary differences is a solution to (1.1) if and only if*

$$X_t = \xi + C_0\sum_{j=1}^{t}\varepsilon_j + \sum_{j=-\infty}^{t} C(t-j)\varepsilon_j, \qquad t \in \mathbb{Z}, \tag{1.3}$$

where:

(i) *ξ is a random vector satisfying $\mathbb{E}[\|\xi\|^2] < \infty$ and $\Pi_0\xi = 0$.*

(ii) $C_0 = \begin{cases} 0 & \textit{if } r = n \\ \beta^\perp\big[(\alpha^\perp)^\top(I_n - \sum_{j=1}^{p-1}\Pi_j)\beta^\perp\big]^{-1}(\alpha^\perp)^\top & \textit{if } r < n \end{cases}.$

(iii) *$C(j)$ is the jth coefficient in the Taylor expansion of $z \mapsto \Gamma(z)^{-1} - (1-z)^{-1}C_0$ at $z = 1$ for $j \geq 0$.*

(We use the conventions $\sum_{j=1}^{0} = 0$ and $\sum_{j=1}^{t} = -\sum_{j=t+1}^{0}$ when $t < 0$, and $\|\cdot\|$ denotes the Euclidean norm on $\mathbb{R}^n$.) The representation (1.3) has several immediate consequences: (i) any solution with stationary differences can be decomposed into an initial value, a unique stationary part and a unique non-stationary part, (ii) if $r = n$ the solution is stationary and unique, and (iii) if $r < n$ the process $(\gamma^\top X_t)_{t\in\mathbb{Z}}$ is stationary if and only if $\gamma \in \mathbb{R}^n$ belongs to the row space of $\Pi_0 = -\Gamma(1)$. In particular, cointegration is present in the VAR model when Π_0 has rank $r \in (0,n)$, and the cointegration space is spanned by the rows of Π_0. There exists a massive literature on (cointegrated) VAR models, which have been applied in various fields. We refer to [9, 13, 14, 15, 20, 24] for further details.

In many ways, the multivariate stochastic delay differential equation (MSDDE)

$$X_t - X_s = \int_s^t \eta * X(u)\,\mathrm{d}u + Z_t - Z_s, \qquad s < t, \tag{1.4}$$

may be viewed as a continuous-time version of the (possibly infinite order) VAR equation (1.1). Here $Z_t = [Z_t^1,\ldots,Z_t^n]^\top$, $t \in \mathbb{R}$, is a Lévy process with $Z_0 = 0$ and

$\mathbb{E}[\|Z_1\|^2] < \infty$, η is an $n \times n$ matrix such that each entry η_{ij} is a signed measure on $[0, \infty)$ satisfying

$$\int_{[0,\infty)} e^{\delta t} |\eta_{ij}|(\mathrm{d}t) < \infty \qquad (1.5)$$

for some $\delta > 0$, and $*$ denotes convolution. (For more on the notation used in this paper, see Section 2.) Moreover, $(X_t)_{t \in \mathbb{R}}$ will be required to satisfy $\mathbb{E}[\|X_t\|^2] < \infty$ and be given such that $(X_t, Z_t)_{t \in \mathbb{R}}$ has stationary increments. The precise meaning of (1.4) is that

$$X_t^i - X_s^i = \sum_{j=1}^{n} \int_s^t \int_{[0,\infty)} X_{u-v}^j \, \eta_{ij}(\mathrm{d}v) \, \mathrm{d}u + Z_t^i - Z_s^i, \qquad i = 1, \ldots, n,$$

almost surely for any $s < t$. The model (1.4) results in the multivariate Ornstein-Uhlenbeck process when choosing $\eta = A\delta_0$, δ_0 being the Dirac measure at 0 and $A \in \mathbb{R}^{n \times n}$, and stationary invertible multivariate CARMA (MCARMA) processes with a non-trivial moving average component can be represented as an MSDDE with infinite delay. Stationary solutions to equations of the type (1.4), MCARMA processes and their relations have been studied in [4, 5, 12, 17, 18]. Similarly to Γ for the VAR model, questions concerning solutions to (1.4) are tied to the function

$$h_\eta(z) = zI_n - \int_{[0,\infty)} e^{-zt} \eta(\mathrm{d}t), \qquad \mathrm{Re}(z) > -\delta. \qquad (1.6)$$

In particular, it was shown that if $\det h_\eta(z) = 0$ implies $\mathrm{Re}(z) < 0$, then the unique stationary solution $(X_t)_{t \in \mathbb{R}}$ to (1.4) with $\mathbb{E}[\|X_t\|^2] < \infty$ takes the form

$$X_t = \int_{-\infty}^t C(t-u) \, \mathrm{d}Z_u, \qquad t \in \mathbb{R},$$

where $C \colon [0, \infty) \to \mathbb{R}^{n \times n}$ is characterized by its Laplace transform:

$$\int_0^\infty e^{-zt} C(t) \, \mathrm{d}t = h_\eta(z)^{-1}, \qquad \mathrm{Re}(z) \geq 0.$$

It follows that this result is an analogue to Theorem 1.1 when $r = n$. To the best of our knowledge, there is no literature on solutions to (1.4) which are non-stationary, and hence no counterpart to Theorem 1.1 exists for the case $r < n$.

The main result of this paper is a complete analogue of Theorem 1.1. In the following we will set

$$\Pi_0 = \eta([0,\infty)) \quad \text{and} \quad \pi(t) = \eta([0,t]) - \eta([0,\infty)), \qquad t \geq 0. \qquad (1.7)$$

In Proposition 3.1 we show that (1.4) admits the following error correction form:

$$X_t - X_s = \Pi_0 \int_s^t X_u \, \mathrm{d}u + \int_0^\infty \pi(u)(X_{t-u} - X_{s-u}) \, \mathrm{d}u + Z_t - Z_s, \qquad s < t. \qquad (1.8)$$

To make (1.8) comparable to (1.2), one can formally apply the derivative operator D to the equation and obtain

$$DX_t = \Pi_0 X_t + \Pi * (DX)(t) + DZ_t, \qquad t \in \mathbb{R}, \qquad (1.9)$$

with $\Pi(\mathrm{d}t) = \pi(t) \, \mathrm{d}t$. We can now formulate the counterpart to Theorem 1.1. In the following, r refers to the rank of Π_0 and in case $r < n$, then $\alpha^\perp, \beta^\perp \in \mathbb{R}^{n \times (n-r)}$ are matrices of rank $n - r$ which satisfy $\Pi_0^\top \alpha^\perp = \Pi_0 \beta^\perp = 0$.

Theorem 1.2. *Suppose that* $\det h_\eta(z) = 0$ *implies* $\mathrm{Re}(z) < 0$ *or* $z = 0$. *Moreover, suppose either that the rank* r *of* Π_0 *is* n, *or strictly less than* n *and* $(\alpha^\perp)^\top(I_n - \Pi([0,\infty)))\beta^\perp$ *is invertible. Then a process* $(X_t)_{t \in \mathbb{R}}$ *is a solution to* (1.4) *if and only if*

$$X_t = \xi + C_0 Z_t + \int_{-\infty}^t C(t-u)\,\mathrm{d}Z_u, \qquad t \in \mathbb{R}, \tag{1.10}$$

where the following holds:

(i) ξ *is a random vector satisfying* $\mathbb{E}[\|\xi\|^2] < \infty$ *and* $\Pi_0\xi = 0$.

(ii) $C_0 = \begin{cases} 0 & \text{if} \quad r = n, \\ \beta^\perp\big[(\alpha^\perp)^\top(I_n - \Pi([0,\infty)))\beta^\perp\big]^{-1}(\alpha^\perp)^\top & \text{if} \quad r < n. \end{cases}$

(iii) $C\colon [0,\infty) \to \mathbb{R}^{n\times n}$ *is characterized by*

$$\int_0^\infty e^{-zt}C(t)\,\mathrm{d}t = h_\eta(z)^{-1} - z^{-1}C_0, \qquad \mathrm{Re}(z) \geq 0.$$

Similarly to the VAR model, Theorem 1.2 shows that cointegration occurs in the MSDDE model when Π_0 is of reduced rank $r \in (0,n)$, and the rows of Π_0 span the cointegration space. It follows as well that we always have uniqueness up to the discrepancy term ξ, and the restrictions on ξ depend ultimately on the rank of Π_0. Since an invertible MCARMA equation may be rephrased as an MSDDE, the notion of cointegrated invertible MCARMA processes can be studied in the MSDDE framework by relying on Theorem 1.2 (see Section 4 for details).

In Section 2 we will introduce some notation which will be used throughout the paper, and which already has been used in the introduction. The purpose of Section 3 is to develop a general theory for non-stationary solutions to MSDDEs with stationary increments, some of which will later be used to prove Theorem 1.2. In this section we will also put some emphasis on the implications of the representation (1.10), both in terms of stationary properties and concrete examples. Section 4 discusses how one can rely on the relation between invertible MCARMA equations and MSDDEs to define cointegrated MCARMA processes. In particular, under conditions similar to those imposed in [10, Theorem 4.6], we show existence and uniqueness of a cointegrated solution to the MSDDE associated to the MCARMA$(p, p-1)$ equation. This complements the result of [10], which ensures existence of cointegrated MCARMA(p, q) processes when $p > q + 1$. Finally, Section 5 contains the proofs of all the statements presented in the paper together with a few technical results.

2 Preliminaries

Let $f = [f_{ij}]\colon \mathbb{R} \to \mathbb{C}^{m\times k}$ be a measurable function and $\mu = [\mu_{ij}]$ a $k \times n$ matrix where each μ_{ij} is a measure on $\mathbb{R}$. Then, provided that

$$\int_{\mathbb{R}} |f_{il}(t)|\,\mu_{lj}(\mathrm{d}t) < \infty$$

for $l = 1, \ldots, k$, $i = 1, \ldots, m$ and $j = 1, \ldots, n$, we set

$$\int_{\mathbb{R}} f(t)\,\mu(\mathrm{d}t) = \sum_{l=1}^{k} \begin{bmatrix} \int_{\mathbb{R}} f_{1l}(t)\,\mu_{l1}(\mathrm{d}t) & \cdots & \int_{\mathbb{R}} f_{1l}(t)\,\mu_{ln}(\mathrm{d}t) \\ \vdots & \ddots & \vdots \\ \int_{\mathbb{R}} f_{ml}(t)\,\mu_{l1}(\mathrm{d}t) & \cdots & \int_{\mathbb{R}} f_{ml}(t)\,\mu_{ln}(\mathrm{d}t) \end{bmatrix}. \tag{2.1}$$

The integral $\int_{\mathbb{R}} f(t)\mu(\mathrm{d}t)$ is defined in a similar manner when either f or μ is one-dimensional. Moreover, we will say that μ is a signed measure if it takes the form $\mu = \mu^+ - \mu^-$ for two mutually singular measures μ^+ and μ^-, where at least one of them is finite. The definition of the integral (2.1) extends naturally to signed matrix measures provided that the integrand is integrable with respect to the variation measure $|\mu| := \mu^+ + \mu^-$ (simply referred to as being integrable with respect to μ). For a given point $t \in \mathbb{R}$, if $f(t - \cdot)$ is integrable with respect to μ, we define the convolution as

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

For a measurable function $f : \mathbb{R} \to \mathbb{C}^{k \times m}$ and μ an $n \times k$ signed matrix measure, if $f^\top(t - \cdot)$ is integrable with respect to $\mu^\top$, we set $\mu * f(t) := (f^\top * \mu^\top)(t)^\top$. Also, if μ is a given signed matrix measure and $z \in \mathbb{C}$ is such that $\int_{\mathbb{R}} e^{-\mathrm{Re}(z)t}\,|\mu_{ij}|(\mathrm{d}t) < \infty$ for all i and j, the (i,j)-th entry of the Laplace transform $\mathcal{L}[\mu](z)$ of μ at z is defined by

$$\mathcal{L}[\mu]_{ij}(z) = \int_{\mathbb{R}} e^{-zt}\,\mu_{ij}(\mathrm{d}t).$$

Eventually, if $|\mu|$ is finite, we will also use the notation $\mathcal{F}[\mu](y) = \mathcal{L}[\mu](iy)$, $y \in \mathbb{R}$, referring to the Fourier transform of μ. When $\mu(\mathrm{d}t) = f(t)\,\mathrm{d}t$ for some measurable function f we write $\mathcal{L}[f]$ and $\mathcal{F}[f]$ instead.

Finally, a stochastic process $Y_t = [Y_t^1, \ldots, Y_t^n]^\top$, $t \in \mathbb{R}$, is said to be stationary, respectively have stationary increments, if the finite dimensional marginal distributions of $(Y_{t+h})_{t \in \mathbb{R}}$, respectively $(Y_{t+h} - Y_h)_{t \in \mathbb{R}}$, do not depend on $h \in \mathbb{R}$.

3 General results on existence, uniqueness and representations of solutions to MSDDEs

Suppose that $Z_t = [Z_t^1, \ldots, Z_t^n]^\top$, $t \in \mathbb{R}$, is an n-dimensional measurable process with $Z_0 = 0$, stationary increments and $\mathbb{E}[\|Z_t\|^2] < \infty$, and let $\eta = [\eta_{ij}]$ be a signed $n \times n$ matrix measure which satisfies (1.5) for some $\delta > 0$. We will say that a stochastic process $X_t = [X_t^1, \ldots, X_t^n]^\top$, $t \in \mathbb{R}$, is a solution to the corresponding multivariate stochastic delay differential equation (MSDDE) if it meets the following requirements:

(i) $(X_t)_{t \in \mathbb{R}}$ is measurable and $\mathbb{E}[\|X_t\|^2] < \infty$ for all $t \in \mathbb{R}$.

(ii) $(X_t, Z_t)_{t \in \mathbb{R}}$ has stationary increments.

(iii) The relations

$$X_t^i - X_s^i = \sum_{j=1}^{n} \int_s^t \int_{[0,\infty)} X_{u-v}^j\,\eta_{ij}(\mathrm{d}v)\,\mathrm{d}u + Z_t^i - Z_s^i, \qquad i = 1, \ldots, n,$$

hold true almost surely for each $s < t$.

As indicated in the introduction, (iii) may be compactly written as

$$dX_t = \eta * X(t)\, dt + dZ_t, \qquad t \in \mathbb{R}. \tag{3.1}$$

We start with the observation that (3.1) can always be written in an error correction form (as noted in (1.8)):

Proposition 3.1. *Let $\Pi_0 \in \mathbb{R}^{n \times n}$ and $\pi \colon [0, \infty) \to \mathbb{R}^{n \times n}$ be defined by (1.7), and suppose that $\delta > 0$ is given such that (1.5) is satisfied. Then $\sup_{t \geq 0} e^{\varepsilon t} \|\pi(t)\| < \infty$ for all $\varepsilon < \delta$, and (3.1) can be written as*

$$X_t - X_s = \Pi_0 \int_s^t X_u\, du + \int_0^\infty \pi(u)(X_{t-u} - X_{s-u})\, du + Z_t - Z_s, \qquad s < t, \tag{3.2}$$

so if $(X_t)_{t \in \mathbb{R}}$ is a solution to (3.1), then $(\Pi_0 X_t)_{t \in \mathbb{R}}$ is stationary.

Remark 3.2. Using the notation $\Pi(dt) = \pi(t)\, dt$, we do the following observations in relation to Proposition 3.1:

   (i) If Π_0 is invertible, a solution $(X_t)_{t \in \mathbb{R}}$ must be stationary itself.

  (ii) If $\Pi_0 = 0$ the statement does not provide any further insight. Observe, however, the equation (3.2) depends in this case only on the increments of $(X_t)_{t \in \mathbb{R}}$ so a solution needs not to be stationary in this case.

 (iii) If the rank r of Π_0 satisfies $0 < r < n$, there exist non-trivial linear combinations of the entries of $(X_t)_{t \in \mathbb{R}}$ which are stationary

At this point we have not argued whether or not $(X_t)_{t \in \mathbb{R}}$ can be stationary even when $r < n$ and, ultimately, it depends on the structure of the noise process $(Z_t)_{t \in \mathbb{R}}$. However, it is not too difficult to verify from Theorem 3.5 that if $(Z_t)_{t \in \mathbb{R}}$ is a Lévy process such that $\mathbb{E}[Z_1 Z_1^\top]$ is invertible and $A \in \mathbb{R}^{m \times n}$, $(AX_t)_{t \in \mathbb{R}}$ is stationary if and only if $A = B\Pi_0$ for some $B \in \mathbb{R}^{m \times n}$. In case of (iii), one often considers a rank factorization of Π_0; that is, one chooses $\alpha, \beta \in \mathbb{R}^{n \times r}$ of rank r such that $\Pi_0 = \alpha \beta^\top$. In this way one can identify the columns of β as cointegrating vectors spanning the cointegration space, and α as the adjustment matrix determining how deviations from a long run equilibrium affect short run dynamics. This type of intuition is well-known for the cointegrated VAR models, so we refer to [9] for details.

In the following we will search for a solution to (3.1). To this end, let $\delta > 0$ be chosen such that (1.5) holds, set $\mathbb{H}_\delta := \{z \in \mathbb{C} : \text{Re}(z) > -\delta\}$ and define $h_\eta \colon \mathbb{H}_\delta \to \mathbb{C}^{n \times n}$ by

$$h_\eta(z) = z I_n - \mathcal{L}[\eta](z), \qquad z \in \mathbb{H}_\delta. \tag{3.3}$$

Since h_η is analytic on $\mathbb{H}_\delta$ and $|\det h_\eta(z)| \to \infty$ as $|z| \to \infty$, $z \mapsto h_\eta(z)^{-1}$ is meromorphic on $\mathbb{H}_\delta$. Recall that if z_0 is a pole of $z \mapsto h_\eta(z)^{-1}$, there exists $n \in \mathbb{N}$ such that $z \mapsto (z - z_0)^n h_\eta(z)^{-1}$ is analytic and non-zero in a neighborhood of z_0. If $n = 1$ the pole is called simple.

Condition 3.3. *For the function h_η in (3.3) it holds that*

   (i) $\det(h_\eta(z)) \neq 0$ *for all $z \in \mathbb{H}_\delta \setminus \{0\}$ and*

(ii) $z \longmapsto h_\eta(z)^{-1}$ has either no poles at all or a simple pole at 0.

For convenience, we have chosen to work with Condition 3.3 rather than the assumptions of Theorem 1.2. The following result shows that they are essentially the same.

Proposition 3.4. *Suppose that, for some $\varepsilon > 0$,*

$$\int_{[0,\infty)} e^{\varepsilon t} |\eta_{ij}|(\mathrm{d}t) < \infty, \qquad i, j = 1, \dots, n.$$

The following two statements are equivalent:

(i) *There exists $\delta \in (0, \varepsilon]$ such that (1.5) and Condition 3.3 are satisfied.*

(ii) *The assumptions of Theorem 1.2 hold true.*

We will construct a solution $(X_t)_{t \in \mathbb{R}}$ to (3.1) in a similar way as in [4], namely by applying a suitable filter (i.e., a finite signed $n \times n$ matrix measure) μ to $(Z_t)_{t \in \mathbb{R}}$. Theorem 3.5 reveals that the appropriate filter to apply is $\mu(\mathrm{d}t) = \delta_0(\mathrm{d}t) - f(t)\,\mathrm{d}t$ for a suitable function $f : \mathbb{R} \to \mathbb{R}^{n \times n}$. This result may be viewed as a Granger type representation theorem for solutions to MSDDEs and as a general version of Theorem 1.2.

Theorem 3.5. *Suppose that Condition 3.3 holds. Then there exists a unique function $f : [0, \infty) \to \mathbb{R}^{n \times n}$ satisfying*

$$\mathcal{L}[f](z) = I_n - z h_\eta(z)^{-1}, \qquad z \in \mathbb{H}_\delta, \tag{3.4}$$

and the function $u \mapsto f(u)Z_{t-u}$ belongs to L^1 almost surely for each $t \in \mathbb{R}$. Moreover, a process $(X_t)_{t \in \mathbb{R}}$ is a solution to (3.1) if and only if

$$X_t = \xi + C_0 Z_t + \int_0^\infty f(u)[Z_t - Z_{t-u}]\,\mathrm{d}u, \qquad t \in \mathbb{R}, \tag{3.5}$$

where $\Pi_0 \xi = 0$, $\mathbb{E}[\|\xi\|^2] < \infty$ and $C_0 = I_n - \int_0^\infty f(t)\,\mathrm{d}t$.

Concerning the function f of Theorem 3.5, it can also be obtained as a solution to a certain multivariate delay differential equation; we refer to Lemma 5.1 for more on its properties.

Remark 3.6. Let the situation be as described in Theorem 3.5 and note that

$$C_0 = I_n - \mathcal{L}[f](0) = z h_\eta(z)^{-1}\big|_{z=0}.$$

Hence, if the rank r of Π_0 is equal to n we have that $C_0 = 0$, and if r is strictly less than n, C_0 can be computed by the residue formula given in [23]. Specifically, $C_0 = \beta^\perp[(\alpha^\perp)^\top(I_n - \Pi([0, \infty)))\beta^\perp]^{-1}(\alpha^\perp)^\top$, where $\alpha^\perp, \beta^\perp \in \mathbb{R}^{n \times (n-r)}$ are matrices of rank $n - r$ satisfying $\Pi_0^\top \alpha^\perp = \Pi_0 \beta^\perp = 0$ (note that the inverse matrix in the expression of C_0 does indeed exist by Proposition 3.4).

In the special case where $z \mapsto h_\eta(z)^{-1}$ has no poles at all, it was shown in [4, Theorem 3.1] that there exists a unique stationary solution to (3.1). The same conclusion can be reached by Theorem 3.5 using that Π_0 is invertible. Indeed, in this case any solution is stationary, $C_0 = 0$ and $\xi = 0$ (the first two implications follow from Remarks 3.2 and 3.6). While there exist several solutions when Π_0 is singular, Theorem 3.5 shows that any two solutions always have the same increments. The term ξ reflects how much solutions may differ and its possible values are determined by the relation $\Pi_0 \xi = 0$.

In view of Proposition 3.4 and Remark 3.6, Theorem 1.2 is an obvious consequence of Theorem 3.5 if

$$\int_0^\infty f(u)[Z_t - Z_{t-u}]\,\mathrm{d}t = \int_{-\infty}^t C(t-u)\,\mathrm{d}Z_u, \qquad t \in \mathbb{R}. \tag{3.6}$$

Clearly, the right-hand side of (3.6) requires that we can define integration with respect to $(Z_t)_{t \in \mathbb{R}}$. Although this is indeed possible if $(Z_t)_{t \in \mathbb{R}}$ is a Lévy process (for instance, in the sense of [19]), we will here put the less restrictive assumption that $(Z_t)_{t \in \mathbb{R}}$ is a regular integrator as defined in [4, Proposition 4.1]:

Corollary 3.7. *Suppose that Condition 3.3 holds. Assume also that, for each $i = 1, \ldots, n$, there exists a linear map $\mathcal{I}_i \colon L^1 \cap L^2 \to L^1(\mathbb{P})$ which satisfies*

(i) $\mathcal{I}_i(\mathbb{1}_{(s,t]}) = Z_t^i - Z_s^i$ *for all $s < t$, and*

(ii) *for all finite measures μ on $\mathbb{R}$ with $\int_{\mathbb{R}} |r|\,\mu(\mathrm{d}r) < \infty$,*

$$\mathcal{I}_i\left(\int_{\mathbb{R}} f_r(t - \,\cdot\,)\,\mu(\mathrm{d}r)\right) = \int_{\mathbb{R}} \mathcal{I}_i(f_r(t - \,\cdot\,))\,\mu(\mathrm{d}r), \qquad t \in \mathbb{R},$$

where $f_r = \mathbb{1}_{[0,\infty)}(\,\cdot\, - r) - \mathbb{1}_{[0,\infty)}$.

Then the statement of Theorem 1.2 holds true with

$$\left[\int_{-\infty}^t C(t-u)\,\mathrm{d}Z_u\right]_i = \sum_{j=1}^n \mathcal{I}_j(C_{ij}(t - \,\cdot\,)), \qquad i = 1, \ldots, n. \tag{3.7}$$

In Theorem 1.2 the function C is characterized through its Laplace transform $\mathcal{L}[C]$, but one can also obtain it as a solution to a certain multivariate delay differential equation. This follows by using the similar characterization given for f in Lemma 5.1; the details are discussed in Remark 5.3. It should also be stressed that the conditions for being a regular integrator (i.e., for $\mathcal{I}_1, \ldots, \mathcal{I}_n$ to exist) are mild; many semimartingales with stationary increments (in particular, Lévy processes) and fractional Lévy processes, as studied in [16], are regular integrators. For more on regular integrators, see [4, Section 4.1].

Remark 3.8. Suppose that Condition 3.3 is satisfied, let $(Z_t)_{t \in \mathbb{R}}$ be a regular integrator, and let $(X_t)_{t \in \mathbb{R}}$ be a solution to (3.1). Since $\Pi_0 \xi = 0$ and $\Pi_0 C_0 = 0$ (the latter by Remark 3.6), Corollary 3.7 implies that the stationary process $(\Pi_0 X_t)_{t \in \mathbb{R}}$ is unique and given by

$$\Pi_0 X_t = \Pi_0 \int_{-\infty}^t C(t-u)\,\mathrm{d}Z_u, \qquad t \in \mathbb{R}. \tag{3.8}$$

If $(Z_t)_{t \in \mathbb{R}}$ is not a regular integrator one can instead rely on Theorem 3.5 to replace $\int_{-\infty}^t C(t-u)\,\mathrm{d}Z_u$ by $\int_0^\infty f(u)[Z_t - Z_{t-u}]\,\mathrm{d}u$ in (3.8).

We end this section by giving two examples. In both examples we suppose for convenience that $(Z_t)_{t\in\mathbb{R}}$ is a regular integrator.

Example 3.9 (The univariate case). Consider the case where $n = 1$ and η is a measure which admits an exponential moment in the sense of (1.5) and satisfies $h_\eta(z) \neq 0$ for all $z \in \mathbb{H}_\delta \setminus \{0\}$. In this setup Condition 3.3 can be satisfied in two ways which ultimately determine the class of solutions characterized in Corollary 3.7:

(i) If $\Pi_0 \neq 0$. In this case, the solution to (3.1) is unique and given by

$$X_t = \int_{-\infty}^t C(t-u)\,\mathrm{d}Z_u, \qquad t \in \mathbb{R},$$

where $\mathcal{L}[C](z) = 1/h_\eta(z)$ for $z \in \mathbb{C}$ with $\mathrm{Re}(z) \geq 0$. This is consistent with the literature on stationary solutions to univariate SDDEs (see [3, 12]).

(ii) If $\Pi_0 = 0$ and $\Pi([0,\infty)) \neq 1$. In this case, a process $(X_t)_{t\in\mathbb{R}}$ is a solution to (3.1) if and only if

$$X_t = \xi + (1 - \Pi([0,\infty)))Z_t + \int_{-\infty}^t C(t-u)\,\mathrm{d}Z_u, \qquad t \in \mathbb{R},$$

where ξ can be any random variable with $\mathbb{E}[\xi^2] < \infty$ and $\mathcal{L}[C](z) = 1/h_\eta(z) - (1 - \Pi([0,\infty)))/z$ for $z \in \mathbb{C}$ with $\mathrm{Re}(z) \geq 0$.

Suppose that we are in case (ii) and fix $h > 0$. Using the notation $\Delta_h Y_t := Y_t - Y_{t-h}$, it follows from Proposition 3.1 that $(\Delta_h X_t)_{t\in\mathbb{R}}$ is a stationary solution to the equation

$$Y_t = \int_0^\infty Y_{t-u}\,\Pi(\mathrm{d}u) + \Delta_h Z_t, \qquad t \in \mathbb{R}. \tag{3.9}$$

Existence and uniqueness of stationary solutions to equations of the type (3.9) were studied in [3, Section 3] (when $(\Delta_h Z_t)_{t\in\mathbb{R}}$ is a suitable Lévy-driven moving average), and it was shown how these sometimes can be used to construct stationary increment solutions to univariate SDDEs.

Example 3.10 (Ornstein–Uhlenbeck). Suppose that $\eta = A\delta_0$ for some $A \in \mathbb{R}^{n\times n}$, for which its spectrum $\sigma(A)$ satisfies

$$\sigma(A) \setminus \{0\} \subseteq \{z \in \mathbb{C} : \mathrm{Re}(z) < 0\}. \tag{3.10}$$

With this specification, the MSDDE (3.1) reads

$$\mathrm{d}X_t = AX_t\,\mathrm{d}t + \mathrm{d}Z_t, \qquad t \in \mathbb{R}. \tag{3.11}$$

Under the assumption (3.10) we have that

$$h_\eta(z)^{-1} = \int_0^\infty e^{(A-I_n z)t}\,\mathrm{d}t = \mathcal{L}\big[t \longmapsto \mathbb{1}_{[0,\infty)}(t)e^{At}\big](z), \qquad \mathrm{Re}(z) > 0.$$

Since the set of zeroes of h_η coincides with $\sigma(A)$, it follows immediately that Condition 3.3 is satisfied for some $\delta > 0$ if $0 \notin \sigma(A)$. This is the stationary case where the solution to (3.11) takes the well-known form

$$X_t = \int_{-\infty}^t e^{A(t-u)}\,\mathrm{d}Z_u, \qquad t \in \mathbb{R}.$$

If instead $0 \in \sigma(A)$, let $r < n$ be the rank of A and choose $\alpha^\perp, \beta^\perp \in \mathbb{R}^{n \times (n-r)}$ of rank $n-r$ such that $A^\top \alpha^\perp = A \beta^\perp = 0$. We can now rely on Proposition 3.4 and the observation that $\Pi \equiv 0$ to conclude that Condition 3.3 is satisfied if $(\alpha^\perp)^\top \beta^\perp$ is invertible. This is the cointegrated case where the solution takes the form

$$X_t = \xi + C_0 Z_t + \int_{-\infty}^t \left[e^{A(t-u)} - C_0 \right] dZ_u, \qquad t \in \mathbb{R},$$

with $\mathbb{E}[\|\xi\|^2] < \infty$, $A\xi = 0$ and $C_0 = \beta^\perp [(\alpha^\perp)^\top \beta^\perp]^{-1} (\alpha^\perp)^\top$. In particular, the stationary process $(AX_t)_{t \in \mathbb{R}}$ takes the form

$$AX_t = \int_{-\infty}^t A e^{A(t-u)} dZ_u, \qquad t \in \mathbb{R}.$$

Stationary Ornstein-Uhlenbeck processes have been widely studied in the literature (see, e.g., [1, 21, 22]). Cointegrated solutions to (3.11) have also received some attention, for instance, in [6].

4   Cointegrated multivariate CARMA processes

In [4, Theorem 4.8] it was shown that any stationary MCARMA process satisfying a certain invertibility assumption can be characterized as the unique solution to a suitable MSDDE. This may be viewed as the continuous-time analogue of representing a discrete-time ARMA process as an infinite order AR equation. In this section we will rely on this idea and the results obtained in Section 3 to define cointegrated MCARMA processes. The focus will only be on MCARMA$(p, p-1)$ processes for a given $p \in \mathbb{N}$. However, the analysis should also be doable for MCARMA(p, q) processes for a general $q \in \mathbb{N}_0$ with $q < p$ by extending the theory developed in the former sections to higher order MSDDEs. This was done in [4] in the stationary case. For convenience we will also assume that $(Z_t)_{t \in \mathbb{R}}$ is a regular integrator in the sense of Corollary 3.7.

We start by introducing some notation. Define $P, Q : \mathbb{C} \to \mathbb{C}^{n \times n}$ by

$$P(z) = I_n z^p + P_1 z^{p-1} + \cdots + P_p$$
$$\text{and} \quad Q(z) = I_n z^{p-1} + Q_1 z^{p-2} + \cdots + Q_{p-1}$$

for $P_1, \ldots, P_p, Q_1, \ldots, Q_{p-1} \in \mathbb{R}^{n \times n}$. Essentially, any definition of the MCARMA process $(X_t)_{t \in \mathbb{R}}$ aims at rigorously defining the solution to the formal differential equation

$$P(D)X_t = Q(D)DZ_t, \qquad t \in \mathbb{R}. \tag{4.1}$$

Since $P(D)\xi = P_p \xi$ for any random vector ξ, one should only expect solutions to be unique up to translations belonging to the null space of P_p. To solve (4.1) it is only necessary to impose assumptions on P, but since we will be interested in an autoregressive representation of the equation, we will also impose an invertibility assumption on Q:

Condition 4.1 (Stationary case). *If $\det P(z) = 0$ or $\det Q(z) = 0$, then $\mathrm{Re}(z) < 0$.*

Under Condition 4.1 it was noted in [17, Remark 3.23] that one can find $g : [0, \infty) \to \mathbb{R}^{n \times n}$ which belongs to $L^1 \cap L^2$ with

$$\mathcal{F}[g](y) = P(iy)^{-1} Q(iy), \qquad y \in \mathbb{R}. \tag{4.2}$$

Consequently, by heuristically applying the Fourier transform to (4.1) and rearranging terms, one arrives at the conclusion

$$X_t = \int_{-\infty}^{t} g(t-u)\,\mathrm{d}Z_u, \qquad t \in \mathbb{R}. \tag{4.3}$$

As should be the case, any definition used in the literature results in this process (although $(Z_t)_{t\in\mathbb{R}}$ is sometimes restricted to being a Lévy process). In Proposition 4.2 we state two characterizations without proofs; these are consequences of [17, Definition 3.20] and [4, Theorem 4.8], respectively.

Proposition 4.2. *Suppose that Condition 4.1 is satisfied and let $(X_t)_{t\in\mathbb{R}}$ be defined by (4.2)–(4.3).*

(i) *Choose $B_1,\dots,B_p \in \mathbb{R}^{n\times n}$ such that $z \mapsto P(z)[B_1 z^{p-1} + \cdots + B_p] - Q(z)z^p$ is at most of order $p-1$, and set*

$$A = \begin{bmatrix} 0 & I_n & 0 & \cdots & 0 \\ 0 & 0 & I_n & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & I_n \\ -P_p & -P_{p-1} & \cdots & -P_2 & -P_1 \end{bmatrix} \quad and \quad B = \begin{bmatrix} B_1 \\ \vdots \\ B_p \end{bmatrix}.$$

Then $X_t = CG_t$, where $C = [I_n, 0, \dots, 0]^\top \in \mathbb{R}^{np\times n}$ and $(G_t)_{t\in\mathbb{R}}$ is the unique stationary process satisfying

$$\mathrm{d}G_t = AG_t\,\mathrm{d}t + B\,\mathrm{d}Z_t, \qquad t \in \mathbb{R}.$$

(ii) *Set $\eta_0 = Q_1 - P_1$ and let $\eta_1 \colon [0,\infty) \to \mathbb{R}^{n\times n}$ be characterized by*

$$\mathcal{F}[\eta_1](y) = I_n iy - \eta_0 - Q(iy)^{-1}P(iy), \qquad y \in \mathbb{R}. \tag{4.4}$$

Then $(X_t)_{t\in\mathbb{R}}$ is the unique stationary process satisfying

$$\mathrm{d}X_t = \eta_0 X_t\,\mathrm{d}t + \int_0^\infty \eta_1(u)X_{t-u}\,\mathrm{d}u\,\mathrm{d}t + \mathrm{d}Z_t, \qquad t \in \mathbb{R}.$$

It follows from Proposition 4.2 that $(X_t)_{t\in\mathbb{R}}$ can either be defined in terms of a state-space model using the triple (A, B, C) or by an MSDDE of the form (3.1) with

$$\eta(\mathrm{d}t) = \eta_0 \delta_0(\mathrm{d}t) + \eta_1(t)\,\mathrm{d}t. \tag{4.5}$$

While $(X_t)_{t\in\mathbb{R}}$ given by (4.3) is stationary by definition, it does indeed make sense to search for non-stationary, but cointegrated, processes satisfying (i) or (ii) of Proposition 4.2 also when Condition 4.1 does not hold. Fasen-Hartmann and Scholz [10] follow this idea by first characterizing cointegrated solutions to state-space equations and, next, define the cointegrated MCARMA process as a cointegrated solution corresponding to the specific triple (A, B, C). Their definition applies to any MCARMA(p,q) process and they give sufficient conditions on P and Q for the cointegrated MCARMA process to exist when $q < p-1$. We will use the results from the former sections to define the cointegrated MCARMA$(p, p-1)$ process as the solution to an MSDDE.

Condition 4.3 (Cointegrated case). *The following statements are true:*

(i) *If* $\det P(z) = 0$, *then either* $\mathrm{Re}(z) < 0$ *or* $z = 0$.

(ii) *The rank* r *of* $P(0) = P_p$ *is reduced* $r \in (0, n)$.

(iii) *The matrix* $(\alpha^\perp)^\top P_{p-1} \beta^\perp$ *is invertible, where* $\alpha^\perp, \beta^\perp \in \mathbb{R}^{n \times (n-r)}$ *are of rank* $n - r$ *and satisfy* $P_p^\top \alpha^\perp = P_p \beta^\perp = 0$.

(iv) *If* $\det Q(z) = 0$ *then* $\mathrm{Re}(z) < 0$.

The assumptions (i)–(iii) of Condition 4.3 are also imposed in [10], and (iv) is imposed to ensure that (4.1) admits an MSDDE representation. In [10] they impose an additional assumption, namely that the polynomials P and Q are so-called left coprime, which is used to ensure that the pole of $z \mapsto P(z)^{-1}$ at 0 is also a pole of $z \mapsto P(z)^{-1}Q(z)$. However, in our case this is implied by (iv).

Theorem 4.4. *Suppose that Condition 4.3 holds. Then the measure in (4.5) is well-defined and satisfies (1.5) as well as Condition 3.3 for a suitable $\delta > 0$, and the rank of $\Pi_0 = \eta([0, \infty))$ is r. In particular, a process $(X_t)_{t \in \mathbb{R}}$ is a solution to the corresponding MSDDE if and only if*

$$X_t = \xi + C_0 Z_t + \int_{-\infty}^t C(t - u) \, \mathrm{d}Z_u, \qquad t \in \mathbb{R}, \tag{4.6}$$

where $\mathbb{E}[\|\xi\|^2] < \infty$, $P_p \xi = 0$, $C_0 = \beta^\perp[(\alpha^\perp)^\top P_{p-1} \beta^\perp]^{-1}(\alpha^\perp)^\top Q_{p-1}$ *and*

$$\mathcal{L}[C](z) = P(z)^{-1}Q(z) - z^{-1}C_0, \qquad \mathrm{Re}(z) \geq 0.$$

Remark 4.5. Suppose that Condition 4.3 is satisfied and define η by (4.5). In this case, Theorem 4.4 shows that $(X_t)_{t \in \mathbb{R}}$ given by (4.6) defines a solution to the corresponding MSDDE. As noted right after the formal CARMA equation (4.1), the initial value ξ should not affect whether $(X_t)_{t \in \mathbb{R}}$ can be thought of as a solution (since $P_p \xi = 0$). Hence, suppose that $\xi = 0$. By heuristically computing $\mathcal{F}[X]$ from (4.6) we obtain

$$\mathcal{F}[X](y) = (iy)^{-1}C_0 \mathcal{F}[DZ](y) + \mathcal{F}[C](y)\mathcal{F}[DZ](y) = P(iy)^{-1}Q(iy)\mathcal{F}[DZ](y)$$

for $y \in \mathbb{R}$ which, by multiplication of $P(iy)$, shows that $(X_t)_{t \in \mathbb{R}}$ solves (4.1).

5 Proofs

Proof of Proposition 3.1. We start by arguing that

$$\sup_{t \geq 0} e^{\varepsilon t} \|\pi(t)\| < \infty \tag{5.1}$$

for a given $\varepsilon \in (0, \delta)$. Note that, for any given finite signed matrix-valued measure μ on $[0, \infty)$,

$$\mathcal{L}[\mu](z) = \int_{[0,\infty)} e^{-zu} \mu(\mathrm{d}u) = z \int_0^\infty e^{-zu} \mu([0, u]) \, \mathrm{d}u \tag{5.2}$$

for all $z \in \mathbb{C}$ with $\mathrm{Re}(z) > 0$ using integration by parts. Consequently,

$$\mathcal{L}[\pi](z) = z^{-1}\mathcal{L}[\bar{\eta}](z), \qquad \mathrm{Re}(z) > 0, \tag{5.3}$$

using the notation $\tilde{\eta} = \eta - \Pi_0 \delta_0$. On the other hand, $z \mapsto \mathcal{L}[\tilde{\eta}](z)$ is analytic on $\mathbb{H}_\delta$ (by (1.5)), and since $\mathcal{L}[\tilde{\eta}](0) = \mathcal{L}[\eta](0) - \eta([0, \infty)) = 0$, $z \mapsto z^{-1}\mathcal{L}[\tilde{\eta}](z)$ is also analytic on $\mathbb{H}_\delta$, and we deduce that

$$C := \sup_{\mathrm{Re}(z) \geq -\tilde{\varepsilon}} \|\mathcal{L}[\tilde{\eta}](z)\| + \sup_{\mathrm{Re}(z) \geq -\tilde{\varepsilon}} \|z^{-1}\mathcal{L}[\tilde{\eta}](z)\| < \infty$$

for an arbitrary $\tilde{\varepsilon} \in (\varepsilon, \delta)$. Hence, we find that

$$\sup_{\mathrm{Re}(z) > -\tilde{\varepsilon}} \int_{\mathbb{R}} \|z^{-1}\mathcal{L}[\tilde{\eta}](z)\|^2 \, \mathrm{dIm}(z) \leq \left(2 + \int_{[-1,1]^c} y^{-2} \, \mathrm{d}y\right) C^2 < \infty,$$

and it follows by [3, Lemma 4.1] (or a slight modification of [8, Theorem 1 (Section 3.4)]) and (5.3) that $t \mapsto e^{\tilde{\varepsilon}t}\|\pi(t)\|$ belongs to L^2. For (5.1) to be satisfied it suffices to argue that $\sup_{t \geq 0} e^{\varepsilon t}|\pi_{ij}(t)| < \infty$, where π_{ij} refers to an arbitrarily chosen entry of π. Using integration by parts we find that

$$e^{\varepsilon t}|\pi_{ij}(t)| \leq |\pi_{ij}(0)| + \int_0^\infty e^{\varepsilon u}|\tilde{\eta}_{ij}|(\mathrm{d}u) + \varepsilon \int_0^\infty e^{\varepsilon u}|\pi_{ij}(u)| \, \mathrm{d}u. \tag{5.4}$$

It is clear that the first term on the right-hand side of (5.4) is finite, and the same holds for the second term by (1.5). For the last term we use the Cauchy–Schwarz inequality and the fact that $(u \mapsto e^{\tilde{\varepsilon}u}\pi_{ij}(u)) \in L^2$ to deduce

$$\left(\int_0^\infty e^{\varepsilon u}|\pi_{ij}(u)| \, \mathrm{d}u\right)^2 \leq \int_0^\infty e^{-2(\tilde{\varepsilon}-\varepsilon)u} \, \mathrm{d}u \int_0^\infty \left(e^{\tilde{\varepsilon}u}\pi_{ij}(u)\right)^2 \, \mathrm{d}u < \infty$$

and this ultimately allows us to conclude that (5.1) holds. To show (3.2) it suffices to argue that

$$\left(\int_s^t X^\top * \tilde{\eta}^\top(u) \, \mathrm{d}u\right)^\top = \int_0^\infty \pi(u)[X_{t-u} - X_{s-u}] \, \mathrm{d}u \tag{5.5}$$

almost surely for each $s < t$. Using that $\tilde{\eta}$ coincides with the Lebesgue–Stieltjes measure of π, together with integration by parts on the functions $v \mapsto \pi(v)$ and $v \mapsto \int_{s-v}^{t-v} X_u \, \mathrm{d}u$, we obtain

$$\begin{aligned}
\left(\int_s^t X^\top * \tilde{\eta}^\top(u) \, \mathrm{d}u\right)^\top &= \lim_{N \to \infty} \left(\int_{[0,N]} \left(\int_{s-v}^{t-v} X_u \, \mathrm{d}u\right)^\top \pi^\top(\mathrm{d}v)\right)^\top \\
&= \lim_{N \to \infty} \left(\pi(N) \int_{s-N}^{t-N} X_u \, \mathrm{d}u + \int_0^N \pi(u)[X_{t-u} - X_{s-u}] \, \mathrm{d}u\right).
\end{aligned} \tag{5.6}$$

By [2, Corollary A.3], since $(X_t)_{t \in \mathbb{R}}$ has stationary increments and $\mathbb{E}[\|X_t\|] < \infty$, there exist $\alpha, \beta > 0$ such that $\mathbb{E}[\|X_u\|] \leq \alpha + \beta|u|$ for all $u \in \mathbb{R}$. Consequently, we may as well find $\alpha^*, \beta^* > 0$ (depending on s and t) which satisfy

$$\mathbb{E}\left[\left\|\int_{s-N}^{t-N} X_u \, \mathrm{d}u\right\|\right] \leq \alpha^* + \beta^* N.$$

From this inequality, and due to (5.1), each entry of $\pi(N)\int_{s-N}^{t-N} X_u \, \mathrm{d}u$ converges to 0 in $L^1(\mathbb{P})$ as $N \to \infty$. The same type of reasoning gives that

$$\mathbb{E}\left[\int_0^\infty \|\pi(u)(X_{t-u} - X_{s-u})\| \, \mathrm{d}u\right] < \infty,$$

showing that each entry of $u \mapsto \pi(u)(X_{t-u} - X_{s-u})$ is almost surely integrable with respect to the Lebesgue measure and, hence, (5.6) implies (5.5). Finally, we need to argue that if $(X_t)_{t\in\mathbb{R}}$ is a solution to (3.1), $(\Pi_0 X_t)_{t\in\mathbb{R}}$ is stationary. Since $(X_t, Z_t)_{t\in\mathbb{R}}$ has stationary increments it follows immediately from (3.2) that $V_t^\lambda := \lambda^{-1} \int_t^{t+\lambda} \Pi_0 X_u \, \mathrm{d}u$, $t \in \mathbb{R}$, is a stationary process for any $\lambda > 0$. Since $(X_t)_{t\in\mathbb{R}}$ has stationary increments and $\mathbb{E}[\|X_t\|] < \infty$, it is continuous in $L^1(\mathbb{P})$ (see [2, Corollary A.3]), and hence V_t^λ converges to $\Pi_0 X_t$ in $L^1(\mathbb{P})$ as $\lambda \downarrow 0$ for any $t \in \mathbb{R}$. Consequently, $(\Pi_0 X_t)_{t\in\mathbb{R}}$ is stationary as well, and this finishes the proof. $\qquad\square$

Proof of Proposition 3.4. Assume that we are in case (i). If $z \mapsto h_\eta(z)^{-1}$ has no poles at all, then $\det h_\eta(z) = 0$ implies $\mathrm{Re}(z) < 0$ and the rank of Π_0 is n, and thus case (ii) is satisfied as well. If $z \mapsto h_\eta(z)^{-1}$ has a simple pole at 0, the rank r of $\Pi_0 = -h_\eta(0)$ is strictly less than n, and the residue formula in [23] implies that $(\alpha^\perp)^\top M \beta^\perp$ is invertible, where

$$M := \left.\frac{h_\eta(z) + \eta([0,\infty))}{z}\right|_{z=0} = I_n - \Pi([0,\infty))$$

is the derivative of h_η at 0, and $\alpha^\perp, \beta^\perp \in \mathbb{R}^{n\times(n-r)}$ are any two matrices of rank $n-r$ satisfying $\Pi_0^\top \alpha^\perp = \Pi_0 \beta^\perp = 0$. Conversely, if we are in case (ii), the facts that the zeroes of $z \mapsto \det h_\eta(z)$ are isolated points in $\{z \in \mathbb{C} : \mathrm{Re}(z) > -\varepsilon\}$ and $|\det h_\eta(z)| \neq 0$ for $|z|$ sufficiently large ensure the existence of a $\delta \in (0, \varepsilon)$ such that $\det h_\eta(z) \neq 0$ for all $z \in \mathbb{H}_\delta \setminus \{0\}$. If the rank r of Π_0 is n, $z \mapsto h_\eta(z)^{-1}$ has no poles at all on $\mathbb{H}_\delta$, and if $r < n$ and $(\alpha^\perp)^\top M \beta^\perp$ is invertible, the residue formula in [23] implies that $z \mapsto h_\eta(z)^{-1}$ has a simple pole at 0. $\qquad\square$

We will now turn to the construction of a solution to (3.1). Lemma 5.1 concerns the existence of the function f introduced in Theorem 3.5 and its properties.

Lemma 5.1. *Suppose that Condition 3.3 holds. Then there exists a unique function $f : \mathbb{R} \to \mathbb{R}^{n\times n}$ enjoying the following properties:*

(i) $\sup_{t\geq 0} e^{\varepsilon t}\|f(t)\| < \infty$ *for all $\varepsilon < \delta$.*

(ii) $\mathcal{L}[f](z) = I_n - z h_\eta(z)^{-1}$ *for all $z \in \mathbb{H}_\delta$.*

(iii) $f(t) = 0$ *for $t < 0$ and $f(t) = \int_0^t f * \eta(u) \, \mathrm{d}u - \eta([0,t])$ for $t \geq 0$.*

Proof. First note that, by assumption, $z \mapsto I_n - z h_\eta(z)^{-1}$ is an analytic function on $\mathbb{H}_\delta$. For any $\varepsilon \in (0, \delta)$ we will argue that

$$\sup_{\mathrm{Re}(z)>-\varepsilon} \int_\mathbb{R} \|I_n - z h_\eta(z)^{-1}\|^2 \, \mathrm{dIm}(z) < \infty. \tag{5.7}$$

If this is the case, a slight extension of the characterization of Hardy spaces (see [3, Lemma 4.1] or [8, Theorem 1 (Section 3.4)]) ensures the existence of a function $f : \mathbb{R} \to \mathbb{C}^{n\times n}$, vanishing on $(-\infty, 0)$, such that each entry of $t \mapsto e^{\varepsilon t} f(t)$ belongs to L^2 and $\mathcal{L}[f](z) = I_n - z h_\eta(z)^{-1}$ for all $z \in \mathbb{C}$ with $\mathrm{Re}(z) > -\varepsilon$. Since ε was arbitrary and, by uniqueness of the Laplace transform, the relation holds true for all $z \in \mathbb{H}_\delta$. Moreover, since $\overline{\mathcal{F}[f](-y)} = \mathcal{F}[f](y)$ for all $y \in \mathbb{R}$ ($\bar{z}$ denoting the complex conjugate of $z \in \mathbb{C}$), f takes values in $\mathbb{R}^{n\times n}$. To show (5.7) observe initially that

$$C_1 := \sup_{\mathrm{Re}(z)\geq-\varepsilon} \|\mathcal{L}[\eta](z)\| < \infty,$$

since $e^{\varepsilon t}|\eta_{ij}|(dt)$ is a finite measure for all $i,j = 1,\dots,n$. The same fact ensures that

(i) the absolute value of the determinant of $h_\eta(z)$ behaves as $|z|^n$ as $|z| \to \infty$, and

(ii) the dominating cofactors of $h_\eta(z)$ as $|z| \to \infty$ are those on the diagonal (the (i,i)-th cofactor, $i = 1,\dots,n$) and their absolute values behave as $|z|^{n-1}$ as $|z| \to \infty$.

In particular, $\|h_\eta(z)^{-1}\|$ behaves as $|z|^{-1}$ as $|z| \to \infty$ and, hence,

$$C_2 := \sup_{\text{Re}(z) \geq -\varepsilon} \|z h_\eta(z)^{-1}\| < \infty. \tag{5.8}$$

Consequently, for any $z \in \mathbb{C}$ with $\text{Re}(z) \geq -\varepsilon$,

$$\int_{[-1,1]} \|I_n - z h_\eta(z)^{-1}\|^2 \, d\text{Im}(z) \leq 2(\sqrt{n} + C_2)^2$$

and

$$\int_{[-1,1]^c} \|I_n - z h_\eta(z)^{-1}\|^2 \, d\text{Im}(z) \leq C_1^2 \int_{[-1,1]^c} \|h_\eta(z)^{-1}\|^2 \, d\text{Im}(z)$$

$$\leq (C_1 C_2)^2 \int_{[-1,1]^c} |x|^{-2} \, dx$$

using that $I_n - z h_\eta(z)^{-1} = -h_\eta(z)^{-1} \mathcal{L}[\eta](z)$ and that $\|\cdot\|$ is a submultiplicative norm. This verifies (5.7) and, hence, proves the existence of a function $f \colon \mathbb{R} \to \mathbb{R}^{n \times n}$ with $f(t) = 0$ for $t < 0$ and $\mathcal{L}[f](z) = I_n - z h_\eta(z)^{-1}$ for $z \in \mathbb{H}_\delta$ (in particular, verifying (ii)). To show (iii), note that

$$\mathcal{L}[f * \eta](z) - \mathcal{L}[\eta](z) = -z h_\eta(z)^{-1} \mathcal{L}[\eta](z) = z \mathcal{L}[f](z), \qquad z \in \mathbb{H}_\delta. \tag{5.9}$$

By using the observation in (5.2) on the measures $f * \eta(u) \, du$ and η together with (5.9) we establish that

$$f(t) = \int_0^t f * \eta(u) \, du - \eta([0,t]) \tag{5.10}$$

for almost all $t \geq 0$. Since we can choose f to satisfy (5.10) for all $t \geq 0$ without modifying its Laplace transform, we have established (iii). By the càdlàg property of f, the uniqueness part follows as well. Finally, we need to argue that (i) holds, and for this it suffices to argue that $\sup_{t \geq 0} e^{\varepsilon t}|f_{ij}(t)| < \infty$ for all $\varepsilon \in (0,\delta)$ where f_{ij} refers to an arbitrarily chosen entry of f. From (5.10) it follows that the Lebesgue–Stieltjes measure of f_{ij} is given by $\sum_{k=1}^n f_{ik} * \eta_{kj}(t) \, dt - \eta_{ij}(dt)$. Therefore, integration by parts yields

$$e^{\varepsilon t}|f_{ij}(t)| \leq |f_{ij}(0)| + \sum_{k=1}^n \int_0^\infty e^{\varepsilon u}|f_{ik}| * |\eta_{kj}|(u) \, du + \int_0^\infty e^{\varepsilon u} |\eta_{ij}|(du)$$
$$+ \varepsilon \int_0^\infty e^{\varepsilon u}|f_{ij}(u)| \, du, \tag{5.11}$$

so to prove the result we only need to argue that each term on right-hand side of (5.11) is finite. The assumption (1.5) implies immediately that $\int_0^\infty e^{\varepsilon u}|\eta_{ij}|(du) < \infty$.

As noted in the beginning of the proof, $u \mapsto e^{\varepsilon' u} f_{ij}(u)$ belongs to L^2 for an arbitrary $\varepsilon' \in (0, \delta)$. In particular, for $\varepsilon' \in (\varepsilon, \delta)$,

$$\int_0^\infty e^{\varepsilon u} |f_{ij}(u)| \, \mathrm{d}u \le \left(\int_0^\infty e^{-2(\varepsilon'-\varepsilon)u} \, \mathrm{d}u \int_0^\infty \left(e^{\varepsilon' u} f_{ij}(u) \right)^2 \, \mathrm{d}u \right)^{1/2} < \infty.$$

Finally, since

$$\int_0^\infty e^{\varepsilon u} |f_{ik}| * |\eta_{kj}|(u) \, \mathrm{d}u = \int_{[0,\infty)} e^{\varepsilon u} |\eta_{kj}|(\mathrm{d}u) \int_0^\infty e^{\varepsilon u} |f_{ik}(u)| \, \mathrm{d}u,$$

it follows by the former arguments that this term is finite as well, and this concludes the proof. □

Remark 5.2. Suppose that $\det h_\eta(z) \ne 0$ for all $z \in \mathbb{H}_\delta$ so that Condition 3.3 is satisfied and $z \mapsto h_\eta(z)^{-1}$ has no poles. Under this assumption it was argued in [4, Proposition 5.1] that there exists a function $g \colon \mathbb{R} \to \mathbb{R}^{n \times n}$, which is vanishing on $(-\infty, 0)$, is absolutely continuous on $[0, \infty)$ and decays exponentially fast at ∞, such that $\mathcal{L}[g](z) = h_\eta(z)^{-1}$ for $z \in \mathbb{H}_\delta$. Since property (ii) implies $\mathcal{L}[f](z) = -h_\eta(z)^{-1} \mathcal{L}[\eta](z)$, it must be the case that $f = -g * \eta$.

Proof of Theorem 3.5. The existence of f is covered by Lemma 5.1. According to [2, Corollary A.3] and by equivalence of matrix norms, we may choose $\alpha, \beta, \gamma > 0$ such that $\mathbb{E}[\|Z_t\|] \le \alpha + \beta|t|$ for all $t \in \mathbb{R}$ and $\sum_{i,j=1}^n |a_{ij}| \le \gamma \|A\|$ for all $A = [a_{ij}] \in \mathbb{R}^{n \times n}$. Using this together with property Lemma 5.1(i), we obtain that

$$\mathbb{E}\left[\int_0^\infty \|f(u) Z_{t-u}\| \, \mathrm{d}u \right] \le (\alpha + \beta|t|)\gamma \int_0^\infty \|f(u)\| \, \mathrm{d}u + \beta\gamma \int_0^\infty \|f(u)\| |u| \, \mathrm{d}u < \infty.$$

In particular, this shows that $u \mapsto f(u) Z_{t-u}$ belongs to L^1 almost surely and, hence, $(X_t)_{t \in \mathbb{R}}$ given by (3.5) is a well-defined process. We will now split the proof in two parts: first, we argue that $(X_t)_{t \in \mathbb{R}}$ given by (3.5) is indeed a solution to (3.1) (existence) and, next, we show that any other solution necessarily admits this representation (uniqueness).

Existence: Note that $\mathbb{E}[\|Z_t\|^2] \le \gamma_1 + \gamma_2 t^2$ for all t and suitable $\gamma_1, \gamma_2 > 0$ by [2, Corollary A.3], so we may use similar reasoning as above to deduce that $\mathbb{E}[\|X_t\|^2] < \infty$ for all $t \in \mathbb{R}$. Moreover, since $(X_t)_{t \in \mathbb{R}}$ solves (3.1) if and only if it solves (3.2), we may and do assume $\xi = 0$ so that

$$X_t = Z_t - \int_0^\infty f(u) Z_{t-u} \, \mathrm{d}u, \qquad t \in \mathbb{R}.$$

To show that $(X_t)_{t \in \mathbb{R}}$ satisfies (3.1), we need to argue that

$$X_t - X_s - (Z_t - Z_s) = \int_s^t \eta * X(u) \, \mathrm{d}u, \qquad s < t. \tag{5.12}$$

To this end, note that

$$X_t - X_s - (Z_t - Z_s) = \int_{\mathbb{R}} \eta((s-u, t-u]) Z_u \, \mathrm{d}u - \int_{\mathbb{R}} \int_{s-u}^{t-u} f * \eta(v) \, \mathrm{d}v \, Z_u \, \mathrm{d}u \tag{5.13}$$

and

$$\int_s^t \eta * X(u) \, du$$

$$= \int_{\mathbb{R}} \eta((s-u, t-u]) X_u \, du \tag{5.14}$$

$$= \int_{\mathbb{R}} \eta((s-u, t-u]) Z_u \, du - \int_{\mathbb{R}} \eta((s-u, t-u]) \int_{\mathbb{R}} f(v) Z_{u-v} \, dv \, du$$

using Lemma 5.1(iii) and (3.5), respectively. Moreover, by comparing their Laplace transforms, one can verify that $\eta * f := (f^\top * \eta^\top)^\top = f * \eta$ and, thus,

$$\int_{\mathbb{R}} \int_{s-u}^{t-u} f * \eta(v) \, dv \, Z_u \, du = \int_{\mathbb{R}} \int_{\mathbb{R}} \eta((s-u-v, t-u-v]) f(v) \, dv \, Z_u \, du$$

$$= \int_{\mathbb{R}} \eta((s-u, t-u]) \int_{\mathbb{R}} f(v) Z_{u-v} \, dv \, du \tag{5.15}$$

It follows by combining (5.13)–(5.15) that (5.12) is satisfied. Recall that, for $(X_t)_{t \in \mathbb{R}}$ to be a solution, we need to argue that $(X_t, Z_t)_{t \in \mathbb{R}}$ has stationary increments. However, since

$$X_{t+h} - X_h = (Z_{t+h} - Z_h) - \int_0^\infty f(u)[(Z_{t-u+h} - Z_h) - (Z_{-u+h} - Z_h)] \, du, \quad t \in \mathbb{R},$$

and the distribution of $(Z_{t+h} - Z_h)_{t \in \mathbb{R}}$ does not depend on h, it follows that the distribution of $(X_{t+h} - X_h, Z_{t+h} - Z_h)_{t \in \mathbb{R}}$ does not depend on h. A rigorous argument can be carried out by approximating the above Lebesgue integral by Riemann sums in $L^1(\mathbb{P})$; since this procedure is similar to the one used in the proof of [4, Theorem 3.1], we omit the details here.

Uniqueness: Suppose that $(Y_t)_{t \in \mathbb{R}}$ satisfies (3.1), $\mathbb{E}[\|Y_t\|^2] < \infty$ for all $t \in \mathbb{R}$, and $(Y_t, Z_t)_{t \in \mathbb{R}}$ has stationary increments. In addition, suppose for the moment that we have already shown that

$$Y_t - Y_s = X_t - X_s, \qquad s, t \in \mathbb{R}. \tag{5.16}$$

Then it follows from (3.2) that $V_\lambda := \lambda^{-1} \int_0^\lambda \Pi_0(Y_u - X_u) \, du = 0$ almost surely for all $\lambda > 0$. On the other hand, since $(X_t)_{t \in \mathbb{R}}$ and $(Y_t)_{t \in \mathbb{R}}$ have stationary increments, they are continuous in $L^1(\mathbb{P})$ and, hence, $V_\lambda \to \Pi_0(Y_0 - X_0)$ in $L^1(\mathbb{P})$ as $\lambda \downarrow 0$. This shows that $Y_0 - X_0$ belongs to the null space of Π_0 almost surely and, consequently, $(Y_t)_{t \in \mathbb{R}}$ is necessarily of the form (3.5). The remaining part of the proof concerns showing (5.16) or, equivalently, the process $\Delta_h Y_t := Y_t - Y_{t-h}$, $t \in \mathbb{R}$, is unique for any $h > 0$. We will rely on the same type of ideas as in the proof of [6, Proposition 7] and [10, Proposition 4.5]. Suppose first that Π_0 has reduced rank $r \in (0, n)$ and let $\alpha, \beta \in \mathbb{R}^{n \times r}$ be a rank decomposition of Π_0 as in Remark 3.2. Moreover, let $\alpha^\perp, \beta^\perp \in \mathbb{R}^{n \times (n-r)}$ be matrices of rank $n - r$ such that $\alpha^\top \alpha^\perp = \beta^\top \beta^\perp = 0$. Then it follows from Theorem 3.1 that

$$\alpha^\top \Delta_h Y_t = \alpha^\top \alpha \beta^\top \int_0^h Y_{t-u} \, du + \alpha^\top \int_0^\infty \pi(u) \Delta_h Y_{t-u} \, du + \alpha^\top \Delta_h Z_t$$

$$\text{and} \qquad (\alpha^\perp)^\top \Delta_h Y_t = (\alpha^\perp)^\top \int_0^\infty \pi(u) \Delta_h Y_{t-u} \, du + (\alpha^\perp)^\top \Delta_h Z_t \tag{5.17}$$

175

for each $t \in \mathbb{R}$. Define the stationary processes

$$U_t = (\beta^\top \beta)^{-1} \beta^\top Y_t \quad \text{and} \quad V_t = ((\beta^\perp)^\top \beta^\perp)^{-1} (\beta^\perp)^\top \Delta_h Y_t, \qquad t \in \mathbb{R}.$$

By using that $\Delta_h Y_t = \beta \Delta_h U_t + \beta^\perp V_t$ and rearranging terms, (5.17) can be written as

$$\mu * \begin{bmatrix} U_t \\ V_t \end{bmatrix} = \tilde{Z}_t, \qquad t \in \mathbb{R}, \tag{5.18}$$

where

$$\mu = \begin{bmatrix} \alpha^\top \big[(\delta_0 - \delta_h) I_n - (\alpha\beta^\top \mathbb{1}_{(0,h]} + \Delta_h \pi) \cdot \lambda \big] \beta & \alpha^\top [\delta_0 I_n - \pi \cdot \lambda] \beta^\perp \\ (\alpha^\perp)^\top [(\delta_0 - \delta_h) I_n - (\Delta_h \pi) \cdot \lambda] \beta & (\alpha^\perp)^\top [\delta_0 I_n - \pi \cdot \lambda] \beta^\perp \end{bmatrix}$$

and $\tilde{Z}_t = [\Delta_h Z_t^\top \alpha, \Delta_h Z_t^\top \alpha^\perp]^\top$. (For brevity, we have used the notation $f \cdot \lambda(\mathrm{d}u) = f(u)\,\mathrm{d}u$.) Now, note that the Fourier transform $\mathcal{F}[\mu]$ of μ takes the form

$$\mathcal{F}[\mu](y)$$
$$= \begin{bmatrix} \alpha^\top \big[(1 - e^{-ihy})[I_n - \mathcal{F}[\pi](y)] - \alpha\beta^\top \mathcal{F}[\mathbb{1}_{(0,h]}](y) \big] \beta & \alpha^\top [I_n - \mathcal{F}[\pi](y)]\beta^\perp \\ (\alpha^\perp)^\top (1 - e^{-ihy})[I_n - \mathcal{F}[\pi](y)]\beta & (\alpha^\perp)^\top [I_n - \mathcal{F}[\pi](y)]\beta^\perp \end{bmatrix}.$$

In particular, it follows that

$$\det \mathcal{F}[\mu](0) = \det \begin{bmatrix} -\alpha^\top \alpha \beta^\top \beta h & \alpha^\top [I_n - \mathcal{F}[\pi](0)]\beta^\perp \\ 0 & (\alpha^\perp)^\top [I_n - \mathcal{F}[\pi](0)]\beta^\perp \end{bmatrix}$$
$$= (-h)^r \det(\alpha^\top \alpha) \det(\beta^\top \beta) \det\big((\alpha^\perp)^\top [I_n - \Pi([0,\infty))]\beta^\perp\big),$$

which is non-zero by Proposition 3.4. Consequently, it follows from (5.18) that the means of $(U_t)_{t\in\mathbb{R}}$ and $(V_t)_{t\in\mathbb{R}}$ are uniquely determined by the one of $(\tilde{Z}_t)_{t\in\mathbb{R}}$; namely $[\mathbb{E}[U_0]^\top, \mathbb{E}[V_0]^\top]^\top = \mu([0,\infty))^{-1}\mathbb{E}[\tilde{Z}_0]$. For this reason we may without loss of generality assume that $(U_t)_{t\in\mathbb{R}}$, $(V_t)_{t\in\mathbb{R}}$ and $(\tilde{Z}_t)_{t\in\mathbb{R}}$ are all zero mean processes so that they admit spectral representations. Recall that the spectral representation of a stationary, square integrable and zero mean process $(S_t)_{t\in\mathbb{R}}$ is given by $S_t = \int_{\mathbb{R}} e^{ity} \Lambda_S(\mathrm{d}y)$, $t \in \mathbb{R}$, where $(\Lambda_S(t))_{t\in\mathbb{R}}$ is a complex-valued spectral process which is square integrable and continuous in $L^2(\mathbb{P})$, and which has orthogonal increments. (Integration with respect to Λ_S can be defined as in [11, pp. 388–390] for all functions in $L^2(F_S)$, F_S being the spectral distribution of $(S_t)_{t\in\mathbb{R}}$.) Consequently, by letting Λ_U, Λ_V and $\Lambda_{\tilde{Z}}$ be the spectral processes corresponding to $(U_t)_{t\in\mathbb{R}}$, $(V_t)_{t\in\mathbb{R}}$ and $(\tilde{Z}_t)_{t\in\mathbb{R}}$, equation (5.18) can be rephrased as

$$\int_{\mathbb{R}} e^{ity} \mathcal{F}[\mu](y) \begin{bmatrix} \Lambda_U \\ \Lambda_V \end{bmatrix}(\mathrm{d}y) = \int_{\mathbb{R}} e^{ity} \Lambda_{\tilde{Z}}(\mathrm{d}y), \qquad t \in \mathbb{R}. \tag{5.19}$$

Here we have used a stochastic Fubini result for spectral processes, e.g., [7, Proposition A.1]. Since the functions $y \mapsto e^{ity}$, $t \in \mathbb{R}$, are dense in $L^2(F)$ for any finite measure F (cf. [25, p. 150]), the relation (5.19) remains true when $y \mapsto e^{ity}$ is replaced by any measurable and, say, bounded function $g \colon \mathbb{R} \to \mathbb{C}^{n\times n}$. In particular, we will choose

$$g(y) = e^{ity}(iy)h_\eta(iy)^{-1} \begin{bmatrix} \alpha^\top \\ (\alpha^\perp)^\top \end{bmatrix}^{-1}, \qquad y \neq 0,$$

and $g(0) = [0_{n \times r}, \beta^\perp] \mathcal{F}[\mu](0)^{-1}$. Note that by (5.8), g is indeed bounded. After observing that

$$\mathcal{F}[\mu](y) = \begin{bmatrix} \alpha^\top \\ (\alpha^\perp)^\top \end{bmatrix} \begin{bmatrix} (1 - e^{-ihy})[I_n - \mathcal{F}[\pi](y) + (iy)^{-1}\alpha\beta^\top] & I_n - \mathcal{F}[\pi](y) \\ (1 - e^{-ihy})[I_n - \mathcal{F}[\pi](y)] & I_n - \mathcal{F}[\pi](y) \end{bmatrix} \begin{bmatrix} \beta & \beta^\perp \end{bmatrix}$$

$$= (iy)^{-1} \begin{bmatrix} \alpha^\top \\ (\alpha^\perp)^\top \end{bmatrix} h_\eta(iy) \begin{bmatrix} (1 - e^{-ihy})\beta & \beta^\perp \end{bmatrix}$$

for $y \neq 0$, it is easy to verify that $g(y)\mathcal{F}[\mu](y) = [\beta(e^{ity} - e^{i(t-h)y}), \beta^\perp e^{ity}]$ for all $y \in \mathbb{R}$. Consequently, it follows from (5.19) that

$$\Delta_h Y_t = \int_{\mathbb{R}} \begin{bmatrix} \beta(e^{ity} - e^{i(t-h)y}) & \beta^\perp e^{ity} \end{bmatrix} \begin{bmatrix} \Lambda_U \\ \Lambda_V \end{bmatrix} (\mathrm{d}y) = \int_{\mathbb{R}} g(y) \Lambda_{\tilde{Z}}(\mathrm{d}y),$$

showing that the process $(\Delta_h Y_t)_{t \in \mathbb{R}}$ is uniquely determined by $(\tilde{Z}_t)_{t \in \mathbb{R}}$. Now we only need to ague that this type of uniqueness also holds when Π_0 is invertible and $\Pi_0 = 0$. If Π_0 is invertible, $(Y_t)_{t \in \mathbb{R}}$ must in fact be stationary (cf. Remark 3.2), and by [4, Theorem 3.1] there is only one process enjoying this property. If $\Pi_0 = 0$, the case is simpler than if $r \in (0, n)$, since here we only need to consider the second equation of (5.17) with $\alpha^\perp = I_n$ and the spectral representation of $(\Delta_h Y_t)_{t \in \mathbb{R}}$. To avoid too many repetitions we leave out the details. □

Proof of Corollary 3.7. As noted right before the statement we only need to argue that (3.6) is satisfied with respect to the definition (3.7). In order to do so, note that

$$\left[\int_0^\infty f(u)(Z_t - Z_{t-u}) \, \mathrm{d}u \right]_i = \sum_{j=1}^n \int_0^\infty \mathcal{I}_j(\mathbb{1}_{(t-u,t]}) f_{ij}(u) \, \mathrm{d}u$$

$$= \sum_{j=1}^n \mathcal{I}_j \left(\mathbb{1}_{[0,\infty)}(t - \cdot) \int_{t-\cdot}^\infty f_{ij}(u) \, \mathrm{d}u \right)$$

$$= \left[\int_{-\infty}^t C(t-u) \, \mathrm{d}Z_u \right]_i,$$

where $C(t) = 0$ for $t < 0$ and $C(t) = \int_t^\infty f(u) \, \mathrm{d}u$ for $t \geq 0$. Now observe that, for $z \in \mathbb{C}$ with $\mathrm{Re}(z) < 0$,

$$\mathcal{L}[C](z) = z^{-1} \left(\int_0^\infty f(t) \, \mathrm{d}t - \mathcal{L}[f](z) \right) = h_\eta(z)^{-1} - z^{-1} C_0 \tag{5.20}$$

using Remark 3.6 and Lemma 5.1(ii). Since both sides of (5.20) are analytic functions on $\mathbb{H}_\delta$, the equality holds true on $\mathbb{H}_\delta$. This proves that C can be characterized as in the statement of Theorem 1.2 and, thus, finishes the proof. □

Remark 5.3. As was the case for the function f of Lemma 5.1, C can also be obtained as a solution to a multivariate delay differential equation. Specifically, the shifted function $\tilde{C}(t) = C_0 + C(t)$, $t \geq 0$, satisfies

$$\tilde{C}(t) - \tilde{C}(s) = \int_s^t \tilde{C} * \eta(u) \, \mathrm{d}u, \qquad 0 \leq s < t. \tag{5.21}$$

By Theorem 3.5 the initial condition is $\tilde{C}(0) = I_n$. To see that (5.21) holds note that, for fixed $0 \le s < t$, Lemma 5.1(iii) implies

$$\tilde{C}(t) - \tilde{C}(s) = -\int_s^t f(u)\,\mathrm{d}u = \int_s^t \Big(\eta([0,u]) - \int_0^u f * \eta(v)\,\mathrm{d}v\Big)\,\mathrm{d}u,$$

and

$$\int_0^u f * \eta(v)\,\mathrm{d}v = \int_{[0,\infty)} \int_0^{u-r} f(v)\,\mathrm{d}v\,\eta(\mathrm{d}r) = \eta([0,u]) - \tilde{C} * \eta(u)$$

by Fubini's theorem. In the same way as in the proof of Theorem 3.1, one can rely on integration by parts to write (5.21) in error correction form:

$$\tilde{C}(t) - \tilde{C}(s) = \int_s^t \tilde{C}(u)\Pi_0\,\mathrm{d}u + \int_0^\infty [\tilde{C}(t-u) - \tilde{C}(s-u)]\Pi(\mathrm{d}u), \qquad 0 \le s < t.$$

Proof of Theorem 1.2. In view of Proposition 3.4 we may assume that Condition 3.3 is satisfied. Consequently, by using [4, Example 4.2], which states that an n-dimensional Lévy process with finite first moments is a regular integrator (that is, there exist $\mathcal{I}_1,\ldots,\mathcal{I}_n$ satisfying Corollary 3.7(i)–(ii)), the result is an immediate consequence of Corollary 3.7. $\qquad\square$

Proof of Theorem 4.4. Note that, by Condition 4.3(iv), we can choose $\varepsilon > 0$ such that $\det Q(z) \ne 0$ whenever $\mathrm{Re}(z) \ge -\varepsilon$. To show that (4.5) is well-defined and satisfies (1.5) for some $\delta > 0$ it suffices to establish

$$\sup_{\mathrm{Re}(z) > -\varepsilon} \int_{\mathbb{R}} \|I_n z - \eta_0 - Q(z)^{-1}P(z)\|^2\,\mathrm{d}\mathrm{Im}(z) < \infty. \tag{5.22}$$

(See, e.g., the beginning of the proof of Lemma 5.1.) It is straightforward to verify that $\eta_0 = Q_1 - P_1$ is chosen such that $z \mapsto Q(z)(I_n z - \eta_0) - P(z)$ is a polynomial of at most order $p - 2$. Consequently, the integrand in (5.22) is of the form $\|Q(z)^{-1}R(z)\|^2$, where Q is of strictly larger degree than R, and hence it follows by sub-multiplicativity of $\|\cdot\|$ that it decays at least as fast as $|z|^{-2}$ when $|z| \to \infty$. Since the integrand is also bounded on compact subsets of $\{z \in \mathbb{C} : \mathrm{Re}(z) \le \varepsilon\}$ we conclude that (5.22) is satisfied.

Next, we will show that the assumptions of Theorem 1.2 are satisfied (which, by Proposition 3.4, is equivalent to showing that Condition 3.3 holds). Observe that $h_\eta(z) = Q(z)^{-1}P(z)$ when $\mathrm{Re}(z) > -\varepsilon$, so by (i) and (iv) in Condition 4.3 it follows that $\det h_\eta(z) = 0$ implies $\mathrm{Re}(z) < 0$ or $z = 0$. Now, a Taylor expansion of $z \mapsto Q(z)^{-1}$ around 0 yields

$$\mathcal{L}[\eta](z) = \eta([0,\infty)) + \big(I_n + Q_{p-1}^{-1}Q_{p-2}Q_{p-1}^{-1}P_p - Q_{p-1}^{-1}P_{p-1}\big)z + O(z^2), \quad |z| \to 0,$$

and hence

$$\Pi([0,\infty)) = \frac{\mathcal{L}[\eta](z) - \eta([0,\infty))}{z}\bigg|_{z=0} = I_n - Q_{p-1}^{-1}[P_{p-1} - Q_{p-2}Q_{p-1}^{-1}P_p].$$

Let $\tilde{\alpha} = Q_{p-1}^\top \alpha^\perp$ and $\tilde{\beta} = \beta^\perp$, and note that these matrices are of rank $n - r$ and satisfy $\Pi_0^\top \tilde{\alpha} = \Pi_0 \tilde{\beta} = 0$. Thanks to Condition 4.3(iii), the matrix

$$\tilde{\alpha}^\top(I_n - \Pi([0,\infty)))\tilde{\beta} = (\alpha^\perp)^\top P_{p-1}\beta^\perp$$

is invertible, so the assumptions of Theorem 1.2 are satisfied. The remaining statements are now simply consequences of Corollary 3.7.

$\qquad\square$

Acknowledgments

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] 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.

[5] 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.

[6] Comte, F. (1999). Discrete and continuous time cointegration. *J. Econometrics* 88(2), 207–226. DOI: 10.1016/S0304-4076(98)00025-6.

[7] Davis, R.A., M.S. Nielsen and V. Rohde (2019). Stochastic differential equations with a fractionally filtered delay: a semimartingale model for long-range dependent processes. *Bernoulli.* Forthcoming.

[8] Dym, H. and H.P McKean (2016). *Séries et intégrales de Fourier.* Vol. 13. Nouvelle Bibliothèque Mathématique [New Mathematics Library]. Translated from the 1972 English original by Éric Kouris. Cassini, Paris.

[9] Engle, R.F. and C.W.J. Granger (1987). Co-integration and error correction: representation, estimation, and testing. *Econometrica* 55(2), 251–276. DOI: 10.2307/1913236.

[10] Fasen-Hartmann, V. and M. Scholz (2016). Cointegrated Continuous-time Linear State Space and MCARMA Models. arXiv: 1611.07876.

[11] Grimmett, G. and D. Stirzaker (2001). *Probability and random processes.* Oxford University Press.

[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]   Hansen, P.R. (2005). Granger's representation theorem: a closed-form expression for $I(1)$ processes. *Econom. J.* 8(1), 23–38. DOI: 10.1111/j.1368-423X.2005.00149.x.

[14]   Johansen, S. (1991). Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica* 59(6), 1551–1580. DOI: 10.2307/2938278.

[15]   Johansen, S. (2009). "Cointegration: Overview and development". *Handbook of financial time series*. Springer, 671–693.

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

[17]   Marquardt, T. and R. Stelzer (2007). Multivariate CARMA processes. *Stochastic Process. Appl.* 117(1), 96–120. DOI: 10.1016/j.spa.2006.05.014.

[18]   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.

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

[20]   Runkle, D.E. (2002). Vector autoregressions and reality. *J. Bus. Econom. Statist.* 20(1), 128–133. DOI: 10.1198/073500102753410435.

[21]   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. DOI: 10.2969/jmsj/04620245.

[22]   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.

[23]   Schumacher, J.M (1991). "System-theoretic trends in econometrics". *Mathematical system theory*. Springer, Berlin, 559–577.

[24]   Sims, C.A. (1980). Macroeconomics and reality. *Econometrica*, 1–48.

[25]   Yaglom, A.M (1987). *Correlation theory of stationary and related random functions. Vol. I*. Springer Series in Statistics. Basic results. New York: Springer-Verlag.

Low Frequency Estimation of Lévy-Driven Moving Averages

Mikkel Slot Nielsen

Abstract

In this paper we consider least squares estimation of the driving kernel of a moving average and argue that, under mild regularity conditions and a decay condition on the kernel, the suggested estimator is consistent and asymptotically normal. On one hand this result unifies scattered results of the literature on low frequency estimation of moving averages, and on the other hand it emphasizes the validity of inference also in cases where the moving average is not strongly mixing. We assess the performance of the estimator through a simulation study.

Keywords: Least squares estimation; Lévy-driven moving averages; Long memory processes

1 Introduction

The class of continuous time Lévy-driven moving averages of the form

$$X_t = \int_{\mathbb{R}} \varphi(t-s)\,\mathrm{d}L_s, \qquad t \in \mathbb{R}, \tag{1.1}$$

where $(L_t)_{t\in\mathbb{R}}$ is a Lévy process with $\mathbb{E}[L_1] = 0$ and $\mathbb{E}[L_1^2] < \infty$ and $\varphi \in L^2$, is large and has received much attention in earlier literature. Part of the reason for this popularity might be explained by the celebrated discrete time counterpart (in particular, ARMA processes) as well as the Wold–Karhunen decomposition. The latter states that, up to a drift term, essentially any centered and square integrable stationary process may be written in the form (1.1) with $(L_t)_{t\in\mathbb{R}}$ replaced by a process with second order stationary and orthogonal increments ([2, 16]). While φ may be specified directly, one often characterizes it in the spectral domain in terms of its Fourier transform,

$$\mathcal{F}[\varphi](y) = \int_0^\infty e^{-iyt}\varphi(t)\,\mathrm{d}t, \qquad y \in \mathbb{R}.$$

One class in the framework of (1.1) is the continuous time ARMA (CARMA) processes, where $\mathcal{F}[\varphi](y) = Q(iy)/P(iy)$ for $y \in \mathbb{R}$ and some monic polynomials $P, Q \colon \mathbb{C} \to \mathbb{C}$ with real coefficients, $p := \deg(P) > \deg(Q) =: q$, and $P(z) \neq 0$ for all $z \in \mathbb{C}$ with $\mathrm{Re}(z) \geq 0$. One may regard a CARMA process as the solution to the formal equation

$$P(D)X_t = Q(D)DL_t, \qquad t \in \mathbb{R}, \tag{1.2}$$

where D denotes the derivative with respect to time. Indeed, by heuristically applying the Fourier transform to (1.2) and rearranging terms one reaches the conclusion that $(X_t)_{t \in \mathbb{R}}$ is the convolution between φ and $(DL_t)_{t \in \mathbb{R}}$. The simplest CARMA process, which has been particularly popular, is the Ornstein–Uhlenbeck process which corresponds to $p = 1$ and $q = 0$. CARMA processes have been used as models for various quantities including stochastic volatility, electricity spot prices and temperature dynamics ([5, 12, 24]), and there exists a vast amount of literature on their existence, uniqueness and representations as well as generalizations to the multivariate and fractional noise setting ([6, 19, 20]). Another class consists of affine stochastic delay differential equations (SDDEs) of the form

$$\mathrm{d}X_t = \int_{[0,\infty)} X_{t-s}\,\eta(\mathrm{d}s)\,\mathrm{d}t + \mathrm{d}L_t, \qquad t \in \mathbb{R}. \tag{1.3}$$

Here η is a suitable finite signed measure satisfying $z - \int_{[0,\infty)} e^{-zt}\,\eta(\mathrm{d}t) \neq 0$ for all $z \in \mathbb{C}$ with $\mathrm{Re}(z) \geq 0$. In this case, the solution of (1.3) is a moving average and the kernel φ is determined by the relation

$$\mathcal{F}[\varphi](y) = \left(iy - \int_{[0,\infty)} e^{-iyt}\,\eta(\mathrm{d}t)\right)^{-1}, \qquad y \in \mathbb{R}. \tag{1.4}$$

The choice $\eta = -\lambda\delta_0$, $\lambda > 0$, results in the Ornstein–Uhlenbeck process; a related example is considered in Example 3.2. (We use the notation δ_x for the Dirac measure at x.) Some relevant references on SDDEs are [4, 14].

Estimation of P and Q, given a sample $X_{n:1} = [X_{n\Delta}, X_{(n-1)\Delta}, \ldots, X_\Delta]^\top$ of equidistant observations of a CARMA process sampled at some frequency $\Delta > 0$, has received some attention. For instance, Brockwell et al. [8] show that a sampled CARMA process $(X_{t\Delta})_{t \in \mathbb{Z}}$ is a weak ARMA process. By combining this with the fact that CARMA processes are strongly mixing ([20, Proposition 3.34]), they can rely on general results of Francq and Zakoïan [11] to prove strong consistency and asymptotic normality for an estimator of least squares type. Other papers dealing with low frequency estimation of CARMA processes are [10, 22]. Küchler and Sørensen [18] studied low frequency parametric estimation of the measure η in (1.3) in case the support of the measure is known to be contained in some compact set and $(L_t)_{t \in \mathbb{R}}$ is a Brownian motion. They used results about strong mixing properties of Gaussian processes to obtain consistency and asymptotic normality of a maximum pseudo likelihood estimator. Generally, these results for CARMA processes and solutions to SDDEs cannot be extended to other parametric classes of φ in (1.1), since they use specific properties of the subclass in question. Indeed, strong mixing conditions may be difficult to verify and there exist several non-trivial examples of processes which are not strongly mixing (see the discussion and the corresponding examples in [1]). There exist results on strong mixing properties for discrete time moving averages,

such as [13], but to the best of our knowledge, no version for the continuous time counterpart (1.1) has been proven (not even when it is sampled on a discrete grid).

In this paper we provide a result (Theorem 2.4) concerning consistency and asymptotic normality of an estimator of least squares type when parametrically estimating φ in (1.1) from a sample of low frequency observations $X_{n:1}$. To be more concrete, let Θ be a compact subset of $\mathbb{R}^d$, let $\varphi_\theta \in L^2$ for $\theta \in \Theta$, and suppose that $(X_t)_{t \in \mathbb{R}}$ follows the model (1.1) with $\varphi = \varphi_{\theta_0}$ for some unknown parameter $\theta_0 \in \Theta$. Then we will be interested in the estimator $\hat{\theta}_n$ obtained as a point, which minimizes

$$\sum_{t=k+1}^{n} (X_{t\Delta} - \pi_k(X_{t\Delta}; \theta))^2, \qquad \theta \in \Theta, \tag{1.5}$$

where $\pi_k(X_{t\Delta}; \theta)$ denotes the projection of $X_{t\Delta}$ onto the linear $L^2(\mathbb{P})$ subspace spanned by $X_{(t-1)\Delta}, \ldots, X_{(t-k)\Delta}$ under the model (1.1) with $\varphi = \varphi_\theta$. Besides the usual identifiability and smoothness conditions, the conditions given here to ensure asymptotic normality of the estimator concern the decay of the kernel. This ensures that we can apply our result in situations where the process is not, or cannot be verified to be, strongly mixing. In cases where φ_θ can be specified directly, e.g., when it belongs to the class of CARMA processes or fractional noise processes, it is a straightforward task to check the decay condition, but even when the kernel is not explicitly known (e.g., when it can only be specified through its Fourier transform as in the SDDE case) one can sometimes still assess its decay properties. In Example 2.3 we consider some situations where the imposed decay condition is satisfied. Section 3 demonstrates the properties of the estimator through a simulation study.

2 Estimators of interest and asymptotic results

Let $(L_t)_{t \in \mathbb{R}}$ be a centered Lévy process with $\mathbb{E}[L_1] = 0$ and $\mathbb{E}[L_1^4] < \infty$, and suppose that $\mathbb{E}[L_1^2] = 1$. Moreover, let Θ be a compact subset of $\mathbb{R}^d$ and, for each $\theta \in \Theta$, suppose that $\varphi_\theta \in L^2$ and define the corresponding stationary process $(X_t^\theta)_{t \in \mathbb{R}}$ by

$$X_t^\theta = \int_{\mathbb{R}} \varphi_\theta(t - s) \, dL_s, \qquad t \in \mathbb{R}. \tag{2.1}$$

To avoid trivial cases we assume that $\{t : \varphi_\theta(t) \neq 0\}$ is not a Lebesgue null set. Let γ_θ be the autocovariance function of $(X_t^\theta)_{t \in \mathbb{R}}$, that is,

$$\gamma_\theta(h) := \mathbb{E}[X_h^\theta X_0^\theta] = \int_{\mathbb{R}} \varphi_\theta(h + t) \varphi_\theta(t) \, dt, \qquad h \in \mathbb{R}. \tag{2.2}$$

It will be assumed throughout that $\theta \mapsto \gamma_\theta(h)$ is twice continuously differentiable for all h. Recall that, for fixed $\Delta > 0$ and any $t \in \mathbb{Z}$, the projection of $X_{t\Delta}^\theta$ onto the linear span of $X_{(t-1)\Delta}^\theta, \ldots, X_{(t-k)\Delta}^\theta$ is given by $\alpha_k(\theta)^\top X_{t-1:t-k}^\theta$ where $\alpha_k(\theta) = \Gamma_k(\theta)^{-1} \gamma_k(\theta)$, $\Gamma_k(\theta) = [\gamma_\theta((i-j)\Delta)]_{i,j=1,\ldots,k}$ is the covariance matrix of $X_{t-1:t-k}^\theta$, and $\gamma_k(\theta) = [\gamma_\theta(\Delta), \ldots, \gamma_\theta(k\Delta)]^\top$. (Here we use the notation $Y_{t:s} = [Y_{t\Delta}, Y_{(t-1)\Delta}, \ldots, Y_{s\Delta}]^\top$ for $s, t \in \mathbb{Z}$ with $s < t$.) Note that by [7, Proposition 5.1.1], $\Gamma_k(\theta)$ is always invertible. Now suppose that $X_t = X_t^{\theta_0}$ for all $t \in \mathbb{R}$ and some unknown parameter θ_0 belonging to the interior of Θ, and consider n equidistant observations $X_{n:1} = [X_{n\Delta}, \ldots, X_\Delta]^\top$. We will estimate θ_0

by the least squares estimator $\hat{\theta}_n$, which is chosen to minimize (1.5). Thus, with the introduced notation,

$$\hat{\theta}_n \in \arg\min_{\theta \in \Theta} \sum_{t=k+1}^{n} (X_{t\Delta} - \alpha_k(\theta)^\top X_{(t-1):(t-k)})^2. \tag{2.3}$$

The estimator (2.3) can be seen as a truncated version of

$$\tilde{\theta}_n \in \arg\min_{\theta \in \Theta} \sum_{t=2}^{n} (X_{t\Delta} - \alpha_{t-1}(\theta)^\top X_{(t-1):1})^2. \tag{2.4}$$

From an implementation point of view, while evaluation of the objective function in (2.4) will demand computing $\alpha_1(\theta), \ldots, \alpha_{n-1}(\theta)$ (usually obtained recursively by the Durbin–Levinson algorithm [7, Proposition 5.2.1]), one only needs to compute $\alpha_k(\theta)$ in order to evaluate the objective function in (2.3). As discussed in [18], in short-memory models where the projection coefficients are rapidly decaying it is reasonable to use $\hat{\theta}_n$ with a suitably chosen depth k as a proxy for (2.4).

To show strong consistency and asymptotic normality of $\hat{\theta}_n$ we impose the following set of conditions:

Condition 2.1.

(a) $\gamma_\theta(j\Delta) = \gamma_{\theta_0}(j\Delta)$ for $j = 0, 1, \ldots, k$ if and only if $\theta = \theta_0$.

(b) $\gamma_k'(\theta_0) - \Gamma_k'(\theta_0)[\alpha_k(\theta_0) \otimes I_d]$ has full rank.

(c) $\left(t \longmapsto \sum_{s \in \mathbb{Z}} |\varphi_{\theta_0}(t + s\Delta)|^\beta\right) \in L^2([0, \Delta])$ for $\beta = 4/3, 2$.

Remark 2.2. Concerning Condition 2.1, (a)–(b) are standard assumptions ensuring that θ_0 is identifiable from the autocovariances and that the (suitably scaled version of the) second derivative of the objective function in (2.3) converges to an invertible deterministic matrix. The difference between Condition 2.1 and the typical set of conditions for proving asymptotic normality is that an assumption on the strong mixing coefficients of $(X_{t\Delta})_{t \in \mathbb{Z}}$ is replaced by (c), a rather explicit condition on the driving kernel. In fact, according to [21, Theorem 1.2], sufficient conditions for (c) to be satisfied are that

$$\varphi_{\theta_0} \in L^4 \quad \text{and} \quad \sup_{t \in \mathbb{R}} |t|^\beta |\varphi_{\theta_0}(t)| < \infty \tag{2.5}$$

for a suitable $\beta \in (3/4, 1)$.

Example 2.3. In view of Remark 2.2 the key condition to check is if we are in a subclass of moving average processes, where (2.2) (or, more generally, Condition 2.1(c)) holds true. In the following we consider a few popular classes of kernels φ.

(i) *CARMA and gamma:* It is clear that the gamma kernel $\varphi(t) \propto t_+^\beta e^{-\gamma t}$ meets (2.2) when $\beta \in (-1/4, \infty)$ and $\gamma \in (0, \infty)$. The CARMA kernel characterized in Section 1 can always be bounded by a sum of gamma kernels (see, e.g., [6, Equation (36)]), and hence (2.2) is satisfied for this choice as well.

(ii) *SDDE:* If the variation $|\eta|$ of η satisfies $\int_{[0,\infty)} t^2 |\eta|(\mathrm{d}t) < \infty$, it follows by [21, Example 3.10] that the kernel φ associated to the solution of (1.3) meets (2.2).

(iii) *Fractional noise:* If $\varphi(t) \propto t_+^d - (t-\tau)_+^d$ for some $d \in (0, 1/4)$ and $\tau \in (0, \infty)$, then φ is continuous and the mean value theorem implies that $\varphi(t)$ is asymptotically proportional to t^{d-1} as $t \to \infty$. These properties establish the validity of (2.2). Note that the corresponding discretely sampled moving average $(X_{t\Delta})_{t\in\mathbb{Z}}$ is not strongly mixing in this setup (cf. [9, Theorem A.1]).

Before stating and proving consistency and asymptotic normality of $\hat{\theta}_n$ in (2.3) we introduce some notation. For a twice continuously differentiable function f, defined on some open subset of $\mathbb{R}^d$ and with values in $\mathbb{R}^m$, the gradient and Hessian of f at θ are denoted by $f'(\theta)$ and $f''(\theta)$, respectively:

$$f'(\theta) = \left[\frac{\partial f}{\partial \theta_1}(\theta), \dots, \frac{\partial f}{\partial \theta_d}(\theta)\right] \in \mathbb{R}^{m\times d}, \quad f''(\theta) = \begin{bmatrix} \frac{\partial^2 f}{\partial \theta_1 \partial \theta_1} & \cdots & \frac{\partial^2 f}{\partial \theta_1 \partial \theta_d} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial \theta_d \partial \theta_1} & \cdots & \frac{\partial^2 f}{\partial \theta_d \partial \theta_d} \end{bmatrix} \in \mathbb{R}^{dm\times d}.$$

Moreover, with $v_1(\theta)^\top = [1, -\alpha_k(\theta)^\top]$, $v_2(\theta)^\top = [0, \alpha_k'(\theta)^\top]$ and $F_s(t;\theta) = [\varphi_\theta(t - (i-1)\Delta)\varphi_\theta(t-(j-s-1)\Delta)]_{i,j=1,\dots,k+1}$ we define

$$V_s^{ij}(t;\theta) = v_i(\theta)^\top F_s(t;\theta)v_j(\theta) \qquad \text{for } i,j = 1,2 \text{ and } s \in \mathbb{Z}. \tag{2.6}$$

Finally, we set $\sigma^2 = \mathbb{E}[L_1^2]$ and $\kappa_4 = \mathbb{E}[L_1^4] - 3\sigma^4$.

Theorem 2.4. *Suppose that θ_0 belongs to the interior of Θ and that Condition 2.1 is in force. Let $\hat{\theta}_n$ be the estimator given in (2.3). Then $\hat{\theta}_n \to \theta_0$ almost surely and $\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{\mathscr{D}} N(0, H^{-1}AH^{-1})$ as $n \to \infty$, where $H = 2\alpha_k'(\theta_0)^\top \Gamma_k(\theta_0)\alpha_k'(\theta_0)$ and*

$$\begin{aligned} A = \sum_{s\in\mathbb{Z}} \Big(&\kappa_4 \int_\mathbb{R} V_s^{11}(t;\theta_0)V_s^{22}(t;\theta_0)\, dt + \sigma^4 \int_\mathbb{R} V_s^{11}(t;\theta_0)\, dt \int_\mathbb{R} V_s^{22}(t;\theta_0)\, dt \\ &+ \sigma^4 \int_\mathbb{R} V_s^{21}(t;\theta_0)\, dt \int_\mathbb{R} V_s^{12}(t;\theta_0)\, dt \Big). \end{aligned} \tag{2.7}$$

Proof. Set $\ell_n(\theta) = \sum_{t=k+1}^n (X_{t\Delta} - \alpha_k(\theta)^\top X_{(t-1):(t-k)})^2$, and let ℓ_n' and ℓ_n'' be the first and second order derivative of ℓ_n, respectively. As usual, the consistency and part of the asymptotic normality rely on an application of a suitable (uniform) ergodic theorem to ensure almost sure convergence of the sequences $(n^{-1}\ell_n)_{n\in\mathbb{N}}$ and $(n^{-1}\ell_n'')_{n\in\mathbb{N}}$. The difference lies in the proof of a central limit theorem for $(n^{-1/2}\ell_n'(\theta_0))_{n\in\mathbb{N}}$.

Consistency: Note that $\mathbb{E}[\sup_{\theta\in\Theta}(X_{k\Delta} - \alpha_k(\theta)^\top X_{(k-1):0})^2] < \infty$, since the vector of projection coefficients $\alpha_k(\theta)$ is bounded due to the continuity of $\theta \mapsto \gamma_\theta(h)$. Thus, we find by the ergodic theorem for Banach spaces ([23, Theorem 2.7]) that $n^{-1}\ell_n(\theta) \to \mathbb{E}[(X_{k\Delta} - \alpha_k(\theta)^\top X_{(k-1):0})^2] =: \ell^*(\theta)$ almost surely and uniformly in θ as $n \to \infty$. Thus, strong consistency follows immediately if ℓ^* is uniquely minimized at θ_0. Since $\alpha_k(\theta_0)^\top X_{(k-1):0}$ is the projection of $X_{k\Delta}$ onto the linear span of $X_0, \dots, X_{(k-1)\Delta}$, it must be the case that $\ell^*(\theta_0) \le \ell^*(\theta)$ for all $\theta \in \Theta$. If $\theta \ne \theta_0$, Condition 2.1(a) implies that $\gamma_\theta(j\Delta) \ne \gamma_{\theta_0}(j\Delta)$ for at least one j, and hence $\ell^*(\theta_0) < \ell^*(\theta)$ by uniqueness of the projection coefficients.

Asymptotic normality: It suffices to show that (i) $n^{-1}\ell_n''(\theta)$ converges almost surely and uniformly in θ as $n \to \infty$ and $H := \lim_{n\to\infty} n^{-1}\ell_n''(\theta_0)$ is a deterministic positive definite matrix, and (ii) $n^{-1/2}\ell_n'(\theta_0)$ converges in distribution to a Gaussian random variable. Concerning (i), note that

$$\ell_n''(\theta) = 2 \sum_{t=k+1}^{n} \Big[\alpha_k'(\theta)^\top X_{(t-1):(t-k)} X_{(t-1):(t-k)}^\top \alpha_k'(\theta)$$
$$- (X_{t\Delta} - \alpha_k(\theta)^\top X_{(t-1):(t-k)})[X_{(t-1):(t-k)}^\top \otimes I_d]\,\alpha_k''(\theta)\Big],$$

where I_d is the $d \times d$ identity matrix and the jth row of α_k' (resp. the jth $d \times d$ block of α_k'') is the gradient (resp. Hessian) of the jth entry of α_k. Thus, it follows by [23, Theorem 2.7] that $n^{-1}\ell_n''(\theta) \to 2\alpha_k'(\theta)^\top \Gamma_k(\theta_0)\alpha_k'(\theta) =: H(\theta)$ almost surely and uniformly in θ as $n \to \infty$. Since $\Gamma_k(\theta_0)$ is positive definite and

$$\alpha_k'(\theta_0) = \Gamma_k(\theta_0)^{-1}\big(\gamma_k'(\theta_0) - \Gamma_k'(\theta_0)[\alpha_k(\theta_0) \otimes I_d]\big),$$

it follows from Condition 2.1(b) that $H = H(\theta_0)$ is positive definite. To show (ii), observe that $\ell_n'(\theta_0)$ takes the form

$$\ell_n'(\theta_0) = \sum_{t=k+1}^{n} \int_{\mathbb{R}} \psi_1(t\Delta - s)\,\mathrm{d}L_s \int_{\mathbb{R}} \psi_2(t\Delta - s)\,\mathrm{d}L_s$$

with $\psi_i(t) = v_i(\theta_0)^\top \varphi_{\theta_0,k}(t)$, using the notation $\varphi_{\theta_0,k}(t) = [\varphi_{\theta_0}(t), \varphi_{\theta_0}(t-\Delta), \ldots, \varphi_{\theta_0}(t-k\Delta)]^\top$. Since the space of functions f satisfying

$$\Big(t \longmapsto \sum_{s\in\mathbb{Z}} |f(t+s\Delta)|^\beta\Big) \in L^2([0,\Delta]) \qquad \text{for } \beta = 4/3, 2 \tag{2.8}$$

forms a vector space, and φ_{θ_0} satisfies (2.8) by Condition 2.1(c), ψ_1 and (each entry of) ψ_2 satisfy (2.8) as well. Moreover, as $\int_{\mathbb{R}} \psi_1(t\Delta - s)\,\mathrm{d}L_s = X_{t\Delta} - \alpha_k(\theta_0)^\top X_{(t-1):(t-k)}$ is orthogonal to $\int_{\mathbb{R}} \psi_2(t\Delta - s)\,\mathrm{d}L_s = X_{(t-1):(t-k)}^\top \alpha_k'(\theta_0)$ in $L^2(\mathbb{P})$ (entrywise), we have that $\mathbb{E}[\ell_n'(\theta_0)] = 0$. Consequently, by [21, Theorem 1.2], $n^{-1/2}\ell_n'(\theta_0)$ converges in distribution to a centered Gaussian vector with covariance matrix given by

$$\sum_{s\in\mathbb{Z}} \Big(\kappa_4 \int_{\mathbb{R}} \psi_1(t)\psi_1(t+s\Delta)\psi_2(t)\psi_2(t+s\Delta)^\top \,\mathrm{d}t + \sigma^4 \int_{\mathbb{R}} \psi_1(t)\psi_1(t+s\Delta)\,\mathrm{d}t$$
$$\cdot \int_{\mathbb{R}} \psi_2(t)\psi_2(t+s\Delta)^\top \,\mathrm{d}t + \sigma^4 \int_{\mathbb{R}} \psi_1(t)\psi_2(t+s\Delta)\,\mathrm{d}t \int_{\mathbb{R}} \psi_2(t)^\top \psi_1(t+s\Delta)\,\mathrm{d}t\Big),$$

which is equal to A given in (2.7). This concludes the proof. □

3 Examples

In this section we give two examples where Theorem 2.4 is applicable and accompany these by simulating the properties of the estimator $\hat{\theta}_n$. In both examples we fix the sample frequency $\Delta = 1$ as well as the depth $k = 10$. We have checked (by simulation) that the estimator is rather insensitive to the choice of k; this is supported by the fact that both models result in geometrically decaying projection coefficients.

Example 3.1. Suppose that $(L_t)_{t\in\mathbb{R}}$ is a standard Brownian motion and, for $\theta = (\nu, \lambda) \in (3/4, \infty) \times (0, \infty)$, set

$$\varphi_\theta(t) = \Gamma(\nu)^{-1} t^{\nu-1} e^{-\lambda t}, \qquad t > 0. \tag{3.1}$$

The moving average model (2.1) with gamma kernel (3.1) has received some attention in the literature and has, e.g., been used to model the timewise behavior of the velocity in turbulent regimes (see [3] and references therein). Moreover, particular choices of ν result in special cases of well-known and widely studied models. To be concrete, if $\nu = 1$ then $(X_t^\theta)_{t\in\mathbb{R}}$ is an Ornstein–Uhlenbeck process with parameter $\lambda > 0$ and, more generally, if $\nu \in \mathbb{N}$ then $(X_t^\theta)_{t\in\mathbb{R}}$ is a CAR(ν) process with polynomial $P(z) = (z + \lambda)^\nu$. The autocovariance function γ_θ of $(X_t^\theta)_{t\in\mathbb{R}}$ under the model specification (2.1) and (3.1) takes the form

$$\gamma_\theta(h) = \begin{cases} \Gamma(2\nu - 1)(2\lambda)^{1-2\nu} & \text{if} \quad h = 0, \\ \Gamma(\nu)(2\pi^{-1})^{1/2} 2^{-\nu}(\lambda^{-1}|h|)^{\nu-1/2} K_{\nu-1/2}(\lambda|h|) & \text{if} \quad h \neq 0, \end{cases}$$

where $K_{\nu-1/2}$ denotes the modified Bessel function of the third kind of order $\nu - 1/2$ (cf. [3]). The corresponding autocorrelation function $\gamma_\theta/\gamma_\theta(0)$ is known as the Whittle–Matérn correlation function ([15]). In Figure 1 we have simulated $X_{400:1}$ and plotted the corresponding sample and theoretical autocorrelation function for $\theta_0 = (1.3, 1.1)$. To demonstrate the ability to infer the true parameter $\theta_0 = (\nu_0, \lambda_0)$ from $X_{n:1}$ using
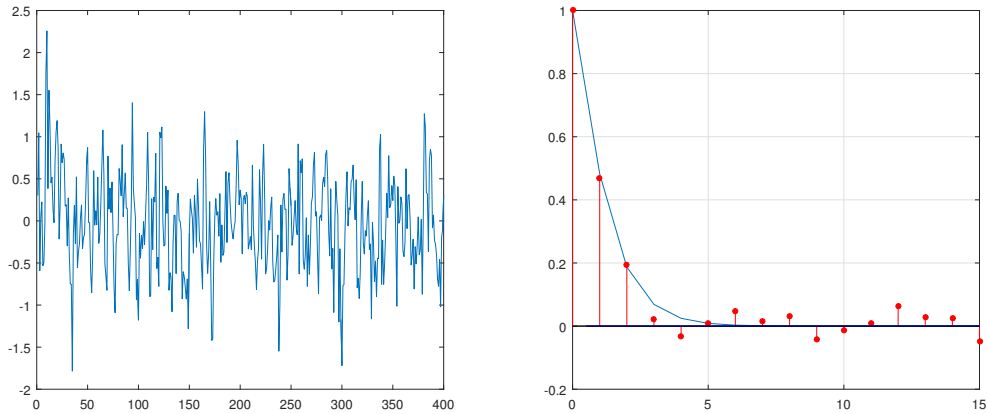
Figure 1: Left: simulation of $X_{400:1}$ under the model specification (2.1) and (3.1) with $\theta_0 = (1.3, 1.1)$ when $(L_t)_{t\in\mathbb{R}}$ is a Brownian motion. Right: the corresponding sample autocorrelation function and its theoretical counterpart.

the least squares estimator (2.3) we simulate $X_{n:100}$ under the model corresponding to θ_0 for $n = 400, 1600, 6400$, obtain the associated realizations of $\hat{\theta}_n = (\hat{\nu}_n, \hat{\lambda}_n)$ for truncation lag $k = 10$ and repeat the experiment 500 times. We perform this study for different choices of θ_0. In Table 1 we have, for each n, summarized the sample mean, bias and variance for the realizations of the least squares estimator. To show the robustness regarding the choice of the underlying noise we did the same analysis in the case where $(L_t)_{t\in\mathbb{R}}$ is a centered gamma Lévy process with both shape and scale parameter equal to one. In other words, $(L_t)_{t\in\mathbb{R}}$ was chosen to be the unique Lévy process where L_1 has density $t \mapsto \mathbb{1}_{\{t\geq -1\}} e^{-t-1}$. The findings, which are reported in Table 2, are seen to be similar to those of Table 1. To illustrate the asymptotic

normality of $\hat{\theta}_n$ we have plotted histograms based on the 500 realizations of $\hat{v}_n$ and $\hat{\lambda}_n$ when $n = 6400$ in the situation where $(v_0, \lambda_0) = (1.3, 1.1)$, see Figure 2.

Example 3.2. As in the last part of Example 3.1 let $(L_t)_{t\in\mathbb{R}}$ be a centered gamma Lévy process with both shape and scale parameter equal to one, and consider the model (1.3) where $\eta = \alpha\delta_0 + \beta\delta_1$ for some $\alpha, \beta \in \mathbb{R}$:

$$\mathrm{d}X_t = (\alpha X_t + \beta X_{t-1})\,\mathrm{d}t + \mathrm{d}L_t, \qquad t \in \mathbb{R}. \tag{3.2}$$

We will perform a simulation study similar to that of [18], except that they consider a Brownian motion as the underlying noise and use a certain pseudo (Gaussian) likelihood rather than the least squares estimator in (2.3). In [17] it is argued that a stationary solution to (3.2) exists if $\alpha < 1$ and

$$\beta \in \begin{cases} \left(-\frac{\alpha}{\cos(\xi(\alpha))}, -\alpha\right) & \text{if} \quad \alpha \neq 0, \\ \left(-\frac{\pi}{2}, 0\right) & \text{if} \quad \alpha = 0. \end{cases}$$

The function ξ is characterized by $\xi(0) = \pi/2$ and $\xi(t) = t\tan(\xi(t))$ for $t \neq 0$. We will compute (2.3) by using that

$$\gamma_\theta(h) = 2 \int_0^\infty \frac{\cos(hy)}{|iy + \alpha + \beta e^{iy}|^2}\,\mathrm{d}y, \qquad h \in \mathbb{R},$$

which follows from (1.4), (2.2) and Plancherel's theorem. We let $(\alpha_0, \beta_0) = (-1, -0.1353)$ in line with [18], and in Table 3 we provide statistics similar to those of Tables 1–2.
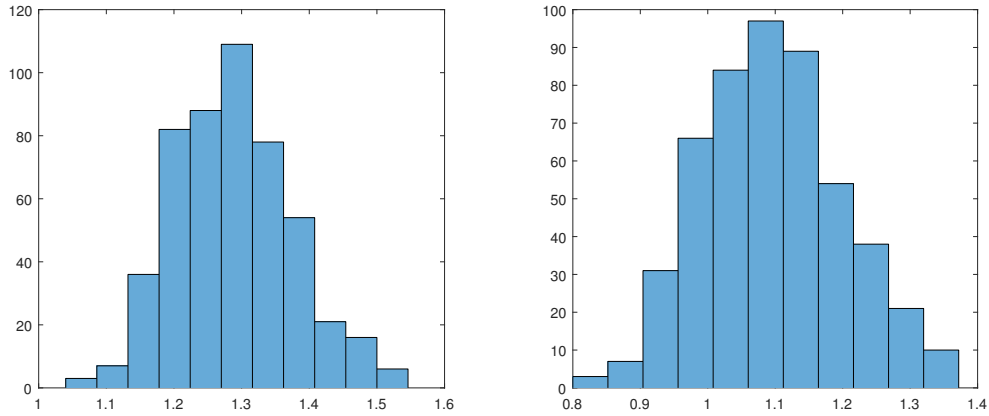
Table 1: Sample mean, bias and variance based on 500 realizations of $\hat{\theta}_n = (\hat{v}_n, \hat{\lambda}_n)$ various choices of n. The noise is a Brownian motion.

n		400		1600		6400	
		$\hat{v}_n$	$\hat{\lambda}_n$	$\hat{v}_n$	$\hat{\lambda}_n$	$\hat{v}_n$	$\hat{\lambda}_n$
$v_0 = 1.3$ $\lambda_0 = 1.1$	Mean	1.3869	1.1613	1.3353	1.1271	1.3008	1.0982
	Bias	0.0869	0.0613	0.0353	0.0271	0.0008	−0.0018
	Var.×10	1.2143	1.5452	0.3553	0.4501	0.0749	0.1039
$v_0 = 0.9$ $\lambda_0 = 1.1$	Mean	1.1460	1.4244	0.9867	1.2151	0.9092	1.1079
	Bias	0.2460	0.3244	0.0857	0.1151	0.0092	0.0079
	Var.×10	2.0205	4.7529	0.6148	1.6267	0.0851	0.3354
$v_0 = 1.3$ $\lambda_0 = 0.5$	Mean	1.3202	0.5166	1.3079	0.5060	1.2989	0.4987
	Bias	0.0202	0.0166	0.0079	0.0060	−0.0011	−0.0013
	Var.×10	0.1910	0.1424	0.0417	0.0333	0.0099	0.0079

Table 2: Sample mean, bias and variance based on 500 realizations of $\hat{\theta}_n = (\hat{v}_n, \hat{\lambda}_n)$ various choices of n. The noise is a centered gamma Lévy process.

n		400		1600		6400	
		$\hat{v}_n$	$\hat{\lambda}_n$	$\hat{v}_n$	$\hat{\lambda}_n$	$\hat{v}_n$	$\hat{\lambda}_n$
$v_0 = 1.3$	Mean	1.3638	1.1537	1.3158	1.1234	1.2870	1.0969
$\lambda_0 = 1.1$	Bias	0.0638	0.0537	0.0158	0.0234	−0.0130	−0.0031
	Var. ×10	1.1162	1.5069	0.3358	0.4505	0.0729	0.1061
$v_0 = 0.9$	Mean	1.1339	1.3813	1.0049	1.2249	0.9323	1.1262
$\lambda_0 = 1.1$	Bias	0.2339	0.2813	0.1049	0.1249	0.0323	0.0262
	Var. ×10	1.8303	4.3999	0.5879	1.5714	0.0900	0.3446
$v_0 = 1.3$	Mean	1.3017	0.5095	1.2902	0.5000	1.2871	0.4964
$\lambda_0 = 0.5$	Bias	0.0017	0.0095	−0.0098	0.0000	−0.0129	−0.0036
	Var. ×10	0.1615	0.1352	0.0401	0.0319	0.0097	0.0079

Figure 2: Histograms of 500 realizations of $(\hat{v}_{6400}, \hat{\gamma}_{6400})$ when $(v_0, \lambda_0) = (1.3, 1.1)$ and the noise is a gamma Lévy process.

Table 3: Sample mean, bias and variance based on 500 realizations of $\hat{\theta}_n = (\hat{\alpha}_n, \hat{\beta}_n)$ various choices of n when the true parameters are $\alpha_0 = -1$ and $\beta_0 = -0.1353$. The noise is a centered gamma Lévy process.

n	400		1600		6400	
	$\hat{\alpha}_n$	$\hat{\beta}_n$	$\hat{\alpha}_n$	$\hat{\beta}_n$	$\hat{\alpha}_n$	$\hat{\beta}_n$
Mean	−0.9980	−0.1654	−1.0127	−0.1508	−1.0132	−0.1459
Bias	0.0020	−0.0301	−0.0127	−0.0155	−0.0132	−0.0106
Var. ×10	0.3022	0.0979	0.1165	0.0498	0.0379	0.0189

Acknowledgments

References

[1] Ango Nze, P., P. Bühlmann and P. Doukhan (2002). Weak dependence beyond mixing and asymptotics for nonparametric regression. *Ann. Statist.* 30(2), 397–430. DOI: 10.1214/aos/1021379859.

[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. et al. (2012). Notes on the gamma kernel. *Thiele Research Reports*, *Department of Mathematics*, *Aarhus University*.

[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. *Stochastic Process. Appl.* Forthcoming. DOI: 10.1016/j.spa.2018.11.011.

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

[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] Cohen, S. and A. Lindner (2013). A central limit theorem for the sample autocorrelations of a Lévy driven continuous time moving average process. *J. Statist. Plann. Inference* 143(8), 1295–1306. DOI: 10.1016/j.jspi.2013.03.022.

[10] Fasen-Hartmann, V. and S. Kimmig (2018). Robust estimation of continuous-time ARMA models via indirect inference. arXiv: 1804.00849.

[11] Francq, C. and J.-M. Zakoïan (1998). Estimating linear representations of nonlinear processes. *J. Statist. Plann. Inference* 68(1), 145–165.

[12] 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.

[13] Gorodetskii, V. (1978). On the strong mixing property for linear sequences. *Theory Probab. Appl.* 22(2), 411–413.

[14] 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.

[15] Guttorp, P. and T. Gneiting (2005). On the Whittle-Matérn correlation family. *National Research Center for Statistics and the Environment-Technical Report Series, Seattle, Washington.*

[16] Karhunen, K. (1950). Über die Struktur stationärer zufälliger Funktionen. *Ark. Mat.* 1, 141–160. DOI: 10.1007/BF02590624.

[17] 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.

[18] 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.

[19] Marquardt, T. (2007). Multivariate fractionally integrated CARMA processes. *Journal of Mult. Anal.* 98(9), 1705–1725.

[20] Marquardt, T. and R. Stelzer (2007). Multivariate CARMA processes. *Stochastic Process. Appl.* 117(1), 96–120. DOI: 10.1016/j.spa.2006.05.014.

[21] Nielsen, M.S. and J. Pedersen (2019). Limit theorems for quadratic forms and related quantities of discretely sampled continuous-time moving averages. *ESAIM: Probab. Stat.* Forthcoming. DOI: 10.1051/ps/2019008.

[22] Schlemm, E. and R. Stelzer (2012). Quasi maximum likelihood estimation for strongly mixing state space models and multivariate Lévy-driven CARMA processes. *Electron. J. Stat.* 6, 2185–2234. DOI: 10.1214/12-EJS743.

[23] Straumann, D. and T. Mikosch (2006). Quasi-maximum-likelihood estimation in conditionally heteroscedastic time series: a stochastic recurrence equations approach. *Ann. Statist.* 34(5), 2449–2495. DOI: 10.1214/009053606000000803.

[24] 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.

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.

MSC: 62P30; 65C20; 91B68

Keywords: Extreme event estimation; Wind turbines; Surrogate model

1 Introduction

Suppose that we are interested in the distributional properties of a certain one-dimensional random variable Y. For instance, one may want to know the probability of the occurrence of large values of Y as they could 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 $y_1,\ldots,y_n$ of Y and then fit a suitable distribution (for instance, 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, for example, 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, for instance, 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 $\widetilde{\mathbb{P}}$ such that

$$\mathbb{P} = \widetilde{\mathbb{P}}(\cdot \mid Z = 0) \quad \text{and} \quad \mathbb{Q} = \widetilde{\mathbb{P}}(\cdot \mid Z = 1).$$

Effectively, Z indicates whether we are using the simulation tool or not, and $\widetilde{\mathbb{P}}(Z = 1) \in (0, 1)$ 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

$$\widetilde{\mathbb{P}}(Z = 1 \mid X, Y) = \widetilde{\mathbb{P}}(Z = 1 \mid X). \qquad (2.2)$$

In words, (2.2) means that Y and Z are conditionally independent under $\widetilde{\mathbb{P}}$ given X. The assumption (2.2) was introduced in Rosenbaum and Rubin [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 $b \colon \mathbb{R}^d \to \mathbb{R}^m$, $m < d$, maintaining the property

$$\mathbb{P}(Y \in \cdot \mid b(X) = b(x)) = \mathbb{Q}(Y \in \cdot \mid b(X) = b(x)) \qquad (2.3)$$

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 Rosenbaum and Rubin [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 $\widetilde{\mathbb{P}}$ (cf. Remark 2.2), $\pi(x) = \widetilde{\mathbb{P}}(Z = 1 \mid X = x)$, and next arguing that a general function b is a balancing function if and only if

$$f(b(x)) = \pi(x) \qquad \text{for some function } f. \qquad (2.4)$$

2.1 Estimation of the conditional probability

The ultimate goal is to estimate $\mu_x = \mathbb{Q}(Y \in \cdot \mid X = x)$, for instance, in terms of the cumulative distribution function (CDF) or density function, from a sample $(x_1^s, y_1^s), \ldots, (x_m^s, y_m^s)$ 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 $k \in \mathbb{N}, k \leq m$, let $I_k(x) \subseteq \{1, \ldots, m\}$ denote the k indices corresponding to the k points in $\{x_1^s, \ldots, x_m^s\}$ 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

$$p_j(x) = \mathbb{Q}(C = j \mid X = x), \qquad j = 0,1,\ldots,k-1.$$

Given these estimates, say $\widehat{p_0},\widehat{p_1},\ldots,\widehat{p_{k-1}}$, the CDF of μ_x is estimated by

$$\widehat{\mu_x}((-\infty,\alpha_i]) = \sum_{j=1}^{k} \widehat{p_{j-1}}(x)\Phi\left(\frac{\alpha_i - \alpha_j}{\sigma}\right), \qquad i = 1,\ldots,k,$$

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

$$d(x,y) = \sqrt{(x-y)^\top M(x-y)}, \qquad x,y \in \mathbb{R}^d,$$

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 (that is, 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 Hodge and Austin [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 $x_1^s, \ldots, x_m^s$ 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 (such as the Euclidean or Mahalanobis distance). Then, for observations $x_1, \ldots, x_n$ under $\mathbb{P}$, the algorithm goes as follows:

(1) For $i = 1, \ldots, n$ 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 $\bar{d}_{(1)} \leq \cdots \leq \bar{d}_{(n)}$ of $\bar{d}_1, \ldots, \bar{d}_n$ 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 $\bar{d}_i \geq \bar{d}_{(j)}$ 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)
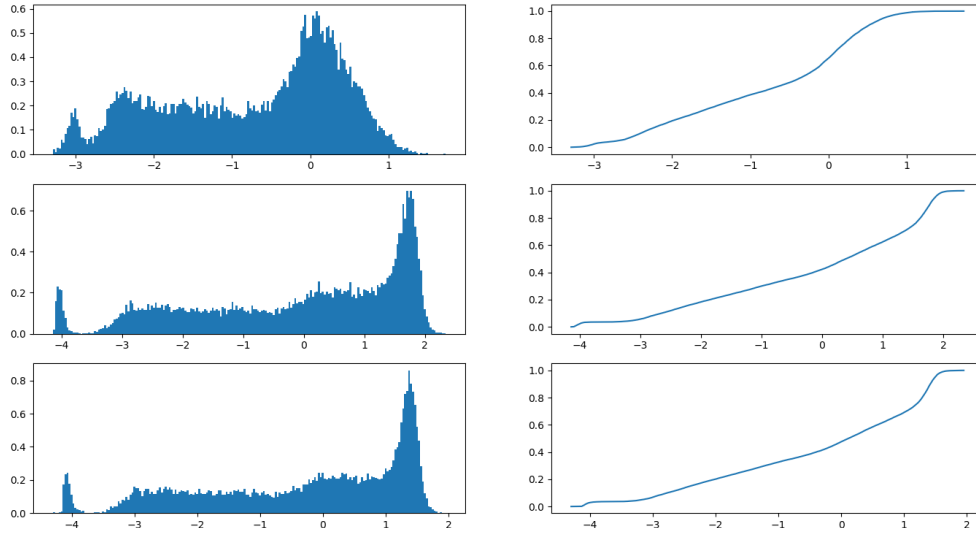
Figure 1: Measured load distributions. Left and right plots correspond to histograms and CDFs, respectively, based on 19 976 observations of the tower top (first row), middle (second row) and base (third row) down-wind bending moments.

The selection of covariates is based on a physical interpretation of the problem and by leaving out covariates which from a visual inspection (that is, 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 τth smallest average distance, where τ is the point that minimizes the L^1 distance

$$d_1(f, f_\tau) := \int_1^m |f(x) - f_\tau(x)| \, \mathrm{d}x \tag{3.1}$$

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 4506 (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).
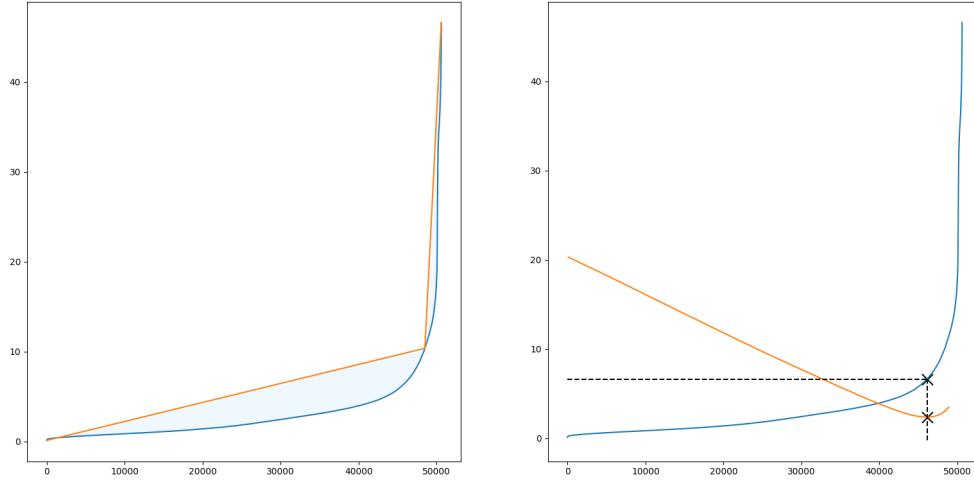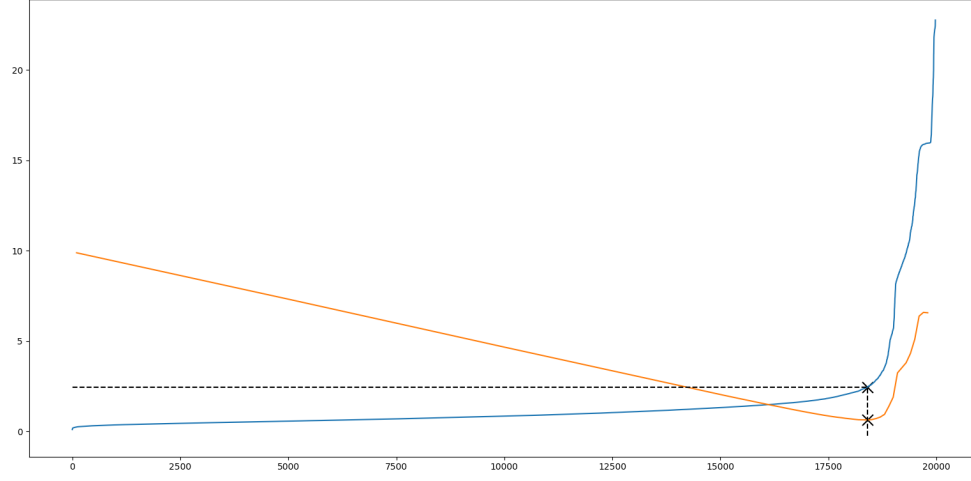

Figure 2: Blue curve: sorted distance from simulated covariates to nearest measured covariates. Left: linear interpolation of $(1, d_{(1)}), (\tau, d_{(\tau)}), (m, d_{(m)})$ with shaded region representing the corresponding L^1 error for $\tau = 48\,500$. Right: the orange curve is the normalised L^1 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 1577 (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.
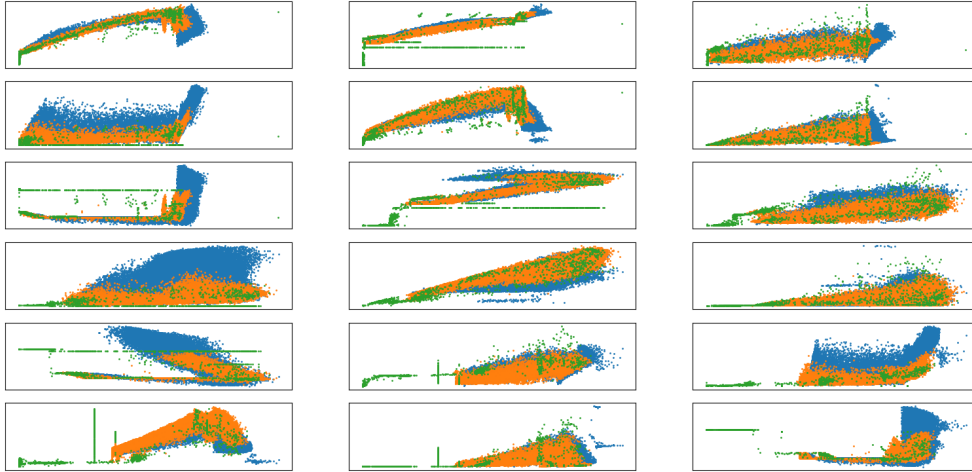
Figure 4: Some of the two-dimensional projections of the covariates. Blue dots are simulations, orange dots are inliers and green dots are potential outliers.

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, that is, 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.
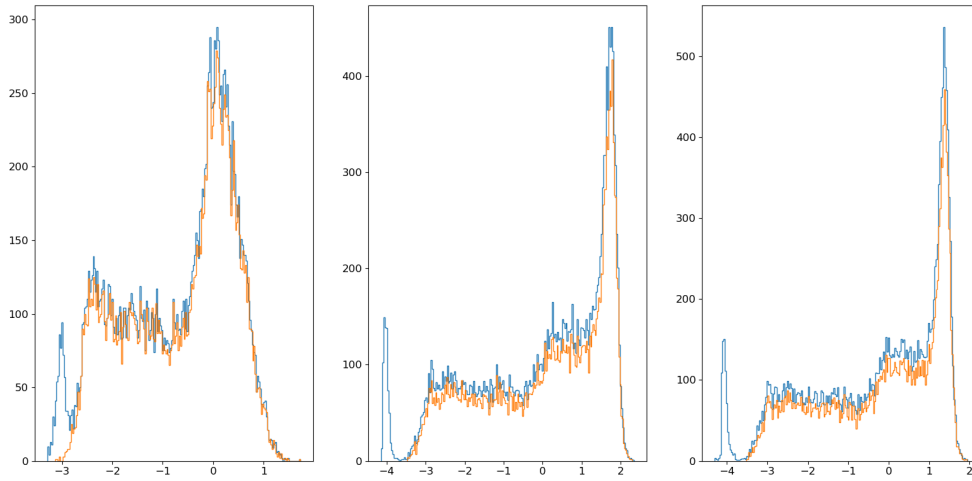
Figure 5: Histograms of measurements on tower top (left), middle (mid) and base (right) down-wind bending moments. Measurements including and excluding outliers are represented in blue and orange, respectively.

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 (that is, $\sigma = 0$) 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) (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 (that is, 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 % to 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.
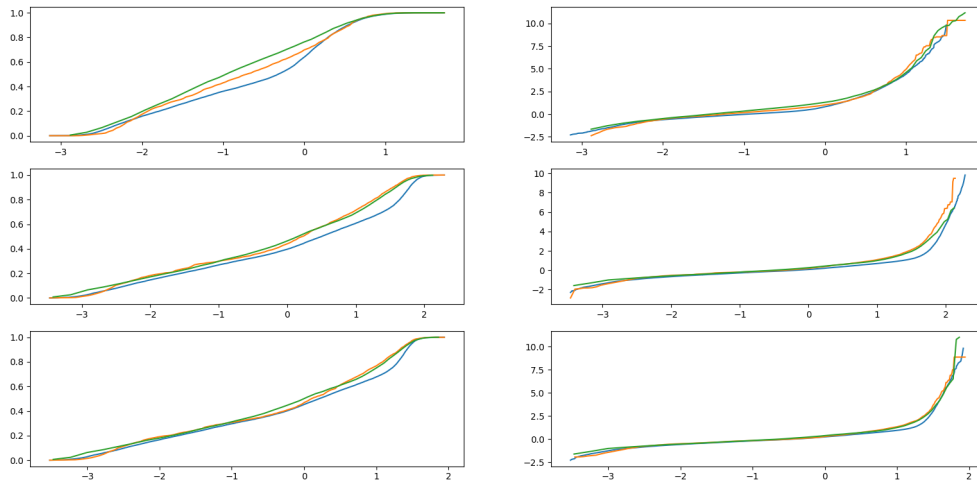
Figure 6: Plots of CDFs (first column) and the corresponding $-\log(-\log(\,\cdot\,))$ transforms (second column) of tower top (first row), middle (second row) and base (third row) down-wind bending moments. The blue curve is the empirical distribution of the measurements, and the orange and green curves 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 the curves.
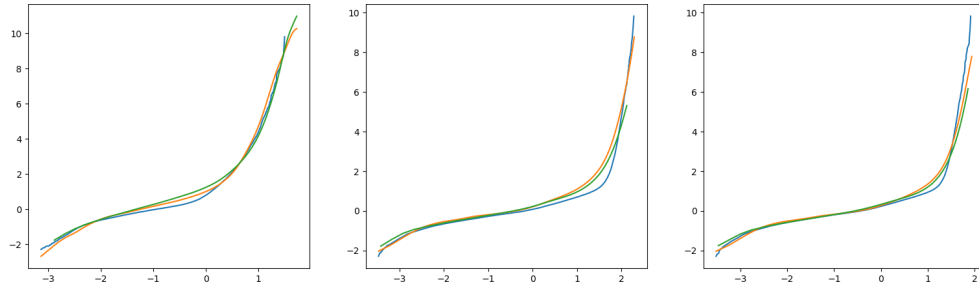
Figure 7: Plots of $-\log(-\log(\cdot))$ transforms of CDFs of tower top (left), middle (center) and base (right) down-wind bending moments. The blue curve is the empirical distribution of the measurements, and the orange and green curves are the smoothed kNN and random forest predictions, respectively, using Silverman's rule of thumb.

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.

Table 1: Some quantiles of the empirical load distributions and of the corresponding kNN and random forest 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

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

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. *J. Sol. Energy Eng.* 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.