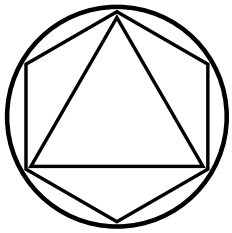
Multivariate Continuous Time Stochastic Volatility Models Driven by a Lévy Process

Robert Josef Stelzer



Centre for Mathematical Sciences

Munich University of Technology

D-85747 Garching

2007

Zentrum Mathematik
Lehrstuhl für Mathematische Statistik der
Technischen Universität München

Multivariate Continuous Time Stochastic Volatility Models Driven by a Lévy Process

Robert Josef Stelzer

Vollständiger Abdruck der von der Fakultät für Mathematik der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat)

genehmigten Dissertation.

Vorsitzender:	Univ.-Prof. Dr. Herbert Spohn
Prüfer der Dissertation:	1. Univ.-Prof. Dr. Claudia Klüppelberg
	2. Univ.-Prof. Dr. Alexander Lindner, Philipps-Universität Marburg
	3. Prof. Dr. Jean Jacod, Université Pierre et Marie Curie Paris, Frankreich

Die Dissertation wurde am 12. Juli 2007 bei der Technischen Universität München eingereicht und durch die Fakultät für Mathematik am 28. September 2007 angenommen.

Zusammenfassung

Es werden verschiedene multivariate stochastische Modelle in stetiger Zeit eingeführt und aus probabilistischer und statistischer Sicht im Detail untersucht. Alle diese Modelle werden von Lévyprozessen getrieben und können allgemein für die Modellierung mehrdimensionaler Beobachtungsreihen verwendet werden. Hierbei liegt in dieser Arbeit der Schwerpunkt auf verschiedenen stochastischen Volatilitätsmodellen für Finanzmarktdaten.

Zunächst werden multivariate zeitstetige autoregressive Moving-Average (CARMA) Prozesse betrachtet und darauf aufbauend ein multivariates zeitstetiges exponentielles GARCH Modell (ECOGARCH). Danach werden positiv semidefinite Ornstein-Uhlenbeck-Prozesse eingeführt und es wird allgemein das Verhalten stochastischer Prozesse endlicher Variation mit Werten in den positiv semidefiniten Matrizen, die als Summe eines Integrales bezüglich der Zeit und eines Integrales bezüglich eines erweiterten Poissonzufallsmaßes dargestellt werden können, bei der Wurzelbildung (und ähnlichen Transformationen) untersucht. Die positiv semidefiniten Ornstein-Uhlenbeck-Prozesse bilden die Grundlage für die Definition einer multivariaten Erweiterung des populären stochastischen Volatilitätsmodells von Barndorff-Nielsen und Shephard. Nach einer detaillierten theoretischen Untersuchung wird dieses Modell für beobachtete Aktienkurse geschätzt. Als weiteres Modell mit stochastischer Volatilität werden multivariate zeitstetige GARCH (COGARCH) Prozesse eingeführt und deren probabilistischen und statistischen Eigenschaften näher betrachtet.

Abstract

Several multivariate stochastic models in continuous time are introduced and their probabilistic and statistical properties are studied in detail. All models are driven by Lévy processes and can generally be used to model multidimensional time series of observations. In this thesis the focus is on various stochastic volatility models for financial data.

Firstly, multidimensional continuous-time autoregressive moving-average (CARMA) processes are considered and, based upon them, a multivariate continuous time exponential GARCH model (ECOGARCH). Thereafter, positive semi-definite Ornstein-Uhlenbeck type processes are introduced and the behaviour of the square root (and similar transformations) of stochastic processes of finite variation, which take values in the positive semi-definite matrices and can be represented as the sum of an integral with respect to time and another integral with respect to an extended Poisson random measure, is analysed in general. The positive semi-definite Ornstein-Uhlenbeck type processes form the basis for the definition of a multivariate extension of the popular stochastic volatility model of Barndorff-Nielsen and Shephard. After a detailed theoretical study this model is estimated for some observed stock price series. As a further model with stochastic volatility multivariate continuous time GARCH (COGARCH) processes are introduced and their probabilistic and statistical properties are analysed.

Acknowledgements

Taking this opportunity I would like to thank all those people without whom the time I have been working on this thesis would not have been what it has been.

First of all my thanks go to Prof. Claudia Klüppelberg for all her very helpful advice and encouragement throughout the years, as well as the excellent working conditions and climate. I am especially grateful to her for enabling me to travel to various scientific meetings and to present my work there and for bringing me into contact with many distinguished researchers.

It is also a particular pleasure for me to thank Prof. Jean Jacod and Prof. Alexander Lindner for accepting to act as referees of this thesis. Furthermore, I would like to express my gratitude to Prof. Alexander Lindner for his continued helpful advice.

My special thanks go also to Prof. Ole Barndorff-Nielsen for many inspiring discussions, very kind invitations to most pleasant stays in Århus and especially for the very fruitful collaboration.

Likewise, I would like to especially thank Dr. Stephan Haug, Dr. Tina Marquardt and Dr. Christian Pigorsch for our fruitful and pleasant collaboration. Furthermore, I am grateful to my colleagues at the Chair of Mathematical Statistics and the GK for their support and for the enjoyable company during the last years.

The privilege of having been financially supported by the Deutsche Forschungsgemeinschaft through the graduate programme "Angewandte Algorithmische Mathematik" at the Munich University of Technology is also gratefully acknowledged by me.

Last but not at all least I would like to thank especially my family for simply everything and, of course, Christine for all her loving support.

Contents

1. Introduction

Today sophisticated probabilistic and statistical models play an important role in finance and related areas. The widespread availability of high frequency or tick-by-tick data being unequally spaced calls for the use of continuous time models. Furthermore, it has become clear over the years that the classical model of Black and Scholes specifying the logarithmic asset price as a Brownian motion with drift fails to reproduce important features of empirically observed price series, the so-called "stylized facts" (cf. Barndorff-Nielsen and Shephard (2001a), Cont and Tankov (2004) or Guillaume, Dacorogna, Davé, Müller, Olsen and Pictet (1997), for instance).

The "stylized facts" of price data of financial assets (stocks, currencies, etc.) are that the volatility is not constant over time but stochastic and exhibits jumps, that there are clusters of extremes in the volatility, that the marginal distributions are asymmetric and heavy-tailed and that the returns exhibit a (quasi)long-range dependent behaviour although they are uncorrelated. Moreover, one often adds certain scaling and self-similarity properties.

In recent years many models reproducing at least some of these stylized facts have been proposed and applied. Such models are typically stochastic volatility models of some kind, i.e. the price process is determined by a stochastic differential equation involving a stochastic process describing the instantaneous volatility. In order to account for the jumps observed, this stochastic volatility process is often defined using a Lévy process. Two prominent examples of such models are the Ornstein-Uhlenbeck stochastic volatility model introduced in Barndorff-Nielsen and Shephard (2001b) and studied and applied in various papers thereafter, and the continuous time GARCH (generalized autoregressive conditional heteroskedasticity) model presented first in Klüppelberg, Lindner and Maller (2004) which is a continuous time analogue of the GARCH process popular in discrete time and inherits most of the appealing features of this discrete time process. Likewise, Haug and Czado (2007) introduced recently a continuous time analogue of the exponential GARCH (EGARCH) process modelling the logarithm of the volatility as a Lévy-driven CARMA process.

The models mentioned above are all univariate and thus can only model the price of a single asset. However, for portfolio optimization, risk management and the pricing of multi-asset options, which are becoming more and more important, one needs a model for the joint evolution of the prices of several assets. From empirical observations and economic reasoning it is clear that in general the prices of different assets exhibit highly non-trivial interdependencies. Clearly, it is essential to cover these in multivariate models. So far there are only rather few continuous time multivariate stochastic volatility models available which model the whole covariance matrix and not only the variances. Hubalek and Nicolato (2005) and Lindberg (2005) present different factor models extending the univariate model of Barndorff-Nielsen and Shephard, and Gourieroux (2006) uses Gaussian vector Ornstein-Uhlenbeck processes to specify the stochastic volatility process, which in a multivariate setting is the instantaneous covariance matrix and thus has to be a stochastic

process in the positive semi-definite matrices.

In this thesis we introduce and study several multivariate stochastic volatility models which are generalizations of univariate ones. Since this involves defining suitable stochastic processes taking values in the positive semi-definite matrices, the generalizations are not straightforward. Moreover, it should be noted that the processes we introduce in the definition of the multivariate stochastic volatility models are of considerable interest in their own right and should not only be of use in finance, but also in very different areas. Moreover, our analysis of the processes presented often goes further than what is needed for the definition and understanding of the stochastic volatility models.

Extending the univariate work of Brockwell (2001b) in particular, we introduce and analyse multivariate Lévy-driven continuous time autoregressive moving-average (CARMA) processes. As these processes are the continuous time analogues of the discrete time ARMA processes, they form a very elementary class of processes with wide-ranging applications. In particular, for Lévy processes with finite variance we construct associated random orthogonal measures and use them to obtain spectral representations of multivariate CARMA processes. Based upon these processes we define our first stochastic volatility model, the multivariate continuous time exponential GARCH(p, q) (ECOGARCH) model generalizing the univariate model presented in Haug and Czado (2007). A major motivation for considering exponential GARCH models is that a negative shock in the price process may be accompanied by a larger positive shock in the stochastic volatility than a positive one, which resembles the "leverage effect" often encountered empirically. After analysing some elementary properties we show that a multivariate ECOGARCH(1,1) process can be approximated by a sequence of piecewise constant processes determined by discrete time multivariate E-GARCH(1,1) processes arbitrarily well in the Skorokhod topology in probability. Actually, we establish in general that a multidimensional and infinite time extension of the first jump approximation of a Lévy process introduced in Szimayer and Maller (2007) satisfies a "uniform tightness" condition important for the convergence of solutions of sequences of stochastic differential equations in the Skorokhod topology.

Furthermore, Ornstein-Uhlenbeck type processes taking values in the positive semi-definite matrices are introduced using matrix subordinators (see Barndorff-Nielsen and Pérez-Abreu (2007)) and appropriate linear operators, and for a certain related class of stochastic processes taking values in the positive semi-definite matrices the behaviour of the square root (and similar transformations) is analysed. Along the way several elementary but important results for processes in the positive semi-definite matrices are obtained. The positive semi-definite Ornstein-Uhlenbeck processes are used to define another multivariate stochastic volatility model, namely a multivariate generalization of the Barndorff-Nielsen and Shephard model. This model is analysed in detail focusing in particular on the second order structure of discretely observed returns, a state space representation and the behaviour of the realized quadratic variation, where we state the results for a general class of stochastic volatility models whenever possible. The analysis is completed by presenting some further results for positive semi-definite Ornstein-Uhlenbeck type processes, in particular regarding the stationary distribution and relating them to superpositions of univariate Ornstein-Uhlenbeck type processes. After the theoretical study the model is successfully applied to empirical data. Our results show that this multivariate stochastic volatility model is highly tractable and that basically all results obtained for the univariate model can be extended. It generates dependent returns with zero autocorrelation and thus reproduces this

important stylized fact. The dependence is seen by showing that the squared returns have a non-zero autocorrelation function decaying exponentially after lag one. Moreover, like in the univariate model one can get more flexibility (and long-range dependence) by specifying the volatility process as a superposition of positive semi-definite Ornstein-Uhlenbeck processes.

The final model with stochastic volatility, which we introduce, is the multivariate continuous time GARCH(1,1) (MUCOGARCH(1,1)) process extending the univariate COGARCH(1,1) process of Klüppelberg et al. (2004). For the definition we use some insight from the construction of the univariate COGARCH(p, q) process in Brockwell, Chadraa and Lindner (2006). One reason why COGARCH processes, like their discrete time counterparts, are particularly interesting is that their stationary distribution is usually heavy-tailed and that they normally exhibit extremal clusters (cf. Fasen, Klüppelberg and Lindner (2006)), whereas for most other models with stochastic volatility this is only the case under some special assumptions. To analyse the MUCOGARCH(1,1) process and also to establish its well-definedness we need to present some results on the existence and uniqueness of solutions of multidimensional stochastic differential equations being defined only on an open set and satisfying only a local Lipschitz property, and some results on their Markovian properties. Again we present a thorough study of the probabilistic and statistical properties, especially the second order structure, and show that the returns are uncorrelated, whereas the squared returns have a non-vanishing autocorrelation function. The analysis of the MUCOGARCH(1,1) process is concluded by considering several examples along with some simulations.

Finally, we address the issue of scaling laws in a univariate set-up extending the work of Barndorff-Nielsen and Prause (2001). We obtain series representations of certain (absolute) moments of generalized hyperbolic distributions and normal inverse Gaussian (NIG) Lévy processes. Based upon this some asymptotic scaling results for the moments of NIG Lévy processes are obtained and some graphs are presented which indicate that the behaviour of moments of an NIG Lévy process over time may optically look very close to scaling.

1.1. Outline of the thesis

As every of the following chapters and appendices of this thesis is based on a paper, they are basically self-contained and the notation is only unified within the individual chapters. However, some general abbreviations and notation are also explained at the end of the thesis.

In the following brief abstracts of the contents of the individual chapters are presented. More detailed information on the content of an individual chapter is to be found in the introductory section of each chapter.

Chapter 2 is based on Marquardt and Stelzer (2007). Here a multivariate continuous time autoregressive moving average (CARMA) model of order (p, q), $q < p$, driven by a Lévy process is introduced. It extends the well-known univariate CARMA and multivariate discrete time ARMA models. We give an explicit construction using a state space representation and a spectral representation of the driving Lévy process. Furthermore, various probabilistic properties of the state space model and the multivariate CARMA process itself are discussed in detail.

Thereafter, a multivariate extension of the exponential continuous time GARCH(p, q)

model (ECOGARCH) is introduced and studied in *Chapter 3* being based on Haug and Stelzer (2007). Stationarity and mixing properties of the new model are investigated and ways to model a component-wise leverage effect are presented. Further it is shown that there exists a sequence of piecewise constant processes determined by discrete time multivariate EGARCH$(1,1)$ processes which converge in probability in the Skorokhod topology to the multivariate ECOGARCH$(1,1)$ process. To obtain this result the first jump approximation of a Lévy process is used and it is shown to satisfy an important uniform tightness condition regarding the convergence of stochastic differential equations in general.

In the following *Chapter 4* based on Barndorff-Nielsen and Stelzer (2007) processes of finite variation, which take values in the positive semi-definite matrices and are representable as the sum of an integral with respect to time and one with respect to an extended Poisson random measure, are considered. For such processes we derive conditions for the square root (and the r-th power with $0 < r < 1$) to be of finite variation and obtain integral representations of the square root. Our discussion is based on a variant of the Itô formula for finite variation processes. Moreover, Ornstein-Uhlenbeck type processes taking values in the positive semi-definite matrices are introduced and some of their probabilistic properties are studied.

Chapter 5 is an extended version of Pigorsch and Stelzer (2007). This chapter generalizes the univariate Ornstein-Uhlenbeck (OU) stochastic volatility model proposed by Barndorff-Nielsen and Shephard (2001b) to the multivariate case using positive semi-definite processes of Ornstein-Uhlenbeck type. We show that most of the properties of the univariate model are still available in closed form for the multivariate model. After deriving the second order properties of the returns and their outer product ("squared returns"), we establish a state-space representation for the joint sequence which can be used to estimate the model via quasi maximum likelihood. Moreover, the state-space representation provides a straightforward way to filter out the unobservable, time varying covariance using the Kalman filter. Likewise we analyse the realized quadratic variation and superpositions of positive semi-definite OU type processes, for which we also obtain state-space representations. Furthermore, conditions ensuring the identifiability of our model are presented and additionally some further results on positive semi-definite OU type processes, especially regarding the marginal dynamics and the stationary distribution. Noteworthy, many of our results are shown to be valid for more general multivariate stochastic volatility models which are driven by a stationary and square-integrable instantaneous covariance process.

Furthermore, a multivariate extension of the COGARCH(1,1) process introduced in Klüppelberg et al. (2004) is presented and shown to be well-defined in *Chapter 6* which contains the results of Stelzer (2007a). The definition generalizes the idea of Brockwell et al. (2006) for the definition of the univariate COGARCH(p,q) process and is in a natural way related to multivariate discrete time GARCH processes as well as positive definite Ornstein-Uhlenbeck type processes. Furthermore, we establish important Markovian properties and sufficient conditions for the existence of a stationary distribution for the volatility process, which lives in the positive semi-definite matrices, by bounding it by a univariate COGARCH(1,1) process in a special norm. Moreover, criteria ensuring the finiteness of moments of both the multivariate COGARCH process as well as its volatility process are given. Under certain assumptions on the moments of the driving Lévy process, explicit expressions for the first and second order moments and (asymptotic) second order stationarity are obtained. As a necessary prerequisite we study the existence of solutions and some

other properties of stochastic differential equations being only defined on a subset of $\mathbb{R}^d$ and satisfying only local Lipschitz conditions.

The main part of this thesis is concluded by *Chapter 7* which has been published as Barndorff-Nielsen and Stelzer (2005). In this chapter expressions for (absolute) moments of generalized hyperbolic (GH) and normal inverse Gaussian (NIG) laws are given in terms of moments of the corresponding symmetric laws. For the (absolute) moments centred at the location parameter μ explicit expressions as series containing Bessel functions are provided. Furthermore, the derivatives of the logarithms of absolute μ–centred moments with respect to the logarithm of time are calculated explicitly for NIG Lévy processes. Computer implementation of the formulae obtained is briefly discussed. Finally, some further insight into the apparent scaling behaviour of NIG Lévy processes is gained.

Finally, there are two general appendices based on Stelzer (2006, 2007b). In *Appendix A* the linear operators $\mathbf{B}$ on the real symmetric matrices whose exponential $\exp(\mathbf{B}t)$ leaves the inertia invariant and maps the positive (semi-)definite matrices onto themselves for all $t \in \mathbb{R}$ are fully characterized. *Appendix B* addresses the question which multivariate GARCH models in the vec form are representable in the BEKK form. Using results from linear algebra, it is established that all vec models not representable in the simplest BEKK form contain matrices as parameters which map the vectorized positive semi-definite matrices into a strict subset of themselves. Moreover, a general result from linear algebra is presented implying that in dimension two the models are equivalent and in dimension three a simple analytically tractable example for a vec model having no BEKK representation is given.

2. Multivariate CARMA Processes[1]

2.1. Introduction

Being the continuous time analogue of the well-known autoregressive moving-average (ARMA) processes (see e.g. Brockwell and Davis (1991)), continuous time ARMA (CARMA) processes, dating back to Doob (1944), have been extensively studied over the recent years (see e.g. Brockwell (2001a,b), Todorov and Tauchen (2006) and references therein) and widely used in various areas of application like engineering, finance and the natural sciences (e.g. Jones and Ackerson (1990), Mossberg and Larsson (2004) and Todorov and Tauchen (2006)). The advantage of continuous time modelling is that it allows handling irregularly spaced time series and in particular high-frequency data often appearing in finance. Originally, driving processes of CARMA models were restricted to Brownian motion, however, Brockwell (2001b) allowed for Lévy processes which have a finite r-th moment for some $r > 0$.

Since CARMA processes are short memory moving average processes, Brockwell and Marquardt (2005) developed fractionally integrated CARMA (FICARMA, for short) processes, which exhibit long range dependence. So far only univariate CARMA processes have been defined and investigated. However, in order to model the joint behaviour of several time series (e.g. prices of various stocks) multivariate models are required. Thus, we develop multivariate CARMA processes and study their probabilistic properties in this chapter.

Unfortunately, it is not straightforward to define the multivariate CARMA processes analogously to the univariate ones, as the state space representation (see Section 3.1) relies on the ability to exchange the autoregressive and moving average operators, which is only possible in one dimension. Simply taking this approach would lead to a spectral representation which does not reflect the autoregressive moving average structure. Our approach leads to a model which can be interpreted as a solution to the formal differential equation $P(D)Y(t) = Q(D)DL(t)$, where D denotes the differential operator with respect to t, L a Lévy process and P and Q the autoregressive and moving average polynomial, respectively. Moreover, it is the continuous time analogue of the multivariate ARMA model.

The chapter is organized as follows. In Section 2.2 we review elementary properties of multidimensional Lévy processes and the stochastic integration theory for deterministic functions with respect to them. A brief summary of univariate Lévy-driven CARMA processes forms the first part of the third section and is followed by the development of what will turn out to be the state space representation of multivariate CARMA (MCARMA) processes. We start by constructing a random orthogonal measure allowing for a spectral representation of the driving Lévy process and continue by studying a stochastic differential equation. Analysing the spectral representation of its solution shows that it can be used to define multivariate CARMA processes. After spending a closer look on the probabilistic properties of this SDE (second moments, Markov property, stationary and limiting distribu-

[1]The contents of this chapter appeared in Marquardt, T. and Stelzer, R. (2007), Multivariate CARMA Processes, *Stochastic Process. Appl.*, **117**(1), 96–120

tions and path behaviour), we state the definition of MCARMA processes in Section 2.3.3. Furthermore, we establish a kernel representation, which enables us to derive some further probabilistic properties of MCARMA models. In particular, we characterize the stationary distribution and the path behaviour and give conditions for the existence of moments, the existence of a C_b^∞ density as well as for strong mixing.

Throughout this chapter we use the following notation. We call the space of all real or complex $m \times m$-matrices $M_m(\mathbb{R})$ or $M_m(\mathbb{C})$, respectively, and the space of all complex invertible $m \times m$-matrices $\mathcal{G}l_m(\mathbb{C})$. Furthermore, A^* denotes the adjoint of the matrix A and Ker$A$ its kernel. $I_m \in M_m(\mathbb{C})$ is the identity matrix and $\|A\|$ is the operator norm corresponding to the norm $\|x\|$ for $x \in \mathbb{C}^m$. Finally, $I_B(\cdot)$ is the indicator function of the set B and $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$.

2.2. Multivariate Lévy processes

2.2.1. Basic facts on multivariate Lévy processes

We state some elementary properties of multivariate Lévy processes that will be needed. For a more general treatment and proofs we refer to Applebaum (2004), Protter (2004) or Sato (1999).

We consider a Lévy process $L = \{L(t)\}_{t\geq 0}$ (where $L(0) = 0$ a.s.) in $\mathbb{R}^m$ without Brownian component determined by its characteristic function in the Lévy-Khintchine form $E\left[e^{i\langle u, L(t)\rangle}\right] = \exp\{t\psi_L(u)\}$, $t \geq 0$, with

$$\psi_L(u) = i\langle \gamma, u\rangle + \int_{\mathbb{R}^m} (e^{i\langle u,x\rangle} - 1 - i\langle u, x\rangle I_{[0,1]}(\|x\|))\, \nu(dx), \quad u \in \mathbb{R}^m, \tag{2.2.1}$$

where $\gamma \in \mathbb{R}^m$ and ν is a measure on $\mathbb{R}^m$ that satisfies $\nu(\{0\}) = 0$ and $\int_{\mathbb{R}^m} (\|x\|^2 \wedge 1)\, \nu(dx) < \infty$. The measure ν is referred to as the Lévy measure of L. It is a well-known fact that to every càdlàg Lévy process L on $\mathbb{R}^m$ one can associate a random measure J on $\mathbb{R}^+ \times \mathbb{R}^m \setminus \{0\}$ describing the jumps of L. For any measurable set $B \subset \mathbb{R}^+ \times \mathbb{R}^m \setminus \{0\}$,

$$J(B) = \sharp\{s \geq 0 :\ (s, L_s - L_{s-}) \in B\}.$$

The jump measure J is a Poisson random measure on $\mathbb{R}^+ \times \mathbb{R}^m \setminus \{0\}$ (see e.g. Definition 2.18 in Cont and Tankov (2004)) with intensity measure $n(ds, dx) = ds\, \nu(dx)$. By the Lévy-Itô decomposition we can rewrite L almost surely as

$$L(t) = \gamma t + \int_{\|x\|\geq 1, s\in[0,t]} x\, J(ds, dx) + \lim_{\varepsilon \downarrow 0} \int_{\varepsilon \leq \|x\| \leq 1,\, s\in[0,t]} x\tilde{J}(ds, dx), \quad t \geq 0. \tag{2.2.2}$$

Here $\tilde{J}(ds, dx) = J(ds, dx) - ds\nu(dx)$ is the compensated jump measure, the terms in (2.2.2) are independent and the convergence in the last term is a.s. and locally uniform in $t \geq 0$.

In the sequel we will work with a two-sided Lévy process $L = \{L(t)\}_{t\in\mathbb{R}}$, constructed by taking two independent copies $\{L_1(t)\}_{t\geq 0}$, $\{L_2(t)\}_{t\geq 0}$ of a one-sided Lévy process and setting

$$L(t) = \begin{cases} L_1(t) & \text{if} \quad t \geq 0 \\ -L_2(-t-) & \text{if} \quad t < 0. \end{cases} \tag{2.2.3}$$

Assuming that ν satisfies additionally

$$\int_{\|x\|>1} \|x\|^2 \, \nu(dx) < \infty, \tag{2.2.4}$$

L has finite mean and covariance matrix Σ_L given by

$$\Sigma_L = \int_{\mathbb{R}^m} xx^* \, \nu(dx). \tag{2.2.5}$$

Furthermore, if we suppose that $E[L(1)] = \gamma + \int_{\|x\|>1} x \, \nu(dx) = 0$, then it follows that (2.2.1) can be written in the form

$$\psi_L(u) = \int_{\mathbb{R}^m} (e^{i\langle u,x\rangle} - 1 - i\langle u, x\rangle) \, \nu(dx), \quad u \in \mathbb{R}^m, \tag{2.2.6}$$

and (2.2.2) simplifies to

$$L(t) = \int_{x\in\mathbb{R}^m\setminus\{0\},\, s\in[0,t]} x\tilde{J}(ds, dx), \quad t \in \mathbb{R}. \tag{2.2.7}$$

In this case $L = \{L(t)\}_{t\geq 0}$ is a martingale.

2.2.2. Stochastic integrals with respect to Lévy processes

In this section we consider the stochastic process $X = \{X(t)\}_{\in\mathbb{R}}$ given by

$$X(t) = \int_{\mathbb{R}} f(t, s) \, L(ds), \qquad t \in \mathbb{R}, \tag{2.2.8}$$

where $f : \mathbb{R}\times\mathbb{R} \to M_m(\mathbb{R})$ is a measurable function and $L = \{L(t)\}_{t\in\mathbb{R}}$ is an m-dimensional Lévy process without Brownian component. For integration with respect to Brownian motion we refer to any of the standard books.

We first assume that the process L in (2.2.8) is an m-dimensional Lévy process without a Gaussian component satisfying $E[L(1)] = 0$ and $E[L(1)L(1)^*] < \infty$, i.e., L can be represented as in (2.2.7).

In this case it follows from (2.2.7) that the process X can be represented by

$$X(t) = \int_{\mathbb{R}\times\mathbb{R}^m} f(t, s)x \, \tilde{J}(ds, dx), \qquad t \in \mathbb{R}, \tag{2.2.9}$$

where $\tilde{J}(ds, dx) = J(ds, dx) - ds\nu(dx)$ is the compensated jump measure of L. A sufficient condition for the existence of the stochastic integral (2.2.9) in $L^1(\Omega, P)$ (see e.g. Rajput and Rosinski (1989) or Marcus and Rosinski (2005)) is that

$$\int_{\mathbb{R}} \int_{\mathbb{R}^m} (\|f(t, s)x\|^2 \wedge \|f(t, s)x\|) \, \nu(dx) \, ds < \infty, \quad \forall \, t \in \mathbb{R}.$$

Then the law of $X(t)$ is for all $t \in \mathbb{R}$ infinitely divisible with characteristic function

$$E\left[\exp\left\{i\langle u, X(t)\rangle\right\}\right] = \exp\left\{\int_{\mathbb{R}}\int_{\mathbb{R}^m}\left(e^{i\langle u, f(t,s)x\rangle} - 1 - i\langle u, f(t,s)x\rangle\right)\nu(dx)\,ds\right\}.$$

Furthermore, if $f(t,\cdot) \in L^2(\mathbb{R}; M_m(\mathbb{R}))$, the integral (2.2.9) exists in $L^2(\Omega, P)$ and

$$E\left[X(t)X(t)^*\right] = \int_{\mathbb{R}} f(t,s)\Sigma_L f^*(t,s)\,ds. \tag{2.2.10}$$

If additionally

$$\int_{\mathbb{R}}\int_{\mathbb{R}^m}\left(\|f(t,s)x\| \wedge 1\right)\nu(dx)\,ds < \infty, \quad \forall\, t \in \mathbb{R},$$

the stochastic integral (2.2.8) exists without a compensator and we can write

$$X(t) = \int_{\mathbb{R}\times\mathbb{R}^m} f(t,s)x\,J(ds,dx), \qquad t \in \mathbb{R}, \tag{2.2.11}$$

still assuming $E(L_1) = 0$.

Finally, in the general case, where condition (2.2.4) is not satisfied, necessary and sufficient conditions for the integral (2.2.8) to exist are (see Rajput and Rosinski (1989) and Sato (2006))

$$\int_{\mathbb{R}}\int_{\mathbb{R}^m}\left(\|f(t,s)x\|^2 \wedge 1\right)\nu(dx)\,ds < \infty, \quad \forall\, t \in \mathbb{R}, \tag{2.2.12}$$

and

$$\int_{\mathbb{R}}\left\|f(t,s)\gamma + \int_{\mathbb{R}^m} f(t,s)x\left(I_{[0,1]}(\|f(t,s)x\|) - I_{[0,1]}(\|x\|)\right)\nu(dx)\right\|ds < \infty. \tag{2.2.13}$$

Then we represent X as

$$X(t) = \int_{\mathbb{R}}\left(\int_{\mathbb{R}^m} f(t,s)x\left[J(ds,dx) - I_{[0,1]}(\|f(t,s)x\|)\,\nu(dx)\right] + f(t,s)\gamma\right)ds, \;\; t \in \mathbb{R}.$$

Moreover, if the integral in (2.2.8) is well-defined, the distribution of $X(t)$ is infinitely divisible with characteristic triplet $(\gamma_X^t, 0, \nu_X^t)$ given by

$$\gamma_X^t = \int_{\mathbb{R}}\left(f(t,s)\gamma + \int_{\mathbb{R}^m} f(t,s)x\left[I_{[0,1]}(\|f(t,s)x\|) - I_{[0,1]}(\|x\|)\right]\nu(dx)\right)ds, \tag{2.2.14}$$

$$\nu_X^t(B) = \int_{\mathbb{R}}\int_{\mathbb{R}^m} 1_B(f(t,s)x)\,\nu(dx)\,ds. \tag{2.2.15}$$

It follows that the characteristic function of $X(t)$ can be written as

$$E\left[e^{i\langle u,X(t)\rangle}\right] = \exp\left\{i\langle\gamma_X^t, u\rangle + \int_{\mathbb{R}^m}\left[e^{i\langle u,x\rangle} - 1 - i\langle u,x\rangle I_{[0,1]}(\|x\|)\right]\,\nu_X^t(dx)\right\}$$
$$= \exp\left\{\int_{\mathbb{R}}\psi_L(f(t,s)^*u)\,ds\right\}, \tag{2.2.16}$$

where ψ_L is given as in (2.2.1). These facts follow from Sato (2006, Theorem 3.1, Proposition 2.17 and Corollary 2.19).

2.3. Multivariate CARMA processes

In this section we discuss CARMA processes driven by general Lévy processes, i.e., the Lévy processes may have a Brownian component and does not need to have finite variance, if not stated otherwise. We start with a brief review of the well-known one-dimensional case.

2.3.1. Univariate Lévy-driven CARMA processes

Continuous-time ARMA (CARMA) processes constitute a special class of moving average (MA) processes (see, for instance, Applebaum (2004, Section 4.3.5)) with short memory and are the continuous time analogues of the well-known autoregressive moving average (ARMA) processes. We give here a short summary of their definition and properties. For further details see Brockwell (2001a,b, 2004).

Definition 2.3.1 (CARMA Process)**.** *Let $\{L(t)\}_{t\in\mathbb{R}}$ be a Lévy process satisfying*

$$\int_{|x|\geq 1}\log|x|\nu(dx) < \infty,$$

p, q be in $\mathbb{N}_0$ with $p > q$ and $a_1, \ldots, a_p, b_0, \ldots, b_q \in \mathbb{R}$, $a_p, b_0 \neq 0$ such that

$$A := \left[\begin{array}{c|ccc} 0 & & I_{p-1} & \\ \hline -a_p & -a_{p-1} & \ldots & -a_1 \end{array}\right]$$

has only eigenvalues with strictly negative real part. Furthermore, let $\{X(t)\}_{t\in\mathbb{R}}$ denote the unique stationary solution to

$$dX(t) = AX(t)dt + e\,L(dt), \quad t \in \mathbb{R}, \tag{2.3.1}$$

where $e^T = [0, \ldots, 0, 1]$. Then the process

$$Y(t) = b^T X(t), \tag{2.3.2}$$

with $b^T = \left[\ b_q, b_{q-1}, \ldots, b_{q-p+1}\ \right]$, is called a Lévy-driven continuous time autoregressive moving average process of order (p, q) (CARMA(p, q), for short). If $q < p-1$, we set $b_{-1} = \ldots = b_{q-p+1} = 0$.

The CARMA(p,q) process can be interpreted as the stationary solution of the p-th order linear differential equation,

$$p(D)Y(t) = q(D)DL(t), \quad t \geq 0, \tag{2.3.3}$$

where D denotes differentiation with respect to t and

$$p(z) := z^p + a_1 z^{p-1} + ... + a_p \text{ and } q(z) := b_0 z^q + b_1 z^{q-1} + ... + b_q$$

are the so-called autoregressive and moving-average polynomials, respectively. To see this note first that in the case $q(z) = 1$ (i.e. $q = 0$ and $b^T = (1, 0, \ldots, 0)$) rewriting (2.3.3) as a system of first-order differential equations in the standard way gives (2.3.1) and (2.3.2) with $X_t^T = (Y_t, DY_t, \ldots, D^{p-1}Y_t)$. In the general case we transform (2.3.3) to $Y(t) = p(D)^{-1}q(D)DL(t) = q(D)p(D)^{-1}DL(t)$ (note that we may commute $p^{-1}(D)$ and $q(D)$, since the real coefficients and the operator D all commute). From the previous case we infer that the process in (2.3.1) is formed by $p(D)^{-1}DL(t)$ and the first $p-1$ derivatives of this process. Now one can immediately see that $Y_t = b^T X_t = q(D)p(D)^{-1}DL_t$.

Remark 2.3.2. *Observe that the process $\{X(t)\}_{t\in\mathbb{R}}$ can be represented as*

$$X(t) = \int_{-\infty}^{t} e^{A(t-u)} e \, L(du), \; t \in \mathbb{R}, \tag{2.3.4}$$

and is a multivariate Ornstein-Uhlenbeck-type process (Jurek and Mason (1993), Sato and Yamazato (1984), Wolfe (1982)). Hence, we have

$$Y(t) = \int_{-\infty}^{t} b^T e^{A(t-u)} e \, L(du), \quad t \in \mathbb{R}. \tag{2.3.5}$$

From (2.3.5) it is obvious that $\{Y(t)\}_{t\in\mathbb{R}}$ is a causal short memory moving average process, since it has the form

$$Y(t) = \int_{-\infty}^{\infty} g(t-u) \, L(du), \quad t \in \mathbb{R}, \tag{2.3.6}$$

with kernel $g(t) = b^T e^{At} e I_{[0,\infty)}(t)$. Replacing e^{At} by its spectral representation, the kernel g can be expressed as

$$g(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{it\lambda} \frac{q(i\lambda)}{p(i\lambda)} d\lambda, \quad t \in \mathbb{R}. \tag{2.3.7}$$

Note that the representation of $\{Y(t)\}_{t\in\mathbb{R}}$ given by (2.3.6) together with (2.3.7) defines a strictly stationary process even if there are eigenvalues of A with strictly positive real part. However, if there are eigenvalues with positive real part, the CARMA process will be no longer causal. Henceforth we focus on causal CARMA processes.

Proposition 2.3.3 (Brockwell (2004, Section 2))**.** *If $E[L(1)^2] < \infty$, then the spectral density f_Y of $Y = \{Y(t)\}_{t\in\mathbb{R}}$ is given by*

$$f_Y(\lambda) = \frac{\text{var}(L(1))}{2\pi} \frac{|q(i\lambda)|^2}{|p(i\lambda)|^2}, \quad \lambda \in \mathbb{R}.$$

Consequently, the autocovariance function γ_Y of the CARMA process Y can be expressed as

$$\gamma_Y(h) = \operatorname{cov}(Y(t+h), Y(t)) = \frac{\operatorname{var}(L(1))}{2\pi} \int_{-\infty}^{\infty} e^{ih\lambda} \left| \frac{q(i\lambda)}{p(i\lambda)} \right|^2 d\lambda, \quad h \in \mathbb{R}.$$

Moreover, for a causal CARMA process an application of the residue theorem leads to

$$\gamma_Y(h) = \operatorname{var}(L(1)) \sum_{r=1}^{p} \frac{q(\lambda_r)q(-\lambda_r)}{p'(\lambda_r)p(-\lambda_r)} e^{\lambda_r |h|}, \quad h \in \mathbb{R},$$

provided all eigenvalues $\lambda_1, \ldots, \lambda_p$ of the matrix A are algebraically simple.

2.3.2. State Space Representation of Multivariate CARMA Processes

This section contains the necessary results and insights enabling us to define multivariate CARMA processes in the next section. As we shall heavily make use of spectral representations of stationary processes (see Doob (1953), Gihman and Skorohod (1974), Rozanov (1967) for comprehensive treatments), let us briefly recall the notions and results we shall employ.

Definition 2.3.4. *Let $\mathcal{B}(\mathbb{R})$ denote the Borel-σ-algebra over $\mathbb{R}$.*

A family $\{\zeta(\Delta)\}_{\Delta \in \mathcal{B}(\mathbb{R})}$ of $\mathbb{C}^m$-valued random variables is called an m-dimensional random orthogonal measure, if

(i) $\zeta(\Delta) \in L^2$ *for all bounded* $\Delta \in \mathcal{B}(\mathbb{R})$,

(ii) $\zeta(\emptyset) = 0$,

(iii) $\zeta(\Delta_1 \cup \Delta_2) = \zeta(\Delta_1) + \zeta(\Delta_2)$ *a.s., if* $\Delta_1 \cap \Delta_2 = \emptyset$, *and*

(iv) $F : \mathcal{B}(\mathbb{R}) \to M_m(\mathbb{C})$, $\Delta \mapsto E[\zeta(\Delta)\zeta(\Delta)^*]$ *defines a σ-additive positive definite matrix measure (i.e., a σ-additive set function that assumes values in the positive semi-definite matrices) and it holds that* $E[\zeta(\Delta_1)\zeta(\Delta_2)^*] = F(\Delta_1 \cap \Delta_2)$ *for all* $\Delta_1, \Delta_2 \in \mathcal{B}(\mathbb{R})$.

F is referred to as the spectral measure of ζ.

The definition above obviously implies $E[\zeta(\Delta_1)\zeta(\Delta_2)^*] = 0$ for disjoint Borel sets Δ_1, Δ_2.

Stochastic integrals $\int_\Delta f(t)\zeta(dt)$ of deterministic Lebesgue-measurable functions $f : \mathbb{R} \to M_m(\mathbb{C})$ with respect to a random orthogonal measure ζ are now as usually defined in an L^2-sense (see, in particular, Rozanov (1967, Ch. 1) for details). Note that the integration can be understood component-wise: Denoting the coordinates of ζ by ζ_i, i.e. $\zeta = (\zeta_1, \ldots, \zeta_m)^*$, the i-th element $\left(\int_\Delta f(t)\zeta(dt)\right)_i$ of $\int_\Delta f(t)\zeta(dt)$ is given by $\sum_{k=1}^m \int_\Delta f_{ik}(t)\zeta_k(dt)$, where the integrals are standard one-dimensional stochastic integrals in an L^2-sense and $f_{ik}(t)$ denotes the element in the i-th row and k-th column of $f(t)$. The above integral is defined whenever the integral

$$\int_\Delta f(t)F(dt)f(t)^* := \left(\sum_{k,l=1}^{m} \int_{\mathbb{R}} f_{ik}(t)\overline{f}_{jl}(t)F_{kl}(dt) \right)_{1 \le i,j \le m}$$

exists. Functions satisfying this condition are said to be in $L^2(F)$. For two functions $f, g \in L^2(F)$ we have

$$E\left[\int_\Delta f(t)\zeta(dt)\left(\int_\Delta g(t)\zeta(dt)\right)^*\right] = \int_\Delta f(t)F(dt)g(t)^*. \tag{2.3.8}$$

In the following we will only encounter random orthogonal measures, whose associated spectral measures have constant density with respect to the Lebesgue measure λ on $\mathbb{R}$, i.e. $F(dt) = C\lambda(dt) := C\,dt$ for some positive definite $C \in M_m(\mathbb{C})$, which simplifies the integration theory considerably. In this case it is easy to see that it is sufficient for $\int_\Delta f(t)F(dt)f(t)^*$ to exist that $\int_\Delta \|f(t)\|^2\,dt$ is finite, where $\|\cdot\|$ is some norm on $M_m(\mathbb{C})$. To ease notation we define the space of square-integrable matrix-valued functions

$$L^2(\mathbb{R}; M_m(\mathbb{C})) := \left\{f : \mathbb{R} \to M_m(\mathbb{C}),\ \int_{\mathbb{R}} \|f(t)\|^2 dt < \infty\right\}. \tag{2.3.9}$$

In the following we abbreviate $L^2(\mathbb{R}; M_m(\mathbb{C}))$ by $L^2(M_m(\mathbb{C}))$. This space is independent of the norm $\|\cdot\|$ on $M_m(\mathbb{C})$ used in the definition and is equal to the space of functions $f = (f_{ij}) : \mathbb{R} \to M_m(\mathbb{C})$ where all components f_{ij} are in the usual space $L^2(\mathbb{R}; \mathbb{C})$.

$$\|f\|_{L^2(M_m(\mathbb{C}))} = \left(\int_{\mathbb{R}} \|f(t)\|^2 dt\right)^{1/2} \tag{2.3.10}$$

defines a norm on $L^2(M_m(\mathbb{C}))$ and again it is immaterial, which norm we use, as all norms $\|\cdot\|$ on $M_m(\mathbb{C})$ lead to equivalent norms $\|\cdot\|_{L^2(M_m(\mathbb{C}))}$. With this norm $L^2(M_m(\mathbb{C}))$ is a Banach space and even a Hilbert space, provided the original norm $\|\cdot\|$ on $M_m(\mathbb{C})$ is induced by a scalar product. Observe that as usual we do not distinguish between functions and equivalence classes in $L^2(\cdot)$. The integrals $\int_\Delta f(t)\zeta(dt)$ and $\int_\Delta g(t)\zeta(dt)$ agree (in L^2), if f and g are identical in $L^2(M_m(\mathbb{C}))$, and a sequence of integrals $\int_\Delta \|f_n(t)\|^2\,dt$ converges (in L^2) to $\int_\Delta \|f(t)\|^2\,dt$ for $n \to \infty$, if $\|f_n(t) - f(t)\|_{L^2(M_m(\mathbb{C}))} \to 0$ as $n \to \infty$. Moreover,

$$E\left[\int_\Delta f(t)\zeta(dt)\left(\int_\Delta g(t)\zeta(dt)\right)^*\right] = \int_\Delta f(t)Cg(t)^*dt. \tag{2.3.11}$$

Our first step in the construction of multivariate CARMA processes is the following theorem extending the well-known fact that

$$W(t) = \int_{-\infty}^{\infty} \frac{e^{i\mu t} - 1}{i\mu}\,\phi(d\mu), \qquad t \in \mathbb{R},$$

is an m-dimensional standard Wiener process, if ϕ is an m-dimensional Gaussian random orthogonal measure satisfying $E[\phi(A)] = 0$ and $E[\phi(A)\phi(A)^*] = \frac{I_m}{2\pi}\lambda(A)$ for all $A \in \mathcal{B}(\mathbb{R})$ (see e.g. Arató (1982, Section 2.1, Lemma 5)).

Theorem 2.3.5. *Let $L = \{L(t)\}_{t\in\mathbb{R}}$ be a two-sided m-dimensional square integrable Lévy process with $E[L(1)] = 0$ and $E[L(1)L(1)^*] = \Sigma_L$. Then there exists an m-dimensional random orthogonal measure Φ_L with spectral measure F_L such that $E[\Phi_L(\Delta)] = 0$ for any bounded Borel set Δ,*

$$F_L(dt) = \frac{\Sigma_L}{2\pi}dt \tag{2.3.12}$$

and

$$L(t) = \int_{-\infty}^{\infty} \frac{e^{i\mu t} - 1}{i\mu} \, \Phi_L(d\mu). \tag{2.3.13}$$

The random measure Φ_L is uniquely determined by

$$\Phi_L([a,b)) = \int_{-\infty}^{\infty} \frac{e^{-i\mu a} - e^{-i\mu b}}{2\pi i\mu} \, L(d\mu) \tag{2.3.14}$$

for all $-\infty < a < b < \infty$.

Proof. Observe that setting $\tilde{\Phi}([a,b)) = L(b) - L(a)$ defines a random orthogonal measure on the semi-ring of intervals $[a;b)$, with $-\infty < a < b < \infty$. Using an obvious multidimensional extension of Rozanov (1967, Theorem 2.1), we extend $\tilde{\Phi}_L$ to a random orthogonal measure on the Borel sets. It is immediate that the associated spectral measure $\tilde{F}_L$ satisfies $\tilde{F}_L(dt) = \Sigma_L \, dt$ and that integrating with respect to $\tilde{\Phi}_L$ is the same as integrating with respect to the Lévy process L.

Now define $\Phi_L([a,b))$ for $-\infty < a < b < \infty$ by (2.3.14) which is equivalent to

$$\Phi_L([a,b)) = \int_{-\infty}^{\infty} \frac{e^{-i\mu a} - e^{-i\mu b}}{2\pi i\mu} \, \tilde{\Phi}_L(d\mu). \tag{2.3.15}$$

Using (2.3.11) we obtain for any two intervals $[a,b)$ and $[a',b')$

$$\begin{aligned} E[\Phi_L([a,b))\Phi_L([a',b'))^*] &= \int_{-\infty}^{\infty} \frac{e^{-i\mu a} - e^{-i\mu b}}{2\pi i\mu} \Sigma_L \overline{\left(\frac{e^{-i\mu a'} - e^{-i\mu b'}}{2\pi i\mu} \right)} \, d\mu \qquad (2.3.16) \\ &= \int_{-\infty}^{\infty} \frac{e^{-i\mu a} - e^{-i\mu b}}{2\pi i\mu} \Sigma_L^{1/2} \left(\frac{e^{-i\mu a'} - e^{-i\mu b'}}{2\pi i\mu} \Sigma_L^{1/2} \right)^* \, d\mu, \end{aligned}$$

where $\Sigma_L^{1/2}$ denotes the unique square root of Σ_L defined by spectral calculus. The crucial point is now to observe that the function $\hat{\phi}_{a,b}(\mu) = \frac{e^{-i\mu a} - e^{-i\mu b}}{\sqrt{2\pi} i\mu} \Sigma_L^{1/2}$ is the Fourier transform of the function $I_{[a,b)}(t)\Sigma_L^{1/2}$, i.e.,

$$\hat{\phi}_{a,b}(\mu) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-i\mu t} I_{[a,b)}(t) \Sigma_L^{1/2} dt.$$

The standard theory of Fourier-Plancherel transforms $\mathcal{F}$ (see e.g. Chandrasekharan (1989, Chapter II) or Yosida (1965, Chapter 6)) extends immediately to $L^2(M_m(\mathbb{C}))$ by setting

$$\mathcal{F}_m : L^2(M_m(\mathbb{C})) \to L^2(M_m(\mathbb{C})), \; f(t) \mapsto \hat{f}(\mu) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-i\mu t} f(t) dt$$

where $\int_{-\infty}^{\infty} e^{-i\mu t} f(t) dt$ is the limit in $L^2(M_m(\mathbb{C}))$ of $\int_{-R}^{R} e^{-i\mu t} f(t) dt$ as $R \to \infty$, because this can be interpreted as a component-wise Fourier-Plancherel transformation and, as stated

before, a function f is in $L^2(M_m(\mathbb{C}))$, if and only if all components f_{ij} are in $L^2(\mathbb{R};\mathbb{C})$. In particular, $\mathcal{F}_m$ is an invertible continuous linear operator on $L^2(M_m(\mathbb{C}))$ with

$$\mathcal{F}_m^{-1} : L^2(M_m(\mathbb{C})) \to L^2(M_m(\mathbb{C})),\ \hat{f}(\mu) \mapsto f(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{i\mu t} \hat{f}(\mu) d\mu,$$

and Plancherel's identity generalises to:

$$\int_{\mathbb{R}} f(t) g(t)^* dt = \int_{\mathbb{R}} \hat{f}(\mu) \hat{g}(\mu)^* d\mu. \tag{2.3.17}$$

Combining (2.3.16) with (2.3.17) gives

$$E[\Phi_L([a,b))\Phi_L([a',b'))^*] \;=\; \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{\phi}_{a,b}(\mu) \left(\hat{\phi}_{a',b'}(\mu)\right)^* d\mu = \frac{\Sigma_L}{2\pi} \int_{-\infty}^{\infty} I_{[a,b)}(t) I_{[a',b')}(t)\, dt.$$

This implies immediately that $E[\Phi_L([a,b))\Phi_L([a',b'))^*] = 0$, if $[a,b)$ and $[a',b')$ are disjoint, $E[\Phi_L([a,b))\Phi_L([a,b))^*] = \frac{\Sigma_L \lambda([a,b))}{2\pi}$ and that Φ_L is a random orthogonal measure on the semi-ring of intervals $[a,b)$, which we extend to one on all Borel sets. Therefore, (2.3.15) extends to

$$\int_{-\infty}^{\infty} I_\Delta(t) \Phi_L(dt) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \hat{\phi}_\Delta(\mu)\, \tilde{\Phi}_L(d\mu) \tag{2.3.18}$$

for all Borel sets Δ, where $\hat{\phi}_\Delta = \mathcal{F}_m(I_\Delta)$ is the Fourier transform of I_Δ.

For any function $\varphi \in L^2(M_m(\mathbb{C}))$ there is a sequence of elementary functions $\varphi_k(t)$, $k \in \mathbb{N}$, (i.e., matrix-valued functions of the form $\sum_{i=1}^N C_i I_{\Delta_i}(t)$ with appropriate $N \in \mathbb{N}, C_i \in M_m(\mathbb{C})$ and Borel sets Δ_i) which converges to φ in $L^2(M_m(\mathbb{C}))$. As the Fourier-Plancherel transform is a topological isomorphism that maps $L^2(M_m(\mathbb{C}))$ onto itself, the Fourier-Plancherel transforms $\hat{\varphi}_k(t)$ converge to the Fourier-Plancherel transform $\hat{\varphi}(t)$ in $L^2(M_m(\mathbb{C}))$, which allows us to extend (2.3.18), exchanging the roles of μ and t, to

$$\int_{-\infty}^{\infty} \varphi(\mu)\, \Phi_L(d\mu) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \hat{\varphi}(t)\, \tilde{\Phi}_L(dt) \tag{2.3.19}$$

for all functions φ in $L^2(M_m(\mathbb{C}))$ and their Fourier-Plancherel transforms $\hat{\varphi}$. Now choose $\varphi(\mu) = \frac{e^{i\mu b} - e^{i\mu a}}{i\mu}$, then $\hat{\varphi}(t) = \sqrt{2\pi} I_{[a,b)}(t)$. This shows that

$$\int_{-\infty}^{\infty} \frac{e^{i\mu b} - e^{i\mu a}}{i\mu} \Phi_L(d\mu) = L(b) - L(a)$$

and thus (2.3.13) is shown.

The uniqueness of Φ_L follows easily, as (2.3.13) implies (2.3.19) using arguments analogous to the above ones. □

Note that for one-dimensional random orthogonal measures such results can already be found in Doob (1953, Section IX.4).

Remark 2.3.6. *If we formally differentiate (2.3.13), we obtain*

$$\frac{dL(t)}{dt} = \int_{-\infty}^{\infty} e^{i\mu t}\Phi_L(d\mu),$$

as in the spectral representation differentiation is the transform given by

$$\int_{-\infty}^{\infty} e^{i\mu t}\Phi(d\mu) \mapsto \int_{-\infty}^{\infty} i\mu e^{i\mu t}\Phi(d\mu).$$

Thus, a univariate CARMA processes should have the representation

$$Y(t) = \int_{-\infty}^{\infty} e^{i\mu t}\frac{q(i\mu)}{p(i\mu)}\Phi_L(d\mu), \tag{2.3.20}$$

as this reflects the differential equation (2.3.3). Later, in Theorem 2.3.22, we will see that this is indeed the case. The square integrability necessary for (2.3.20) to be defined, explains why one can only consider CARMA processes with $q < p$ (cf. Lemma 2.3.11).

The next lemma deals with the spectral representation of integrals of processes.

Lemma 2.3.7. *Let Φ be an m-dimensional random orthogonal measure with spectral measure $F(dt) = C\,dt$ for some positive definite $C \in M_m(\mathbb{C})$ and $g \in L^2(M_m(\mathbb{C}))$. Define the m-dimensional random process $G = \{G(t)\}_{t\in\mathbb{R}}$ by*

$$G(t) = \int_{-\infty}^{\infty} e^{i\mu t} g(i\mu)\,\Phi(d\mu).$$

Then G is weakly stationary,

$$\int_0^t G(s)\,ds \;<\; \infty \quad a.s. \quad \text{for every} \quad t > 0 \qquad \text{and}$$

$$\int_0^t G(s)\,ds \;=\; \int_{-\infty}^{\infty} \frac{e^{i\mu t}-1}{i\mu} g(i\mu)\,\Phi(d\mu), \quad t > 0.$$

Proof. Weak stationarity follows immediately from (2.3.11), which implies

$$E[G(t)G(s)^*] = \int_{-\infty}^{\infty} e^{i\mu(t-s)} g(i\mu) C g(i\mu)^* d\mu.$$

The weak stationarity implies that

$$\|G(s)\|_{L^2} := E[\|G(s)\|_2^2]^{1/2} = E[G(s)^*G(s)]^{1/2}$$

is finite and constant, where $\|\cdot\|_2$ denotes the Euclidean norm. Thus an elementary Fubini argument and using $\|\cdot\|_{L^1} \le \|\cdot\|_{L^2}$ gives:

$$E\left\|\int_0^t G(s)ds\right\|_2 \;\le\; E\left[\int_0^t \|G(s)\|_2 ds\right] = \int_0^t E\left[\|G(s)\|_2\right] ds \le \int_0^t \|G(s)\|_{L^2} ds < \infty.$$

In particular, $\int_0^t G(s)ds$ is almost surely finite. Finally, we obtain

$$\begin{aligned}\int_0^t G(s)ds &= \int_0^t \int_{-\infty}^{\infty} e^{i\mu s} g(i\mu)\,\Phi(d\mu)ds = \int_{-\infty}^{\infty} \int_0^t e^{i\mu s} ds\, g(i\mu)\,\Phi(d\mu) \\ &= \int_{-\infty}^{\infty} \frac{e^{i\mu t}-1}{i\mu} g(i\mu)\,\Phi(d\mu),\end{aligned}$$

using a stochastic version of Fubini's theorem (e.g. the obvious multidimensional extension of Gihman and Skorohod (1974, Section IV.4, Lemma 4)). □

Before turning to a theorem enabling us to define MCARMA processes we establish three lemmata and one corollary which contain necessary technical results relating the zeros of what is to become the autoregressive polynomial to the spectrum of a particular matrix A. The first lemma contains furthermore some additional insight into the eigenvectors of A.

Lemma 2.3.8. *Let $A_1, \ldots, A_p \in M_m(\mathbb{C})$, $p \in \mathbb{N}$, define $P : \mathbb{C} \to M_m(\mathbb{C})$, $z \mapsto I_m z^p + A_1 z^{p-1} + A_2 z^{p-2} + \ldots + A_p$ and set*

$$\mathcal{N}(P) = \{z \in \mathbb{C} : \det(P(z)) = 0\}, \tag{2.3.21}$$

i.e., $\mathcal{N}(P)$ is the set of all $z \in \mathbb{C}$ such that $P(z) \notin \mathcal{G}l_m(\mathbb{C})$. Furthermore, set

$$A = \begin{pmatrix} 0 & I_m & 0 & \ldots & 0 \\ 0 & 0 & I_m & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & \ldots & \ldots & 0 & I_m \\ -A_p & -A_{p-1} & \ldots & \ldots & -A_1 \end{pmatrix} \in M_{mp}(\mathbb{C}) \tag{2.3.22}$$

and denote the spectrum of A by $\sigma(A)$. Then $\mathcal{N}(P) = \sigma(A)$ and $\bar{x} \in \mathbb{C}^{mp} \setminus \{0\}$ is an eigenvector of A with corresponding eigenvalue λ, if and only if there is an $\tilde{x} \in \mathrm{Ker}P(\lambda) \setminus \{0\}$, such that $\bar{x} = (\tilde{x}^*, (\lambda\tilde{x})^*, \ldots, (\lambda^{p-1}\tilde{x})^*)^*$. Moreover, $0 \in \sigma(A)$, if and only if $0 \in \sigma(A_p)$.

Proof. It is immediate from the structure of A, that A is of full rank, if and only if A_p is of full rank.

Let λ be an eigenvalue of A and $\bar{x} = (x_1^*, \ldots x_p^*)^* \in \mathbb{R}^{mp}$, $x_i \in \mathbb{R}^m$, a corresponding eigenvector, i.e., $A\bar{x} - \lambda\bar{x} = 0$ from which $\lambda x_1 = x_2$, $\lambda x_2 = x_3, \ldots, \lambda x_{p-1} = x_p$, $\lambda x_p + A_1 x_p + A_2 x_{p-1} + \ldots + A_p x_1 = 0$ follows. Hence, $x_i = \lambda^{i-1} x_1$, $i = 1, 2, \ldots, p$ and

$$\lambda^p x_1 + A_1\lambda^{p-1}x_1 + A_2\lambda^{p-2}x_1 + \ldots + A_p x_1 = (I_m\lambda^p + A_1\lambda^{p-1} + \ldots + A_p)x_1 = 0. \tag{2.3.23}$$

As $\bar{x} \neq 0$, we have $x_1 \neq 0$ and (2.3.23) gives $x_1 \in \mathrm{Ker}P(\lambda)$. Hence, we can set $\tilde{x} = x_1$. Furthermore the non-triviality of the kernel of $P(\lambda)$ implies $\det(P(\lambda)) = 0$. Thus $\mathcal{N}(P) \supseteq \sigma(A)$ has been established.

Now we turn to the converse implication. Let $\lambda \in \mathcal{N}(P)$, then $P(\lambda)$ has a non-trivial kernel. Take any $\tilde{x} \in \mathrm{Ker}P(\lambda) \setminus \{0\}$ and set

$$\bar{x} = (\tilde{x}^*, (\lambda\tilde{x})^*, \ldots, (\lambda^{p-1}\tilde{x})^*)^*.$$

Then (2.3.23) shows that $A\bar{x} = \lambda\bar{x}$ and thus $\lambda \in \sigma(A)$. Therefore $\mathcal{N}(P) \subseteq \sigma(A)$ and $\bar{x}$ is an eigenvector of A to the eigenvalue λ. □

Corollary 2.3.9. *$\sigma(A) \subseteq (-\infty, 0) + i\mathbb{R}$ if and only if $\mathcal{N}(P) \subseteq (-\infty, 0) + i\mathbb{R}$.*

Lemma 2.3.10. *If $\mathcal{N}(P) \subseteq \mathbb{R}\backslash\{0\} + i\mathbb{R}$, then $P(iz) \in \mathcal{G}l_m(\mathbb{C})$ for all $z \in \mathbb{R}$.*

Proof. As all zeros of $\det(P(z))$ have non-vanishing real part, all zeros of $\det(P(iz))$ must have non-vanishing imaginary part and thus $P(iz)$ is invertible for all $z \in \mathbb{R}$. □

Lemma 2.3.11. *Let $C_0, C_1, \ldots, C_{p-1} \in M_m(\mathbb{C})$ and $R(z) = \sum_{i=0}^{p-1} C_i z^i$. Assume that $\mathcal{N}(P) \subseteq \mathbb{R}\backslash\{0\} + i\mathbb{R}$, then*

$$\int_{-\infty}^{\infty} \|P(iz)^{-1}R(iz)\|^2 \, dz < \infty,$$

where $P(z) = I_m z^p + A_1 z^{p-1} + ... + A_p$.

Proof. As $\det(P(iz))$, $z \in \mathbb{R}$, has no zeros, $\|P(iz)^{-1}R(iz)\|$ is finite for all $z \in \mathbb{R}$, continuous and thus bounded on any compact set. Hence,

$$\int_{-K}^{K} \|P(iz)^{-1}R(iz)\|^2 \, dz$$

exists for all $K \in \mathbb{R}$. For any $x \in \mathbb{R}^m$ we have

$$\begin{aligned}\|P(z)x\| &= \left\| \left(I_m z^p + \sum_{k=0}^{p-1} A_{p-k} z^k \right) x \right\| \geq \|z^p x\| - \left\| \sum_{k=0}^{p-1} A_{p-k} z^k x \right\| \\ &\geq \left(|z|^p - \sum_{k=0}^{p-1} \|A_{p-k}\| |z^k| \right) \|x\|.\end{aligned}$$

Thus, there is $K > 0$ such that $\|P(z)x\| \geq |z|^p \|x\|/2$ for all z such that $|z| \geq K$, $x \in \mathbb{R}^m$. This implies $\|P(z)^{-1}\| \leq 2|z|^{-p} \, \forall \, |z| \geq K$ and thus for all $z \in \mathbb{R}$, $|z| \geq K$,

$$\|P(iz)^{-1}R(iz)\|^2 \leq \|P(iz)^{-1}\|^2 \|R(iz)\|^2 \leq \frac{4}{|z|^{2p}} \left(\sum_{i=0}^{p-1} \|C_i\| |z|^i \right)^2,$$

which gives that $\int_{-\infty}^{-K} \|P(iz)^{-1}R(iz)\|^2 \, dz$ and $\int_{K}^{\infty} \|P(iz)^{-1}R(iz)\|^2 \, dz$ are finite. □

The following result provides the key to be able to define multivariate CARMA processes.

Theorem 2.3.12. *Let $L = \{L(t)\}_{t\in\mathbb{R}}$ be an m-dimensional square-integrable Lévy process with zero mean and corresponding m-dimensional random orthogonal measure Φ as in Theorem 2.3.5 and $p, q \in \mathbb{N}_0$, $q < p$ (i.e., $p \geq 1$). Let further $A_1, A_2, \ldots, A_p, B_0, B_1, \ldots, B_q \in M_m(\mathbb{R})$, where $B_0 \neq 0$ and define $\beta_1 = \beta_2 = \ldots = \beta_{p-q-1} = 0$ (if $p > q+1$) and $\beta_{p-j} = -\sum_{i=1}^{p-j-1} A_i \beta_{p-j-i} + B_{q-j}$ for $j = 0, 1, 2, \ldots, q$. (Alternatively, $\beta_{p-j} = -\sum_{i=1}^{p-j-1} A_i \beta_{p-j-i} + B_{q-j}$ for $j = 0, 1, \ldots, p-1$, setting $B_i = 0$ for $i < 0$.) Assume that A as defined in (2.3.22) satisfies $\sigma(A) \subseteq (-\infty, 0) + i\mathbb{R}$, which implies $A_p \in \mathcal{G}l_m(\mathbb{R})$.*

Denote by

$$G = (G_1^*(t), \ldots, G_p^*(t))^*$$

an mp-dimensional process and set $\beta^* = \left(\beta_1^*, \ldots, \beta_p^*\right)$. *Then the stochastic differential equation*

$$dG(t) = AG(t)dt + \beta dL(t) \tag{2.3.24}$$

is uniquely solved by the process G *given by*

$$\begin{aligned} G_j(t) &= \int_{-\infty}^{\infty} e^{i\lambda t} w_j(i\lambda)\, \Phi(d\lambda), \quad j = 1, 2, \ldots, p, \quad t \in \mathbb{R}, \text{ where} \qquad (2.3.25)\\ w_j(z) &= \frac{1}{z}(w_{j+1}(z) + \beta_j), \quad j = 1, 2, \ldots, p-1 \quad \text{and} \\ w_p(z) &= \frac{1}{z}\left(-\sum_{k=0}^{p-1} A_{p-k} w_{k+1}(z) + \beta_p\right). \end{aligned}$$

The strictly stationary process G *can also be represented as*

$$G(t) = \int_{-\infty}^{t} e^{A(t-s)} \beta\, L(ds), \quad t \in \mathbb{R}. \tag{2.3.26}$$

Moreover, $G(0)$ *and* $\{L(t)\}_{t\geq 0}$ *are independent, in particular,*

$$E[G_j(0)L(t)^*] = 0 \quad \text{for all} \quad t \geq 0, \;\; j = 1, 2, \ldots, p.$$

Finally, it holds that

$$\begin{aligned} w_p(z) &= P(z)^{-1}\left(\beta_p z^{p-1} - \sum_{j=0}^{p-2}\sum_{k=0}^{j} A_{p-k}\beta_{p+k-j-1} z^j\right), \qquad (2.3.27)\\ w_1(z) &= (P(z))^{-1} Q(z), \qquad (2.3.28) \end{aligned}$$

where

$$\begin{aligned} P(z) &= I_m z^p + A_1 z^{p-1} + \ldots + A_p \text{ (“autoregressive polynomial”)}, \\ Q(z) &= B_0 z^q + B_1 z^{q-1} + \ldots + B_q \text{ (“moving-average polynomial”)} \end{aligned}$$

and $\int_{-\infty}^{\infty} \|w_j(i\lambda)\|^2\, d\lambda < \infty$ *for all* $j \in \{1, 2, \ldots, p\}$.

Proof. $A_p \in \mathcal{G}l_m(\mathbb{R})$ follows from Lemma 2.3.8. That (2.3.26) is the strictly stationary solution of (2.3.24) is a standard result, since all elements of $\sigma(A)$ have strictly negative real part, and a simple application of Gronwall's Lemma shows that the solution of (2.3.24) is a.s. unique for all $t \in \mathbb{R}$ (see e.g. Ikeda and Watanabe (1989), Theorem 3.1). Since $G(0) = \int_{-\infty}^{0} e^{-As}\beta\, L(ds)$ and the processes $\{L(t)\}_{t<0}$ and $\{L(t)\}_{t\geq 0}$ are independent according to our definition (2.2.3) of L, $G(0)$ and $\{L(t)\}_{t\geq 0}$ are independent.

To prove (2.3.27) and (2.3.28) we first show

$$w_j(z) = \frac{1}{z^{p-j}} \left(w_p(z) + \sum_{i=1}^{p-j} \beta_{p-i} z^{i-1} \right) \quad \text{for} \quad j = 1, \ldots, p-1. \tag{2.3.29}$$

In fact, for $p - j = 1$ (2.3.29) becomes $w_{p-1} = \frac{1}{z}(w_p(z) + \beta_{p-1})$ which proves the identity for $j = p - 1$ immediately. Assume the identity holds for $j + 1 \in \{2, 3, \ldots, p-1\}$, then

$$\begin{aligned} w_j(z) &= \frac{1}{z}(w_{j+1}(z) + \beta_j) = \frac{1}{z} \left[\frac{1}{z^{p-j-1}} \left(w_p(z) + \sum_{i=1}^{p-j-1} \beta_{p-i} z^{i-1} \right) + \beta_j \right] \\ &= \frac{1}{z^{p-j}} \left(w_p(z) + \sum_{i=1}^{p-j} \beta_{p-i} z^{i-1} \right), \end{aligned}$$

which proves (2.3.29). Now we turn to (2.3.27):

$$\begin{aligned} w_p(z) &= \frac{1}{z} \left(-\sum_{k=0}^{p-1} A_{p-k} w_{k+1}(z) + \beta_p \right) \\ &= \frac{1}{z} \left[-\sum_{k=0}^{p-1} A_{p-k} \left(\frac{1}{z^{p-k-1}} \left(w_p(z) + \sum_{i=1}^{p-k-1} \beta_{p-i} z^{i-1} \right) \right) \right] + \frac{\beta_p}{z}. \end{aligned}$$

It follows,

$$\left(I_m z^p + \sum_{k=0}^{p-1} A_{p-k} z^k \right) w_p(z) = \beta_p z^{p-1} - \sum_{k=0}^{p-1} \sum_{i=1}^{p-k-1} A_{p-k} \beta_{p-i} z^{k+i-1}.$$

Set $j = k + i - 1$, then

$$\begin{aligned} w_p(z) &= (P(z))^{-1} \left(\beta_p z^{p-1} - \sum_{k=0}^{p-2} \sum_{j=k}^{p-2} A_{p-k} \beta_{p+k-j-1} z^j \right) \\ &= (P(z))^{-1} \left(\beta_p z^{p-1} - \sum_{j=0}^{p-2} \sum_{k=0}^{j} A_{p-k} \beta_{p+k-j-1} z^j \right), \end{aligned}$$

which proves (2.3.27).

Let now $l \in \{1, 2, \ldots, p-1\}$. Then setting $A_0 = I_m$,

$$\begin{aligned} w_l(z) &= \frac{1}{z^{p-l}} \left(w_p(z) + \sum_{i=1}^{p-l} \beta_{p-i} z^{i-1} \right) \\ &= \frac{1}{z^{p-l}} \left[(P(z))^{-1} \left(\beta_p z^{p-1} - \sum_{j=0}^{p-2} \sum_{k=0}^{j} A_{p-k} \beta_{p+k-j-1} z^j \right) + \sum_{i=1}^{p-l} \beta_{p-i} z^{i-1} \right] \\ &= \frac{(P(z))^{-1}}{z^{p-l}} \left[\beta_p z^{p-1} - \sum_{j=0}^{p-2} \sum_{k=0}^{j} A_{p-k} \beta_{p+k-j-1} z^j + \left(\sum_{i=1}^{p-l} \beta_{p-i} z^{i-1} \right) \right] \\ &= \frac{(P(z))^{-1}}{z^{p-l}} \left[\beta_p z^{p-1} - \sum_{j=0}^{p-2} \sum_{k=0}^{j} A_{p-k} \beta_{p+k-j-1} z^j + \sum_{k=0}^{p} \sum_{i=0}^{p-l-1} A_{p-k} \beta_{p-i-1} z^{i+k} \right]. \end{aligned}$$

Setting $j = k + l$ we obtain,

$$\begin{aligned} w_l(z) &= \frac{(P(z))^{-1}}{z^{p-l}} \left[\beta_p z^{p-1} - \sum_{j=0}^{p-2} \sum_{k=0}^{j} A_{p-k} \beta_{p+k-j-1} z^j + \sum_{k=0}^{p} \sum_{j=k}^{k+p-l-1} A_{p-k} \beta_{p+k-j-1} z^j \right] \\ &= \frac{(P(z))^{-1}}{z^{p-l}} \left[- \sum_{j=0}^{p-2} \sum_{k=0}^{j} A_{p-k} \beta_{p+k-j-1} z^j + \sum_{k=0}^{p-l-1} \sum_{j=k}^{p-l-1} A_{p-k} \beta_{p+k-j-1} z^j \right. \\ &\quad \left. + \beta_p z^{p-1} + \sum_{k=p-l}^{p} \sum_{j=k}^{k+p-l-1} A_{p-k} \beta_{p+k-j-1} z^j + \sum_{k=1}^{p-l-1} \sum_{j=p-l}^{k+p-l-1} A_{p-k} \beta_{p+k-j-1} z^j \right]. \end{aligned}$$

It follows,

$$\begin{aligned} w_l(z) &= (P(z))^{-1} \left[\beta_p z^{l-1} - \sum_{j=p-l}^{p-2} \sum_{k=0}^{j} A_{p-k} \beta_{p+k-j-1} z^{j-p+l} \right. \\ &\quad \left. + \sum_{k=p-l}^{p} \sum_{j=k}^{k+p-l-1} A_{p-k} \beta_{p+k-j-1} z^{j-p+l} + \sum_{k=1}^{p-l-1} \sum_{j=p-l}^{k+p-l-1} A_{p-k} \beta_{p+k-j-1} z^{j-p+l} \right]. \end{aligned}$$

The last term in the bracket appears only if $p - l - 1 \geq 1$. Therefore, the whole term in the bracket is a polynomial of at most order $p - 1$. Fixing $l = 1$ and setting $i = j - p + 1$ we obtain,

$$\begin{aligned} w_1(z) &= P(z)^{-1} \left[\beta_p + \sum_{k=p-1}^{p} \sum_{i=k-p+1}^{k-1} A_{p-k} \beta_{k-i} z^i + \sum_{k=1}^{p-2} \sum_{i=0}^{k-1} A_{p-k} \beta_{k-i} z^i \right] \\ &= P(z)^{-1} \left[\beta_p + \sum_{k=1}^{p-1} \sum_{i=0}^{k-1} A_{p-k} \beta_{k-i} z^i + A_0 \sum_{i=1}^{p-1} \beta_{p-i} z^i \right] \\ &= P(z)^{-1} \left[\sum_{i=0}^{p-1} \beta_{p-i} z^i + \sum_{i=0}^{p-2} \sum_{k=i+1}^{p-1} A_{p-k} \beta_{k-i} z^i \right]. \end{aligned}$$

Using the fact that $\beta_1 = B_{q-p+1}$ and setting $j = p - k$, we finally get

$$\begin{aligned} w_1(z) &= (P(z))^{-1} \left[B_{q-p+1} z^{p-1} + \sum_{i=0}^{p-2} \left(\beta_{p-i} + \sum_{j=1}^{p-i-1} A_j \beta_{p-j-i} \right) z^i \right] \\ &= P(z)^{-1} \left[B_{q-p+1} z^{p-1} + \sum_{i=0}^{p-2} B_{q-i} z^i \right] = P(z)^{-1} \sum_{i=0}^{p-1} B_{q-i} z^i \\ &= P(z)^{-1} \sum_{i=0}^{q} B_{q-i} z^i = P(z)^{-1} Q(z). \end{aligned}$$

The finiteness of $\int_{-\infty}^{\infty} \|w_j(i\lambda)\|^2 \, d\lambda$ for all $j = 1, 2, \ldots, p$ is now a direct consequence of Lemmata 2.3.10, 2.3.11 and Corollary 2.3.9.

It remains to show that the process defined in (2.3.25) solves (2.3.24): For $j = 1, \ldots, p$ we have as a consequence of (2.3.25),

$$G_j(t) - G_j(0) = \int_{-\infty}^{\infty} (e^{i\lambda t} - 1) w_j(i\lambda)\, \Phi(d\lambda). \tag{2.3.30}$$

For $j = 1, \ldots, p-1$ the recursion for w_j together with Lemma 2.3.7 gives

$$\begin{aligned} G_j(t) - G_j(0) &= \int_{-\infty}^{\infty} \frac{e^{i\lambda t} - 1}{i\lambda} w_{j+1}(i\lambda)\, \Phi(d\lambda) + \beta_j \int_{-\infty}^{\infty} \frac{e^{i\lambda t} - 1}{i\lambda}\, \Phi(d\lambda) \\ &= \int_0^t \int_{-\infty}^{\infty} w_{j+1}(i\lambda) e^{i\lambda s}\, \Phi(d\lambda)\, ds + \beta_j L(t) \\ &= \int_0^t G_{j+1}(s)\, ds + \beta_j L(t). \end{aligned}$$

Hence,

$$dG_j(t) = G_{j+1}(t)dt + \beta_j dL(t). \tag{2.3.31}$$

Analogously we obtain for G_p,

$$\begin{aligned} G_p(t) - G_p(0) &= \int_{-\infty}^{\infty} (e^{i\lambda t} - 1) w_p(i\lambda)\, \Phi(d\lambda) \\ &= -\sum_{k=0}^{p-1} \int_0^t \int_{-\infty}^{\infty} e^{i\lambda s} A_{p-k} w_{k+1}(i\lambda)\, \Phi(d\lambda)\, ds + \beta_p L(t) \\ &= -\left(\int_0^t A_p G_1(s) + \cdots + A_1 G_p(s)\, ds \right) + \beta_p L(t). \end{aligned}$$

Therefore,

$$dG_p(t) = -(A_p G_1(t) + \ldots + A_1 G_p(t))dt + \beta_p dL(t).$$

Together with (2.3.31) this gives that the process G defined by (2.3.25) solves (2.3.24). □

Obviously, $E[G(t)] = 0$ for the process $G = \{G(t)\}_{t\in\mathbb{R}}$ which solves (2.3.24). Noting that G is a multivariate Ornstein-Uhlenbeck process, the second-order structure follows immediately.

Proposition 2.3.13. *Let $G = \{G(t)\}_{t\in\mathbb{R}}$ be the process that solves (2.3.24). Then its autocovariance matrix function has the form*

$$\Gamma(h) = E[G(t+h)G(t)^*] = e^{Ah}\Gamma(0), \quad h \geq 0, \tag{2.3.32}$$

with $\Gamma(0) = \int_0^\infty e^{Au} \beta \Sigma_L \beta^* e^{A^* u}\, du$ satisfying $A\Gamma(0) + \Gamma(0)A^* = -\beta\Sigma_L\beta^*$.

Proof. (2.3.32) follows from (2.3.11) and the last identity is a standard result from matrix theory (see e.g. Bhatia (1997, Theorem VII.2.3)). □

From Chojnowska-Michalik (1987), Jurek and Mason (1993), Sato and Yamazato (1984) or Wolfe (1982) we know that (2.3.26) is the unique stationary solution to (2.3.24) whenever the Lévy measure ν of the driving process L satisfies

$$\int_{\|x\|\geq 1} \log \|x\| \nu(dx) < \infty.$$

This condition is sufficient (and necessary, provided β is injective) for the stochastic integral in (2.3.26) to exist, as can be seen from (2.2.12) and (2.2.13) with $f(t,s) = e^{A(t-s)}\beta I_{[0,\infty)}(t-s)$. As we shall use this fact later on to define CARMA processes driven by Lévy processes with infinite second moment, we state the following two results on the process G in a general manner.

Proposition 2.3.14. *For any driving Lévy process $L = \{L(t)\}_{t\in\mathbb{R}}$, the process $G = \{G(t)\}_{t\in\mathbb{R}}$ solving (2.3.24) in Theorem 2.3.12 is a temporally homogeneous strong Markov process with an infinitely divisible transition probability $P_t(x, dy)$ having characteristic function*

$$\int_{\mathbb{R}^{mp}} e^{i\langle u,y\rangle} P_t(x,dy) = \exp\left\{ i\langle x, e^{A^*t}u\rangle + \int_0^t \psi_L((e^{Av}\beta)^*u)\, dv \right\}, \quad u \in \mathbb{R}^{mp}. \tag{2.3.33}$$

Proof. See Sato and Yamazato (1984, Th. 3.1) and additionally Protter (2004, Theorem V.32) for the strong Markov property. □

Proposition 2.3.15. *Consider the unique solution $G = \{G(t)\}_{t\geq 0}$ of (2.3.24) with initial value $G(0)$ independent of $L = \{L(t)\}_{t\geq 0}$, where L is a Lévy process on $\mathbb{R}^m$ satisfying $\int_{\|x\|\geq 1} \log\|x\|\nu(dx) < \infty$.*

Let $\mathcal{L}(G(t))$ denote the marginal distribution of the process $G = \{G(t)\}_{t\geq 0}$ at time t. Then there exists a limit distribution F such that $\mathcal{L}(G(t)) \to F$ in distribution as $t \to \infty$. This F is infinitely divisible with characteristic function

$$E\left[e^{i\langle u,F\rangle}\right] = \exp\left\{ \int_0^\infty \psi_L((e^{As}\beta)^*u)\, ds \right\}, \quad u \in \mathbb{R}^{mp}. \tag{2.3.34}$$

Proof. See Sato and Yamazato (1984, Theorem 4.1). □

Remark 2.3.16. *Obviously F is also the marginal distribution of the stationary solution considered in Theorem 2.3.12.*

The sample path behaviour of the process $G = \{G(t)\}_{t\in\mathbb{R}}$ is described below.

Proposition 2.3.17. *If the driving Lévy process $L = \{L(t)\}_{t\in\mathbb{R}}$ of the process $G = \{G(t)\}_{t\in\mathbb{R}}$ in Theorem 2.3.12 is Brownian motion, the sample paths of G are continuous. Otherwise the process G has a jump, whenever L has one. In particular, $\Delta G(t) = \beta\Delta L(t)$.*

2.3.3. Multivariate CARMA processes

We are now in a position to define an m-dimensional CARMA (MCARMA) process by using the spectral representation for square-integrable driving Lévy processes and to extend this definition making use of the insight obtained in Theorem 2.3.12.

Definition 2.3.18 (MCARMA Process)**.** *Let $L = \{L(t)\}_{t\in\mathbb{R}}$ be a two-sided square integrable m-dimensional Lévy-process with*

$$E[L(1)] = 0 \text{ and } E[L(1)L(1)^*] = \Sigma_L.$$

*An m-dimensional Lévy-driven continuous time autoregressive moving average process $\{Y(t)\}_{t\in\mathbb{R}}$ of order (p,q), $p > q$, (*MCARMA(p,q) process*) is defined as*

$$\begin{aligned} Y(t) &= \int_{-\infty}^{\infty} e^{i\lambda t} P(i\lambda)^{-1} Q(i\lambda)\, \Phi(d\lambda), \quad t \in \mathbb{R}, \qquad \textit{where} \qquad (2.3.35)\\ P(z) : &= I_m z^p + A_1 z^{p-1} + ... + A_p, \\ Q(z) : &= B_0 z^q + B_1 z^{q-1} + ... + B_q \quad \textit{and} \end{aligned}$$

Φ is the Lévy orthogonal random measure of Theorem 2.3.5 satisfying $E[\Phi(d\lambda)] = 0$ and $E[\Phi(d\lambda)\Phi(d\lambda)^*] = \frac{d\lambda}{2\pi}\Sigma_L$. Here $A_j \in M_m(\mathbb{R})$, $j = 1, ..., p$ and $B_j \in M_m(\mathbb{R})$ are matrices satisfying $B_q \neq 0$ and $\mathcal{N}(P) := \{z \in \mathbb{C} : det(P(z)) = 0\} \subset \mathbb{R}\backslash\{0\} + i\mathbb{R}$.

The process G defined as in Theorem 2.3.12 is called the state space representation of the MCARMA process Y.

Remark 2.3.19. *(i) There are several reasons why the name "multivariate continuous time ARMA process" is indeed appropriate. The same arguments as in Remark 2.3.6 show that an MCARMA process Y can be interpreted as a solution to the p-th order m-dimensional differential equation*

$$P(D)Y(t) = Q(D)DL(t),$$

where D denotes the differentiation operator. Moreover, the upcoming Theorem 2.3.22 shows that for $m = 1$ the well-known univariate CARMA processes are obtained (an alternative proof for some special cases needing no finite variances is to be found in Lemma 6.6.1 in Chapter 6) and, finally, the spectral representation (2.3.35) is the obvious continuous time analogue of the spectral representation of multivariate discrete time ARMA processes (see, for instance, Brockwell and Davis (1991, Section 11.8)).

(ii) The well-definedness is ensured by Lemma 2.3.11. Observe, moreover, that, if the determinant $\det(P(z))$ *has zeros with positive real part, all assertions of Theorem 2.3.12 except the alternative representation (2.3.26) and the independence of $G(0)$ and $\{L(t)\}_{t\geq 0}$ remain still valid interpreting the stochastic differential equation as an integral equation as in the proof of the theorem. However, in this case the process is no longer causal, i.e. adapted to the natural filtration of the driving Lévy process.*

(iii) Assuming $E[L(1)] = 0$ is actually no restriction. If $E[L(1)] = \mu_L \neq 0$, one simply observes that $\tilde{L}(t) = L(t) - \mu_L t$ has zero expectation and

$$P(D)^{-1}Q(D)DL(t) = P(D)^{-1}Q(D)D\tilde{L}(t) + P(D)^{-1}Q(D)\mu_L.$$

The first term is the MCARMA process driven by $\tilde{L}(t)$ and the second an ordinary differential equation having the unique "stationary" solution $-A_p^{-1}B_q\mu_L$, as simple calculations show. Thus, the definition can be immediately extended to $E[L(1)] \neq 0$. Moreover, it is easy to see that the SDE representation given in Theorem 2.3.12 still holds and one can also extend the spectral representation by adding an atom with mass μ_L to $\Phi_{\tilde{L}}$ at 0.

(iv) Furthermore, observe that the representation of MCARMA processes by the stochastic differential equation (2.3.24) is a continuous time version of state space representations for (multivariate) ARMA processes as given in Brockwell and Davis (1991, Example 12.1.5) or Wei (1990, p. 387). In the univariate Gaussian case it can already be found in Arató (1982, Lemma 3, Chapter 2.2).

As already noted before, we extend the definition of MCARMA processes to driving Lévy processes L with finite logarithmic moment using Theorem 2.3.12. As they agree with the above defined MCARMA processes, when L is square-integrable, and are always causal, we call them *causal MCARMA processes*.

Definition 2.3.20 (Causal MCARMA Process)**.** *Let $L = \{L(t)\}_{t\in\mathbb{R}}$ be an m-dimensional Lévy process satisfying*

$$\int_{\|x\|\geq 1} \log\|x\|\,\nu(dx) < \infty, \tag{2.3.36}$$

$p, q \in \mathbb{N}_0$ with $q < p$, and further $A_1, A_2, \ldots, A_p, B_0, B_1, \ldots, B_q \in M_m(\mathbb{R})$, where $B_0 \neq 0$. Define the matrices A, β and the polynomial P as in Theorem 2.3.12 and assume $\sigma(A) = \mathcal{N}(P) \subseteq (-\infty, 0) + i\mathbb{R}$. Then the m-dimensional process

$$Y(t) = \left(I_m, 0_{M_m(\mathbb{C})}, \ldots, 0_{M_m(\mathbb{C})}\right) G(t) \tag{2.3.37}$$

where G is the unique stationary solution to $dG(t) = AG(t)dt + \beta dL(t)$ is called causal MCARMA(p, q) *process. Again G is referred to as the state space representation.*

Remark 2.3.21. *In the following we will write "MCARMA" when referring to Definition 2.3.18, "causal MCARMA" when referring to Definition 2.3.20 and "(causal) MCARMA" when referring to both Definitions 2.3.18 and 2.3.20.*

Let us now state a result extending the short memory moving average representation of univariate CARMA processes to our MCARMA processes and showing that our definition is in line with univariate CARMA processes.

Theorem 2.3.22. *Analogously to a one-dimensional CARMA process (confer Equation (2.3.7)), the MCARMA process (2.3.35) can be represented as a moving average process*

$$Y(t) \;=\; \int_{-\infty}^{\infty} g(t-s)\,L(ds), \quad t \in \mathbb{R}, \tag{2.3.38}$$

where the kernel matrix function $g : \mathbb{R} \to M_m(\mathbb{R})$ is given by

$$g(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i\mu t} P(i\mu)^{-1} Q(i\mu)\, d\mu. \tag{2.3.39}$$

Proof. Using the notation of the proof of Theorem 2.3.5 we obtain this immediately from (2.3.19):

$$\begin{aligned} Y(t) &= \int_{-\infty}^{\infty} e^{i\mu t} P(i\mu)^{-1} Q(i\mu)\, \Phi(d\mu) \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{i\mu(t-s)} P(i\mu)^{-1} Q(i\mu)\, d\mu\, \tilde{\Phi}_L(ds) \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{i\mu(t-s)} P(i\mu)^{-1} Q(i\mu)\, d\mu\, L(ds) = \int_{-\infty}^{\infty} g(t-s)\, L(ds). \end{aligned}$$

□

Remark 2.3.23. *For causal MCARMA processes an analogous result holds with the kernel function g replaced by*

$$\tilde{g}(s) = (I_m, 0_{M_m(\mathbb{C})}, \ldots, 0_{M_m(\mathbb{C})}) e^{As} \beta I_{[0,\infty)}(s).$$

Moreover, the function g simplifies in the square-integrable causal case as the following extension of a well-known result for univariate CARMA processes shows.

Lemma 2.3.24. *Assume that $\sigma(A) = \mathcal{N}(P) \subseteq (-\infty, 0) + i\mathbb{R}$. Then the function g given in (2.3.39) vanishes on the negative real line.*

Proof. We need the following consequence of the residue theorem from complex analysis (cf., for instance, Lang (1993, Section VI.2, Theorem 2.2)):

Let q and $p : \mathbb{C} \mapsto \mathbb{C}$ be polynomials where p is of higher degree than q. Assume that p has no zeros on the real line. Then

$$\int_{-\infty}^{\infty} \frac{q(t)}{p(t)} \exp(i\alpha t) dt = 2\pi i \sum_{z \in \mathbb{C}: \Im(z) > 0, p(iz) = 0} Res\,(f, z)\ \forall\, \alpha > 0, \tag{2.3.40}$$

$$\int_{-\infty}^{\infty} \frac{q(t)}{p(t)} \exp(i\alpha t) dt = -2\pi i \sum_{z \in \mathbb{C}: \Im(z) < 0, p(iz) = 0} Res\,(f, z)\ \forall\, \alpha < 0 \tag{2.3.41}$$

with $f : \mathbb{C} \mapsto \mathbb{C}$, $z \mapsto \frac{q(z)}{p(z)} \exp(i\alpha z)$ and $Res(f, a)$ denoting the residual of the function f at point a.

Turning to our function g, we have from elementary matrix theory that

$$P(iz)^{-1} Q(iz) = \frac{S(z)}{\det(P(iz))}$$

where $S : \mathbb{C} \mapsto M_m(\mathbb{C})$ is some matrix-valued polynomial in z. Observe that $\det(P(iz))$ is a complex-valued polynomial in z and that Lemma 2.3.11 applied to $R = Q$ implies that $\det(P(iz))$ is of higher degree than $S(z)$. Thus, we can apply the above stated results from complex function theory component-wise to (2.3.39). But as all zeros of $\det(P(z))$ are in the left half plane $(-\infty, 0) + i\mathbb{R}$, all zeros of $\det(P(iz))$ are in the upper half plane $\mathbb{R} + i(0, \infty)$ and therefore (2.3.41) shows that

$$g(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i\mu t} P(i\mu)^{-1} Q(i\mu)\, d\mu = 0 \text{ for all } t < 0.$$

□

Remark 2.3.25. *The above result reflects the causality, i.e., that the MCARMA process $Y(t)$ only depends on the past of the driving Lévy process, i.e., on $\{L(s)\}_{s \le t}$. Similarly g vanishes on the positive half line, if $\mathcal{N}(P) \subset (0, \infty) + i\mathbb{R}$. In this case the MCARMA process $Y(t)$ depends only on the future of the driving Lévy process, i.e., on $\{L(s)\}_{s \ge t}$. In all other non-causal cases the MCARMA process depends on the driving Lévy process at all times.*

Using the kernel representations, strict stationarity of MCARMA processes is obtained by applying Applebaum (2004, Theorem 4.3.16).

Proposition 2.3.26. *The (causal) MCARMA process is strictly stationary.*

Furthermore, we can characterize the stationary distribution by applying representation (2.3.38) and the results of Sato (2006) mentioned at the end of Section 2.2.2.

Proposition 2.3.27. *If (γ, σ, ν) is the characteristic triplet of the driving Lévy process L, then the distribution of the MCARMA process $Y(t)$ is infinitely divisible for $t \in \mathbb{R}$ and the characteristic triplet of the stationary distribution is $(\gamma_Y^\infty, \sigma_Y^\infty, \nu_Y^\infty)$, where*

$$\begin{aligned}
\gamma_Y^\infty &= \int_{\mathbb{R}} g(s)\gamma\, ds + \int_{\mathbb{R}} \int_{\mathbb{R}^m} g(s)x[I_{[0,1]}(\|g(s)x\|) - I_{[0,1]}(\|x\|)]\, \nu(dx)\, ds, \\
\sigma_Y^\infty &= \int_{\mathbb{R}} g(s)\sigma g^*(s) ds, \\
\nu_Y^\infty(B) &= \int_{\mathbb{R}} \int_{\mathbb{R}^m} I_B(g(s)x)\, \nu(dx)\, ds.
\end{aligned} \tag{2.3.42}$$

For a causal MCARMA process the same result holds with g replaced by $\tilde{g}$.

2.3.4. Further properties of MCARMA processes

Having defined multivariate CARMA processes above, we analyse their probabilistic behaviour further in this section. First we turn to the second order properties.

Proposition 2.3.28. *Let $\{Y(t)\}_{t \in \mathbb{R}}$ be the MCARMA process defined by (2.3.35). Then its autocovariance matrix function is given by*

$$\Gamma_Y(h) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i\lambda h} P(i\lambda)^{-1} Q(i\lambda) \Sigma_L Q(i\lambda)^* (P(i\lambda)^{-1})^*\, d\lambda, \quad h \in \mathbb{R}.$$

Proof. It follows directly from the spectral representation (2.3.35) that the MCARMA process $Y = \{Y(t)\}_{t\in\mathbb{R}}$ has the spectral density

$$f_Y(\lambda) = \frac{1}{2\pi} P(i\lambda)^{-1} Q(i\lambda) \Sigma_L Q(i\lambda)^* (P(i\lambda)^{-1})^*, \quad \lambda \in \mathbb{R}. \tag{2.3.43}$$

The autocovariance function is the Fourier transform of (2.3.43). □

Remark 2.3.29. *Note that in Proposition 2.3.13 we already obtained an expression for the autocovariance matrix function of the process $\{G(t)\}_{t\in\mathbb{R}}$ of Theorem 2.3.12. The upper left $m \times m$ block of (2.3.32) is also equal to Γ_Y.*

Regarding the general existence of moments, it is mainly the driving Lévy process that matters.

Proposition 2.3.30. *Let Y be a causal MCARMA process and assume that the driving Lévy process L is in $L^r(\Omega, P)$ for some $r > 0$. Then Y and its state space representation G are in $L^r(\Omega, P)$. Provided β is injective, the converse is true as well for G.*

Proof. We use the general fact that an infinitely divisible distribution with characteristic triplet (γ, σ, ν) has finite r-th moment, if and only if

$$\int_{\|x\|\geq C} \|x\|^r \nu(dx) < \infty$$

for one and hence all $C > 0$ (see Sato (1999, Corollary 25.8)). Using the Kernel representation (2.3.38) with

$$\tilde{g}(s) = (I_m, 0_{M_m(\mathbb{C})}, \ldots, 0_{M_m(\mathbb{C})}) e^{As} \beta I_{[0,\infty)}(s),$$

(2.3.42) and the fact that there are $C, c > 0$ such that

$$\|(I_m, 0_{M_m(\mathbb{C})}, \ldots, 0_{M_m(\mathbb{C})}) e^{As} \beta\| \leq C e^{-cs}$$

we obtain for the stationary distribution of Y

$$\begin{aligned}
\int_{\|x\|\geq 1} \|x\|^r \nu_Y^\infty(dx) &= \int_0^\infty \int_{\mathbb{R}^m} I_{[1,\infty)} \left(\left\| (I_m, 0_{M_m(\mathbb{C})}, \ldots, 0_{M_m(\mathbb{C})}) e^{As} \beta x \right\| \right) \\
&\quad \times \left\| (I_m, 0_{M_m(\mathbb{C})}, \ldots, 0_{M_m(\mathbb{C})}) e^{As} \beta x \right\|^r \nu(dx) ds \\
&\leq \int_0^\infty \int_{\mathbb{R}^m} I_{[1,\infty)} \left(C e^{-cs} \|x\| \right) C^r e^{-rcs} \|x\|^r \nu(dx) ds \\
&= \int_{\|x\|\geq 1/C} \int_0^{\frac{\log(1/(C\|x\|))}{-c}} C^r e^{-rcs} \|x\|^r \, ds \nu(dx) \\
&= \frac{C^r}{rc} \int_{\|x\|\geq 1/C} (\|x\|^r - 1/C^r) \, \nu(dx),
\end{aligned}$$

which is finite, if and only if L has a finite r-th moment.

Basically the same arguments apply to $G(t) = \int_{-\infty}^t e^{A(t-s)} \beta L(ds)$. Provided β is injective, there are $D, d > 0$ such that $\|e^{As}\beta\| \geq D e^{-ds}$ and calculations analogous to the above ones lead to a lower bound which establishes the necessity of $L \in L^r$ for $G \in L^r$. □

Since the characteristic function of $Y(t)$ for each t is explicitly given, we can investigate the existence of a C_b^∞ density, where C_b^∞ denotes the space of bounded, continuous, infinitely often differentiable functions whose derivatives are bounded.

Proposition 2.3.31. *Suppose that there exists an $\alpha \in (0,2)$ and a constant $C > 0$ such that*

$$\int_{\mathbb{R}} \int_{\mathbb{R}^m} |\langle u, g(t-s)x\rangle|^2 \, 1_{\{|\langle u, g(t-s)x\rangle| \leq 1\}} \, \nu(dx)\, ds \geq C\|u\|^{2-\alpha} \tag{2.3.44}$$

for any vector u such that $\|u\| \geq 1$. Then the MCARMA process $Y(t)$ has a C_b^∞ density.

The same holds for a causal MCARMA $Y(t)$ process with g replaced by $\tilde{g}$.

Proof. It is sufficient to show that $\int \|u\|^k \|\Phi(u)\|\, du < \infty$ for any non-negative integer k, where Φ denotes the characteristic function of $Y(t)$ (see e.g. Picard (1996, Proposition 0.2)).

The characteristic function of the (causal) MCARMA process $Y(t)$ is given by

$$\Phi(u) = \exp\left\{ \int_{\mathbb{R}} \int_{\mathbb{R}^m} \left[e^{i\langle u, g(t-s)x\rangle} - 1 - i\langle u, g(t-s)x\rangle I_{\{|\langle u, g(t-s)x\rangle| \leq 1\}} \right] \nu(dx)\, ds \right\},$$

where g stands for either g or $\tilde{g}$. Thus,

$$\begin{aligned}
\|\Phi(u)\| &= \left(\exp\left\{ \int_{\mathbb{R}} \int_{\mathbb{R}^m} \left[e^{i\langle u, g(t-s)x\rangle} + e^{-i\langle u, g(t-s)x\rangle} - 2 \right] \nu(dx)\, ds \right\} \right)^{1/2} \\
&= \exp\left\{ \int_{\mathbb{R}} \int_{\mathbb{R}^m} \left(\cos\langle u, g(t-s)x\rangle - 1 \right) \nu(dx)\, ds \right\} \\
&\leq \exp\left\{ \int_{\mathbb{R}} \int_{\mathbb{R}^m} \left(\cos\langle u, g(t-s)x\rangle - 1 \right) I_{\{|\langle u, g(t-s)x\rangle| \leq 1\}} \, \nu(dx)\, ds \right\},
\end{aligned}$$

as $\cos\langle u, g(t-s)x\rangle - 1 \leq 0$. Then, using the inequality $1 - \cos(z) \geq 2(z/\pi)^2$ for $|z| \leq \pi$ and Assumption (2.3.44) we have

$$\begin{aligned}
\|\Phi(u)\| &\leq \exp\left\{ -\tilde{C} \int_{\mathbb{R}} \int_{\mathbb{R}^m} |\langle u, g(t-s)x\rangle|^2 I_{\{|\langle u, g(t-s)x\rangle| \leq 1\}} \, \nu(dx)\, ds \right\} \\
&\leq \exp\{-C\|u\|^{2-\alpha}\},
\end{aligned}$$

where $C, \tilde{C} > 0$ are generic constants and the proof is complete. The inequality $1 - \cos(z) \geq 2(z/\pi)^2$ for $|z| \leq \pi$ can be easily shown: Define $f(z) = 1 - \cos(z) - 2(z/\pi)^2$. Then $f(0) = f(\pi) = 0$ and there is $y \in (0, \pi)$ such that $f'(z) > 0$, $z \in [0, y)$ and $f'(z) < 0$, $z \in (y, \pi]$. Hence, $f(z) > 0$ for all $z \in (0, \pi)$. □

We summarize the sample path behaviour of the MCARMA(p,q) process $Y = \{Y(t)\}_{t \in \mathbb{R}}$, which is immediate from the state space representation (2.3.24) and the proof of Theorem 2.3.12.

Proposition 2.3.32. *If $p > q+1$, then the (causal) MCARMA(p,q) process $Y = \{Y(t)\}_{t\in\mathbb{R}}$ is $(p-q-1)$-times differentiable. Using the state space representation $G = \{G(t)\}_{t\in\mathbb{R}}$ we have $\frac{d^i}{dt^i}Y(t) = G_{i+1}(t)$ for $i = 1, 2, \ldots, p-q-1$.*
If $p = q+1$, then $\Delta Y(t) = \beta_1 \Delta L(t)$, i.e., Y has a jump, whenever L has one.
If the driving Lévy process $L = \{L(t)\}_{t\in\mathbb{R}}$ of the MCARMA(p,q) process is Brownian motion, the sample paths of Y are continuous and $(p-q-1)$-times continuously differentiable, provided $p > q+1$.

Ergodicity and mixing properties (see, for instance, Doukhan (1994) for a comprehensive treatment) have far reaching implications. We thus conclude the analysis of MCARMA processes with a result on their mixing behaviour. Recall the following notions:

Definition 2.3.33 (cf. Davydov (1973))**.** *A continuous time stationary stochastic process $X = \{X_t\}_{t\in\mathbb{R}}$ is called* strongly (or α-) mixing*, if*

$$\alpha_l := \sup\left\{|P(A\cap B) - P(A)P(B)| : A \in \mathcal{F}_{-\infty}^0, B \in \mathcal{F}_l^\infty\right\} \to 0$$

as $l \to \infty$, where $\mathcal{F}_{-\infty}^0 := \sigma\left(\{X_t\}_{t\le 0}\right)$ and $\mathcal{F}_l^\infty = \sigma\left(\{X_t\}_{t\ge l}\right)$.
It is said to be β- mixing (or completely regular)*, if*

$$\beta_l := E\left(\sup\left\{\left|P(B|\mathcal{F}_{-\infty}^0) - P(B)\right| : B \in \mathcal{F}_l^\infty\right\}\right) \to 0$$

as $l \to \infty$.

Note that $\alpha_l \le \beta_l$ and thus any β-mixing process is strongly mixing.

Proposition 2.3.34. *Let Y be a causal MCARMA process and G be its state space representation. If the driving Lévy process L satisfies*

$$\int_{\|x\|\ge 1} \|x\|^r \nu(dx) < \infty \tag{2.3.45}$$

for some $r > 0$, then G is β-mixing with mixing coefficients $\beta_l = O(e^{-al})$ for some $a > 0$ and Y is strongly mixing. In particular, both G and Y are ergodic.

Proof. Since $G(t) = \int_{-\infty}^t e^{A(t-s)}\beta L(ds)$ is a multidimensional Ornstein-Uhlenbeck process driven by the Lévy process βL, we may apply Masuda (2004, Theorem 4.3) noting that (2.3.45) together with Proposition 2.3.30 ensure that all conditions are satisfied. Hence, the β-mixing of G with exponentially decaying coefficients is shown. But this implies that $G = (G_1^*, G_2^*, \ldots, G_p^*)^*$ is also strongly mixing, which in turn shows the strong mixing property for Y, since Y is equal to G_1 and it is obvious from the definition of strong mixing that strong mixing of a multidimensional process implies strong mixing of its components. Note that we also obtain $\alpha_l \le \beta_l$ for the mixing coefficients α_l of Y. Using the well-known result that mixing implies ergodicity concludes the proof. □

3. A Multivariate Exponential Continuous Time GARCH Process

3.1. Introduction

GARCH type processes have become very popular in financial econometrics to model returns of stocks, exchange rates and other series observed at equidistant time points. They have been designed (see Engle (1982) and Bollerslev (1986)) to capture so-called *stylized facts* of such data, which are e.g. volatility clustering, dependence without correlation and tail heaviness. Another characteristic is that stock returns seem to be negatively correlated with changes in the volatility, i.e. that volatility tends to increase after negative shocks and to fall after positive ones. This effect is called *leverage effect* and cannot be modelled by a GARCH type process without further extensions. This finding led Nelson (1991) to introduce the exponential GARCH process, which is able to model this asymmetry. In that paper the log-volatility of the EGARCH(p, q) process was modelled as an ARMA$(q, p-1)$ process.

Starting with Nelson (1990) continuous time models related to GARCH processes have been investigated for a long time. As several important characteristic features of GARCH processes get lost in the originally studied diffusion limits of GARCH processes, Klüppelberg et al. (2004) introduced the COGARCH process as a continuous time analogue of the GARCH process, which inherits many of the characteristic features of GARCH processes. Likewise, Haug and Czado (2007) recently defined and analysed an EGARCH process in continuous time and Haug (2007) presented first estimation results. Moreover, Haug and Czado (2006) extended this model allowing for long range dependence using fractionally integrated processes.

In this chapter we develop a multivariate version of the continuous time EGARCH process of Haug and Czado (2007). Note that in discrete time matrix exponential GARCH processes have for the first time been studied by Kawakatsu (2006) in a truly multivariate sense, whereas before only the variances, but not the whole covariance matrix, have been modelled as EGARCH processes (cf. Östermark (2001), Tse and Hackard (2004) or Yang and Doong (2004) for some typical examples).

In our EGARCH specification we model the logarithm of the covariance matrix process as a CARMA process in the symmetric matrices using the multivariate CARMA processes introduced in Chapter 2. Taking the exponential then automatically ensures positive definiteness of the covariance matrix process. The standard mathematical fact that the exponential of a symmetric matrix is positive definite seems to have been used only very rarely in order to model covariance matrices so far (the recent paper Kawakatsu (2006), for instance, does not credit any references for this idea). To the best of our knowledge the first appearance in the statistics literature is Chiu, Leonard and Tsui (1996).

This chapter is organized as follows. At the end of this section some notations used throughout the chapter are given. In Section 3.2 we first introduce a general specification

of the discrete time multivariate EGARCH process and propose two ways of modelling asymmetric behaviour in the vectorized log-volatility process. Afterwards we recall some basic facts on multivariate Lévy processes and on the multivariate Lévy-driven CARMA process, as defined in Chapter 2. We further give a sufficient condition for the existence of the α-th exponential moment of a CARMA process. In the first part of Section 3.3 the multivariate exponential continuous time GARCH process (ECOGARCH) is defined and stationarity conditions are discussed. In the second part we show a strong mixing property of the volatility and the return process and shortly consider the mean and autocovariance function of the return process. In the last section it is shown that for an ECOGARCH(1,1) process there exists a sequence of EGARCH(1,1) processes converging to the ECOGARCH process in a strong sense, viz. in the Skorokhod topology in probability. To this end we give a multivariate and infinite time extension of the first jump approximation of a Lévy process presented in Szimayer and Maller (2007) and show that it satisfies an important general condition of Kurtz and Protter (1991) regarding the convergence of solutions of stochastic differential equations.

3.1.1. Notation

Throughout this chapter we write $\mathbb{R}^+$ for the positive real numbers including zero, $\mathbb{R}^{++}$ when zero is excluded, and we denote the set of real $m \times n$ matrices by $M_{m,n}(\mathbb{R})$. If $m = n$ we simply write $M_n(\mathbb{R})$ and denote the group of invertible $n \times n$ matrices by $GL_n(\mathbb{R})$, the linear subspace of symmetric matrices by $\mathbb{S}_n$, the (closed) positive semi-definite cone by $\mathbb{S}_n^+$ and the open (in $\mathbb{S}_n$) positive definite cone by $\mathbb{S}_n^{++}$. I_n stands for the $n \times n$ identity matrix, $\det(A)$ for the determinant and $\sigma(A)$ for the spectrum (the set of all eigenvalues) of a matrix $A \in M_n(\mathbb{R})$. Moreover, $\operatorname{vech} : \mathbb{S}_d \to \mathbb{R}^{d(d+1)/2}$ denotes the "vector-half" operator that stacks the columns of the lower triangular part of a symmetric matrix below another. Finally, A^* is the adjoint of a matrix $A \in M_n(\mathbb{R})$.

Norms of vectors and matrices are denoted by $\|\cdot\|$. If the norm is not specified then it is irrelevant which particular norm is used.

The exponential of a matrix A is denoted by $\exp(A)$ or e^A (see Horn and Johnson (1991, Ch. 6) for a detailed discussion). Recall that for symmetric matrices it is defined by functional calculus and it holds that

$$\exp(A) = \sum_{k=0}^{\infty} \frac{A^k}{k!}.$$

From functional calculus it is immediately clear that the matrix exponential maps the symmetric $d \times d$ matrices to the positive definite ones. Likewise, we can define the logarithm $\log : \mathbb{S}_d^{++} \to \mathbb{S}_d$ by functional calculus and log is the inverse function of $\exp : \mathbb{S}_d \to \mathbb{S}_d^{++}$. Moreover, we denote by $A^{1/2}$ the unique positive semi-definite square root of a matrix $A \in \mathbb{S}_d^+$.

For a matrix A we denote by A_{ij} the element in the i-th row and j-th column and this notation is extended to processes in a natural way.

Regarding all random variables and processes we assume that they are defined on a given appropriate filtered probability space $(\Omega, \mathcal{F}, P, (\mathcal{F}_t)_{t\in\mathbb{R}^+})$ satisfying the usual hypotheses (complete and right continuous filtration). L^p denotes as usual the space of all random variables with a finite p-th moment, i.e. all random variables X with $E(\|X\|^p) < \infty$ in a multivariate setting.

The indicator function of some set A is denoted by 1_A. A sequence $(x_n)_{n\in\mathbb{N}}$ in a normed space is called an ℓ_1-sequence if $\sum_{n\in\mathbb{N}} \|x_n\| < \infty$.

3.2. Related processes

3.2.1. Multivariate EGARCH processes in discrete time

In this section we discuss a multivariate EGARCH model in discrete time providing the basis for our continuous time modelling. We will throughout use a very general specification and briefly indicate some possible specifications to be used in practice later on.

Definition 3.2.1 (Multivariate Discrete Time EGARCH)**.** *Let $d \in \mathbb{N}$, $\mu \in \mathbb{S}_d$, $(\beta_k)_{k\in\mathbb{N}}$ be an ℓ^1-sequence in $M_m(\mathbb{R})$ with $m = \frac{d(d+1)}{2}$, $\epsilon = (\epsilon_n)_{n\in\mathbb{Z}}$ an i.i.d. sequence of $\mathbb{R}^d$-valued random variables with $E(\epsilon_1) = 0$ and $\mathrm{var}(\epsilon_1) = I_d$ and $f : \mathbb{R}^d \to \mathbb{R}^m$ a measurable function such that $f(\epsilon_1) \in L^2$. Then the process $Y = (Y_t)_{t\in\mathbb{Z}}$ with*

$$Y_t = \exp((\mu + H_t)/2)\epsilon_t$$

and vectorized logarithmic volatility process $H = (H_t)_{t\in\mathbb{Z}}$ given by

$$\mathrm{vech}(H_t) = \sum_{k=1}^{\infty} \beta_k f(\epsilon_{t-k})$$

*for all $t \in \mathbb{Z}$ is called a d-*dimensional EGARCH process*.*

As in the univariate case one usually is less interested in the log-volatility being an infinitely moving average process, but more in it being an ARMA process. This leads to the following definition of multivariate EGARCH(p, q) processes, where we refer to Brockwell and Davis (1991) for the necessary background on multivariate ARMA processes.

Definition 3.2.2 (Multivariate Discrete Time EGARCH(p,q))**.** *Let $d, p, q \in \mathbb{N}, \mu \in \mathbb{S}_d$, $\alpha_1, \ldots, \alpha_q, \beta_1, \ldots, \beta_p \in M_m(\mathbb{R})$ with $m = \frac{d(d+1)}{2}$, $\epsilon = (\epsilon_n)_{n\in\mathbb{Z}}$ an i.i.d. sequence of $\mathbb{R}^d$-valued random variables with $E(\epsilon_1) = 0$ and $\mathrm{var}(\epsilon_1) = I_d$ and $f : \mathbb{R}^d \to \mathbb{R}^m$ a measurable function such that $f(\epsilon_1) \in L^2$. Suppose $\alpha_q \neq 0\,, \beta_p \neq 0$ and that*

$$\det(1 - \alpha_1 z - \cdots - \alpha_q z^q) \neq 0$$

on $\{z \in \mathbb{C} \mid |z| \leq 1\}$. Then the process $Y = (Y_t)_{t\in\mathbb{Z}}$, where

$$Y_t = \exp((\mu + H_t)/2)\epsilon_t$$

and the vectorized log volatility H is given by

$$\mathrm{vech}(H_t) = \sum_{k=1}^{p} \beta_k f(\epsilon_{t-k}) + \sum_{k=1}^{q} \alpha_k \mathrm{vech}(H_{t-k})$$

for all $t \in \mathbb{Z}$, is called an EGARCH(p, q) process*.*

So far we have considered a general transformation f of the noise sequence ϵ. Concrete specifications should be made in such a way that the model exhibits some desired properties, e.g. a leverage effect. In the univariate case the "standard choice" introduced originally in Nelson (1991) is

$$f(\eta) = \theta\eta + \gamma(|\eta| - E(|\epsilon_1|))$$

with some real parameters θ, γ. This choice allows for a leverage effect, is at the same time of a simple structure and ensures $E(f(\epsilon_1)) = 0$. The logarithmic volatility models put forth in Kawakatsu (2006) can all be transformed into our above model using appropriate choices of f. However, all of them lead to functional forms involving only the individual components $\epsilon_{i,t}$, $i = 1, \dots, d$, of the innovation sequence ϵ and their absolute values $|\epsilon_{i,t}|$ in a linear manner. In particular, crossproducts of the form $\epsilon_{i,t}\epsilon_{j,t}$ do not enter the specification of f. Dependence on these crossproducts seems, however, desirable, especially when comparing things to multivariate GARCH specifications. We thus suggest two new possible choices for f now. The first possible choice

$$f(\eta) = \Theta\eta + \Gamma\left(\text{vech}\left((\eta\eta^*)^{1/2}\right) - E\left(\text{vech}\left((\epsilon_1\epsilon_1^*)^{1/2}\right)\right)\right) \tag{3.2.1}$$

with $\Theta \in M_{m,d}(\mathbb{R})$ and $\Gamma \in M_m(\mathbb{R})$ is a straightforward multivariate extension of the standard choice. Note that $(\eta\eta^*)^{1/2}$ can be interpreted as an extension of the absolute value to a multidimensional setting and that $\left((\eta\eta^*)^{1/2}\right)_{ij} = \eta_i\eta_j/\|\eta\|_2$ with $\|\cdot\|_2$ denoting the Euclidean norm on $\mathbb{R}^d$. The second possibility we suggest is to use a generalized standard choice component-wise, viz.

$$\begin{aligned}
f(\eta) &= \text{vech}(g(\eta) - E(g(\epsilon_1))) \text{ with} \\
g &: \mathbb{R}^d \to \mathbb{S}_d, (\eta_1, \eta_2, \cdots, \eta_d) \mapsto (f_{ij}(\eta_i, \eta_j))_{1\le i,j\le d} \\
f_{ii}(\eta_i, \eta_i) &:= \theta_{i,i}\eta_i + \gamma_{i,i}|\eta_i| \text{ for } i = 1, 2, \dots, d \\
f_{ij}(\eta_i, \eta_j) &:= \theta_{i,j}\frac{\eta_i\eta_j}{\sqrt{\eta_i\eta_j}} + \gamma_{i,j}\sqrt{|\eta_i\eta_j|} \text{ for } i = 1, 2, \dots, d,\ j = 1, 2, \dots, i-1 \\
f_{ij}(\eta_i, \eta_i) &:= f_{ji}(\eta_i, \eta_j) \text{ for } i = 1, 2, \dots, d,\ j = i+1, i+2, \dots, d
\end{aligned} \tag{3.2.2}$$

where $\theta_{i,j}, \gamma_{i,j}$ with $i = 1, 2, \dots, d$, $j = 1, 2 \dots, i$ are real parameters.

The following proposition shows that f as specified in (3.2.1) or (3.2.2) satisfies the required conditions for EGARCH processes.

Proposition 3.2.3. *Let ϵ_1 be an $\mathbb{R}^d$-valued random variable with $\epsilon_1 \in L^2$ and $f : \mathbb{R}^d \to \mathbb{R}^m$ as specified in Equation (3.2.1) or (3.2.2). Then f is well-defined and $f(\epsilon_1) \in L^2$.*

Proof. If f is specified by (3.2.2) this follows from an element-wise application of the Cauchy-Schwartz inequality.

If f is given by (3.2.1) we are free to choose any norm for the proof. Thus we work in the following with the Euclidean norm $\|\cdot\|_2$ on $\mathbb{R}^d$, resp. $\mathbb{R}^m$, and the induced operator norm on matrix spaces. Elementary calculations give $\left\|(\epsilon_1\epsilon_1{}^*)^{1/2}\right\|_2 = \|\epsilon_1\|_2$, which implies the well-definedness. Likewise, we use the operator norm $\|\cdot\|$ induced by these choices for the vech operator. We have

$$\|f(\epsilon_1)\|_2 \le \|\Theta\|_2\|\epsilon_1\|_2 + \|\Gamma\|_2\left(\|\text{vech}\|\left\|(\epsilon_1\epsilon_1{}^*)^{1/2}\right\|_2 + \left\|E\left(\text{vech}\left((\epsilon_1\epsilon_1{}^*)^{1/2}\right)\right)\right\|_2\right).$$

Using Jensen's inequality one obtains $\left\|E\left(\text{vech}\left((\epsilon_1\epsilon_1{}^*)^{1/2}\right)\right)\right\|_2 \le \|\text{vech}\|E(\|\epsilon_1\|_2)$. Thus

$$\|f(\epsilon_1)\|_2 \le (\|\Theta\|_2 + \|\Gamma\|_2\|\text{vech}\|)\,\|\epsilon_1\|_2 + \|\Gamma\|_2\|\text{vech}\|E(\|\epsilon_1\|_2).$$

Since $\epsilon_1 \in L^2$ this immediately implies $f(\epsilon_1) \in L^2$. □

3.2.2. Multivariate Lévy and Lévy-driven CARMA processes

Before defining multivariate ECOGARCH processes in the next section, we briefly review Lévy processes and multivariate CARMA processes as introduced in Chapter 2.

3.2.2.1. Basic facts on multivariate Lévy processes

We state some elementary properties of multivariate Lévy processes that will be needed. For a more general treatment and proofs we refer to Sato (1999), Applebaum (2004) or Protter (2004).

We consider a Lévy process $L = (L_t)_{t\in\mathbb{R}^+}$ (where $L_0 = 0$ a.s.) in $\mathbb{R}^d$ determined by its characteristic function in the Lévy-Khintchine form $E\left[e^{i\langle u, L_t\rangle}\right] = \exp\{t\psi_L(u)\}, \quad t \geq 0$, where

$$\psi_L(u) = i\langle\gamma_L, u\rangle - \frac{1}{2}\langle u, C_L u\rangle + \int_{\mathbb{R}^d} \left(e^{i\langle u,x\rangle} - 1 - i\langle u, x\rangle 1_{[0,1]}(\|x\|)\right) \nu_L(dx), \quad u \in \mathbb{R}^d, \tag{3.2.3}$$

where the drift $\gamma_L \in \mathbb{R}^d$, $C_L \in \mathbb{S}_d^+$ and ν_L is a measure on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ that satisfies $\nu_L(\{0\}) = 0$ and $\int_{\mathbb{R}^d} (\|x\|^2 \wedge 1)\, \nu_L(dx) < \infty$. The measure ν_L is referred to as the Lévy measure of L. Here $\|\cdot\|$ may be any fixed norm (not only the Euclidean one). Different norms simply correspond to different truncation functions and thus for a given Lévy process γ_L changes when the norm is changed. Implicitly we presume throughout this chapter that given a Lévy process γ_L is set to the value such that (3.2.3) holds with the currently employed norm.

It is a well-known fact that to every càdlàg Lévy process L on $\mathbb{R}^d$ one can associate a random measure N_L on $\mathbb{R}^+ \times \mathbb{R}^d \setminus \{0\}$ describing the jumps of L (see e.g. Jacod and Shiryaev (2003, Section II.1)). For any measurable set $B \subset \mathbb{R}^+ \times \mathbb{R}^d \setminus \{0\}$,

$$N_L(B) = \sharp\{s \geq 0 : \ (s, L_s - L_{s-}) \in B\}.$$

The jump measure N_L is an extended Poisson random measure (as defined in Jacod and Shiryaev (2003, Definition II.1.20)) on $\mathbb{R}^+ \times \mathbb{R}^d \setminus \{0\}$ with intensity measure $n_L(ds, dx) = ds\, \nu_L(dx)$. By the Lévy-Itô decomposition we can rewrite L almost surely as

$$L_t = \gamma_L t + B_t + \int_{\|x\|\geq 1, s\in[0,t]} x\, N_L(ds, dx) + \lim_{\varepsilon\downarrow 0} \int_{\varepsilon\leq\|x\|\leq 1,\, s\in[0,t]} x\tilde{N}_L(ds, dx), \quad t \geq 0. \tag{3.2.4}$$

Here B is a d-dimensional Brownian motion with covariance matrix C_L, $\tilde{N}(ds, dx) = N(ds, dx) - ds\nu_L(dx)$ is the compensated jump measure, the terms in (3.2.4) are independent and the convergence in the last term is a.s. and locally uniform in $t \geq 0$.

In the sequel we will sometimes work with a two-sided Lévy process $L = (L_t)_{t\in\mathbb{R}}$, constructed by taking two independent copies $(L_{1,t})_{t\in\mathbb{R}^+}$, $(L_{2,t})_{t\in\mathbb{R}^+}$ of a one-sided Lévy process and setting

$$L_t = \begin{cases} L_{1,t} & \text{if} \quad t \geq 0 \\ -L_{2,-t-} & \text{if} \quad t < 0. \end{cases}$$

Assuming that ν_L satisfies additionally

$$\int_{\|x\|>1} \|x\|^2 \, \nu_L(dx) < \infty,$$

L has finite mean and covariance matrix Σ_L given by

$$\Sigma_L = C_L + \int_{\mathbb{R}^d} xx^* \, \nu_L(dx).$$

For the stochastic integration theory (with respect to Lévy processes and/or random measures) we refer to the brief overview in Section 2.2.2 in Chapter 2 or any of the standard texts, e.g. Jacod and Shiryaev (2003), Protter (2004) or Applebaum (2004).

3.2.2.2. Multivariate Lévy-driven CARMA processes

As the name "continuous time ARMA" (CARMA) already suggests, these processes are the continuous time analogue of the well-known ARMA processes. A d-dimensional CARMA(q,p) process Y with $q,p \in \mathbb{N}_0$ can be viewed as the stationary solution to the formal differential equation:

$$Q(\mathcal{D})Y_t = P(\mathcal{D})\mathcal{D}L_t$$

where $L = (L_t)_{t\in\mathbb{R}}$ is a d-dimensional Lévy process and $\mathcal{D}$ the differential operator with respect to t.

$$\begin{aligned} Q(z) &= z^q + A_1 z^{q-1} + A_2 z^{q-2} + \ldots + A_q \\ P(z) &= B_0 z^p + B_1 z^{p-1} + \ldots + B_p \end{aligned}$$

with $B_0, \ldots, B_p, A_1, \ldots, A_q \in M_d(\mathbb{R})$, $A_q \in GL_d(\mathbb{R})$ and $B_0 \neq 0$ are referred to as the autoregressive and moving average polynomial, respectively. In order to be able to define CARMA processes properly one needs $q > p$ and that the zeros of $\det(Q(z))$ have all strictly negative real parts. Then the CARMA(q,p) process Y is defined as the unique stationary solution of

$$Y_t = (I_d, 0, \ldots, 0) X_t \tag{3.2.5}$$

$$dX_t = \begin{pmatrix} 0 & I_d & 0 & \cdots & 0 \\ 0 & 0 & I_d & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & I_d \\ -A_q & -A_{q-1} & -A_{q-2} & \cdots & -A_1 \end{pmatrix} X_t dt + \beta dL_t, \tag{3.2.6}$$

where $\beta = (\beta_1^T, \beta_2^T, \ldots, \beta_q^T)^T$ is a $qd \times d$ matrix with elements $\beta_{q-j} = -\sum_{i=1}^{q-j-1} A_i \beta_{q-j-i} + B_{p-j}$ (setting $B_i = 0$ for $i < 0$). The process X is usually called state space representation.

In the univariate case $d = 1$ the representation by (3.2.5), (3.2.6) can be replaced by

$$\begin{aligned} Y_t &= (B_p, B_{p-1}, \ldots, B_{p-q+1}) \tilde{X}_t \\ d\tilde{X}_t &= \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ -A_q & -A_{q-1} & -A_{q-2} & \cdots & -A_1 \end{pmatrix} \tilde{X}_t dt + \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix} dL_t \,, \end{aligned}$$

setting $B_i = 0$ for $i < 0$. This representation was used in the definition of the univariate ECOGARCH processes in Haug and Czado (2007).

For CARMA$(q, 0)$ processes the equivalence of the two representation for univariate CARMA processes is clear. In general, the equivalence in L^2 (i.e. the driving Lévy process L has to have a finite second moment) has been shown in Theorem 2.3.22 in Chapter 2 using Fourier methods. For CARMA(2,1) processes a proof of the equivalence not needing a finite variance is presented in Lemma 6.6.1 of Chapter 6.

Later on we need the following result on the existence of exponential moments. By $\mathrm{Ei} : \mathbb{R}\backslash\{0\} \to \mathbb{R}$ we denote the exponential integral, i.e.

$$\mathrm{Ei}(x) = \int_{-\infty}^{x} \frac{e^t}{t} dt = \gamma + \ln|x| + \sum_{k=1}^{\infty} \frac{x^k}{k \cdot k!} \text{ for all } x \in \mathbb{R}\backslash\{0\} \tag{3.2.7}$$

taking the Cauchy principal value of the integral for $x > 0$ and γ being the Euler constant.

Proposition 3.2.4. *Let Y be a stationary d-dimensional CARMA(q, p) process satisfying $\sigma(A) \subset (-\infty, 0) + i\mathbb{R}$, $\|\cdot\|$ a norm on $\mathbb{R}^d$ and its induced operator norm, $\alpha \in \mathbb{R}^{++}$ and $C, b \in \mathbb{R}^{++}$ such that $\|(I_d, 0, \ldots, 0)e^{As}\beta\| \leq Ce^{-bs}$ for all $s \in \mathbb{R}^+$. If*

$$\int_{\|x\|\geq 1} \mathrm{Ei}(\alpha C\|x\|)\nu_L(dx) < \infty \tag{3.2.8}$$

then $E\left(e^{\alpha\|Y_0\|}\right) < \infty$.

For all $0 < b < -\max(\Re(\sigma(A)))$ there exists a $C \in \mathbb{R}^{++}$ such that $\|(I_d, 0, \ldots, 0)e^{As}\beta\| \leq Ce^{-bs}$ holds for all $s \in \mathbb{R}^+$. If A is diagonalizable this holds also for $b = -\max(\Re(\sigma(A)))$. Furthermore, (3.2.8) *is implied by*

$$\int_{\|x\|\geq 1} e^{\alpha C\|x\|}\nu_L(dx) < \infty. \tag{3.2.9}$$

Proof. It is elementary to see (using e.g. the Jordan decomposition of A) that $\sigma(A) \subset (-\infty, 0) + i\mathbb{R}$ implies for all $0 < b < -\max(\Re(\sigma(A)))$ that there exists a $C \in \mathbb{R}^{++}$ such that $\|(I_d, 0, \ldots, 0)e^{As}\beta\| \leq Ce^{-bs}$ holds for all $s \in \mathbb{R}^+$. If A is diagonalizable this also shows that one can take $b = -\max(\Re(\sigma(A)))$.

From Proposition 2.3.27 of Chapter 2 we know that the stationary distribution of Y is infinitely divisible. Denote by $(\gamma_Y, \sigma_Y, \nu_Y)$ the characteristic triplet of the stationary distribution of Y. Sato (1999, Theorem 25.3) implies that for all $\alpha > 0$ we have $E(e^{\alpha\|Y_1\|}) < \infty$ if and only if

$$\int_{\|x\|\geq 1} e^{\alpha\|x\|}\nu_Y(dx) < \infty.$$

Proposition 2.3.27 of Chapter 2 implies

$$\begin{aligned}\int_{\|x\|\geq 1} e^{\alpha\|x\|}\nu_Y(dx) &= \int_0^\infty \int_{\mathbb{R}^d} e^{\alpha\|(I_d,0,\ldots,0)e^{As}\beta x\|} 1_{[1,\infty)}(\|(I_d, 0, \ldots, 0)e^{As}\beta x\|)\nu_L(dx)ds \\ &\leq \int_0^\infty \int_{\mathbb{R}^d} e^{\alpha Ce^{-bs}\|x\|} 1_{[1,\infty)}(\alpha Ce^{-bs}\|x\|)\nu_L(dx)ds\end{aligned}$$

$$
\begin{aligned}
&= \int_{\|x\|\geq 1/(\alpha C)} \int_0^{\ln(\alpha C\|x\|)/b} e^{\alpha C e^{-bs}\|x\|} ds \nu_L(dx) \\
&= \frac{1}{b} \int_{\|x\|\geq 1/(\alpha C)} \int_1^{\alpha C\|x\|} \frac{e^z}{z} dz \nu_L(dx) \\
&= \frac{1}{b} \int_{\|x\|\geq 1/(\alpha C)} \left(\mathrm{Ei}(\alpha C\|x\|) - \mathrm{Ei}(1)\right) \nu_L(dx).
\end{aligned}
$$

Since ν_L is a Lévy measure, $\int_{\|x\|\geq 1/(\alpha C)} \mathrm{Ei}(1)\nu_L(dx) < \infty$ for all $\alpha > 0$ and the integral $\int_{\|x\|\geq 1/(\alpha C)} \mathrm{Ei}(\alpha C\|x\|)\nu_L(dx)$ is finite if and only if $\int_{\|x\|\geq 1} \mathrm{Ei}(\alpha C\|x\|)\nu_L(dx) < \infty$. Therefore (3.2.8) is sufficient for $E(e^{\alpha\|Y_1\|}) < \infty$.

From (3.2.7) it follows that for any $c > 0$ there exists a $K(c) \in \mathbb{R}^{++}$ such that $|\mathrm{Ei}(x)| \leq K(c)e^x$ for all $x \geq c$. This shows that (3.2.9) implies (3.2.8). □

If $(q, p) = (1, 0)$, A_1 is diagonal or unitarily diagonalizable, $\|\cdot\|$ is the Euclidean norm and $B_0 = I_d$, then one can take $b = -\max(\Re(\sigma(A)))$ and $C = 1$. So a d-dimensional CARMA(1,0) process (OU process) with unitarily diagonalizable A has at least as many exponential moments as the driving Lévy process.

3.3. Multivariate exponential COGARCH

3.3.1. Definition and stationarity

Now we define the *exponential continuous time* GARCH(p, q) *process* by specifying the vech-transformed log-volatility process as a CARMA$(q, p-1)$ process.

Definition 3.3.1. *Let $L = (L_t)_{t\geq 0}$ be a d-dimensional zero-mean Lévy process with Lévy measure ν_L such that $\int_{\|x\|\geq 1} \|x\|^2 \nu_L(dx) < \infty$ and associated jump measure N_L. Furthermore, let $h : \mathbb{R}^d \to \mathbb{R}^m$ with $m = \frac{d(d+1)}{2}$ be a measurable function satisfying*

$$\int_{\mathbb{R}^d} \|h(x)\|^2 \nu_L(dx) < \infty, \tag{3.3.1}$$

$p, q \in \mathbb{N}$ with $q \geq p$ and $A_1, \ldots, A_q, B_0, \ldots, B_{p-1} \in M_m(\mathbb{R})$ with $A_q \in GL_d(\mathbb{R})$ and $B_0 \neq 0$ such that all zeros of the determinant $\det(Q(z))$ of the autoregressive polynomial $Q(z) := z^q + A_1 z^{q-1} + A_2 z^{q-2} + \ldots + A_q$, $z \in \mathbb{C}$, have strictly negative real part.

Then we define the d-dimensional exponential COGARCH(p, q) process G, *shortly* ECOGARCH(p, q), *as the stochastic process satisfying,*

$$dG_t := \exp((\mu + H_{t-})/2) dL_t, \quad t > 0, \quad G_0 = 0, \tag{3.3.2}$$

where $\mu \in \mathbb{S}_d$ and the vectorized log-volatility process $H = (H_t)_{t\geq 0}$ *is a process in $\mathbb{S}_d$ with mean zero and vectorial state space representation*

$$
\begin{aligned}
\mathrm{vech}(H_t) &:= (I_m, 0, \ldots, 0) X_t, \quad t \geq 0, &(3.3.3)\\
dX_t &= A X_{t-} dt + \beta dM_t, \quad t > 0, &(3.3.4)
\end{aligned}
$$

with the initial value $X_0 \in \mathbb{R}_{qm}$ being independent of the driving Lévy process L and

$$M_t := \int_0^t \int_{\mathbb{R}^d \setminus \{0\}} h(x) \tilde{N}_L(ds, dx), \qquad t \geq 0, \tag{3.3.5}$$

being a zero-mean Lévy process. The matrices $A \in M_{qm}(\mathbb{R})$ *and* $\beta \in M_{qm,m}(\mathbb{R})$ *are defined by*

$$A = \begin{pmatrix} 0 & I_m & 0 & \cdots & 0 \\ 0 & 0 & I_m & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & I_m \\ -A_q & -A_{q-1} & -A_{q-2} & \cdots & -A_1 \end{pmatrix}, \ \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_{q-1} \\ \beta_q \end{pmatrix}$$

with coefficients $\beta_{q-j} = -\sum_{i=1}^{q-j-1} A_i \beta_{q-j-i} + B_{p-1-j}$ *for* $j = 0, 1, \ldots, q-1$ *(setting* $B_i = 0$ *for* $i < 0$*). If* $p = q = 1$*, we have* $A = -A_1$ *and* $\beta = B_0$*.*

Then returns over a time interval of length $r > 0$ ending at time t are described by the increments of G

$$G_t^{(r)} := G_t - G_{t-r} = \int_{(t-r,t]} \exp(\mu + H_{s-}/2)\, dL_s\,, \qquad t \geq r > 0\,. \tag{3.3.6}$$

Thus this gives us the possibility to model ultra high frequency data, which consists of returns over varying time intervals. On the other hand an equidistant sequence of such non-overlapping returns of length r is given by $(G_{nr}^{(r)})_{n\in\mathbb{N}}$. Such a sequence then corresponds to a discrete time multivariate EGARCH process Y.

Remark 3.3.2. *(a) The condition (3.3.1) ensures that the integral defining the Lévy process* M *is indeed well-defined and that* M *has a finite variance.*

(b) The condition that $\det(Q(z))$ *has only zeros with strictly negative real part is equivalent to demanding that all eigenvalues of* A *have strictly negative real part.*

(c) In order for the multivariate CARMA process for the log-volatility to be definable it would actually suffice to demand that $\det(Q(z))$ *has no zero with vanishing real part (see Remark 2.3.19 in Chapter 2). But the resulting multivariate CARMA process would no longer be adapted to the natural filtration of the driving Lévy process* L*. In view of equation (3.3.2) this adaptedness is, however, clearly desired. Thus we refrain from discussing any possible extension of the above model in this direction.*

(d) After extending the Lévy process $(M_t)_{t\in\mathbb{R}^+}$ *to one defined on the whole real line the unique stationary version of* H *can be written as*

$$\operatorname{vech}(H_t) = \int_{-\infty}^{t} (I_m, 0, \ldots, 0) e^{A(t-s)} \beta dM_s\,.$$

(e) In general one can define a continuous time analogue of the infinite moving average discrete time EGARCH process of Definition 3.2.1 by

$$\operatorname{vech}(H_t) = \int_{-\infty}^{t} f(t-s) dM_s$$

with $f : \mathbb{R}^+ \to M_m(\mathbb{R})$ *being an appropriate deterministic function.*

So far we have considered a general transformation h of the jumps of the driving Lévy process L. Concrete specifications should be made in such a way that the model exhibits similar properties, e.g. a leverage effect, as in the discrete time case. The choice

$$h(\eta) = \Theta\eta + \Gamma \operatorname{vech}\left((\eta\eta^*)^{1/2}\right), \tag{3.3.7}$$

with $\Theta \in M_{m,d}(\mathbb{R})$ and $\Gamma \in M_m(\mathbb{R})$, being the continuous time analogue of (3.2.1) clearly is always a valid choice, as an inspection of the proof of Proposition 3.2.3 shows. Again it is notable that this extends the standard choice from the univariate literature.

A choice analogous to (3.2.2) is

$$h(\eta) = \operatorname{vech}(g(\eta)) \tag{3.3.8}$$

with g as in (3.2.2). That $\int_{\mathbb{R}^d} \|h(x)\|^2 \nu_L(dx)$ is finite is elementary to see.

Both specifications (3.3.7) and (3.3.8) obviously allow for asymmetric responses to positive and negative shocks in the logarithmic (co)variance components. For $p = q$ and (3.3.8) the instantaneous leverage effect for the variance components can be shown as in Haug and Czado (2007, Proposition 5.5).

Proposition 3.3.3. *Let H and G be as in Definition 3.3.1 with h satisfying (3.3.1). If the eigenvalues of A all have negative real parts and X_0 has the same distribution as $\int_0^\infty e^{Au}\beta dM_u$, then X, H and $\exp((\mu + H)/2)$ are strictly stationary.*

Proof. The strict stationarity of $X, \operatorname{vech}(H)$ follows from Theorem 2.3.12 in Chapter 2, because X is the state space representation of a CARMA$(q, p-1)$ process. Since strict stationarity is invariant under continuous transformations, $\exp((\mu + H)/2)$ also has this property. □

From (3.3.6) it follows directly that the increments $(G_{nr}^{(r)})_{n\in\mathbb{N}}$ of G are stationary if the volatility $\exp((\mu + H)/2)$ is stationary, since the increments of L are stationary and independent by definition.

Corollary 3.3.4. *If $\exp((\mu + H)/2)$ is strictly stationary, then G has strictly stationary increments.*

Remark 3.3.5. *If $q \geq p + 1$ the log-volatility process is continuous and $(q - p - 1)$ times differentiable, which follows from the state space representation of $\operatorname{vech}(H)$ (cf. Proposition 2.3.32 in Chapter 2). In particular, the volatility will only contain jumps for $p = q$.*

3.3.2. Mixing and second order properties

Mixing properties (see Doukhan (1994) for a comprehensive treatment) are useful for a number of applications. In particular for asymptotic statistics, since consistency results and central limit theorems exist for mixing processes. Thus we will derive mixing properties of the strictly stationary volatility process and the return process. First we recall the definition of strong mixing, which is also called α-mixing for a process with continuous time parameter.

Definition 3.3.6 (Davydov (1973))**.** *For a process $Y = (Y_s)_{s\geq 0}$ define the σ-algebras $\mathcal{F}_{[0,u]}^Y := \boldsymbol{\sigma}((Y_s)_{s\in[0,u]})$ and $\mathcal{F}_{[u+t,\infty)}^Y := \boldsymbol{\sigma}((Y_s)_{s\geq u+t})$ for all $u \geq 0$. Then Y is called strongly or α-mixing, if*

$$\begin{aligned}\alpha(t) &= \sup_{u\geq 0} \alpha(\mathcal{F}_{[0,u]}^Y, \mathcal{F}_{[u+t,\infty)}^Y) \\ &:= \sup_{u\geq 0} \sup\{|P(A\cap B) - P(A)P(B)| : A \in \mathcal{F}_{[0,u]}^Y, B \in \mathcal{F}_{[u+t,\infty)}^Y\} \to 0,\end{aligned}$$

as $t \to \infty$.

Above we denote by $\boldsymbol{\sigma}(\cdot)$ the generated completed σ-algebra. The strong mixing property with exponential rate of the log-volatility and volatility process is the subject of the next proposition. Here strong mixing with exponential rate (exponential α-mixing) means that $\alpha(t)$ decays to zero exponentially fast for $t \to \infty$.

Proposition 3.3.7. *Let* $\mathrm{vech}(H)$ *be defined by (3.3.3) and (3.3.4). Assume that the eigenvalues of* A *all have negative real parts and* X_0 *has the same distribution as* $\int_0^\infty e^{Au}\beta dM_u$*, hence* H *and* $\exp(\mu + H/2)$ *are strictly stationary. Then there exist constants* $K > 0$ *and* $a > 0$ *such that*

$$\alpha_H(t) \leq K \cdot e^{-at} \quad \textit{and} \quad \alpha_{\exp(\mu+H/2)}(t) \leq K \cdot e^{-at}, \qquad \textit{as} \quad t \to \infty,$$

where $\alpha_H(t)$ *and* $\alpha_{\exp(H/2)}(t)$ *are the* α*-mixing coefficients of the log-volatility and volatility process, respectively.*

Proof. Since $\mathrm{vech}(H)$ is a CARMA$(q, p-1)$ process the result follows from Proposition 2.3.34 of Chapter 2 and the fact that α-mixing is preserved under continuous transformations. □

Theorem 3.3.8. *Assume that* L *is a Lévy process with* $\mathbb{E}(\|L_1\|^2) < \infty$*. Let the log-volatility process* H *be strictly stationary and strongly mixing.*

Then for all $r \in \mathbb{R}^{++}$ *the discrete time process* $(G_{nr}^{(r)})_{n\in\mathbb{N}}$*,*

$$G_{nr}^{(r)} = G_{nr} - G_{(n-1)r} = \int_{((n-1)r,nr]} \exp(\mu + H_{s-}/2)\, dL_s\,, n \in \mathbb{N}\,,$$

is strongly mixing with exponential rate and ergodic.

Proof. We define the process $\aleph$ with values in $(\mathbb{S}_m^+)^q$ by

$$\aleph_t := (\exp(\mu + \mathrm{vech}^{-1}(X_t^1)/2), \ldots, \exp(\mu + \mathrm{vech}^{-1}(X_t^q)/2))\,,$$

where $X_t = (X_t^1, \ldots, X_t^q)^*$. Then $\aleph$ is a Markov process and we can define the qm-dimensional process

$$\mathbf{G}_{nr}^{(r)} := \int_{((n-1)r,nr]} \aleph_{s-}\, dL_s\,, n \in \mathbb{N}\,.$$

Now the rest of the proof works along the lines of the proof of Haug and Czado (2007, Theorem 3.10). □

Corollary 3.3.9. *Let* $(t_n)_{n\in\mathbb{N}_0}$ *be a strictly increasing sequence of observation time points with* $\lim_{n\to\infty} t_n = \infty$ *and* $t_n = k_n c$ *for all* $n \in \mathbb{N}_0$*, where* $k_n \in \mathbb{N}_0$ *and* $c > 0$*. Then the discrete time process* $(G_{t_n}^{(\Delta_n)})_{n\in\mathbb{N}}$*,*

$$G_{t_n}^{(\Delta_n)} := G_{t_n} - G_{t_{n-1}}\,,$$

with $\Delta_n = t_n - t_{n-1}$*, is strongly mixing with exponential rate.*

Proof. Simply expand the grid of observation times to an equidistant one with step size c. Then clearly

$$\mathcal{F}^{G^{(\Delta_\cdot)}}_{1,2,\dots,l} \subset \mathcal{F}^{G^{(c)}}_{1,2,\dots,t_l/c} \quad \text{and} \quad \mathcal{F}^{G^{(\Delta_\cdot)}}_{k+l,k+l+1,\dots} \subset \mathcal{F}^{G^{(c)}}_{t_{k+l}/c,t_{k+l+1}/c,\dots},$$

where $\mathcal{F}^{G^{(\Delta_\cdot)}}_{1,2,\dots,l}$ is the σ-algebra generated from the random vectors $G^{(\Delta_1)}_{t_1}, \dots, G^{(\Delta_l)}_{t_l}$ and the other σ-algebras are defined analogously. An application of Theorem 3.3.8 then provides the result. □

Now we derive the second order moment structure of the *return process*

$$G^{(r)}_t := G_t - G_{t-r} = \int_{(t-r,t]} \exp(\mu + H_{s-}/2)\, dL_s\,, \qquad t \geq r > 0\,,$$

considering only the case of a strictly stationary volatility process.

Proposition 3.3.10. *Let L be a Lévy process with $\mathbb{E}(L_1) = 0$ and $\mathbb{E}(\|L_1\|^2) < \infty$. Assume that the log-volatility process H is strictly stationary and $\mathbb{E}(\|\exp(\mu + H_t/2)\|) < \infty$. Then $\mathbb{E}(\|G_t\|^2) < \infty$ for all $t \geq 0$, and for every $t, h \geq r > 0$ it holds that*

$$\begin{aligned} \mathbb{E}G^{(r)}_t &= 0 \\ \mathbb{E}(G^{(r)}_t (G^{(r)}_t)^*) &= \int_0^r \mathbb{E}(\exp(\mu + H_{s-}/2)\mathbb{E}(L_1 L_1^*) \exp(\mu + H_{s-}/2))ds \\ \operatorname{cov}(G^{(r)}_t, G^{(r)}_{t+h}) &= 0. \end{aligned}$$

The results follow analogously to the univariate case in Haug and Czado (2007, Proposition 5.1). Note that the second order moment structure of H is clear from Chapter 2, whereas for the volatility $\exp(\mu + H)$ and the "squared returns" $G^{(r)}_t (G^{(r)}_t)^*$ the formulae obtained in the univariate case are already not explicit. Thus we refrain from stating them in our multivariate setting.

Regarding the finiteness of "exponential moments" of H needed above we have the following result.

Proposition 3.3.11. *(i) Let $\|\cdot\|_*$ be an algebra norm on $\mathbb{S}_d$ and the ECOGARCH log-volatility process H be strictly stationary. Then*

$$E(e^{\alpha_1 \alpha_2 \|H_1\|_*}) < \infty \text{ with } \alpha_1, \alpha_2 \in \mathbb{R}^{++} \tag{3.3.9}$$

implies

$$E\left(\|\exp(\alpha_1(\mu + H_1))\|_*^{\alpha_2}\right) < \infty. \tag{3.3.10}$$

(ii) Let moreover $C \in \mathbb{R}^{++}$ be such that

$$\sup_{x \in \mathbb{R}^m, \|\operatorname{vech}^{-1}(x)\|_* = 1} \left\{ \left\| \operatorname{vech}^{-1}\left((I_m, 0 \dots, 0) e^{As} \beta x\right) \right\|_* \right\} \leq C e^{-bs}$$

for all $s \in \mathbb{R}^+$ and some $b \in \mathbb{R}^{++}$. Then (3.3.9) is in turn implied by

$$\int_{x \in \mathbb{R}^d, \|\operatorname{vech}^{-1}(h(x))\|_* \geq 1} \operatorname{Ei}\left(\alpha_1 \alpha_2 C \|\operatorname{vech}^{-1}(h(x))\|_*\right) \nu_L(dx) < \infty$$

or

$$\int_{x \in \mathbb{R}^d, \|\operatorname{vech}^{-1}(h(x))\|_* \geq 1} \exp\left(\alpha_1 \alpha_2 C \|\operatorname{vech}^{-1}(h(x))\|_*\right) \nu_L(dx) < \infty.$$

Proof. (i) Since $\|\cdot\|$ is an algebra norm, $\|\exp(\alpha_1(\mu + H_1))\|_*^{\alpha_2} \leq e^{\alpha_1\alpha_2\|\mu\|_*} e^{\alpha_1\alpha_2\|H_1\|_*}$. This immediately shows (i).

(ii) The second part follows from Proposition 3.2.4 using the norm $\|\cdot\| = \|\text{vech}^{-1}(\cdot)\|_*$ on $\mathbb{R}^m$ and the definition of M implying $\nu_M(dx) = \nu_L(h^{-1}(dx))$, because $\text{vech}(H)$ is an m-dimensional stationary CARMA process. □

3.4. Approximation of ECOGARCH(1,1) processes by EGARCH(1,1) processes

In this section we show that for any multivariate ECOGARCH(1,1) process where the driving Lévy process L has a vanishing Brownian part there exists a sequence of piecewise constant processes determined by discrete time multivariate EGARCH(1,1) processes which converges in probability in the Skorokhod topology (see Billingsley (1999) or Jacod and Shiryaev (2003, Chapter IV) for a comprehensive introduction) to the ECOGARCH(1,1) process. This result is also new in the univariate case and should be especially useful for statistical purposes (cp. Maller, Müller and Szimayer (2006)).

For a complete and separable normed space $(E, \|\cdot\|_E)$ denote by D_E the set of all functions $f : \mathbb{R}^+ \to E$ that are right continuous and have left limits. Let further Λ be the set of all time change functions, i.e. all continuous and strictly increasing functions $\lambda : \mathbb{R}^+ \to \mathbb{R}^+$ that satisfy $\lambda(0) = 0$ and $\lim_{t\to\infty} \lambda(t) = \infty$. Denote by $e_{\mathbb{R}^+}$ the function $\mathbb{R}^+ \to \mathbb{R}^+, t \mapsto t$ and by $\|x\|_{E,[a,b)} := \sup_{t\in[a,b)} \|x(t)\|_E$ for $x \in D_E$. A sequence $(x_n)_{n\in\mathbb{N}}$ in D_E converges to $x \in D_E$ in the Skorokhod topology if there exists a sequence $(\lambda_n)_{n\in\mathbb{N}}$ in Λ such that $(\lambda_n)_{n\in\mathbb{N}}$ converges uniformly to $e_{\mathbb{R}^+}$, i.e. $\lim_{n\to\infty} \|\lambda_n - e_{\mathbb{R}^+}\|_{\mathbb{R}^+,[0,\infty)} = 0$, and $(x_n \circ \lambda_n)_{n\in\mathbb{N}}$ converges uniformly on compacts to x, i.e. $\lim_{n\to\infty} \|x_n \circ \lambda_n - x\|_{E,[0,N]} = 0$ for all $N \in \mathbb{N}$. In particular uniform convergence on compacts of $(x_n)_{n\in\mathbb{N}}$ implies convergence in the Skorokhod topology. A separable (yet not complete) metric inducing the Skorokhod topology is given by

$$d_E(x, y) = \inf_{\lambda\in\Lambda} d_{\lambda,E}(x, y) \text{ for } x, y \in D_E \tag{3.4.1}$$

where

$$d_{\lambda,E}(x, y) = \|\lambda - e_{\mathbb{R}^+}\|_{\mathbb{R}^+,[0,\infty)} + \sum_{n=1}^{\infty} \frac{1}{2^n} \min\left(1, \sup_{t\leq n} \|x(\min(n, \lambda(t))) - y(\min(n, t))\|_E\right),$$

see e.g. Kurtz and Protter (1996). For equivalent norms on E the above definition leads to equivalent metrics on D_E. As we are only considering finite dimensional normed spaces E in the following it is irrelevant which particular norm on E we use, we may even switch between different norms to use the most convenient one. Convergence of a sequence $(X^{(n)})_{n\in\mathbb{N}}$ of E-valued càdlàg random processes, i.e. random variables in D_E, in probability in the Skorokhod topology to a càdlàg random process X means for us in the following that $\text{plim}_{n\to\infty} d_E(X^{(n)}, X) = 0$ with plim denoting the limit in probability. Again we note that uniform convergence on compacts in probability (ucp convergence, cf. Protter (2004, Chapter II.4)) obviously implies convergence in the Skorokhod topology in probability and that in a metric space convergence in probability is metrizable (follows e.g. by replacing the absolute value with the metric in Loéve (1977, p. 160 (3), p.175 9.)).

In the following we presume that each of the appearing spaces $\mathbb{S}_n, \mathbb{R}^n, M_n(\mathbb{R})$ with $n \in \mathbb{N}$ and products of such spaces are equipped with some norm. For the in the following necessary

theory from stochastic analysis we refer to Protter (2004). Now we can state our main theorem of this section where again $m := d(d+1)/2$.

Theorem 3.4.1. *Let (G, X) in $\mathbb{R}^d \times \mathbb{R}^m$ be a d-dimensional ECOGARCH(1,1) process G and its associated vectorized log-volatility process $X = \mathrm{vech}(H)$ with initial value (G_0, X_0). Assume the driving Lévy process L has no Brownian part and let $(t_i^{(n)})_{i\in\mathbb{N}_0}$ for each $n \in \mathbb{N}$ be a strictly increasing sequence in $\mathbb{R}^+$ with $t_0^{(n)} = 0$ and $\lim_{i\to\infty} t_i^{(n)} = \infty$. Defining $\delta^{(n)} = \sup_{i\in\mathbb{N}} \left\{ t_i^{(n)} - t_{i-1}^{(n)} \right\}$ assume that $\lim_{n\to\infty} \delta^{(n)} = 0$.*

Then there exists for each $n \in \mathbb{N}$ a function $h_n : \mathbb{R}^d \times \mathbb{R}^+ \to \mathbb{R}^m$ and a sequence of independent random variables $(\epsilon_i^{(n)})_{i\in\mathbb{N}}$ in $\mathbb{R}^d$ with finite variance and $E(\epsilon_i^{(n)}) = 0 \,\forall\, i, n \in \mathbb{N}$ such that $h_n\left(\epsilon_i^{(n)}, t_i^{(n)} - t_{i-1}^{(n)}\right)$ has finite variance, $E\left(h_n\left(\epsilon_i^{(n)}, t_i^{(n)} - t_{i-1}^{(n)}\right)\right) = 0$ and

$$\mathrm{plim}_{n\to\infty} d_{\mathbb{R}^d\times\mathbb{R}^m}\left((G^{(n)}, X^{(n)}), (G, X)\right) = 0\,, \tag{3.4.2}$$

where for each $n \in \mathbb{N}$ the process $(G^{(n)}, X^{(n)})$ in $\mathbb{R}^d \times \mathbb{R}^m$ is defined by

$$(G_0^{(n)}, X_0^{(n)}) = (G_0, X_0), \tag{3.4.3}$$

$$G_{t_i^{(n)}}^{(n)} = G_{t_{i-1}^{(n)}}^{(n)} + \exp\left(\left(\mu + \mathrm{vech}^{-1}\left(X_{t_{i-1}^{(n)}}^{(n)}\right)\right)/2\right)\epsilon_i^{(n)}, \tag{3.4.4}$$

$$X_{t_i^{(n)}}^{(n)} = e^{A\left(t_i^{(n)} - t_{i-1}^{(n)}\right)} X_{t_{i-1}^{(n)}}^{(n)} + \beta h_n\left(\epsilon_i^{(n)}, t_i^{(n)} - t_{i-1}^{(n)}\right) \text{ for all } i \in \mathbb{N} \text{ and} \tag{3.4.5}$$

$$(G_t^{(n)}, X_t^{(n)}) = \left(G_{t_{i-1}^{(n)}}^{(n)}, X_{t_{i-1}^{(n)}}^{(n)}\right) \text{ for } t \in (t_{i-1}^{(n)}, t_i^{(n)}),\ i \in \mathbb{N}. \tag{3.4.6}$$

The sequence $(\epsilon_i^{(n)})_{i\in\mathbb{N}}$ can be chosen to be i.i.d. provided $t_i^{(n)} - t_{i-1}^{(n)} = \delta^{(n)}$ for all $i \in \mathbb{N}$.

If h is continuous h_n can be chosen such that the sequence of functions $h_n : \mathbb{R}^d \times \mathbb{R}^+ \to \mathbb{R}^m$ satisfies

$$\lim_{n\to\infty}\left(\sup_{z\in K}\sup_{i\in\mathbb{N}}\left\{\left\|h_n\left(z, t_i^{(n)} - t_{i-1}^{(n)}\right) - h(z)\right\|\right\}\right) = 0 \tag{3.4.7}$$

for all compact $K \subset \mathbb{R}^d$. If h is uniformly continuous, h_n can be chosen such that (3.4.7) holds with $\mathbb{R}^d$ instead of K.

When the time grids are equidistant, i.e. $t_i^{(n)} - t_{i-1}^{(n)} = \delta^{(n)}$ for all $i \in \mathbb{N}$, and $(\epsilon_i^{(n)})_{i\in\mathbb{N}}$ is chosen i.i.d., then the increments $\left(G_{t_i^{(n)}}^{(n)} - G_{t_{i-1}^{(n)}}^{(n)}\right)_{i\in\mathbb{N}} =: \left(Y_i^{(n)}\right)_{i\in\mathbb{N}}$ of $G^{(n)}$ are a discrete time multivariate EGARCH(1,1) process with associated vectorized log-volatility process $\left(X_{t_{i-1}^{(n)}}^{(n)}\right)_{i\in\mathbb{N}} =: \left(\mathrm{vech}(H_i^{(n)})\right)_{i\in\mathbb{N}}$ (apart from the fact that $\mathrm{var}\left(\epsilon_i^{(n)}\right) = I_d$ will usually not be satisfied). We allow for a non-equidistant grid as this may be useful when having irregularly spaced data (cf. Maller et al. (2006)).

We will prove this result in several steps now using the first jump approximation of a Lévy process introduced in Szimayer and Maller (2007) and the theory on the convergence of stochastic integrals and differential equations to be found in Kurtz and Protter(1991, 1996). As similar techniques should be useful elsewhere as well, we state several results in a general manner below.

It should be noted that there is an immediate extension to ECOGARCH(p,q) processes with orders p,q where $q>1$. However, the approximating piecewise constant processes are no longer essentially discrete time multivariate EGARCH processes. Furthermore, it is clear that the result remains valid when considering the processes only on a finite time interval $[0,T]$ and looking at partitions $(t_i^{(n)})_{i\in\{1,2,\dots,N^{(n)}\}}$ of this interval with $N^{(n)}\in\mathbb{N}$.

We start by giving a d-dimensional and infinite time extension of the first jump approximation for a general Lévy process presented in Szimayer and Maller (2007) and a refinement of it for a Lévy process with zero mean. Below the infimum over an empty set is taken to be ∞ as usual. Recall below that for a given Lévy process its drift γ_L is defined such that (3.2.3) is satisfied with the currently employed norm.

Theorem 3.4.2. *Let L be a d-dimensional Lévy process with no Brownian part, drift γ_L and Lévy measure ν_L. Further let $(m^{(n)})_{n\in\mathbb{N}}$ be a positive sequence, which is bounded by 1 and monotonically decreases to 0, and $(t_i^{(n)})_{i\in\mathbb{N}_0}$ be for each $n\in\mathbb{N}$ a strictly increasing sequence with $t_0^{(n)}=0$ and $\lim_{i\to\infty}t_i^{(n)}=\infty$. Setting $\delta^{(n)}:=\sup_{i\in\mathbb{N}}\left\{t_i^{(n)}-t_{i-1}^{(n)}\right\}$, assume further that $\lim_{n\to\infty}\delta^{(n)}=0$ and*

$$\lim_{n\to\infty}\delta^{(n)}\left(\nu_L\left(J^{(n)}\right)\right)^2=0\,, \tag{3.4.8}$$

where $J^{(n)}:=\{x\in\mathbb{R}^d:\|x\|>m^{(n)}\}$.

(a) *Define for all $n\in\mathbb{N}$*

$$\begin{aligned}
\gamma^{(n)}&:=\gamma_L-\int_{m^{(n)}<\|x\|\le 1}x\nu_L(dx),\\
\tau_i^{(n)}&:=\inf\{t:\ t_{(i-1)}^{(n)}<t\le t_i^{(n)},\|\Delta L_t\|>m^{(n)}\}\quad \text{for all } i\in\mathbb{N},\\
\tilde L_t^{(n)}&:=\gamma^{(n)}t+\sum_{i\in\mathbb{N}:\,\tau_i^{(n)}\le t}\Delta L_{\tau_i^{(n)}}\quad \text{for } t\in\mathbb{R}^+,\ \tilde L^{(n)}:=(\tilde L_t^{(n)})_{t\in\mathbb{R}^+},\\
\bar L_t^{(n)}&:=\tilde L_{t_{i-1}^{(n)}}^{(n)},\quad \text{for all } t\in[t_{i-1}^{(n)},t_i^{(n)}),\ i\in\mathbb{N},\quad \bar L^{(n)}:=(\bar L_t^{(n)})_{t\in\mathbb{R}^+}.
\end{aligned}$$

Then it holds that

$$\tilde L^{(n)}\to L \text{ in ucp as } n\to\infty \tag{3.4.9}$$

and

$$\operatorname{plim}_{n\to\infty}d_{\mathbb{R}^d}(\bar L^{(n)},L)=0. \tag{3.4.10}$$

(b) *If L has finite expectation and $E(L_1)=0$, define $L^{(n)}=(L_t^{(n)})_{t\in\mathbb{R}^+}$ by setting*

$$L_t^{(n)}=\sum_{i\in\mathbb{N}:\,t_i^{(n)}\le t}\left(1_{(0,\infty)}(\tau_i^{(n)})\Delta L_{\tau_i^{(n)}}-\frac{1-e^{-\nu_L\left(J^{(n)}\right)\left(t_i^{(n)}-t_{i-1}^{(n)}\right)}}{\nu_L\left(J^{(n)}\right)}\int_{J^{(n)}}x\nu_L(dx)\right). \tag{3.4.11}$$

Then

$$\operatorname{plim}_{n\to\infty}d_{\mathbb{R}^d}(L^{(n)},L)=0. \tag{3.4.12}$$

and $E\left(L_t^{(n)}\right)=0$ for all $t\in\mathbb{R}^+$ as well as $E\left(\Delta L_{t_i^{(n)}}^{(n)}\right)=0$ for all $i\in\mathbb{N}$ and each $n\in\mathbb{N}$.

(c) *Provided $E(\|L_1\|^2)<\infty$, it holds that $E(\|\tilde L_t^{(n)}\|^2)$, $E(\|\bar L_t^{(n)}\|^2)$, $E(\|L_t^{(n)}\|^2)$ as well as $E(\|\Delta\tilde L_t^{(n)}\|^2)$, $E(\|\Delta\bar L_t^{(n)}\|^2)$, $E(\|\Delta L_t^{(n)}\|^2)$ are finite for all $t\in\mathbb{R}^+$ and $n\in\mathbb{N}$.*

The definition of $\tau_i^{(n)}$ above means that it is the first time in the grid interval $(t_{i-1}^{(n)}, t_i^{(n)}]$ at which L has a jump bigger than $m^{(n)}$ in norm. If there is no such jump, $\tau_i^{(n)} = \infty$. $\tilde{L}^{(n)}$ approximates the Lévy process L by a drift and the first jumps of size greater than $m^{(n)}$ in the grid intervals. Since the jumps are left at their original time, we obtain ucp convergence. In the approximation $\bar{L}^{(n)}$ both these jumps and the increment caused by the drift are shifted to the grid points, so that $\bar{L}^{(n)}$ is constant in between the grid times. Due to shifting the jumps we obtain only convergence in the Skorokhod topology. Finally, the approximation $L^{(n)}$ is a modification of $\bar{L}^{(n)}$ when L has a finite and vanishing first moment. It ensures that also the approximation has a vanishing mean, as it will often be desirable to reproduce this property of L.

Proof. (a) An inspection of the proof of Szimayer and Maller (2007, Theorem 3.1) shows that it immediately generalizes to our multivariate set-up with no upper bound on the jump sizes and no binning of the jump sizes. This proves (3.4.9) which implies

$$\operatorname{plim}_{n\to\infty} d_{\mathbb{R}^d}(\tilde{L}^{(n)}, L) = 0.$$

Using the time change λ which is the obvious extension to $[0,\infty)$ of the time change employed in the proof of Szimayer and Maller (2007, Theorem 3.2) and arguments analogous to theirs give $d_{\mathbb{R}^d}(\tilde{L}^{(n)}, \bar{L}^{(n)}) \leq d_{\lambda,\mathbb{R}^d}(\tilde{L}^{(n)}, \bar{L}^{(n)}) \leq \delta^{(n)} + o(\sqrt{\delta^{(n)}}) \to 0$ as $n \to \infty$ a.s. The triangle inequality thus establishes (3.4.10).

(b) Assume now $E(L_1) = 0$. Then $E\left(L_t^{(n)}\right) = 0$ for all $t \in \mathbb{R}^+$ as well as $E\left(\Delta L_{t_i^{(n)}}^{(n)}\right) = 0$ for all $i \in \mathbb{N}$ and each $n \in \mathbb{N}$ follows from

$$E\left(1_{(0,\infty)}(\tau_i^{(n)})\Delta L_{\tau_i^{(n)}}\right) = \frac{1 - e^{-\nu_L\left(J^{(n)}\right)\left(t_i^{(n)} - t_{i-1}^{(n)}\right)}}{\nu_L\left(J^{(n)}\right)} \int_{J^{(n)}} x\nu_L(dx).$$

Since $E(L_1) = 0$, we have $\gamma_L = -\int_{\|x\|>1} x\nu_L(dx)$. Hence, straightforward calculations give for all $t \in \mathbb{R}^+$

$$\begin{aligned}
&\|\bar{L}_t^{(n)} - L_t^{(n)}\| \\
&\quad = \left\| \sum_{i\in\mathbb{N}:\, t_i^{(n)} \leq t} \left(\frac{1 - e^{-\nu_L\left(J^{(n)}\right)\left(t_i^{(n)} - t_{i-1}^{(n)}\right)}}{\nu_L\left(J^{(n)}\right)} \int_{J^{(n)}} x\nu_L(dx) - \left(t_i^{(n)} - t_{i-1}^{(n)}\right) \int_{J^{(n)}} x\nu_L(dx) \right) \right\| \\
&\quad \leq tC^{(n)},
\end{aligned}$$

where

$$C^{(n)} = \sqrt{\delta^{(n)}} \int_{J^{(n)}} \|x\|\nu_L(dx) \sum_{k=1}^{\infty} \frac{\left(\nu_L(J^{(n)})\sqrt{\delta^{(n)}}\right)^k (\delta^{(n)})^{(k-1)/2}}{k+1!}.$$

Since (3.4.8) and $E(\|L_1\|) < \infty$ imply that $\sqrt{\delta^{(n)}} \int_{J^{(n)}} \|x\|\nu_L(dx)$ and $\nu_L(J^{(n)})\sqrt{\delta^{(n)}}$ converge to zero as $n \to \infty$, we have $\lim_{n\to\infty} C^{(n)} = 0$ and hence $\lim_{n\to\infty} \|\bar{L}^{(n)} - L^{(n)}\|_{\mathbb{R}^d,[0,T]} = 0$ a.s. for all $T \in \mathbb{R}^+$. Combining this with (3.4.10) shows (3.4.12).

Finally, (c) is easily seen using

$$E\left(1_{(0,\infty)}(\tau_i)\Delta L_{\tau_i^{(n)}}\left(\Delta L_{\tau_i^{(n)}}\right)^*\right) = \frac{1-e^{-\nu_L\left(J^{(n)}\right)\left(t_i^{(n)}-t_{i-1}^{(n)}\right)}}{\nu_L\left(J^{(n)}\right)}\int_{J^{(n)}} xx^*\nu_L(dx),$$

since $E(\|L_1\|^2)<\infty$ is equivalent to the finiteness of $\int_{\mathbb{R}^d}\|x\|^2\nu_L(dx)$ or of all components of $\int_{\mathbb{R}^d} xx^*\nu_L(dx)$. □

The next Lemma shows that when one of the sequences $(\delta^{(n)})_{n\in\mathbb{N}}$ or $(m^{(n)})_{n\in\mathbb{N}}$ is given one can always choose the other one such that (3.4.8) holds.

Lemma 3.4.3. *Let L be a Lévy process in $\mathbb{R}^d$. Assume that $(\delta^{(n)})_{n\in\mathbb{N}}$ is a monotonically decreasing sequence in $\mathbb{R}^{++}$ with $\lim_{n\to\infty}\delta^{(n)}=0$ or $(m^{(n)})_{n\in\mathbb{N}}$ is a monotonically decreasing sequence in $\mathbb{R}^{++}$ with $\lim_{n\to\infty}m^{(n)}=0$ and $m^{(n)}\leq 1\forall n\in\mathbb{N}$, respectively. Then a monotonically decreasing sequence $(m^{(n)})_{n\in\mathbb{N}}$ in $\mathbb{R}^{++}$ with $\lim_{n\to\infty}m^{(n)}=0$ and $m^{(n)}\leq 1\forall n\in\mathbb{N}$ or a monotonically decreasing sequence $(\delta^{(n)})_{n\in\mathbb{N}}$ in $\mathbb{R}^{++}$ with $\lim_{n\to\infty}\delta^{(n)}=0$, respectively, can be chosen such that (3.4.8) is satisfied for all norms $\|\cdot\|$.*

Proof. We have that $\int_{\mathbb{R}^d}\min(\|x\|^2,1)\nu_L(dx)<\infty$ for all norms $\|\cdot\|$, because ν_L is a Lévy measure. Thus $\lim_{n\to\infty}\delta^{(n)}\nu_L(J^{(n)}\backslash U_1(0))=0$ for all norms where $U_1(0)$ denotes the open ball around zero with radius 1 and $J^{(n)}=\{x\in\mathbb{R}^d:\|x\|>m^{(n)}\}$. Moreover,

$$m(n)^2\nu_L(J^{(n)}\cap U_1(0))\leq\int_{J^{(n)}\cap U_1(0)}\|x\|^2\nu_L(dx)\leq\int_{U_1(0)}\|x\|^2\nu_L(dx)<\infty.$$

This implies that (3.4.8) holds if $\delta^{(n)}=o\left(\left(m^{(n)}\right)^4\right)$. Hence, the lemma is shown by choosing for instance $m^{(n)}=(\delta^{(n)})^{1/5}\wedge 1$ or $\delta^{(n)}=(m^{(n)})^5$, respectively, for all $n\in\mathbb{N}$. □

Next we show that the first jump approximation of a Lévy process has uniformly controlled variations (UCV) which is important, as it ensures convergence of stochastic integrals and solutions to stochastic differential equations (cf. Kurtz and Protter (1991, 1996)). Equivalently one could use a condition called "uniform tightness" (cf. Jacod and Shiryaev (2003, Section IV.6) or Kurtz and Protter (1991, 1996)). This shows in general that the solution of a Lévy-driven stochastic differential equation can be approximated arbitrarily well in the Skorokhod topology in probability by replacing the driving Lévy process with its first jump approximation provided the additional technical assumptions of Kurtz and Protter (1991, 1996) or Jacod and Shiryaev (2003, Section IX.6) are satisfied.

In the following we need to transform any semi-martingale to one with bounded jumps in a suitable way. To this end we define for a semi-martingale Z and $\kappa\in\mathbb{R}^{++}\cup\{\infty\}$ the semi-martingale $Z^{[\kappa]}$ by setting $Z_t^{[\kappa]}=Z_t-\sum_{0<s\leq t}r_\kappa(\Delta Z_s)$ with $r_\kappa(z):=\max(0,1-\kappa/\|z\|)z$ for finite κ and $Z^{[\infty]}=Z$. Furthermore, for a finite variation process A we denote by $(TV(A)_t)_{t\in\mathbb{R}^+}$ the process giving the total variation of A over the interval $[0,t]$ for $t\in\mathbb{R}^+$ and for a d-dimensional martingale $M=(M_1,M_2,\ldots,M_d)^*$ the quadratic variation $[M,M]$ is understood to be defined by $[M,M]_t=\sum_{i=1}^d[M_i,M_i]_t$.

Definition 3.4.4 (Kurtz and Protter (1991))**.** *Let $(Z^{(n)})_{n\in\mathbb{N}}$ be a sequence of semi-martingales in $\mathbb{R}^d$ each defined on its own filtered probability space $(\Omega^{(n)},\mathcal{F}^{(n)},P^{(n)},(\mathcal{F}_t^{(n)})_{t\in\mathbb{R}^+})$*

satisfying the usual hypothesis. If there exists a $\kappa \in \mathbb{R}^{++} \cup \{\infty\}$ *such that for each* $\alpha > 0$ *and* $n \in \mathbb{N}$ *there exist* $(\mathcal{F}_t^{(n)})$*-local martingales* $M^{(n)}$*,* $(\mathcal{F}_t^{(n)})$*-adapted finite variation processes* $A^{(n)}$ *in* $\mathbb{R}^d$ *and* $(\mathcal{F}_t^{(n)})$*-stopping times* $T^{(n,\alpha)}$ *satisfying* $(Z^{(n)})^{[\kappa]} = M^{(n)} + A^{(n)}$*,* $P^{(n)}(T^{(n,\alpha)} \leq \alpha) \leq (1/\alpha)$ *and*

$$\sup_{n\in\mathbb{N}} E_{P^{(n)}}\left([M^{(n)}, M^{(n)}]_{\min(t,T^{(n,\alpha)})} + TV(A^{(n)})_{\min(t,T^{(n,\alpha)})}\right) < \infty \tag{3.4.13}$$

for all $t \in \mathbb{R}^+$ *then the sequence* $(Z^{(n)})_{n\in\mathbb{N}}$ *is said to have* uniformly controlled variations (UCV)*.*

In the following our processes are defined on the same probability space, but the filtrations are different.

Theorem 3.4.5. *Let* L *be a* d*-dimensional Lévy process without a Brownian part and* $(\bar{L}^{(n)})_{n\in\mathbb{N}}$ *the first jump approximation of Theorem 3.4.2 (a). Let* $(\mathcal{F}_t^{(n)})_{t\in\mathbb{R}}$ *be for each* $n \in \mathbb{N}$ *the completed filtration generated by* $\bar{L}^{(n)}$*. Then the usual conditions are satisfied and* $\bar{L}^{(n)}$ *is for each* $n \in \mathbb{N}$ *a semi-martingale on* $(\Omega, \mathcal{F}, P, (\mathcal{F}_t^{(n)})_{t\in\mathbb{R}})$*. Moreover,* $(\bar{L}^{(n)})_{n\in\mathbb{N}}$ *has UCV.*

If L *has finite mean and* $E(L_1) = 0$*, let* $L^{(n)}$ *be the first jump approximation of Theorem 3.4.2 (b). Then* $L^{(n)}$ *is for each* $n \in \mathbb{N}$ *a martingale on* $(\Omega, \mathcal{F}, P, (\mathcal{F}_t^{(n)})_{t\in\mathbb{R}})$ *and* $(L^{(n)})_{n\in\mathbb{N}}$ *has UCV.*

Proof. (1) Since $\bar{L}^{(n)}$ is piecewise constant, it is clear that $(\mathcal{F}_t^{(n)})_{t\in\mathbb{R}}$ is right continuous. Thus the usual conditions are satisfied. The semi-martingale property is also immediate. To see that also $L^{(n)}$ is a semi-martingale with respect to this filtration, provided $E(L_1) = 0$, it suffices to note that $L^{(n)} - \bar{L}^{(n)}$ is a deterministic process of finite variation on compacts. That $L^{(n)}$ is even a martingale is then straightforward as the jumps have zero expectation and are independent of the past.

(2) We now show UCV for $(\bar{L}^{(n)})_{n\in\mathbb{N}}$.

Choose $\kappa \in (2, \infty)$ such that

$$\kappa/2 > \sup_{n\in\mathbb{N}}\left\{\delta^{(n)}\|\gamma^{(n)}\|\right\} + 1 \text{ and } \kappa/2 > \sup_{i,n\in\mathbb{N}}\left\{\left\|\frac{1 - e^{-\nu_L\left(J^{(n)}\right)\left(t_i^{(n)} - t_{i-1}^{(n)}\right)}}{\nu_L\left(J^{(n)}\right)}\int_{J_0^{(n)}} x\nu_L(dx)\right\|\right\}$$

where $J_0^{(n)} := J^{(n)} \cap \{x : \|x\| \leq 1\}$. The finiteness of the first supremum is a consequence of (3.4.8) and the finiteness of the second one follows from

$$\begin{aligned}&\left\|\frac{1 - e^{-\nu_L\left(J^{(n)}\right)\left(t_i^{(n)} - t_{i-1}^{(n)}\right)}}{\nu_L\left(J^{(n)}\right)}\int_{J_0^{(n)}} x\nu_L(dx)\right\| \\ &\qquad \leq \sqrt{\delta^{(n)}}\nu_L(J_0^{(n)})\sum_{k=0}^{\infty}\frac{\left(\sqrt{\delta^{(n)}}\nu_L(J^{(n)})\right)^k(\delta^{(n)})^{(k+1)/2}}{k+1!}\end{aligned}$$

for all $i, n \in \mathbb{N}$, since the right hand side goes to zero as $n \to \infty$.

Define $(M_t^{(n)})_{t\in\mathbb{R}^+}$ by

$$M_t^{(n)} = \sum_{i\in\mathbb{N}:\, t_i^{(n)}\le t} \left(1_{(0,\infty)}(\tau_i^{(n)}) 1_{(0,1]}(\|\Delta L_{\tau_i^{(n)}}\|)\Delta L_{\tau_i^{(n)}} - \frac{1-e^{-\nu_L\left(J^{(n)}\right)\left(t_i^{(n)}-t_{i-1}^{(n)}\right)}}{\nu_L\left(J^{(n)}\right)} \int_{J_0^{(n)}} x\nu_L(dx) \right)$$

and $(A_t^{(n)})_{t\in\mathbb{R}^+}$ by

$$A_t^{(n)} = \sum_{i\in\mathbb{N}:\, t_i^{(n)}\le t} \left(\tilde{r}_\kappa\left(1_{(0,\infty)}(\tau_i^{(n)}) 1_{(1,\infty)}(\|\Delta L_{\tau_i^{(n)}}\|)\Delta L_{\tau_i^{(n)}} + \gamma^{(n)}\left(t_i^{(n)} - t_{i-1}^{(n)}\right)\right) + \frac{1-e^{-\nu_L\left(J^{(n)}\right)\left(t_i^{(n)}-t_{i-1}^{(n)}\right)}}{\nu_L\left(J^{(n)}\right)} \int_{J_0^{(n)}} x\nu_L(dx) \right),$$

where $\tilde{r}_\kappa(x) = x - r_\kappa(x)$. Then $M^{(n)}$ is a $(\mathcal{F}_t^{(n)})$-martingale with expectation zero and $A^{(n)}$ a $(\mathcal{F}_t^{(n)})$-adapted finite variation process for all $n\in\mathbb{N}$. By the choice of κ we have

$$(L^{(n)})^{[\kappa]} = M^{(n)} + A^{(n)}.$$

Since $[M^{(n)}, M^{(n)}]_t = \sum_{i\in\mathbb{N}:\, t_i^{(n)}\le t} \left(\Delta M_{t_i^{(n)}}^{(n)}\right)^* \Delta M_{t_i^{(n)}}^{(n)}$, it follows that

$$\begin{aligned}
E\left([M^{(n)}, M^{(n)}]_t\right) &= \sum_{i\in\mathbb{N}:\, t_i^{(n)}\le t} E\left(1_{(0,\infty)}(\tau_i^{(n)}) 1_{(0,1]}(\|\Delta L_{\tau_i^{(n)}}\|)(\Delta L_{\tau_i^{(n)}})^* \Delta L_{\tau_i^{(n)}}\right) \\
&\quad - \sum_{i\in\mathbb{N}:\, t_i^{(n)}\le t} \left(\frac{1-e^{-\nu_L\left(J^{(n)}\right)\left(t_i^{(n)}-t_{i-1}^{(n)}\right)}}{\nu_L\left(J^{(n)}\right)} \right)^2 \int_{J_0^{(n)}} x^*\nu_L(dx) \int_{J_0^{(n)}} x\nu_L(dx) \\
&\le \sum_{i\in\mathbb{N}:\, t_i^{(n)}\le t} E\left(1_{(0,\infty)}(\tau_i^{(n)}) 1_{(0,1]}(\|\Delta L_{\tau_i^{(n)}}\|)(\Delta L_{\tau_i^{(n)}})^* \Delta L_{\tau_i^{(n)}}\right) \\
&= \sum_{i\in\mathbb{N}:\, t_i^{(n)}\le t} \frac{1-e^{-\nu_L\left(J^{(n)}\right)\left(t_i^{(n)}-t_{i-1}^{(n)}\right)}}{\nu_L\left(J^{(n)}\right)} \int_{J_0^{(n)}} x^* x\nu_L(dx).
\end{aligned}$$

Denoting by $\|\cdot\|_2$ the Euclidean norm and using the elementary inequality $1-e^{-x}\le x$ for all $x\in\mathbb{R}^+$, this implies

$$\sup_{n\in\mathbb{N}} E([M^{(n)}, M^{(n)}]_t) \le t\int_{\|x\|\le 1} \|x\|_2^2 \nu_L(dx) < \infty \tag{3.4.14}$$

for all $t\in\mathbb{R}$. Note that the appearance of the Euclidean norm here does not imply that the Euclidean norm is also used elsewhere, e.g. in the definition of $J^{(n)}$, i.e. $\|\cdot\|$ is not necessarily $\|\cdot\|_2$.

Turning to $A^{(n)}$ we have that

$$\begin{aligned}\|\Delta A^{(n)}_{t_i^{(n)}}\| \leq & 3\left\|\tilde{r}_\kappa\left(1_{(0,\infty)}(\tau_i^{(n)})1_{(1,\infty)}(\|\Delta L_{\tau_i^{(n)}}\|)\Delta L_{\tau_i^{(n)}}\right)\right\| + \|\gamma_L\|\left(t_i^{(n)} - t_{i-1}^{(n)}\right) \\ & + \left\|\left(\frac{1-e^{-\nu_L\left(J^{(n)}\right)\left(t_i^{(n)}-t_{i-1}^{(n)}\right)}}{\nu_L\left(J^{(n)}\right)} - \left(t_i^{(n)} - t_{i-1}^{(n)}\right)\right)\int_{J_0^{(n)}} x\nu_L(dx)\right\|\end{aligned}$$

for all $i, n \in \mathbb{N}$. The inequality is immediate provided

$$\begin{aligned}\left\|1_{(0,\infty)}(\tau_i^{(n)})1_{(1,\infty)}(\|\Delta L_{\tau_i^{(n)}}\|)\Delta L_{\tau_i^{(n)}} + \gamma^{(n)}\left(t_i^{(n)} - t_{i-1}^{(n)}\right)\right\| \leq \kappa \text{ and} \\ \left\|1_{(0,\infty)}(\tau_i^{(n)})1_{(1,\infty)}(\|\Delta L_{\tau_i^{(n)}}\|)\Delta L_{\tau_i^{(n)}}\right\| \leq \kappa.\end{aligned}$$

Otherwise the choice of κ ensures $\left\|1_{(0,\infty)}(\tau_i^{(n)})1_{(1,\infty)}(\|\Delta L_{\tau_i^{(n)}}\|)\Delta L_{\tau_i^{(n)}}\right\| > \kappa/2$ and that $\left\|\Delta A^{(n)}_{t_i^{(n)}}\right\| \leq (3/2)\kappa$ which implies the validity of the inequality. We have

$$\left\|\left(\frac{1-e^{-\nu_L\left(J^{(n)}\right)\left(t_i^{(n)}-t_{i-1}^{(n)}\right)}}{\nu_L\left(J^{(n)}\right)} - \left(t_i^{(n)} - t_{i-1}^{(n)}\right)\right)\int_{J_0^{(n)}} x\nu_L(dx)\right\| \leq C_A^{(n)}\left(t_i^{(n)} - t_{i-1}^{(n)}\right)$$

with

$$C_A^{(n)} := \sqrt{\delta^{(n)}}\nu_L(J_0^{(n)})\sum_{k=1}^{\infty}\frac{\left(\nu_L(J^{(n)})\sqrt{\delta^{(n)}}\right)^k(\delta^{(n)})^{(k-1)/2}}{k+1!},$$

which converges to zero as $n \to \infty$. Hence,

$$TV(A^{(n)})_t \leq 3\int_0^t\int_{\|x\|>1}\min(\|x\|,\kappa)N_L(ds,dx) + t\left(\|\gamma_L\| + \sup_{n\in\mathbb{N}}\left\{C_A^{(n)}\right\}\right)$$

and thus

$$\begin{aligned}&\sup_{n\in\mathbb{N}} E\left(TV(A^{(n)})_t\right) \\ &\qquad \leq t\left(3\int_{\|x\|>1}\min(\|x\|,\kappa)\nu_L(dx) + \|\gamma_L\| + \sup_{n\in\mathbb{N}}\left\{C_A^{(n)}\right\}\right) < \infty\end{aligned} \tag{3.4.15}$$

for all $t \in \mathbb{R}^+$.

Combining (3.4.14) and (3.4.15) and choosing $T^{(n,\alpha)} = \alpha + 1$ for all $n \in \mathbb{N}$ and $\alpha \in \mathbb{R}^{++}$ shows that $(\bar{L}^{(n)})_{n\in\mathbb{N}}$ has UCV.

(3) It remains to verify that $(L^{(n)})_{n\in\mathbb{N}}$ has UCV, provided L has a finite expectation and $E(L_1) = 0$. The arguments presented in the proof of Theorem 3.4.2 (b) imply that

$$TV(L^{(n)} - \bar{L}^{(n)})_t \leq C^{(n)}t.$$

for all $n \in \mathbb{N}$ and $t \in \mathbb{R}^+$ with $\lim_{n\to\infty} C^{(n)} = 0$. This immediately shows that $(L^{(n)} - \bar{L}^{(n)})_{n\in\mathbb{N}}$ has UCV. As $(L^{(n)} - \bar{L}^{(n)})_{n\in\mathbb{N}}$ and $(\bar{L}^{(n)})_{n\in\mathbb{N}}$ converge in probability in the Skorokhod topology, Kurtz and Protter (1996, Theorem 7.6) ensures that both sequences are uniformly tight. Property 6.3 on page 377 of Jacod and Shiryaev (2003) therefore implies that $(L^{(n)})_{n\in\mathbb{N}}$ is uniformly tight and so Kurtz and Protter (1996, Theorem 7.6) shows that it has UCV. □

After these preliminaries we can proof our theorem on the convergence of discrete time EGARCH(1,1) processes to a ECOGARCH(1,1) process.

Proof of Theorem 3.4.1: Let $\|\cdot\|$ be a norm on $\mathbb{R}^{d+m}$. We have that the joint process $\mathbf{L} = (L_t^*, M_t^*)_{t\in\mathbb{R}^d}^*$ is a Lévy process in $\mathbb{R}^{d+m}$ with

$$\mathbf{L}_t = \begin{pmatrix}\gamma_L\\ \gamma_M\end{pmatrix} t + \int_0^t \int_{\|(x^*,h(x)^*)^*\|\le 1} \begin{pmatrix} x\\ h(x)\end{pmatrix} \tilde{N}_L(ds,dx) + \int_0^t \int_{\|(x^*,h(x)^*)^*\|>1} \begin{pmatrix} x\\ h(x)\end{pmatrix} N_L(ds,dx)$$

for all $t \in \mathbb{R}^+$ with $\gamma_L = -\int_{\|(x^*,h(x)^*)^*\|>1} x\nu_L(dx)$ and $\gamma_M = -\int_{\|(x^*,h(x)^*)^*\|>1} h(x)\nu_L(dx)$. Here we used that $\nu_{\mathbf{L}}(W) = \nu_L(f^{-1}(W))$ and $N_{\mathbf{L}}(ds,W) = N_L(ds, f^{-1}(W))$ for all Borel sets $W \subset \mathbb{R}^{d+m}$ where $f : \mathbb{R}^d \to \mathbb{R}^{d+m}$, $x \mapsto (x^*, h(x)^*)^*$. It is clear that $\mathbf{L}$ has a finite expectation and $E(\mathbf{L}_1) = 0$.

First Step: Choice of noise sequences $\epsilon^{(n)}$ and functions h_n
Choose a sequence $(m^{(n)})_{n\in\mathbb{N}}$ such that (3.4.8) is satisfied for $\nu_{\mathbf{L}}$ noting that the existence is ensured by Lemma 3.4.3. Let $(\mathbf{L}^{(n)})_{n\in\mathbb{N}} = \left((L^{(n)})^*, (M^{(n)})^*\right)_{n\in\mathbb{N}}^*$ be the first jump approximation to $\mathbf{L}$ as given in Theorem 3.4.2 (b). Hence, $\operatorname{plim}_{n\to\infty} d_{\mathbb{R}^{d+m}}(\mathbf{L}^{(n)}, \mathbf{L}) = 0$ and due to Theorem 3.4.5 $(\mathbf{L}^{(n)})_{n\in\mathbb{N}}$ has UCV.

Set

$$\epsilon_i^{(n)} = \Delta L^{(n)}_{t_i^{(n)}} = 1_{(0,\infty)}(\tau_i^{(n)})\Delta L_{\tau_i^{(n)}} + \gamma_{L,i}^{(n)} \text{ for all } i,n \in \mathbb{N} \tag{3.4.16}$$

with

$$\gamma_{L,i}^{(n)} = -\frac{1 - e^{-\nu_L\left(J^{(n)}\right)\left(t_i^{(n)} - t_{i-1}^{(n)}\right)}}{\nu_L\left(J^{(n)}\right)} \int_{J^{(n)}} x\nu_L(dx)\,, \tag{3.4.17}$$

where $J^{(n)} = \left\{x \in \mathbb{R}^d : \|(x^*, h(x)^*)^*\| > m^{(n)}\right\}$. Then by construction $(\epsilon_i^{(n)})_{i\in\mathbb{N}}$ is for each $n \in \mathbb{N}$ a sequence of independent random variables having finite variance and zero expectation (cf. Theorem 3.4.2 (b), (c)). If $t_i^{(n)} - t_{i-1}^{(n)} = \delta^{(n)}$ for all $i \in \mathbb{N}$ then $\gamma_{L,i}^{(n)}$ does not depend on $i \in \mathbb{N}$ and $(\epsilon_i^{(n)})_{i\in\mathbb{N}}$ is i.i.d.

Moreover,

$$\Delta M^{(n)}_{t_i^{(n)}} = h\left(1_{(0,\infty)}(\tau_i^{(n)})\Delta L_{\tau_i^{(n)}}\right) + \gamma_{M,i}^{(n)} = h_n\left(\epsilon_i^{(n)}, t_i^{(n)} - t_{i-1}^{(n)}\right) \tag{3.4.18}$$

for all $i, n \in \mathbb{N}$, with

$$\gamma_{M,i}^{(n)} = -\frac{1 - e^{-\nu_L\left(J^{(n)}\right)\left(t_i^{(n)} - t_{i-1}^{(n)}\right)}}{\nu_L\left(J^{(n)}\right)} \int_{J^{(n)}} h(x)\nu_L(dx) \quad \text{and} \tag{3.4.19}$$

$$h_n : \mathbb{R}^d \times \mathbb{R}^+ \to \mathbb{R}^m, \tag{3.4.20}$$

$$(z,t) \mapsto h\left(z + \frac{1 - e^{-\nu_L\left(J^{(n)}\right)t}}{\nu_L\left(J^{(n)}\right)} \int_{J^{(n)}} x\nu_L(dx)\right) - \frac{1 - e^{-\nu_L\left(J^{(n)}\right)t}}{\nu_L\left(J^{(n)}\right)} \int_{J^{(n)}} h(x)\nu_L(dx).$$

Theorem 3.4.2 (b), (c) ensures that $h_n\left(\epsilon_i^{(n)}, t_i^{(n)} - t_{i-1}^{(n)}\right)$ has a finite variance and zero expectation for all $i, n \in \mathbb{N}$.

We have that $\lim_{n\to\infty}\sup_{i\in\mathbb{N}}\left\{\|\gamma_{M,i}^{(n)}\|\right\}=0$ and $\lim_{n\to\infty}\sup_{i\in\mathbb{N}}\left\{\|\gamma_{L,i}^{(n)}\|\right\}=0$. From this it is easy to see that (3.4.7) holds with $\mathbb{R}^d$ instead of K if h is uniformly continuous. If h is only continuous (3.4.7) follows along the same lines noting that any continuous function is uniformly continuous on compacts.

Second Step: Convergence to the ECOGARCH

Define the processes S in $\mathbb{R}^m$ by $S_t=(1,1,\ldots,1)^*t$ for all $t\in\mathbb{R}^+$ and $(\tilde{X}_t^{(n)})_{t\in\mathbb{R}^+}$ for all $n\in\mathbb{N}$ by

$$\begin{aligned}\tilde{X}_0^{(n)}&=X_0,\\ \tilde{X}_{t_i^{(n)}}^{(n)}&=e^{A\left(t_i^{(n)}-t_{i-1}^{(n)}\right)}\tilde{X}_{t_{i-1}^{(n)}}^{(n)}+\beta h_n\left(\epsilon_i^{(n)},t_i^{(n)}-t_{i-1}^{(n)}\right)\quad\text{for all } i\in\mathbb{N}\text{ and}\\ \tilde{X}_t&=e^{A\left(t-t_{i-1}^{(n)}\right)}\tilde{X}_{t_{i-1}^{(n)}}^{(n)}\text{ for } t\in(t_{i-1}^{(n)},t_i^{(n)}),\ i\in\mathbb{N}.\end{aligned}$$

Below $0_{\mathbb{R}^d}$ denotes the zero in $\mathbb{R}^d$.

Then the joint process $(\tilde{X}^{(n)},L^{(n)},M^{(n)},S)$ satisfies the stochastic integral equation

$$\begin{pmatrix}\tilde{X}_t^{(n)}\\ L_t^{(n)}\\ M_t^{(n)}\\ S_t\end{pmatrix}=\begin{pmatrix}X_0\\ 0_{\mathbb{R}^d}\\ 0_{\mathbb{R}^m}\\ 0_{\mathbb{R}^m}\end{pmatrix}+\int_0^t F\left(\tilde{X}_{s-}^{(n)}\right)d\begin{pmatrix}L_s^{(n)}\\ M_s^{(n)}\\ S_s\end{pmatrix}. \tag{3.4.21}$$

with

$$F:\mathbb{R}^m\to M_{d+3m,d+2m}(\mathbb{R}),x\mapsto\begin{pmatrix}0_{M_{m,d}(\mathbb{R})} & \beta & Ax\\ I_d & 0_{M_{d,m}(\mathbb{R})} & 0_{M_{d,m}(\mathbb{R})}\\ 0_{M_{m,d}(\mathbb{R})} & I_m & 0_{M_m(\mathbb{R})}\\ 0_{M_{m,d}(\mathbb{R})} & 0_{M_m(\mathbb{R})} & I_m\end{pmatrix}.$$

Obviously F is globally Lipschitz and hence the stochastic integral equation (3.4.21) has a unique global strong solution. Here we are implicitly using the filtration $(\mathcal{F}_t^{(n)})_{t\in\mathbb{R}}$ for each $n\in\mathbb{N}$ as defined in Theorem 3.4.5.

Likewise (X,L,M,S) is the unique solution to the stochastic integral equation

$$\begin{pmatrix}X_t\\ L_t\\ M_t\\ S_t\end{pmatrix}=\begin{pmatrix}X_0\\ 0_{\mathbb{R}^d}\\ 0_{\mathbb{R}^m}\\ 0_{\mathbb{R}^m}\end{pmatrix}+\int_0^t F\left(X_{s-}\right)d\begin{pmatrix}L_s\\ M_s\\ S_s\end{pmatrix}. \tag{3.4.22}$$

We have that

$$\operatorname{plim}_{n\to\infty}d_{\mathbb{R}^{d+2m}}\left(\left((L^{(n)})^*,(M^{(n)})^*,S^*\right)^*,(L^*,M^*,S^*)^*\right)=0$$

follows from $\operatorname{plim}_{n\to\infty}d_{\mathbb{R}^{d+m}}\left(\mathbf{L}^{(n)},\mathbf{L}\right)=0$. Moreover, the fact that $\mathbf{L}^{(n)}$ has UCV implies that $\left((L^{(n)})^*,(M^{(n)})^*,S^*\right)^*$ has UCV. Hence, Kurtz and Protter (1991, Corollary 5.6) shows that

$$\operatorname{plim}_{n\to\infty}d_{\mathbb{R}^{d+3m}}\left(\left((\tilde{X}^{(n)})^*,(L^{(n)})^*,(M^{(n)})^*,S^*\right)^*,(X^*,L^*,M^*,S^*)^*\right)=0\,, \tag{3.4.23}$$

noting that Kurtz and Protter (1996, Example 8.2) ensures that F satisfies the necessary technical conditions (alternatively one can use Jacod and Shiryaev (2003, Theorem IX.6.9)). Setting $F(\tilde{X}^{(n)}) = \left(F(\tilde{X}_t^{(n)})\right)_{t\in\mathbb{R}^+}$ a continuity argument gives that

$$\text{plim}_{n\to\infty} d_{M_{d+3m,d+2m}(\mathbb{R})\times\mathbb{R}^{d+2m}}\left(\left[F(\tilde{X}^{(n)}),\left((L^{(n)})^*,(M^{(n)})^*,S^*\right)^*\right],\left[F(X),(L^*,M^*,S^*)^*\right]\right)=0$$

and therefore a combination of Theorems 7.7, 7.10, 7.11 of Kurtz and Protter (1996) implies that $\left((\tilde{X}^{(n)})^*,(L^{(n)})^*,(M^{(n)})^*,S^*\right)^*_{n\in\mathbb{N}}$ has UCV.

Next we observe that for all $T\in\mathbb{R}^+$

$$\sup_{t\le T}\left\|X_t^{(n)}-\tilde{X}_t^{(n)}\right\| \le \sup_{s\in[0,\delta^{(n)}]}\left\|e^{-As}-I_m\right\|_* \sup_{t\le T}\left\|\tilde{X}_t^{(n)}\right\|, \tag{3.4.24}$$

where $\|\cdot\|_*$ is the operator norm induced by $\|\cdot\|$. Since $\left((\tilde{X}^{(n)})^*,(L^{(n)})^*,(M^{(n)})^*,S^*\right)^*_{n\in\mathbb{N}}$ has UCV and thus $(\tilde{X}^{(n)})_{n\in\mathbb{N}}$ is uniformly tight (use Kurtz and Protter (1996, Theorem 7.6) and Jacod and Shiryaev (2003, Property 6.3, p. 377)), we have from the definition of uniform tightness (cf. Kurtz and Protter (1996, Definition 7.4)) that $\sup_{t\le T}\left\|\tilde{X}_t^{(n)}\right\|$ is stochastically bounded in $n\in\mathbb{N}$ for all $T\in\mathbb{R}^+$. Combining this with

$$\lim_{n\to\infty}\sup_{s\in[0,\delta^{(n)}]}\left\|e^{-As}-I_m\right\|_*=0$$

and (3.4.24) establishes that

$$X^{(n)}-\tilde{X}^{(n)}\to 0 \text{ in ucp as } n\to\infty. \tag{3.4.25}$$

This implies

$$\text{plim}_{n\to\infty} d_{\mathbb{R}^{d+2m}}\left(\left((X^{(n)})^*,(\tilde{X}^{(n)})^*,(L^{(n)})^*\right)^*,(X^*,X^*,L^*)^*\right)=0$$

and by a continuity argument

$$\text{plim}_{n\to\infty} d_{\mathbb{S}_d\times\mathbb{R}^{d+m}}\left(\left[Z^{(n)},\left((\tilde{X}^{(n)})^*,(L^{(n)})^*\right)^*\right],\left[Z,(X^*,L^*)^*\right]\right)=0 \tag{3.4.26}$$

with Z and $(Z^{(n)})_{n\in\mathbb{N}}$ defined by

$$\begin{aligned} Z_t^{(n)} &= \exp\left(\left(\mu+\text{vech}^{-1}\left(X_t^{(n)}\right)\right)/2\right),\\ Z_t &= \exp\left(\left(\mu+\text{vech}^{-1}\left(X_t\right)\right)/2\right) \text{ for all } t\in\mathbb{R}^+. \end{aligned}$$

Finally, we observe that

$$\begin{aligned} G_t^{(n)} &= G_0+\int_0^t Z_{s-}^{(n)}dL_s^{(n)}, \qquad \tilde{X}_t^{(n)} = X_0+\int_0^t I_m d\tilde{X}_s^{(n)} \text{ and}\\ G_t &= G_0+\int_0^t Z_{s-}dL_s, \qquad X_t = X_0+\int_0^t I_m dX_s \text{ for all } t\in\mathbb{R}^+. \end{aligned}$$

Therefore Kurtz and Protter (1991, Theorem 2.2) (or alternatively Jacod and Shiryaev (2003, Theorem VI.6.22)) shows that

$$\text{plim}_{n\to\infty} d_{\mathbb{R}^d\times\mathbb{R}^m}\left((G^{(n)},\tilde{X}^{(n)}),(G,X)\right)=0,$$

since $((\tilde{X}^{(n)})^*,(L^{(n)})^*)^*$ has UCV. Hence, (3.4.25) establishes

$$\text{plim}_{n\to\infty} d_{\mathbb{R}^d\times\mathbb{R}^m}\left((G^{(n)},X^{(n)}),(G,X)\right)=0.$$

□

In the following we consider one of our special choices for h and thereafter give some variants of our main Theorem 3.4.1.

Proposition 3.4.6. *Assume that h is given by* (3.3.7)*. Then h is Lipschitz and thereby uniformly continuous.*

Proof. To show that h as given in (3.3.7) is Lipschitz it suffices to show that

$$\mathbb{R}^d\to\mathbb{S}_d,\ \eta\mapsto(\eta\eta^*)^{1/2}=\frac{\eta\eta^*}{\|\eta\|_2}$$

is Lipschitz. But this is immediate from

$$\left|\frac{\eta_i\eta_j}{\|\eta\|_2}-\frac{\tilde{\eta}_i\tilde{\eta}_j}{\|\tilde{\eta}\|_2}\right|\le|\eta_i-\tilde{\eta}_i|+|\eta_j-\tilde{\eta}_j|+\|\eta-\tilde{\eta}\|\text{ for all }\eta,\tilde{\eta}\in\mathbb{R}^d,\ i,j\in\{1,2,\dots,d\}.$$

□

Theorem 3.4.7. *Let the set-up of Theorem 3.4.1 be given and assume that* $\text{var}(L_1)=I_d$*. Then in Theorem 3.4.1 the function h_n and the sequence of independent random variables $(\epsilon_i^{(n)})_{i\in\mathbb{N}}$ can for all $n\in\mathbb{N}$ be chosen such that $\epsilon_i^{(n)}=\left(\sqrt{t_i^{(n)}-t_{i-1}^{(n)}}\right)\zeta_i^{(n)}$ for all $i,n\in\mathbb{N}$ where $(\zeta_i^{(n)})_{i\in\mathbb{N}}$ is for all $n\in\mathbb{N}$ a sequence of independent d-dimensional random variables with* $\text{var}(\zeta_i^{(n)})=I_d$ *for all $i,n\in\mathbb{N}$.*

Moreover, if h is continuous h_n can be chosen such that the sequence of functions $h_n:\mathbb{R}^d\times\mathbb{R}^+\to\mathbb{R}^m$ satisfies

$$\lim_{n\to\infty}\left(\sup_{z\in K}\sup_{i\in\mathbb{N}}\left\{\left\|h_n\left(z,t_i^{(n)}-t_{i-1}^{(n)}\right)-h(z)\right\|\right\}\right)=0 \tag{3.4.27}$$

for all compact $K\subset\mathbb{R}^d$.

Proof. Let $\mathbf{L},L^{(n)},M^{(n)},\gamma_{L,i}^{(n)},\gamma_{M,i}^{(n)}$ be as in the proof of Theorem 3.4.1. We have that

$$V_i^{(n)}:=\text{var}\left(1_{(0,\infty)}(\tau_i^{(n)})\Delta L_{\tau_i^{(n)}}\right)=\frac{1-e^{-\nu_L(J^{(n)})(t_i^{(n)}-t_{i-1}^{(n)})}}{\nu_L(J^{(n)})}\int_{J^{(n)}}xx^*\nu_L(dx)$$
$$-\left(\frac{1-e^{-\nu_L(J^{(n)})(t_i^{(n)}-t_{i-1}^{(n)})}}{\nu_L(J^{(n)})}\right)^2\int_{J^{(n)}}x\nu_L(dx)\int_{J^{(n)}}x^*\nu_L(dx)$$

for all $i, n \in \mathbb{N}$. A series expansion shows

$$\lim_{n\to\infty} \sup_{i\in\mathbb{N}} \left\| \left(\frac{1 - e^{-\nu_L(J^{(n)})(t_i^{(n)} - t_{i-1}^{(n)})}}{\left(\sqrt{t_i^{(n)} - t_{i-1}^{(n)}}\right) \nu_L(J^{(n)})} \right)^2 \int_{J^{(n)}} x\nu_L(dx) \int_{J^{(n)}} x^*\nu_L(dx) \right\|_2 = 0. \tag{3.4.28}$$

Combining this with

$$\lim_{n\to\infty} \sup_{i\in\mathbb{N}} \left\| \frac{1 - e^{-\nu_L(J^{(n)})(t_i^{(n)} - t_{i-1}^{(n)})}}{\nu_L(J^{(n)})(t_i^{(n)} - t_{i-1}^{(n)})} \int_{J^{(n)}} xx^*\nu_L(dx) - I_d \right\|_2 \leq \lim_{n\to\infty} \left\| \int_{J^{(n)}} xx^*\nu_L(dx) - I_d \right\|_2$$
$$+ \lim_{n\to\infty} \left(\sum_{k=2}^{\infty} \frac{\left(\nu_L(J^{(n)})\delta^{(n)}\right)^{k-1}}{k!} \int_{J^{(n)}} \|x\|_2^2 \nu_L(dx) \right) = 0$$

establishes

$$\lim_{n\to\infty} \sup_{i\in\mathbb{N}} \left\| \frac{V_i^{(n)}}{t_i^{(n)} - t_{i-1}^{(n)}} - I_d \right\| = 0. \tag{3.4.29}$$

Hence, there exists a $N \in \mathbb{N}$ such that $V_i^{(n)} \in GL_d(\mathbb{R})$ for all $i \in \mathbb{N}$ and $n \geq N$. W.l.o.g. we assume $N = 1$. Then a continuity and compactness argument gives

$$\lim_{n\to\infty} \sup_{i\in\mathbb{N}} \left\| \left(V_i^{(n)}\right)^{-1/2} \sqrt{t_i^{(n)} - t_{i-1}^{(n)}} - I_d \right\| = 0. \tag{3.4.30}$$

Define for each $n \in \mathbb{N}$ the process $(\Sigma_t^{(n)})_{t\in\mathbb{R}^+}$ by

$$\Sigma_t^{(n)} = I_d \text{ for } t \in [0, t_1^{(n)}/2),$$
$$\Sigma_t^{(n)} = \left(V_i^{(n)}\right)^{-1/2} \sqrt{t_i^{(n)} - t_{i-1}^{(n)}} \text{ for } t \in \left[\left(t_i^{(n)} + t_{i-1}^{(n)}\right)/2, \left(t_i^{(n)} + t_{i+1}^{(n)}\right)/2 \right), \ i \in \mathbb{N}.$$

Since (3.4.30) implies that $(\Sigma_t^{(n)})_{t\in\mathbb{R}^+}$ converges uniformly to the identity, it follows that

$$\text{plim}_{n\to\infty} d_{M_{d+m}(\mathbb{R})\times\mathbb{R}^{d+m}} \left(\left(\begin{pmatrix} \Sigma^{(n)} & 0 \\ 0 & I_m \end{pmatrix}, \mathbf{L}^{(n)} \right), (I_{d+m}, \mathbf{L}) \right) = 0. \tag{3.4.31}$$

Setting

$$\tilde{L}_t^{(n)} = \int_0^t \Sigma_{s-}^{(n)} dL_s^{(n)}$$

we thus have from Kurtz and Protter (1991, Theorem 2.2) or Kurtz and Protter (1996, Theorem 7.10) that

$$\text{plim}_{n\to\infty} d_{\mathbb{R}^{d+m}} \left(\left((\tilde{L}^{(n)})^*, (M^{(n)})^* \right), (L^*, M^*)^* \right) = 0 \tag{3.4.32}$$

and from Kurtz and Protter (1996, Theorem 7.11) that $\left((\tilde{L}^{(n)})^*, (M^{(n)})^* \right)_{n\in\mathbb{N}}$ has UCV.

Setting

$$\zeta_i^{(n)} = \left(V_i^{(n)}\right)^{-1/2}\left(1_{(0,\infty)}(\tau_i^{(n)})\Delta L_{\tau_i^{(n)}} + \gamma_{L,i}^{(n)}\right),$$
$$\epsilon_i^{(n)} = \left(\sqrt{t_i^{(n)} - t_{i-1}^{(n)}}\right)\zeta_i^{(n)},$$
$$V_n(t) = \frac{1-e^{-\nu_L(J^{(n)})t}}{\nu_L(J^{(n)})}\int\limits_{J^{(n)}} xx^*\nu_L(dx) - \left(\frac{1-e^{-\nu_L(J^{(n)})t}}{\nu_L(J^{(n)})}\right)^2 \int\limits_{J^{(n)}} x\nu_L(dx)\int\limits_{J^{(n)}} x^*\nu_L(dx),$$
$$h_n : \mathbb{R}^d \times \mathbb{R}^+ \to \mathbb{R}^m,$$
$$(z,t) \mapsto h\left(\frac{(V_n(t))^{1/2}}{\sqrt{t}}z + \frac{1-e^{-\nu_L\left(J^{(n)}\right)t}}{\nu_L\left(J^{(n)}\right)}\int\limits_{J^{(n)}} x\nu_L(dx)\right) - \frac{1-e^{-\nu_L\left(J^{(n)}\right)t}}{\nu_L\left(J^{(n)}\right)}\int\limits_{J^{(n)}} h(x)\nu_L(dx)$$

for all $i, n \in \mathbb{N}$ it is easy to see that $(\epsilon_i^{(n)})_{i\in\mathbb{N}}$, $(\zeta_i^{(n)})_{i\in\mathbb{N}}$ and h_n have for all $n \in \mathbb{N}$ the claimed properties. Finally, noting that

$$\tilde{L}_t^{(n)} = \sum_{i\in\mathbb{N}:\, t_i^{(n)}\le t} \epsilon_i^{(n)} \text{ and}$$
$$M_t^{(n)} = \sum_{i\in\mathbb{N}:\, t_i^{(n)}\le t} h_n\left(\epsilon_i^{(n)}, t_i^{(n)} - t_{i-1}^{(n)}\right) \text{ for all } i, n \in \mathbb{N},\ t \in \mathbb{R}^+$$

the proof continues now as the proof of Theorem 3.4.1 with $\tilde{L}^{(n)}$ in the place of $L^{(n)}$. □

Proposition 3.4.8. *Let the set-up of Theorem 3.4.1 be given and assume that h is linear. Then $h_n(z,t) = h(z)$ can be chosen for all $n \in \mathbb{N}$.*

Proof. Obviously $h_n = h$ for h_n as defined in the proof of Theorem 3.4.1. □

4. Positive Definite Matrix Processes of Finite Variation[1]

4.1. Introduction

The theory of self-decomposability, as developed by Lévy, Urbanik, Sato, Jurek and Mason, and others, has turned out to be of substantial interest for stochastic modelling in finance, turbulence and other fields. See, for instance, Barndorff-Nielsen (1998a), Barndorff-Nielsen and Shephard (2001b) and Barndorff-Nielsen and Schmiegel (2004), where (positive) Lévy driven processes of Ornstein – Uhlenbeck type have a key role.

The focus of the present chapter is on stochastic differential equation representations of square roots of positive definite matrix processes of Lévy or Ornstein – Uhlenbeck type. Such representations are, in particular, of interest in connection with the general theory of multipower variation, cf. Barndorff-Nielsen, Graversen, Jacod, Podolskij and Shephard (2006) and Barndorff-Nielsen, Graversen, Jacod and Shephard (2006).

In the present literature matrix-valued stochastic processes are not commonly used to model multivariate phenomena (see, for instance, the short discussion on multivariate stochastic volatility models at the end of Section 4.4). Our introduction of positive-definite Ornstein-Uhlenbeck processes and the discussion of the representations of square (and other) roots shows that matrix-valued models of considerable generality can be defined in a natural way and univariate results can very often be generalized by using notions and results from matrix analysis. Furthermore, several results of general interest regarding matrix-valued processes (semi-martingales) and matrix analysis are obtained, as we proceed.

This chapter is organized as follows. Section 4.2 establishes some notation, and in Section 4.3 we present a convenient version of Itô's formula for processes of finite variation. In Section 4.4 we introduce positive definite processes of Ornstein – Uhlenbeck type (OU processes), using the concept of matrix subordinators discussed by Barndorff-Nielsen and Pérez-Abreu (2007). The question of establishing tractable stochastic differential equations for roots of positive definite matrix processes is then addressed in Section 4.5, and in Section 4.6 the results are applied to the case of OU processes.

4.2. Notation

Throughout this chapter we write $\mathbb{R}^+$ for the positive real numbers including zero and we denote the set of real $m \times n$ matrices by $M_{m,n}(\mathbb{R})$. If $m = n$ we simply write $M_n(\mathbb{R})$ and denote the group of invertible $n \times n$ matrices by $GL_n(\mathbb{R})$, the linear subspace of symmetric

[1] The contents of this chapter appeared in Barndorff-Nielsen, O. E. and Stelzer, R. (2007), Positive Definite Matrix Processes of Finite Variation, *Probab. Math. Statist.*, **27**(1), 3-43 (Special issue in the memory of K. Urbanik)

matrices by $\mathbb{S}_n$, the (closed) positive semi-definite cone by $\mathbb{S}_n^+$ and the open (in $\mathbb{S}_n$) positive definite cone by $\mathbb{S}_n^{++}$. I_n stands for the $n \times n$ identity matrix and $\sigma(A)$ for the spectrum (the set of all eigenvalues) of a matrix $A \in M_n(\mathbb{R})$. The natural ordering on the symmetric $n \times n$ matrices will be denoted by $\leq$, i.e. for $A, B \in \mathbb{S}_n$ we have that $A \leq B$, if and only if $B - A \in \mathbb{S}_n^+$. The tensor (Kronecker) product of two matrices A, B is written as $A \otimes B$. vec denotes the well-known vectorisation operator that maps the $n \times n$ matrices to $\mathbb{R}^{n^2}$ by stacking the columns of the matrices below one another. Finally, A^* is the adjoint (transposed) of a matrix $A \in M_n(\mathbb{R})$.

For a matrix A we denote by A_{ij} the element in the i-th row and j-th column and this notation is extended to processes in a natural way.

Norms of vectors or matrices are denoted by $\|\cdot\|$. If the specific norm is not specified, then it is irrelevant which particular norm is used.

Regarding all random variables and processes we assume that they are defined on a given appropriate filtered probability space $(\Omega, \mathcal{F}, P, (\mathcal{F}_t))$ satisfying the usual hypotheses. With random functions we usually do not state the dependence on $\omega \in \Omega$ explicitly.

Furthermore, we employ an intuitive notation with respect to the (stochastic) integration with matrix-valued integrators referring to any of the standard texts (e.g. Protter (2004)) for a comprehensive treatment of the theory of stochastic integration. Let $(A_t)_{t\in\mathbb{R}^+}$ in $M_{m,n}(\mathbb{R})$, $(B_t)_{t\in\mathbb{R}^+}$ in $M_{r,s}(\mathbb{R})$ be càdlàg and adapted processes and $(L_t)_{t\in\mathbb{R}^+}$ in $M_{n,r}(\mathbb{R})$ be a semimartingale. Then we denote by $\int_0^t A_{s-} dL_s B_{s-}$ the matrix C_t in $M_{m,s}(\mathbb{R})$ which has ij-th element $C_{ij,t} = \sum_{k=1}^n \sum_{l=1}^r \int_0^t A_{ik,s-} B_{lj,s-} dL_{kl,s}$. Equivalently such an integral can be understood in the sense of Métivier and Pellaumail (1980b), resp. Métivier (1982), by identifying it with the integral $\int_0^t \mathbf{A}_{s-} dL_s$ with $\mathbf{A}_t$ being for each fixed t the linear operator $M_{n,r}(\mathbb{R}) \to M_{m,s}(\mathbb{R})$, $X \mapsto A_t X B_t$. Moreover, we always denote by $\int_a^b$ with $a \in \mathbb{R} \cup \{-\infty\}, b \in \mathbb{R}$ the integral over the half-open interval $(a, b]$ for notational convenience. If $b = \infty$, the integral is understood to be over (a, b).

If $X = (X_t)_{t\in\mathbb{R}^+}$ is some stochastic process and f some function, then $f(X)$ denotes the process $(f(X_t))_{t\in\mathbb{R}^+}$. Extending this notation we also write $\sqrt{X}$ when taking the square root and X^r when taking the r-th power, for instance.

4.3. Itô formulae for finite variation processes in open sets

In this section we provide a univariate and a multivariate version of the Itô formula from stochastic analysis, which is especially suitable for the purposes of this chapter. Actually, our version is a consequence of standard results, but not given in the usual references. Closely related versions for processes taking values in $\mathbb{R}^d$ instead of an open subset C can be found in Protter (2004, Theorem II.31) or Rogers and Williams (2000, pp. 28-29), for example, and for processes in an open subset with not necessarily finite variation in Jacod (1979, Theorem 2.54).

As we are analysing stochastic processes in general open subsets C of $\mathbb{R}^d, M_d(\mathbb{R})$ or $\mathbb{S}_d$, we need an appropriate assumption that the process stays within C and does not hit the boundary, since this causes problems in general. To describe "good" behaviour we thus introduce "local boundedness within C". If C is the whole space it is the same as "local boundedness".

Definition 4.3.1. *Let $(V, \|\cdot\|_V)$ be either $\mathbb{R}^d, M_d(\mathbb{R})$ or $\mathbb{S}_d$ with $d \in \mathbb{N}$ and equipped with the norm $\|\cdot\|_V$. Let $a \in V$ and let $X = (X_t)_{t\in\mathbb{R}^+}$ be a V-valued stochastic process. We say that X is* locally bounded away from *a if there exists a sequence of stopping times $(T_n)_{n\in\mathbb{N}}$ increasing to infinity almost surely and a real sequence $(d_n)_{n\in\mathbb{N}}$ with $d_n > 0$ for all $n \in \mathbb{N}$ such that $\|X_t - a\|_V \geq d_n$ for all $0 \leq t < T_n$.*

Likewise, we say for some open set $C \subset V$ that the process X is locally bounded within *C if there exists a sequence of stopping times $(T_n)_{n\in\mathbb{N}}$ increasing to infinity almost surely and a sequence of compact convex subsets $D_n \subset C$ with $D_n \subset D_{n+1}$ $\forall n \in \mathbb{N}$ such that $X_t \in D_n$ for all $0 \leq t < T_n$.*

Obviously, if a process is locally bounded away from some a or is locally bounded within some C in one norm, then the same holds for all other norms. We will see in the following that these definitions play a central role for our Itô formulae and that they hold for many processes.

Proposition 4.3.2 (Univariate Itô formula for processes of finite variation)**.** *Let $(X_t)_{t\in\mathbb{R}^+}$ be a càdlàg process of finite variation (thus a semi-martingale) with associated jump measure μ_X on $(\mathbb{R}^+ \times \mathbb{R}\backslash\{0\}, \mathcal{B}(\mathbb{R}^+ \times \mathbb{R}\backslash\{0\}))$ (see e.g. Jacod and Shiryaev (2003, Proposition II.1.16)) and let $f : C \to \mathbb{R}$ be continuously differentiable, where C is some open interval $C = (a, b)$ with $a, b \in \mathbb{R} \cup \{\pm\infty\}$, $a < b$. Assume that $(X_t)_{t\in\mathbb{R}^+}$ is locally bounded within C. Then the process X as well as its left limit process $(X_{t-})_{t\in\mathbb{R}^+}$ take values in C at all times $t \in \mathbb{R}^+$, the integral $\int_0^t \int_{\mathbb{R}\backslash\{0\}} (f(X_{s-} + x) - f(X_{s-}))\mu_X(ds, dx)$ exists a.s. for all $t \in \mathbb{R}$ and*

$$f(X_t) = f(X_0) + \int_0^t f'(X_{s-})dX_s^c + \int_0^t \int_{\mathbb{R}\backslash\{0\}} (f(X_{s-} + x) - f(X_{s-}))\mu_X(ds, dx),$$

where $X_t^c = X_t - \int_0^t \int_{\mathbb{R}\backslash\{0\}} x\mu_X(ds, dx)$ is the continuous part of X.
(Strictly speaking $f(X_{s-}+x)$ is not defined for all $x \in \mathbb{R}$, as f is only defined on C. But our assumptions assure that μ_X is concentrated on those x for which $X_{s-} + x \in C$. Therefore we can simply continue f arbitrarily outside of C.)

Proof. As X is locally bounded within C, the process X cannot get arbitrarily close to the boundary of C in finite time and hence X_t and X_{t-} are in C at all times $t \in \mathbb{R}^+$.

Obviously, $\int_0^t \int_{\mathbb{R}\backslash\{0\}} (f(X_{s-} + x) - f(X_{s-}))\mu_X(ds, dx) = \sum_{0<s\leq t} \Delta f(X_s)$. That X_t is locally bounded within C implies the existence of compact intervals $D_n \subset C$ such that $X_t \in D_n$ for all $t \in [0, T_n)$ for some sequence $(T_n)_{n\in\mathbb{N}}$ of stopping times increasing to infinity a.s. However, f' is bounded on D_n, say by c_n, and the mean value theorem gives us that $\Delta f(X_s) = f(X_s) - f(X_{s-}) = f'(\zeta_s)(X_s - X_{s-}) = f'(\zeta_s)\Delta X_s$ with $\zeta_s \in D_n$. Therefore, $\int_0^t \int_{\mathbb{R}\backslash\{0\}} |f(X_{s-} + x) - f(X_{s-})|\mu_X(ds, dx) = \sum_{0<s\leq t} |\Delta f(X_s)| \leq c_n \sum_{0<s\leq t} |\Delta X_s|$ for all $t \in [0, T_n)$, which is finite due to the finite variation of X. Thus the almost sure existence of the integral is shown.

The standard Itô formula (see Bichteler (2002, Theorem 3.9.1 together with Proposition 3.10.10) for an appropriate version) gives

$$\begin{aligned} f(X_t) = f(X_0) &+ \int_0^t f'(X_{s-})dX_s \\ &+ \int_0^t \int_{\mathbb{R}\backslash\{0\}} (f(X_{s-} + x) - f(X_{s-}) - f'(X_{s-})x)\mu_X(ds, dx), \end{aligned}$$

on observing that, since X is a finite variation process, we can move from a twice continuously differentiable f to an only once continuously differentiable one, as in Protter (2004, Theorem II.31). Noting further that

$$\int_0^t f'(X_{s-})dX_s = \int_0^t f'(X_{s-})dX_s^c + \int_0^t \int_{\mathbb{R}\setminus\{0\}} f'(X_{s-})x\mu_X(ds,dx)$$

and that the integral

$$\int_0^t \int_{\mathbb{R}\setminus\{0\}} (f(X_{s-}+x) - f(X_{s-}))\mu_X(ds,dx)$$

exists, we obtain:

$$f(X_t) = f(X_0) + \int_0^t f'(X_{s-})dX_s^c + \int_0^t \int_{\mathbb{R}\setminus\{0\}} (f(X_{s-}+x) - f(X_{s-}))\mu_X(ds,dx). \qquad \square$$

Remark 4.3.3. *a) The assumption that X remains locally bounded within C ensures that $f'(X)$ is locally bounded. This reflects the boundedness of the derivative needed in the proof of Protter (2004, Theorem I.54), which is a special case of the above result.*

b) It is straightforward to see that X is locally bounded within $C = (a,b)$ if and only if X_t is in C at all times and locally bounded away from both a and b, where for $a = -\infty$ or $b = \infty$ this has to be understood as meaning locally bounded. Recall in this context that any finite variation process is locally bounded.

In the multivariate version we use the notion of (total) differentials, sometimes also called Fréchet differentials (see Rudin (1976, Chapter 9), or Bhatia (1997, Section X.4) for an overview focusing on the matrix case), rather than partial derivatives for notational convenience. Recall, however, that a function is continuously differentiable if and only if all partial derivatives exist and are continuous, and that the derivative, which is a linear operator, simply has the partial derivatives as entries. The derivative of a function f at a point x is denoted by $Df(x)$. In particular, we have the following multivariate version of Proposition 4.3.2. We state it only for processes in $\mathbb{R}^d$, but it should be obvious that $\mathbb{R}^d$ can be replaced by $M_d(\mathbb{R})$ or $\mathbb{S}_d$.

Proposition 4.3.4 (Multivariate Itô formula for processes of finite variation)**.** *Let $(X_t)_{t\in\mathbb{R}^+}$ be a càdlàg $\mathbb{R}^d$-valued process of finite variation (thus a semi-martingale) with associated jump measure μ_X on $\left(\mathbb{R}^+\times\mathbb{R}^d\setminus\{0\}, \mathcal{B}\left(\mathbb{R}^+\times\mathbb{R}^d\setminus\{0\}\right)\right)$ and let $f: C\to\mathbb{R}^m$ be continuously differentiable, where $C\subseteq\mathbb{R}^d$ is an open set. Assume that the process $(X_t)_{t\in\mathbb{R}^+}$ is locally bounded within C. Then the process X as well as its left limit process $(X_{t-})_{t\in\mathbb{R}^+}$ take values in C at all times $t\in\mathbb{R}^+$, the integral $\int_0^t\int_{\mathbb{R}^d\setminus\{0\}}(f(X_{s-}+x)-f(X_{s-}))\mu_X(ds,dx)$ exists a.s. for all $t\in\mathbb{R}$ and*

$$f(X_t) = f(X_0) + \int_0^t Df(X_{s-})dX_s^c + \int_0^t\int_{\mathbb{R}^d\setminus\{0\}} (f(X_{s-}+x) - f(X_{s-}))\mu_X(ds,dx),$$

where $X_t^c = X_t - \int_0^t\int_{\mathbb{R}^d\setminus\{0\}} x\mu_X(ds,dx)$ is the continuous part of X.

Proof. The proof is a mere multivariate rephrasing of the one for Proposition 4.3.2 using an appropriate general multidimensional version of Itô's formula (e.g. Bichteler (2002, Proposition 3.10.10), Métivier (1982, Theorem 27.2) or Protter (2004, Theorem 7.33)) and standard results from multivariate calculus. $\square$

4.4. Positive semi-definite matrix processes of OU type

In this section we briefly review one-dimensional processes of Ornstein-Uhlenbeck (OU) type (cf. Applebaum (2004), Cont and Tankov (2004) or Barndorff-Nielsen and Shephard (2001b, 2007) among many others) and then introduce Ornstein-Uhlenbeck processes taking values in the positive semi-definite matrices. For the necessary background on Lévy processes see Protter (2004, Section I.4) or Sato (1999).

In univariate financial modelling, it has become popular in recent years to specify the variance process σ^2 as an Ornstein-Uhlenbeck process (see in particular the works of Barndorff-Nielsen and Shephard). We assume given a Lévy process $(L_t)_{t\in\mathbb{R}^+}$ and consider the SDE

$$d\sigma_t^2 = -\lambda\sigma_{t-}^2 dt + dL_t \tag{4.4.1}$$

with some $\lambda \in \mathbb{R}$ and initial value $\sigma_0^2 \in \mathbb{R}$. The solution can be shown to be

$$\sigma_t^2 = e^{-\lambda t}\sigma_0^2 + \int_0^t e^{-\lambda(t-s)} dL_s \tag{4.4.2}$$

and is referred to as an OU type process. Note that for univariate OU type processes one often applies a time transformation on the Lévy process and then has $dL_{\lambda s}$ instead of dL_s above, but this is not possible in the multivariate case below. Provided the Lévy process L is a subordinator (a.s. non-decreasing Lévy process) and $\sigma_0^2 \geq 0$, the unique strong solution σ^2 is positive and thus can be used as a variance process. After extending the Lévy process to one, $(L_t)_{t\in\mathbb{R}}$, living on the whole real line in the usual way, one can show that (4.4.1) has a unique stationary solution given by

$$\sigma_t^2 = \int_{-\infty}^t e^{-\lambda(t-s)} dL_s$$

provided $\lambda > 0$ and the Lévy process has a finite logarithmic moment, i.e. $E(\log^+(L_1)) < \infty$.

There is a vast literature concerning the extension of OU processes to $\mathbb{R}^d$-valued processes (for instance, Chojnowska-Michalik (1987), Jurek and Mason (1993) or Sato and Yamazato (1984)). By identifying $M_d(\mathbb{R})$ with $\mathbb{R}^{d^2}$ one immediately obtains matrix valued processes. So for a given Lévy process $(L_t)_{t\in\mathbb{R}}$ with values in $M_d(\mathbb{R})$ and a linear operator $\mathbf{A} : M_d(\mathbb{R}) \to M_d(\mathbb{R})$ we call some solution to the SDE

$$dX_t = \mathbf{A}X_{t-}dt + dL_t \tag{4.4.3}$$

a (matrix-valued) process of Ornstein-Uhlenbeck type.

As in the univariate case one can show that for some given initial value X_0 the solution is unique and given by

$$X_t = e^{\mathbf{A}t}X_0 + \int_0^t e^{\mathbf{A}(t-s)} dL_s. \tag{4.4.4}$$

Provided $E(\log^+ \|L_1\|) < \infty$ and $\sigma(\mathbf{A}) \in (-\infty, 0) + i\mathbb{R}$, there exists a unique stationary solution given by

$$X_t = \int_{-\infty}^t e^{\mathbf{A}(t-s)} dL_s.$$

In order to obtain positive semi-definite Ornstein-Uhlenbeck processes we need to consider matrix subordinators as driving Lévy processes. An $M_d(\mathbb{R})$-valued Lévy process L is called "matrix subordinator", if $L_t - L_s \in \mathbb{S}_d^+$ a.s. for all $t \geq s$, see Barndorff-Nielsen and Pérez-Abreu (2002, 2007), Rocha-Arteaga (2006) and the references therein for further details.

Proposition 4.4.1. *Let L be a matrix subordinator, assume that the linear operator $\mathbf{A}$ satisfies $\exp(\mathbf{A}t)(\mathbb{S}_d^+) \subseteq \mathbb{S}_d^+$ for all $t \in \mathbb{R}^+$ and let $X_0 \in \mathbb{S}_d^+$. Then the Ornstein-Uhlenbeck process $(X_t)_{t\in\mathbb{R}^+}$ with initial value X_0 satisfying (4.4.3) takes only values in $\mathbb{S}_d^+$.*

If $E(\log^+ \|L_1\|) < \infty$ and $\sigma(\mathbf{A}) \in (-\infty, 0) + i\mathbb{R}$, then the unique stationary solution $(X_t)_{t\in\mathbb{R}}$ to (4.4.3) takes values in $\mathbb{S}_d^+$ only.

Proof. The first term $e^{\mathbf{A}t}X_0$ in (4.4.4) is obviously positive semi-definite for all $t \in \mathbb{R}^+$ due to the assumption on $\mathbf{A}$. Approximating the integral $\int_0^t e^{\mathbf{A}(t-s)}dL_s$ by sums in the usual way, shows that also the second term is positive semi-definite, since all approximating sums are in $\mathbb{S}_d^+$ due to the assumption on $\mathbf{A}$ and the $\mathbb{S}_d^+$-increasingness of a Lévy subordinator.

The very same argument implies the positive semi-definiteness of the unique stationary solution. □

An important question arises now, namely, which linear operators $\mathbf{A}$ can one actually take to obtain both a unique stationary solution and ensure positive semi-definiteness. The condition $\exp(\mathbf{A}t)(\mathbb{S}_d^+) \subseteq \mathbb{S}_d^+$ means that for all $t \in \mathbb{R}^+$ the exponential operator $\exp(\mathbf{A}t)$ has to preserve positive semi-definiteness. So one seems to need to know first which linear operators on $M_d(\mathbb{R})$ preserve positive semi-definiteness. This problem has been studied for a long time in linear algebra in connection with the general topic "Linear Preserver Problems" (see, for instance, the overview articles Li and Pierce (2001) and Pierce, Lim, Loewy, Li, Tsing, McDonald and Beasley (1992)). We have the following:

Proposition 4.4.2. *Let $\mathbf{A} : \mathbb{S}_d \to \mathbb{S}_d$ be a linear operator. Then $\mathbf{A}(\mathbb{S}_d^+) = \mathbb{S}_d^+$, if and only if there exists a matrix $B \in GL_d(\mathbb{R})$ such that $\mathbf{A}$ can be represented as $X \mapsto BXB^*$.*

Proof. This was initially proved in Schneider (1965). A more general proof in a Hilbert space context may be found in Li, Rodman and Semrl (2003). □

Remark 4.4.3. *No explicit characterization of the linear operators mapping $\mathbb{S}_d^+$ into $\mathbb{S}_d^+$, i.e. $\mathbf{A}(\mathbb{S}_d^+) \subseteq \mathbb{S}_d^+$, is known for general dimension d.*

Naturally, all linear maps on $\mathbb{S}_d$ can be extended to mappings on $M_d(\mathbb{R})$. From this linear algebraic result we obtain the following result, introducing the linear operators preserving positive semi-definiteness which we shall employ.

Proposition 4.4.4. *Assume the operator $\mathbf{A} : M_d(\mathbb{R}) \to M_d(\mathbb{R})$ is representable as $X \mapsto AX + XA^*$ for some $A \in M_d(\mathbb{R})$. Then $e^{\mathbf{A}t}$ has the representation $X \mapsto e^{At}Xe^{A^*t}$ and $e^{\mathbf{A}t}(\mathbb{S}_d^+) = \mathbb{S}_d^+$ for all $t \in \mathbb{R}$.*

Proof. $e^{\mathbf{A}t}X = e^{At}Xe^{A^*t}$ for all $X \in M_d(\mathbb{R})$ follows from Horn and Johnson (1991, pp. 255 and 440) and $e^{\mathbf{A}t}(\mathbb{S}_d^+) = \mathbb{S}_d^+$ for all $t \in \mathbb{R}$ is then implied by Proposition 4.4.2, since e^B is invertible for any matrix $B \in M_d(\mathbb{R})$. □

That all linear operators $\mathbf{A}$ on $\mathbb{S}_d$ satisfying $e^{\mathbf{A}t}(\mathbb{S}_d^+) = \mathbb{S}_d^+$ for all $t \in \mathbb{R}$ are of the above form is shown in Appendix A.

Note the close relation of this kind of operators to Kronecker sums and the so-called "Lyapunov equation" (see Horn and Johnson (1991, Ch. 4)). For a linear operator $\mathbf{A}$ of the type specified in Proposition 4.4.4 formula (4.4.3) becomes

$$dX_t = (AX_{t-} + X_{t-}A^*)dt + dL_t \tag{4.4.5}$$

and the solution is

$$X_t = e^{At}X_0e^{A^*t} + \int_0^t e^{A(t-s)}dL_se^{A^*(t-s)}. \tag{4.4.6}$$

Confer also Horn and Johnson (1991, p. 440) for a related deterministic differential equation.

Using the vec transformation and Horn and Johnson (1991, Theorem 4.4.5) we see that $\sigma(\mathbf{A}) = \sigma(A)+\sigma(A)$, where the addition of two sets $A, B \subseteq \mathbb{R}$ is defined by $A+B = \{a+b : a \in A, b \in B\}$. Thus

Theorem 4.4.5. *Let $(L_t)_{t\in\mathbb{R}}$ be a matrix subordinator with $E(\log^+ \|L_1\|) < \infty$ and $A \in M_d(\mathbb{R})$ such that $\sigma(A) \subset (-\infty, 0)+i\mathbb{R}$. Then the stochastic differential equation of Ornstein-Uhlenbeck type*

$$dX_t = (AX_{t-} + X_{t-}A^*)dt + dL_t$$

has a unique stationary solution

$$X_t = \int_{-\infty}^t e^{A(t-s)}dL_se^{A^*(t-s)}$$

or, in vectorial representation,

$$\mathrm{vec}(X_t) = \int_{-\infty}^t e^{(I_d\otimes A+A\otimes I_d)(t-s)}d\mathrm{vec}(L_s).$$

Moreover, $X_t \in \mathbb{S}_d^+$ for all $t \in \mathbb{R}$.

Recall from Barndorff-Nielsen and Pérez-Abreu (2007) that any matrix subordinator $(L_t)_{t\in\mathbb{R}}$ has paths of finite variation and can be represented as

$$L_t = \gamma t + \int_0^t \int_{\mathbb{S}_d^+\setminus\{0\}} x\mu(ds, dx) \tag{4.4.7}$$

where $\gamma \in \mathbb{S}_d^+$ is a deterministic drift and $\mu(ds, dx)$ an extended Poisson random measure on $\mathbb{R}^+ \times \mathbb{S}_d^+$ (regarding the definitions of random measures and the integration theory with respect to them we refer to Jacod and Shiryaev (2003, Section II.1)). Observe in particular that the integral exists without compensating. Moreover, the expectation of μ factorises, i.e. $E(\mu(ds, dx)) = Leb(ds)\nu(dx)$, Leb denoting the Lebesgue measure and ν the Lévy measure of L. The above equation (4.4.7) can be restated in a differential manner as

$$dL_t = \gamma dt + \int_{\mathbb{S}_d^+\setminus\{0\}} x\mu(dt, dx). \tag{4.4.8}$$

The obvious extension of this to a Lévy process $(L_t)_{t\in\mathbb{R}}$ having been started in the infinite past gives another representation of the above stationary OU process.

Proposition 4.4.6. *The positive semi-definite Ornstein-Uhlenbeck process X defined in Theorem 4.4.5 can equivalently be represented as*

$$\begin{aligned} X_t &= \int_{-\infty}^{t} \int_{\mathbb{S}_d^+\setminus\{0\}} e^{A(t-s)} x e^{A^*(t-s)} \mu(ds, dx) + \int_{-\infty}^{t} e^{A(t-s)} \gamma e^{A^*(t-s)} ds \\ &= \int_{-\infty}^{t} \int_{\mathbb{S}_d^+\setminus\{0\}} e^{A(t-s)} x e^{A^*(t-s)} \mu(ds, dx) - B^{-1}\gamma \end{aligned}$$

where B^{-1} is the inverse of the linear operator $B : M_d(\mathbb{R}) \to M_d(\mathbb{R})$, $X \mapsto AX + XA^*$ which can be represented as $\mathrm{vec}^{-1} \circ ((I_d \otimes A) + (A \otimes I_d))^{-1} \circ \mathrm{vec}$.

Proof. The invertibility of B and the positive semi-definiteness of $-B^{-1}\gamma$ follow immediately from the standard theory on the Lyapunov equations (Horn and Johnson (1991, Th. 2.2.3, 4.4.7). Now only the second equality remains to be shown, but this is immediate as $-B^{-1}\frac{d}{ds} e^{A(t-s)} \gamma e^{A^*(t-s)} = e^{A(t-s)} \gamma e^{A^*(t-s)}$ and $\lim_{s\to-\infty} e^{A(t-s)} = 0$. □

The next proposition provides a characterization of the stationary distribution. To this end observe that $\mathrm{tr}(XY^*)$ (with $X, Y \in M_d(\mathbb{R})$ and tr denoting the usual trace functional) defines a scalar product on $M_d(\mathbb{R})$. Moreover, the vec operator is a Hilbert space isometry between $M_d(\mathbb{R})$ equipped with this scalar product and $\mathbb{R}^{d^2}$ with the usual Euclidean scalar product. This, in particular, implies that the driving Lévy process L at time t has characteristic function (cf. also Barndorff-Nielsen and Pérez-Abreu (2007))

$$\mu_{L_t}(Z) = \exp\left(it\mathrm{tr}(\gamma Z) + t \int_{\mathbb{S}_d^+\setminus\{0\}} (e^{i\mathrm{tr}(XZ)} - 1)\nu(dX) \right),\ Z \in \mathbb{S}_d. \tag{4.4.9}$$

Proposition 4.4.7. *The stationary distribution of the positive semi-definite Ornstein-Uhlenbeck process X is infinitely divisible with characteristic function*

$$\hat{\mu}_X(Z) = \exp\left(i\mathrm{tr}(\gamma_X Z) + \int_{\mathbb{S}_d^+\setminus\{0\}} (e^{i\mathrm{tr}(YZ)} - 1)\nu_X(dY) \right),\ Z \in \mathbb{S}_d \tag{4.4.10}$$

where

$$\gamma_X = -B^{-1}\gamma$$

with B defined as in Proposition 4.4.6 and

$$\nu_X(E) = \int_0^\infty \int_{\mathbb{S}_d^+\setminus\{0\}} I_E(e^{As} x e^{A^*s}) \nu(dx) ds$$

for all Borel sets E in $\mathbb{S}_d^+\setminus\{0\}$.

Assume that the driving Lévy process is square-integrable. Then the second order moment structure is given by

$$E(X_t) = \gamma_X - B^{-1} \int_{\mathbb{S}_d^+\setminus\{0\}} y\nu(dy) = -B^{-1}E(L_1) \tag{4.4.11}$$

$$\begin{aligned} Var(\mathrm{vec}(X_t)) &= \int_0^\infty e^{(A\otimes I_d + I_d \otimes A)t} Var(\mathrm{vec}(L_1)) e^{(A^*\otimes I_d + I_d \otimes A^*)t} dt \\ &= -\mathscr{B}^{-1} Var(\mathrm{vec}(L_1)) \end{aligned} \tag{4.4.12}$$

$$Cov(\mathrm{vec}(X_{t+h}), \mathrm{vec}(X_t)) = e^{(A\otimes I_d + I_d \otimes A)h} Var(\mathrm{vec}(X_t)), \tag{4.4.13}$$

where $t \in \mathbb{R}$ *and* $h \in \mathbb{R}^+$ *and* $\mathscr{B} : M_{d^2}(\mathbb{R}) \to M_{d^2}(\mathbb{R})$, $X \mapsto (A \otimes I_d + I_d \otimes A)X + X(A^* \otimes I_d + I_d \otimes A^*)$. *The linear operator* $\mathscr{B}$ *can be represented as*

$$\mathrm{vec}^{-1} \circ ((I_{d^2} \otimes (A \otimes I_d + I_d \otimes A)) + ((A \otimes I_d + I_d \otimes A) \otimes I_{d^2})) \circ \mathrm{vec}.$$

Above the vec operator is used, as this clarifies the order of the elements of the (co)variance matrix.

Proof. The characteristic function is standard, cf. Barndorff-Nielsen, Pedersen and Sato (2001, p. 178) for instance. Regarding (4.4.11) a general result for infinitely divisible distributions implies that $E(X_t) = \gamma_X + \int_{\mathbb{S}_d^+} y\nu_X(dy)$. Using the explicit representation for ν_X and evaluating the integral as in the proof of the last proposition immediately establishes (4.4.11). The proof of the first equality in (4.4.12) and of (4.4.13) is standard, see e.g. Proposition 2.3.13 in Chapter 2, and the second equality in (4.4.12) follows by an explicit integration as before. □

Remark 4.4.8. *In the existing literature for* $\mathbb{R}^d$*-valued processes only the analogue to the first equality in* (4.4.12) *is stated and an identity is given that becomes*

$$-Var(\mathrm{vec}(L_1)) = (A \otimes I_d + I_d \otimes A)Var(\mathrm{vec}(X_t)) + Var(\mathrm{vec}(X_t))(A^* \otimes I_d + I_d \otimes A^*)$$

in our case. That identity is, of course, equivalent to our second equality in (4.4.12)*, but usually obtained by a very different approach (cf. Arató (1982), for instance). Our version involving* $\mathscr{B}^{-1}$ *stresses that the variance can be calculated by solving a standard linear equation and fits in nicely, as inverse operators of this type appear in many of our results.*

Moreover, conditions ensuring that the stationary OU type process X is almost surely strictly positive definite can be obtained.

Theorem 4.4.9. *If* $\gamma \in \mathbb{S}_d^{++}$ *or* $\nu(\mathbb{S}_d^{++}) > 0$*, then the stationary distribution* P_X *of* X *is concentrated on* $\mathbb{S}_d^{++}$*, i.e.* $P_X(\mathbb{S}_d^{++}) = 1$.

Proof. From Proposition 4.4.6 and its proof we have $X_t \geq -B^{-1}\gamma$. In the case $\gamma \in \mathbb{S}_d^{++}$ this proves the theorem immediately, as then $-B^{-1}\gamma$ is strictly positive definite due to Horn and Johnson (1991, Theorem 2.2.3).

Assume now that $\nu(\mathbb{S}_d^{++}) > 0$. From Proposition 4.4.6 we know that

$$X_0 \geq \sum_{-\infty < s \leq 0} e^{-As}\Delta(L_s)e^{-A^*s} \stackrel{d}{=} \sum_{0 \leq s < \infty} e^{As}\Delta(L_s)e^{A^*s}.$$

Since $Z \mapsto e^{As}Ze^{A^*s}$ preserves positive definiteness for all $s \in \mathbb{R}$, it is obviously sufficient to show that $(L_s)_{s\in\mathbb{R}^+}$ has at least one jump that is positive definite. Choose now $\epsilon > 0$ such that $\nu(\mathbb{S}_d^{++} \cap \{x \in \mathbb{S}_d^+ : \|x\| \geq \epsilon\}) > 0$. Then the process $L_{\epsilon,s} := \sum_{0\leq s\leq t} 1_{\{x\in\mathbb{S}_d^+ : \|x\|\geq\epsilon\}}(\Delta L_s)\Delta L_s$ is a Lévy process with Lévy measure $\nu_\epsilon(\cdot) = \nu(\cdot \cap \{x \in \mathbb{S}_d^+ : \|x\| \geq \epsilon\})$, where we denoted by $1_M(\cdot)$ the indicator function of a set M. L_ϵ is obviously a compound Poisson process and the probability that a jump of L_ϵ is in $\mathbb{S}_d^+ \backslash \mathbb{S}_d^{++}$ is given by $q := \nu_\epsilon(\mathbb{S}_d^+ \backslash \mathbb{S}_d^{++})/\nu_\epsilon(\mathbb{S}_d^+) < 1$. As the individual jump sizes and the jump times are independent and $(L_{\epsilon,s})_{s\in\mathbb{R}^+}$ has a.s. infinitely many jumps in $\mathbb{R}^+$, this implies that with probability zero all jumps of $(L_{\epsilon,s})_{s\in\mathbb{R}^+}$ are in $\mathbb{S}_d^+ \backslash \mathbb{S}_d^{++}$. In other words, $(L_{\epsilon,s})_{s\in\mathbb{R}^+}$ and thus $(L_s)_{s\in\mathbb{R}^+}$ has a.s. at least one jump in $\mathbb{S}_d^{++}$. □

The positive-definite Ornstein-Uhlenbeck processes introduced above can be used as a multivariate stochastic volatility model in finance, as an extension of the one-dimensional approach proposed in Barndorff-Nielsen and Shephard (2001b). A different kind of generalization has been discussed by Hubalek and Nicolato (2005) and Lindberg (2005), who have specified different multivariate stochastic volatility models using factor models, where the individual factors are univariate positive Ornstein-Uhlenbeck type processes. The d-dimensional volatility model of Hubalek and Nicolato is of the form $\Sigma_t^2 = AS_tA^*$ where S is an Ornstein-Uhlenbeck process in $\mathbb{S}_m^+$ (actually only on the diagonal matrices) and $A \in M_{d,m}(\mathbb{R})$. The results for the roots of positive definite processes which we obtain in Section 5 are with a minor obvious adaptation immediately applicable to processes of this type. Another proposal put forth in Gourieroux, Jasiak and Sufana (2004) specifies a $d \times d$ volatility process V as a sum $V_t = \sum_{i=1}^K x_{i,t}x_{i,t}^*$ with the processes x_i being i.i.d. Gaussian Ornstein-Uhlenbeck processes in $\mathbb{R}^d$ and $K \in \mathbb{N}$. These processes are referred to as Wishart autoregressive processes, as the distribution of V is the Wishart distribution (see also Bru (1991)). This specification is not amenable to the type of SDE representations of the root processes that we shall discuss in Section 5, under a general set-up, and in Section 6 for positive definite OU processes. Note also, in this connection, that the Wishart law is not infinitely divisible, hence, in particular, not self-decomposable (see Lévy (1948)).

In stochastic volatility models the integrated variance process is of particular interest (see e.g. Barndorff-Nielsen and Shephard (2001b, 2003)). The same reasoning as in the univariate case (Barndorff-Nielsen (1998b)) leads to the following explicit result for the integrated variance of a positive definite Ornstein-Uhlenbeck stochastic volatility process:

Proposition 4.4.10. *Let X be a positive semi-definite Ornstein-Uhlenbeck process with initial value $X_0 \in \mathbb{S}_d^+$ and driven by the Lévy process L. Then the integrated Ornstein-Uhlenbeck process X^+ is given by*

$$X_t^+ := \int_0^t X_t dt = B^{-1}\left(X_t - X_0 - L_t\right)$$

for $t \in \mathbb{R}^+$, where B is the linear operator defined in Proposition 4.4.6.

The use of positive semi-definite OU type processes in a multivariate stochastic volatility model is discussed at length in Chapter 5.

4.5. Roots of positive semi-definite processes

In this section we obtain stochastic representations of general roots of processes in $\mathbb{R}^+$ and later on of the square root of stochastic processes taking values in $\mathbb{S}_d^+$. Recall that every positive semi-definite matrix A has a unique positive semi-definite square root $A^{1/2}$ defined by functional calculus (see, for instance, Horn and Johnson (1985) and Horn and Johnson (1991) for a comprehensive introduction).

The interest in such representations comes, in particular, from the theoretical works on the properties of multipower variation; see Barndorff-Nielsen, Graversen, Jacod, Podolskij and Shephard (2006), for instance. In that paper the limit theorems are obtained under an hypothesis that the square root of the covariance matrix process is a semi-martingale of a special type. Moreover, in many cases the additional assumption is needed that it takes

values in the strictly positive definite matrices, as this ensures that the covariance matrix process is of the same type (and vice versa). However, as there are no formulae given relating the characteristics of the covariance matrix process with those of its square root, we shall derive the relations explicitly and discuss whether the invertibility assumption is indeed always necessary. Under the invertibility assumption Itô's lemma is the key tool, but as we see later on we can move away from this prerequisite. On the other hand we restrict ourselves to the study of processes of finite variation. The reasons are that the processes we intend to apply our results to are naturally of finite variation and that in the infinite variation case it seems impossible to obtain results for processes that may reach the boundary $\partial\mathbb{S}_d^+ = \mathbb{S}_d^+ \backslash \mathbb{S}_d^{++}$. As a consequence all our "stochastic" integrals coming up can actually be computed path-wise as Lebesgue-Stieltjes integrals.

In the following we start by analysing univariate processes, where we study general r-th powers and then move on to multivariate processes.

4.5.1. The univariate case

Now we shall first present the univariate case, as it involves no advanced matrix analysis, but allows one to understand the behaviour of root processes. Due to the applications we have in mind, we state the following results for finite variation processes, whose discontinuous part is of the special form $\int_0^t \int_{\mathbb{R}^+\backslash\{0\}} g(s-,x)\mu(ds,dx)$ with some extended Poisson random measure μ on $\mathbb{R}^+\backslash\{0\}$ (in the sense of Jacod and Shiryaev (2003, Definition 1.20)). Moreover, $g(s,x) = g(\omega,s,x) : \Omega \times \mathbb{R}^+ \times \mathbb{R}^+\backslash\{0\} \to \mathbb{R}^+\backslash\{0\}$ is a (random) function that is $\mathcal{F}_s \times \mathcal{B}(\mathbb{R}^+)$ measurable in (ω,x) and càdlàg in s. For such a process the jump measure is $\mu_X(ds,dx) = \mu(ds, g^{-1}(s-,\cdot)(dx))$, where $g^{-1}(s-,\cdot)$ is to be understood as taking the preimage of the set dx with respect to the map $\mathbb{R}^+\backslash\{0\} \to \mathbb{R}^+\backslash\{0\}, x \mapsto g(s-,x)$. We frequently refer to the dependence on $\omega \in \Omega$ in the following, but keep suppressing it in the notation.

Theorem 4.5.1. *Let $(X_t)_{t\in\mathbb{R}^+}$ be a given adapted càdlàg process which takes values in $\mathbb{R}^+\backslash\{0\}$, is locally bounded away from zero and can be represented as*

$$dX_t = c_t dt + \int_{\mathbb{R}^+\backslash\{0\}} g(t-,x)\mu(dt,dx)$$

where c is a predictable and locally bounded process, μ an extended Poisson random measure on $\mathbb{R}^+ \times \mathbb{R}^+\backslash\{0\}$ and $g(s,x)$ is $\mathcal{F}_s \times \mathcal{B}(\mathbb{R}^+\backslash\{0\})$ measurable in (ω,x) and càdlàg in s. Moreover, $g(s,x)$ takes only non-negative values.

Then for any $0 < r < 1$ the unique positive process $Y = X^r$ is representable as

$$Y_0 = X_0^r, \qquad dY_t = a_t dt + \int_{\mathbb{R}^+\backslash\{0\}} w(t-,x)\mu(dt,dx),$$

where the drift

$$a_t := r X_{t-}^{r-1} c_t$$

is predictable and locally bounded and where

$$w(s,x) := (X_s + g(s,x))^r - (X_s)^r$$

is $\mathcal{F}_s \times \mathcal{B}(\mathbb{R}^+)$ measurable in (ω,x) and càdlàg in s. Moreover, $w(s,x)$ takes only non-negative values.

Proof. Remark 4.3.3 implies the local boundedness of X within $\mathbb{R}^+$ and restating Proposition 4.3.2 in a differential manner gives

$$dX_t^r = rX_{t-}^{r-1}c_t dt + \int_{\mathbb{R}^+\setminus\{0\}} ((X_{t-}+x)^r - X_{t-}^r)\mu_X(dt,dx).$$

Using the relation between μ_X and μ stated before the theorem, we obtain

$$dX_t^r = rX_{t-}^{r-1}c_t dt + \int_{\mathbb{R}^+\setminus\{0\}} ((X_{t-}+g(t-,x))^r - X_{t-}^r)\mu(dt,dx).$$

The positivity of $w(s,x)$ is a consequence of an elementary inequality recalled in the following lemma and the additional properties stated are now straightforward. □

For the sake of completeness and since it is essential to our results, we recall the following elementary inequality and give a proof.

Lemma 4.5.2. *For $a,x\in\mathbb{R}^+$ and $0<r<1$ we have that $(a+x)^r-a^r$ is monotonically decreasing in a and*

$$(a+x)^r - a^r \le x^r.$$

In particular, for $a,b\in\mathbb{R}^+$ it holds that $|a^r-b^r|\le|a-b|^r$.

Proof. Define for fixed x the function $f:\mathbb{R}^+\to\mathbb{R}$, $a\mapsto(a+x)^r-a^r$. Then $f'(a) = r\left((a+x)^{r-1}-a^{r-1}\right)\le 0$ using that the $r-1$-th power is monotonically decreasing. Hence, f is monotonically decreasing and $f(a)=(a+x)^r-a^r\le f(0)=x^r$. For the second inequality we assume without loss of generality that $a\ge b$. Then $|a^r-b^r| = (b+(a-b))^r - b^r \le (a-b)^r = |a-b|^r$, due to the first inequality. □

Remark 4.5.3. *Actually the representation stated in Theorem 4.5.1 holds for arbitrary powers X^r with $r\in\mathbb{R}$. If $r\ge 1$, the assumption that X is locally bounded away from zero is no longer necessary.*

For processes that start at zero or may become zero, we obviously cannot use Itô's formula in the above manner, since there is no way to extend the r-th power for $0<r<1$ to an open set containing $[0,\infty)$ in a continuously differentiable manner. Likewise, all advanced extensions of Itô's formula we know of (e.g. Bardina and Jolis (1997), Ghomrasni and Peskir (2003), Peskir (2005)), cannot be applied. For instance, the Boleau-Yor formula (Protter (2004, Theorem IV.77)) allows for a non-continuous derivative, but still demands it to be bounded, but for r-th roots it is unbounded at zero. The Meyer-Itô formula (Protter (2004, Theorem IV.70)) needs a left derivative, which again cannot be defined at zero. But by using the very standard Itô formula and applying a tailor-made limiting procedure, we can indeed verify an extension to processes that may become zero:

Theorem 4.5.4. *Let $(X_t)_{t\in\mathbb{R}^+}$ be a given adapted càdlàg process which takes values in $\mathbb{R}^+$ and can be represented as*

$$dX_t = c_t dt + \int_{\mathbb{R}^+\setminus\{0\}} g(t-,x)\mu(dt,dx)$$

where c is a predictable and locally bounded process, μ an extended Poisson random measure on $\mathbb{R}^+\times\mathbb{R}^+\setminus\{0\}$ and $g(s,x)$ is $\mathcal{F}_s\times\mathcal{B}(\mathbb{R}^+\setminus\{0\})$ measurable in (ω,x) and càdlàg in s.

Moreover, $g(s,x)$ takes only non-negative values. Assume that the integrals $\int_0^t rX_{s-}^{r-1}c_s ds$ (in the Lebesgue sense) and $\int_0^t \int_{\mathbb{R}^+\setminus\{0\}} (X_{s-}+g(s-,x))^r - (X_{s-})^r \mu(ds,dx)$ exist a.s. for all $t \in \mathbb{R}^+$.

Then for any $0<r<1$ the unique positive process $Y = X^r$ is representable as

$$Y_0 = X_0^r, \qquad dY_t = a_t dt + \int_{\mathbb{R}^+\setminus\{0\}} w(t-,x)\mu(dt,dx), \tag{4.5.1}$$

where the drift

$$a_t := rX_{t-}^{r-1}c_t$$

is predictable and where

$$w(s,x) = (X_s + g(s,x))^r - (X_s)^r$$

is $\mathcal{F}_s \times \mathcal{B}(\mathbb{R}^+)$ measurable in (ω,x) and càdlàg in s. Moreover, $w(s,x)$ takes only non-negative values and Y_t is a.s. of finite variation.

Note that $c_t = 0$ implies $a_t = 0$ above, even if $X_{t-} = 0$, using the conventions of Lebesgue integration theory.

Proof. We first show that $Y = X^r$ is representable by (4.5.1). Recall below that all integrals can be viewed as path-wise Lebesgue-Stieltjes ones.

For any $\epsilon > 0$ the process $X_{\epsilon,t} := X_t + \epsilon$ is bounded away from zero and

$$X_{\epsilon,t} = X_0 + \epsilon + \int_0^t c_s ds + \int_0^t \int_{\mathbb{R}^+\setminus\{0\}} g(s-,x)\mu(ds,dx).$$

From Theorem 4.5.1 we obtain that

$$\begin{aligned}(X_t+\epsilon)^r = X_{\epsilon,t}^r = {} & (X_0+\epsilon)^r + \int_0^t r\,(X_{s-}+\epsilon)^{r-1} c_s ds \\ & + \int_0^t \int_{\mathbb{R}^+\setminus\{0\}} \left((X_{s-}+\epsilon+g(s-,x))^r - (X_{s-}+\epsilon)^r\right)\mu(ds,dx).\end{aligned} \tag{4.5.2}$$

For $s \in \mathbb{R}^+$ we clearly have that $(X_{s-}+\epsilon)^r \to X_{s-}^r$ point-wise as $\epsilon \to 0$. Moreover, since $r-1 \in (-1,0)$, one has that $(X_{s-}+\epsilon)^{r-1}$ is decreasing in ϵ. Thus,

$$|r\,(X_{s-}+\epsilon)^{r-1} c_s| \le |rX_{s-}^{r-1}c_s| \text{ for all } \epsilon > 0.$$

By assumption $|rX_{s-}^{r-1}c_s|$ is Lebesgue-integrable over $[0,t]$ and so majorized convergence gives that

$$\int_0^t r\,(X_{s-}+\epsilon)^{r-1} c_s ds \to \int_0^t rX_{s-}^{r-1}c_s ds \text{ as } \epsilon \to 0.$$

From Lemma 4.5.2 we see that $(X_{s-}+\epsilon+g(s-,x))^r - (X_{s-}+\epsilon)^r$ is positive and also decreasing in ϵ. So our assumptions and majorized convergence ensure that

$$\begin{aligned}\lim_{\epsilon\to 0} & \int_0^t \int_{\mathbb{R}^+\setminus\{0\}} \left((X_{s-}+\epsilon+g(s-,x))^r - (X_{s-}+\epsilon)^r\right)\mu(ds,dx) \\ & = \int_0^t \int_{\mathbb{R}^+\setminus\{0\}} \left((X_{s-}+g(s-,x))^r - X_{s-}^r\right)\mu(ds,dx).\end{aligned}$$

Combining these results we obtain, from (4.5.2) and by letting $\epsilon \to 0$,

$$X_t^r = X_0^r + \int_0^t r X_{s-}^{r-1} c_s ds + \int_0^t \int_{\mathbb{R}^+\setminus\{0\}} \left((X_{s-} + g(s-,x))^r - X_{s-}^r\right) \mu(ds,dx),$$

which concludes the proof of the representation for Y.

To establish the finite variation of the process Y it suffices now to argue that both integral processes $\int_0^t r X_{s-}^{r-1} c_s ds$ and $\int_0^t \int_{\mathbb{R}^+\setminus\{0\}} (X_{s-} + g(s-,x))^r - (X_{s-})^r \mu(ds,dx)$ are of finite variation. For the second this is immediately clear and for the first we only need to observe that the existence in the Lebesgue sense implies the existence of $\int_0^t |r X_{s-}^{r-1} c_s| ds$. The latter is strictly increasing (thus of finite variation) when viewed as a process in t and its total variation is an upper bound for the total variation of the first integral. □

Remark 4.5.5. *a) Inspecting the proof it is clear that Theorem 4.5.1 remains valid when replacing the square root with any continuously differentiable function $f : \mathbb{R}^+ \to \mathbb{R}$. If additionally $|f'(x+\epsilon)| \le K|f'(x)|$ and $|f(x+\epsilon+y) - f(x+\epsilon)| \le \tilde{K}|f(x+y) - f(x)|$ for all $x, y, \epsilon \in \mathbb{R}^+$, where K and $\tilde{K}$ are some constants, the same is true for Theorem 4.5.4.*

Then $f(X)$ is representable by (4.5.1) with $a_t = f'(X_{t-})c_t$ and $w(t,x) = f(X_t + g(t,x)) - f(X_t)$.

b) In general, r-th powers with $0 < r < 1$ of finite variation processes do not have to be of finite variation, as the following deterministic example exhibits. Let X be given by:

$$X_t = \frac{1}{n^2} - \left(1 + \frac{1}{n}\right)\left(t - 1 + \frac{1}{n}\right) \text{ for } t \in \left[1 - \frac{1}{n}, 1 - \frac{1}{n+1}\right), n \in \mathbb{N},$$
$$X_t = 0 \text{ for } t \in [1, \infty).$$

Then we have that $X_{1-(1/n)} = 1/n^2$ and $X_{(1-\frac{1}{n+1})-} = 0$ for all $n \in \mathbb{N}$ and in each interval $\left[1 - \frac{1}{n}, 1 - \frac{1}{n+1}\right)$ the process X is linearly decreasing. From this it is immediate to see that the total variation of $(X_t)_{t\in\mathbb{R}^+}$ is given by $2\sum_{n=1}^{\infty} \frac{1}{n^2} - 1$, which is finite. Likewise, we see that for $0 < r < 1$ the process X^r has jumps of size $1/n^{2r}$ at the times $1 - (1/n)$. As $\sum_{n=1}^{\infty} \frac{1}{n^\alpha}$ is infinite for all $\alpha \le 1$, this shows that for $r \le 1/2$ the process X^r is not of finite variation. Note, moreover, that X is of the form studied in Theorem 4.5.4 where $c_t = -\left(1 + \frac{1}{n}\right)$ for $t \in [1 - (1/n), 1 - (1/(n+1)))$, which is trivially predictable and locally bounded, $g(s,x) = x$ and $\mu(ds,dx) = \sum_{n=1}^{\infty} \delta_{(1-1/n)}(ds)\delta_{1/n^2}(dx)$ with δ_v denoting the Dirac measure with respect to v.

Naturally, the next step is to give some readily checkable conditions for the existence of the integrals.

Lemma 4.5.6. *The integral $\int_0^t \int_{\mathbb{R}^+\setminus\{0\}} w(s-,x)\mu(ds,dx)$ exists a.s. in the usual sense, if the integral $\int_0^t \int_{\mathbb{R}^+\setminus\{0\}} (g(s-,x))^r \mu(ds,dx)$ exists a.s. or there is some a.s. finite random variable $C > 0$ such that $X_t \ge C$ for all $t \in \mathbb{R}^+$.*

Proof. In the first case the existence follows by a standard majorization argument from $0 \le w(s,x) = (X_s + g(s,x))^r - (X_s)^r \le (g(s,x))^r$ (Lemma 4.5.2). Likewise, we observe in the second case that we can argue ω-wise and the function $x \mapsto x^r$ is Lipschitz on any interval of the form $[a,\infty)$ with $a \in \mathbb{R}^+\setminus\{0\}$. Thus there is a (possibly random) $K \in \mathbb{R}^+$ such that $0 \le (X_s + g(s,x))^r - (X_s)^r \le K g(s,x)$. Hence, the claim follows by a dominated convergence argument, since the integral $\int_0^t \int_{\mathbb{R}^+\setminus\{0\}} g(s-,x)\mu(ds,dx)$ exists. □

The condition $X_t \geq C$ actually means that the previous Theorem 4.5.1 applies.

Lemma 4.5.7. *The integral* $\int_0^t \int_{\mathbb{R}^+\setminus\{0\}} \left((X_{s-} + g(s-,x))^r - X_{s-}^r\right) \mu(ds,dx)$ *exists in the usual sense, provided* $c_t \geq 0$ *for all* $t \in \mathbb{R}^+$. *In particular, the process* X *is monotonically increasing then.*

Proof. The monotonicity of X is obvious. We assume $c_t = 0 \,\forall\, t \in \mathbb{R}^+$ first. As the mapping $x \mapsto x^r$ is monotone, also the process X^r has càdlàg monotonically increasing paths. Thus X^r is necessarily of finite variation (on compacts). Denoting the variation of a function f over a time interval $[t_1,t_2]$ with $0 \leq t_1 \leq t_2$ by $var(f;t_1,t_2)$, one deducts that $var(X_t^r,t_1,t_2) = X_{t_2}^r - X_{t_1}^r = \sum_{t_1<s\leq t_2} \Delta\left(X_s^r\right) = \sum_{t_1<s\leq t_2} \left|\Delta\left(X_s^r\right)\right|$. But obviously, $\sum_{t_1<s\leq t_2} \left|\Delta\left(X_s^r\right)\right| = \int_{t_1}^{t_2} \int_{\mathbb{R}^+\setminus\{0\}} \left|(X_{s-} + g(s-,x))^r - X_{s-}^r\right| \mu(ds,dx)$ and hence the finite variation of X^r implies the existence of the integral.

If c does not vanish, we obtain $X_{t_2}^r - X_{t_1}^r \geq \sum_{t_1<s\leq t_2} \Delta\left(X_s^r\right)$ and can then basically argue as before. □

Lemma 4.5.8. *Suppose the function* $g(s,x) = g(x)$ *is deterministic and independent of* s *and the extended Poisson random measure* μ *is the jump measure of a Lévy subordinator with Lévy measure* ν. *Then the integral*

$$\int_0^t \int_{\mathbb{R}^+\setminus\{0\}} \left((X_{s-} + g(x))^r - X_{s-}^r\right) \mu(ds,dx)$$

is a.s. defined for all $t \in \mathbb{R}^+$ *provided* $\int_{0\leq x\leq 1} g(x)^r \nu(dx)$ *is finite.*

Proof. Recall that $E(\mu(ds,dx)) = ds \times \nu(dx)$ in the given set-up. The existence of the integral follows immediately by combining Lemma 4.5.6 and the fact that $\int_{0\leq x\leq 1} g(x)^r \nu(dx) < \infty$ implies the existence of $\int_0^t \int_{\mathbb{R}^+\setminus\{0\}} g(x)^r \mu(ds,dx)$ for all $t \in \mathbb{R}^+$ (cf. Marcus and Rosinski (2005, p. 113)). □

Regarding the existence of the integral with respect to the Lebesgue measure, we only present the following criterion (a standard consequence of dominated convergence), which is applicable to many processes of interest.

Lemma 4.5.9. *Assume that there exists a (possibly random) function* $f : \mathbb{R}^+ \to \mathbb{R}^+$ *with* $\int_0^t f(t)dt < \infty$ *a.s. such that* $|rX_{t-}^{r-1}c_t| \leq f(t)$ *for all* $t \in \mathbb{R}^+$. *Then the integral* $\int_0^t rX_{t-}^{r-1}c_t dt$ *exists in the Lebesgue sense. The latter is in particular the case if there are (possibly random) constants* $C \geq 0$ *and* $\alpha > -1$ *such that* $|rX_{t-}^{r-1}c_t| \leq Ct^\alpha$.

For positive Lévy processes, i.e. Lévy subordinators, one can immediately apply the above results and obtain the following.

Corollary 4.5.10. *Let* $(L_t)_{t\in\mathbb{R}^+}$ *be a Lévy subordinator with initial value* $L_0 \in \mathbb{R}^+$, *associated drift* γ *and jump measure* μ. *Then for* $0 < r < 1$ *we have that the unique positive process* L^r *is of finite variation and*

$$dL_t^r = r\gamma L_{t-}^{r-1}dt + \int_{\mathbb{R}^+\setminus\{0\}} \left((L_{t-} + x)^r - L_{t-}^r\right) \mu(dt,dx),$$

where the drift $r\gamma L_{t-}^{r-1}$ *is predictable. Moreover, the drift is locally bounded, if and only if* $L_0 > 0$ *or* $\gamma = 0$.

Proof. If γ is zero, the integrability condition imposed on the drift in Theorem 4.5.4 is trivially satisfied and in the case of a non-vanishing γ we know that $L_t \geq \gamma t$ for all $t \in \mathbb{R}^+$. The latter gives $r\gamma L_t^{r-1} \leq r\gamma^r t^{r-1}$ and so an application of Lemma 4.5.9 establishes the existence of $\int_0^t r\gamma L_t^{r-1} dt$ in the Lebesgue sense. Finally, noting that Lévy subordinators are monotonically increasing and using Lemma 4.5.7, the corollary follows immediately from Theorem 4.5.4. The result on the local boundedness of the drift is immediate. □

4.5.2. The multivariate case

The aim of this section is to generalise the above univariate results to processes taking values in the cone of positive semi-definite $d \times d$ matrices. For reasons becoming clear later we only take square roots, but generalizations to general roots are straightforward and we shall indicate them. Before giving rigorous results and proofs, we want to give intuitive but non-rigorous arguments showing what the results should be. The reason is that for the rigorous proof we will need the multidimensional Itô formula and the derivative of the matrix square root, whereas the following two elementary lemmata immediately allow for an intuitive argument implying what the result should be. Though these lemmata are rather elementary, we decided to give complete proofs, as they seem to be unavailable in the standard literature, but should be useful in many situations.

The first result generalizes the representation for the product of two one-dimensional semi-martingales (confer e.g. Protter (2004, p. 68)) to matrix products of semi-martingales and is briefly stated, without proof, in Karandikar (1991) (for the continuous case already in Karandikar (1982a,b)).

Lemma 4.5.11. *Let $m, n, d \in \mathbb{N}$ and A and B be semi-martingales taking values in $M_{d,m}(\mathbb{R})$ and $M_{m,n}(\mathbb{R})$, respectively. Then the matrix product AB in $M_{d,n}(\mathbb{R})$ is a semi-martingale and*

$$A_t B_t = \int_0^t A_{t-} dB_t + \int_0^t dA_t B_{t-} + [A, B]_t^M$$

where $[A, B]_t^M \in M_{d,n}(\mathbb{R})$ is defined by

$$[A, B]_{ij,t}^M = \sum_{k=1}^m [A_{ik}, B_{kj}]_t.$$

If the continuous part of the quadratic covariation of A and B is zero, we have

$$[A, B]_t^M = A_0 B_0 + \sum_{0<s\leq t} \Delta A_s \Delta B_s.$$

Proof. Applying the univariate result component-wise to AB we obtain for $1 \leq i \leq d$, $1 \leq j \leq n$:

$$\begin{aligned}(A_t B_t)_{ij} &= \sum_{k=1}^m A_{ik,t} B_{kj,t} = \sum_{k=1}^m \left(\int_0^t A_{ik,s-} dB_{kj,s} + \int_0^t B_{kj,s-} dA_{ik,s} + [A_{ik}, B_{kj}]_t \right) \\ &= \left(\int_0^t A_{s-} dB_s + \int_0^t dA_s B_{s-} + [A, B]_t^M \right)_{ij}.\end{aligned}$$

In particular, we see immediately that all components of AB are semi-martingales being sums of products of semi-martingales. Thus AB is a matrix-valued semi-martingale.

If the continuous quadratic covariation is zero, we have that

$$\begin{aligned}[A,B]_t^M &= \sum_{k=1}^m [A_{ik}, B_{kj}]_t = \sum_{k=1}^m \left(A_{ik,0}B_{kj,0} + \sum_{0<s\leq t} \Delta A_{ik,s}\Delta B_{kj,s} \right) \\ &= \left(A_0 B_0 + \sum_{0<s\leq t} \Delta A_s \Delta B_s \right)_{ij},\end{aligned}$$

since $\Delta A_s = (\Delta A_{kl,s})_{1\leq k\leq d,\, 1\leq l\leq m}$ and likewise for B. □

Remark 4.5.12. *Obviously the operator $[\cdot,\cdot]^M$ plays the same role for the matrix multiplication of matrix-valued semi-martingales, as the quadratic variation does for ordinary multiplication of one-dimensional semi-martingales. Therefore we call the operator $[\cdot,\cdot]^M$ the* matrix covariation. *Note that in general it can be decomposed into*

$$[A,B]_t^M = A_0B_0 + [A,B]_t^{M,c} + \sum_{0<s\leq t} \Delta A_s \Delta B_s$$

where $[A,B]_{t,ij}^{M,c} = \sum_{k=1}^m [A_{ik}, B_{kj}]_t^c$, i.e. into a continuous part and a pure jump part.

Our next result concerns quadratic equations of positive semi-definite matrices.

Lemma 4.5.13. *Let $A, B \in \mathbb{S}_d^+$. The equation*

$$X^2 + AX + XA - B = 0$$

has a unique positive semi-definite solution given by

$$X = \sqrt{A^2 + B} - A.$$

Proof. We start by establishing the positive semi-definiteness of $\sqrt{A^2+B} - A$. It is clear that $A^2 + B \geq A^2$. Observing that the matrix square root is a matrix monotone function (i.e. preserves the ordering on $\mathbb{S}_d^+$, see e.g. Bhatia (1997, Proposition V.1.8)), we have $\sqrt{A^2+B} \geq A$, which is equivalent to the claim.

Solving the equation can actually be done using the standard trick for complex quadratic equations:

$$X^2 + AX + XA - B = (X+A)^2 - A^2 - B = 0 \Leftrightarrow (X+A)^2 = A^2 + B.$$

Taking any "square root" in the right hand equation would now lead to a solution X. However, we consider only positive semi-definite solutions and thus $X + A$ has to be in $\mathbb{S}_d^+$, which is the case, if and only if we take the unique positive semi-definite square root. Therefore there is one and only one solution in $\mathbb{S}_d^+$ which is given by $X = \sqrt{A^2+B} - A$. □

Let now a positive semi-definite process X be given by

$$dX_t = c_t dt + \int_{\mathbb{S}_d^+ \setminus \{0\}} g(t-, x)\mu(dt, dx)$$

where c is an $\mathbb{S}_d$-valued, predictable and locally bounded process, μ an extended Poisson random measure on $\mathbb{R}^+ \times \mathbb{S}_d^+\backslash\{0\}$ and $g(s,x)$ is $\mathcal{F}_s \times \mathcal{B}(\mathbb{S}_d^+\backslash\{0\})$ measurable in (ω, x) and càdlàg in s. Moreover, $g(s,x)$ assumes only values in $\mathbb{S}_d^+$. Suppose $Y := \sqrt{X}$ is representable as $dY_t = a_t dt + \int_{\mathbb{S}_d^+\backslash\{0\}} w(t-,x)\mu(dt,dx)$ for some appropriate a_t and $w(t,x)$ being of the same type as c_t and $g(t,x)$. Using a differential version of Lemma 4.5.11 we obtain

$$\begin{aligned}
dY_t^2 &= Y_{t-}dY_t + dY_t Y_{t-} + d[Y,Y]_t^M = Y_{t-}dY_t + dY_t Y_{t-} + (\Delta Y_t)^2 \\
&= Y_{t-}\left(a_t dt + \int_{\mathbb{S}_d^+\backslash\{0\}} w(t-,x)\mu(dt,dx)\right) \\
&\quad + \left(a_t dt + \int_{\mathbb{S}_d^+\backslash\{0\}} w(t-,x)\mu(dt,dx)\right)Y_{t-} + \int_{\mathbb{S}_d^+\backslash\{0\}} w^2(t-,x)\mu(dt,dx) \\
&= \left(\sqrt{X_{t-}}a_t + a_t\sqrt{X_{t-}}\right)dt \\
&\quad + \int_{\mathbb{S}_d^+\backslash\{0\}} \left(\sqrt{X_{t-}}w(t-,x) + w(t-,x)\sqrt{X_{t-}} + w^2(t-,x)\right)\mu(dt,dx).
\end{aligned}$$

As one clearly needs to have $dY_t^2 = dX_t$, the equations $c_t = \sqrt{X_{t-}}a_t + a_t\sqrt{X_{t-}}$ and $\sqrt{X_{t-}}w(t-,x) + w(t-,x)\sqrt{X_{t-}} + w^2(t-,x) = g(t-,x)$ have to hold. Assuming the necessary invertibility this gives $a_t = \mathbf{X}_{t-}^{-1}c_t$, where $\mathbf{X}_{t-} : M_d(\mathbb{R}) \to M_d(\mathbb{R})$ is the linear operator $Z \mapsto \sqrt{X_{t-}}Z + Z\sqrt{X_{t-}}$, and $w(s-,x) = \sqrt{X_{s-} + g(s-,x)} - \sqrt{X_{s-}}$ using Lemma 4.5.13. In the following we show that this representation for $\sqrt{X}$ is indeed true. It will also turn out that we implicitly obtained the derivative of the positive definite matrix square root, which is given in the next Lemma. Here and in the following we regard $\mathbb{S}_d^{++}$ as a subset of the vector space $\mathbb{S}_d$, which we identify with $\mathbb{R}^{\frac{d(d+1)}{2}}$ when necessary.

Lemma 4.5.14. *The positive definite square root $\sqrt{\cdot} : \mathbb{S}_d^{++} \to \mathbb{S}_d^{++}$ is continuously differentiable and the derivative $D\sqrt{X}$ is given by the inverse of the linear operator $Z \mapsto \sqrt{X}Z + Z\sqrt{X}$.*

Proof. The square root is the inverse of the bijective function $f : \mathbb{S}_d^{++} \to \mathbb{S}_d^{++}$, $X \mapsto X^2$. It is easy to see that $Df(X)$ is the linear operator $Z \mapsto XZ + ZX$ (see also Bhatia (1997, Example X.4.2)). Using that $\sigma(Df(X)) = \sigma(X) + \sigma(X) \subset \mathbb{R}^+\backslash\{0\}$, we see that $Df(X)$ is invertible for all $X \in \mathbb{S}_d^{++}$. Thus, Rudin (1976, Theorem 9.24) shows that the square root is continuously differentiable and the derivative is given by the claimed linear operator. □

With the above results, we can now generalize our results on the behaviour of univariate square roots in a straightforward manner to the multivariate case.

Theorem 4.5.15. *Let $(X_t)_{t\in\mathbb{R}^+}$ be a given adapted càdlàg process which takes values in $\mathbb{S}_d^{++}$, is locally bounded within $\mathbb{S}_d^{++}$ and can be represented as*

$$dX_t = c_t dt + \int_{\mathbb{S}_d^+\backslash\{0\}} g(t-,x)\mu(dt,dx) \tag{4.5.3}$$

where c is an $\mathbb{S}_d$-valued, predictable and locally bounded process, μ an extended Poisson random measure on $\mathbb{R}^+ \times \mathbb{S}_d^+\backslash\{0\}$, and $g(s,x)$ is $\mathcal{F}_s \times \mathcal{B}(\mathbb{S}_d^+\backslash\{0\})$ measurable in (ω,x) and càdlàg in s. Moreover, $g(s,x)$ takes only values in $\mathbb{S}_d^+$.

Then the integral $\int_0^t \int_{\mathbb{S}_d^+\setminus\{0\}} \big(\sqrt{X_{s-}+g(s-,x)}-\sqrt{X_{s-}}\big)\mu(ds,dx)$ *exists a.s. for all* $t \in \mathbb{R}^+$ *and the unique positive definite square root process* $Y = \sqrt{X}$ *is given by*

$$Y_0 = \sqrt{X_0}, \qquad dY_t = a_t dt + \int_{\mathbb{S}_d^+\setminus\{0\}} w(t-,x)\mu(dt,dx),$$

with

$$a_t = \mathbf{X}_{t-}^{-1} c_t,$$

where $\mathbf{X}_{t-}$ *is the linear operator* $Z \mapsto \sqrt{X_{t-}}Z + Z\sqrt{X_{t-}}$ *on* $M_d(\mathbb{R})$*. The drift process* a *is predictable and locally bounded and*

$$w(s,x) := \sqrt{X_s + g(s,x)} - \sqrt{X_s}$$

is $\mathcal{F}_s \times \mathcal{B}(\mathbb{S}_d^+\setminus\{0\})$ *measurable in* (ω, x) *and càdlàg in* s*. Moreover,* $w(s,x)$ *takes only positive semi-definite values.*

Proof. The representation of Y follows from Proposition 4.3.4 and Lemma 4.5.14 by the same arguments as used for Theorem 4.5.1.

Using the vec-transformation and the Kronecker product, the linear operator $\mathbf{X}_{t-}$ is easily seen to be symmetric (self-adjoint) and to possess a spectrum that is positive and locally bounded away from 0, since $\sigma(\mathbf{X}_{t-}) = \sigma(\sqrt{X_{t-}}) + \sigma(\sqrt{X_{t-}})$, the function $f : \mathbb{S}_d^{++} \to \mathbb{S}_d^{++}, Z \mapsto \min(\sigma(Z))$ is continuous and $\sqrt{X}$ is locally bounded within $\mathbb{S}_d^{++}$. The variational characterizations of the eigenvalues of a self-adjoint operator (cf. Horn and Johnson (1985, Section 4.2) for a matrix formulation) imply that $\min(\sigma(\mathbf{X}_{t-})) = \min_{\|x\|_{\tilde{2}}\neq 0}\big(\frac{\|\mathbf{X}_{t-}x\|_{\tilde{2}}}{\|x\|_{\tilde{2}}}\big)$. Hence, $\|\mathbf{X}_{t-}^{-1}\|_{\tilde{2}} \leq (\min(\sigma(\mathbf{X}_{t-})))^{-1}$ is locally bounded. Here $\|\cdot\|_{\tilde{2}}$ denotes the norm on $M_d(\mathbb{R})$ given by $\|x\|_{\tilde{2}} = \|\mathrm{vec}(x)\|_2 = \sqrt{\mathrm{tr}(xx^T)}$, with $\|\cdot\|_2$ being the Euclidean norm on $\mathbb{R}^{d^2}$, and the associated operator norm on the linear operators over $M_d(\mathbb{R})$. This establishes the local boundedness of a.

That $w(s,x)$ takes only positive semi-definite values follows from Lemma 4.5.13 and the additional properties stated are straightforward. □

Remark 4.5.16. *In principle we could immediately extend the above result to arbitrary* r*-th powers with* $0 < r < 1$ *again. Yet, this would mean that we need to calculate* Df_r *where* f_r *denotes the unique positive definite* r*-th power and* a_t *would become* $Df_r(X_{t-})c_t$*. In general there seems to be no useful formula for* Df_r*. Arguing as in Lemma 4.5.14 was possible for* $r = 1/n$ *with* $n \in \mathbb{N}$*, but then* $Df_r(X)$ *would be characterized as the inverse of the linear operator* $Z \mapsto \sum_{j+k=n-1;j,k\in\mathbb{N}_0} X^{jr} Z X^{kr}$*. Although in principle this can be applied, it appears to be infeasible for general* n*.*

Assuming the existence of the relevant integrals, the strict positivity condition can again be relaxed. To be able to argue as in the univariate case we need two new technical results, the first one involving the so-called trace norm $\|\cdot\|_{tr}$ of matrices. For $A \in M_d(\mathbb{R})$ it is defined as $\|A\|_{tr} = \mathrm{tr}\big((AA^*)^{1/2}\big)$ and it is easy to see that $\|A\|_{tr} = \mathrm{tr}(A)$ for $A \in \mathbb{S}_d^+$.

Lemma 4.5.17. *Let* $A, B \in \mathbb{S}_d^+$ *and* $0 < r < 1$*. Then the function* $\mathbb{R}^+ \to \mathbb{R}^+$, $\epsilon \mapsto \|(A+\epsilon I_d + B)^r - (A+\epsilon I_d)^r\|_{tr}$ *is monotonically decreasing. In particular,*

$$\|(A+\epsilon I_d + B)^r - (A+\epsilon I_d)^r\|_{tr} \leq \|(A+B)^r - A^r\|_{tr}$$

for all $\epsilon \in \mathbb{R}^+$*.*

Proof. Denote for some matrix $Z \in \mathbb{S}_d^+$ by $\lambda_1(Z), \lambda_2(Z), \ldots, \lambda_d(Z)$ the eigenvalues of Z sorted in ascending order.

Choose now some arbitrary $\epsilon, \tilde{\epsilon} \in \mathbb{R}^+$ with $\epsilon \geq \tilde{\epsilon}$. From Horn and Johnson (1985, Corollary 4.3.3) we obtain $\lambda_i(A+B) \geq \lambda_i(A)$ for $i = 1, 2, \ldots, d$. This implies using Lemma 4.5.2 that

$$\begin{aligned}
&\sum_{i=1}^{d} \left((\lambda_i(A+B)+\epsilon)^r - (\lambda_i(A)+\epsilon)^r\right) \\
&\quad = \sum_{i=1}^{d} \left((\lambda_i(A)+\epsilon+\lambda_i(A+B)-\lambda_i(A))^r - (\lambda_i(A)+\epsilon)^r\right) \\
&\quad \leq \sum_{i=1}^{d} \left((\lambda_i(A)+\tilde{\epsilon}+\lambda_i(A+B)-\lambda_i(A))^r - (\lambda_i(A)+\tilde{\epsilon})^r\right) \\
&\quad = \sum_{i=1}^{d} \left((\lambda_i(A+B)+\tilde{\epsilon})^r - (\lambda_i(A)+\tilde{\epsilon})^r\right).
\end{aligned}$$

Noting that the trace of a matrix is the sum of its eigenvalues and that $\lambda_i(Z+\epsilon I_d) = \lambda_i(Z)+\epsilon$ and $\lambda_i(Z)^r = \lambda_i(Z^r)$ for all $Z \in \mathbb{S}_d^+$ and $\epsilon > 0$, we conclude $\operatorname{tr}((A+\epsilon I_d+B)^r) - \operatorname{tr}((A+\epsilon I_d)^r) \leq \operatorname{tr}((A+\tilde{\epsilon} I_d+B)^r) - \operatorname{tr}((A+\tilde{\epsilon} I_d)^r)$. This immediately implies

$$\|(A+\epsilon I_d+B)^r - (A+\epsilon I_d)^r\|_{tr} \leq \|(A+\tilde{\epsilon} I_d+B)^r - (A+\tilde{\epsilon} I_d)^r\|_{tr}.$$

This shows the claimed monotonicity and inequality, choosing $\tilde{\epsilon} = 0$. □

Lemma 4.5.18. *Let $A \in \mathbb{S}_d^+$, $\epsilon \in \mathbb{R}^+$ and denote by $\mathbf{A}_\epsilon$ the linear operator $M_d(\mathbb{R}) \to M_d(\mathbb{R}) : X \mapsto \sqrt{A+\epsilon I_d}X + X\sqrt{A+\epsilon I_d}$. Then we have for every $x \in M_d(\mathbb{R})$ that $\|\mathbf{A}_\epsilon^{-1}x\|_{\tilde{2}}$ is decreasing in ϵ.*

Here $\|\cdot\|_{\tilde{2}}$ denotes again the norm on $M_d(\mathbb{R})$ given by $\|x\|_{\tilde{2}} = \|\operatorname{vec}(x)\|_2 = \sqrt{\operatorname{tr}(xx^T)}$, with $\|\cdot\|_2$ being the Euclidean norm on $\mathbb{R}^{d^2}$, and the associated operator norm on the linear operators over $M_d(\mathbb{R})$.

We understand $\|\mathbf{A}_0^{-1}x\|_{\tilde{2}} = \infty$ in the case $A \in \mathbb{S}_d^+ \backslash \mathbb{S}_d^{++}$ above.

Proof. Note first that $\|\mathbf{A}_\epsilon^{-1}x\|_{\tilde{2}} = \|\left(\sqrt{A+\epsilon I_d} \otimes I_d + I_d \otimes \sqrt{A+\epsilon I_d}\right)^{-1} \operatorname{vec}(x)\|_2$ and that $\sqrt{A+\epsilon I_d}\otimes I_d + I_d \otimes \sqrt{A+\epsilon I_d} \in \mathbb{S}_{d^2}^+$ and in particular self-adjoint. Thus we have $\|\mathbf{A}_\epsilon^{-1}x\|_{\tilde{2}} = \left\langle \operatorname{vec}(x), \left(\sqrt{A+\epsilon I_d} \otimes I_d + I_d \otimes \sqrt{A+\epsilon I_d}\right)^{-2} \operatorname{vec}(x)\right\rangle^{1/2}$. Using that taking the inverse reverses the ordering on $\mathbb{S}_{d^2}^+$, this implies that it is sufficient to show that $(\sqrt{A+\epsilon I_d} \otimes I_d + I_d \otimes \sqrt{A+\epsilon I_d})^2$ is increasing in ϵ in the ordering on $\mathbb{S}_{d^2}$. But let now $U \in M_d(\mathbb{R})$ be a unitary matrix such that U^*AU is diagonal, then $(U^* \otimes U^*)(\sqrt{A+\epsilon I_d} \otimes I_d + I_d \otimes \sqrt{A+\epsilon I_d})^2 (U \otimes U)$ is diagonal and obviously increasing in ϵ. Observing that $U \otimes U$ is again unitary and that such transformations preserve the ordering on $\mathbb{S}_{d^2}^+$ concludes the proof. □

Proposition 4.5.19. *Let $(X_t)_{t\in\mathbb{R}^+}$ be a given adapted càdlàg process which takes values in $\mathbb{S}_d^+$ and can be represented as*

$$dX_t = c_t dt + \int_{\mathbb{S}_d^+ \backslash \{0\}} g(t-, x)\mu(dt, dx)$$

where c is an $\mathbb{S}_d$-valued, predictable and locally bounded process, μ an extended Poisson random measure on $\mathbb{R}^+ \times \mathbb{S}_d^+\backslash\{0\}$ and $g(s,x)$ is $\mathcal{F}_s \times \mathcal{B}(\mathbb{S}_d^+\backslash\{0\})$ measurable in (ω, x) and càdlàg in s. Moreover, $g(s,x)$ takes values in $\mathbb{S}_d^+$. Let $\mathbf{X}_{t-}$ be the linear operator $M_d(\mathbb{R}) \to M_d(\mathbb{R})$, $Z \mapsto \sqrt{X_{t-}}Z + Z\sqrt{X_{t-}}$ and assume that the integrals $\int_0^t \mathbf{X}_{s-}^{-1} c_s ds$ (in the Lebesgue sense) and

$$\int_0^t \int_{\mathbb{S}_d^+\backslash\{0\}} \left(\sqrt{X_{s-} + g(s-,x)} - \sqrt{X_{s-}}\right) \mu(ds, dx)$$

exist a.s. for all $t \in \mathbb{R}^+$.

Then the unique positive semi-definite square root process $Y = \sqrt{X}$ is representable as

$$Y_0 = \sqrt{X_0}, \qquad dY_t = a_t dt + \int_{\mathbb{S}_d^+\backslash\{0\}} w(t-,x)\mu(dt, dx), \tag{4.5.4}$$

where the drift

$$a_t = \mathbf{X}_{t-}^{-1} c_t$$

is predictable and where

$$w(s,x) := \sqrt{X_s + g(s,x)} - \sqrt{X_s}$$

is $\mathcal{F}_s \times \mathcal{B}(\mathbb{S}_d^+\backslash\{0\})$ measurable in (ω, x) and càdlàg in s. Moreover, $w(s,x)$ takes only positive semi-definite values and Y is a.s. of finite variation.

Due to the conventions of Lebesgue integration theory we always have $a_t = 0$ if $c_t = 0$ above.

Proof. We first show that $Y = \sqrt{X}$ is representable by (4.5.4). Recall below that the integral of an $M_d(\mathbb{R})$-valued function exists if and only if the integral of the norm exists for one and hence all norms on $M_d(\mathbb{R})$.

For any $\epsilon > 0$ we define the process $X_{\epsilon,t} := X_t + \epsilon I_d$. Obviously $X_{\epsilon,t} \geq \epsilon I_d$ for all $t \in \mathbb{R}^+$ and the process X_ϵ is of finite variation and hence locally bounded. Observing that for all $\delta, K > 0$ the set $\{x \in \mathbb{S}_d^{++} : x \geq \delta I_d,\ \|x\| \leq K\}$ is convex and compact, this implies that X_ϵ is locally bounded within $\mathbb{S}_d^{++}$ and

$$X_{\epsilon,t} = X_0 + \epsilon I_d + \int_0^t c_s ds + \int_0^t \int_{\mathbb{S}_d^+\backslash\{0\}} g(s-,x)\mu(ds, dx).$$

From Theorem 4.5.15 we obtain that

$$\begin{aligned}\sqrt{X_t + \epsilon I_d} = \sqrt{X_{\epsilon,t}} = \sqrt{X_0 + \epsilon I_d} + \int_0^t \mathbf{X}_{\epsilon,s-}^{-1} c_s ds \\ + \int_0^t \int_{\mathbb{S}_d^+\backslash\{0\}} \left(\sqrt{X_{s-} + \epsilon I_d + g(s-,x)} - \sqrt{X_{s-} + \epsilon I_d}\right) \mu(ds, dx),\end{aligned} \tag{4.5.5}$$

where $\mathbf{X}_{\epsilon,s-}$ denotes the linear operator

$$M_d(\mathbb{R}) \to M_d(\mathbb{R}) : Z \mapsto \sqrt{X_{s-} + \epsilon I_d}Z + Z\sqrt{X_{s-} + \epsilon I_d}.$$

For $s \in \mathbb{R}^+$ we clearly have that $\sqrt{X_{s-} + \epsilon} \to \sqrt{X_{s-}}$ and $\mathbf{X}_{\epsilon,s-} \to \mathbf{X}_{s-}$ point-wise as $\epsilon \to 0$. Moreover, Lemma 4.5.18 ensures $\|\mathbf{X}^{-1}_{\epsilon,s-} c_s\|_{\tilde{2}} \leq \|\mathbf{X}^{-1}_{s-} c_s\|_{\tilde{2}}$ for all $\epsilon > 0$. By assumption $\|\mathbf{X}^{-1}_{\epsilon,s-} c_s\|_{\tilde{2}}$ is Lebesgue-integrable over $[0,t]$ and so majorized convergence gives that

$$\int_0^t \mathbf{X}^{-1}_{\epsilon,s-} c_s ds \to \int_0^t \mathbf{X}^{-1}_{s-} c_s ds \text{ as } \epsilon \to 0.$$

From Lemma 4.5.17 we see that $\|\sqrt{X_{s-} + \epsilon I_d + g(s-,x)} - \sqrt{X_{s-} + \epsilon I_d}\|_{tr}$ is decreasing in ϵ. So our assumptions and majorized convergence ensure that

$$\begin{aligned}\lim_{\epsilon \to 0} \int_0^t \int_{\mathbb{S}_d^+ \setminus \{0\}} \left(\sqrt{X_{s-} + \epsilon I_d + g(s-,x)} - \sqrt{X_{s-} + \epsilon I_d}\right) \mu(ds, dx) \\ = \int_0^t \int_{\mathbb{S}_d^+ \setminus \{0\}} \left(\sqrt{X_{s-} + g(s-,x)} - \sqrt{X_{s-}}\right) \mu(ds, dx).\end{aligned}$$

Combining these results we obtain, from (4.5.5) and by letting $\epsilon \to 0$,

$$\sqrt{X_t} = \sqrt{X_0} + \int_0^t \mathbf{X}^{-1}_{s-} c_s ds + \int_0^t \int_{\mathbb{S}_d^+ \setminus \{0\}} \left(\sqrt{X_{s-} + g(s-,x)} - \sqrt{X_{s-}}\right) \mu(ds, dx),$$

which concludes the proof of the representation for Y.

To establish the finite variation of the process Y it suffices now to argue that both integral processes $\int_0^t \mathbf{X}^{-1}_{s-} c_s ds$ and $\int_0^t \int_{\mathbb{S}_d^+ \setminus \{0\}} \sqrt{X_{s-} + g(s-,x)} - \sqrt{X_{s-}} \mu(ds, dx)$ are of finite variation. For the second this is immediately clear and for the first we only need to observe that the existence in the Lebesgue sense implies the existence of $\int_0^t \|\mathbf{X}^{-1}_{s-} c_s\| ds$ for any norm $\|\cdot\|$. The latter is strictly increasing (thus of finite variation) when viewed as a process in t and its total variation is an upper bound for the total variation of the first integral calculated using the same norm $\|\cdot\|$. □

Remark 4.5.20. *When replacing the square root with an arbitrary continuously differentiable function $f : \mathbb{S}_d^+ \to \mathbb{S}_d$, the above proposition remains valid if $\|Df(x + \epsilon I_d)z\| \leq K\|Df(x)z\|$ and*

$$\|f(x + \epsilon I_d + y) - f(x + \epsilon I_d)\| \leq \tilde{K}\|f(x + y) - f(x)\| \tag{4.5.6}$$

for all $x, y \in \mathbb{S}_d^+$, $z \in \mathbb{S}_d$ and $\epsilon \in \mathbb{R}^+$, where K and $\tilde{K}$ are some constants. Then $f(X)$ is representable by (4.5.4) with $a_t = Df(X_{t-})c_t$ and $w(t,x) = f(X_t + g(t,x)) - f(X_t)$.

For general r-th powers with $0 < r < 1$ condition (4.5.6) holds due to Lemma 4.5.17. In particular, this implies that the above theorem applies immediately to the r-th power if $c_t = 0$ for all $t \in \mathbb{R}^+$. Furthermore, the square root can be replaced by the r-th power in all the following Lemmata 4.5.23, 4.5.24, 4.5.25, 4.5.26 and 4.5.27.

Before giving criteria for the existence of the integrals assumed in the above theorem, we establish some auxiliary results. The first one establishes that $\mathbb{S}_d^+$-increasing functions are always of finite variation.

Lemma 4.5.21. *Let $f : \mathbb{R}^+ \to \mathbb{S}_d^+$ be an $\mathbb{S}_d^+$-increasing function, i.e. $f(a) \leq f(b)$ for all $a, b \in \mathbb{R}^+$ with $a \leq b$. Then f is of finite variation on compacts.*

Proof. Obviously we are free to choose any norm on $M_d(\mathbb{R})$. Let thus $\|\cdot\|_{\mathrm{tr}}$ again denote the trace norm and recall that $\|A\|_{\mathrm{tr}} = \mathrm{tr}(A)$ for all $A \in \mathbb{S}_d^+$. For $s, t \in \mathbb{R}^+$, $t \geq s$ we obtain

$$\|f(t) - f(s)\|_{\mathrm{tr}} = \mathrm{tr}(f(t) - f(s)) = \mathrm{tr}(f(t)) - \mathrm{tr}(f(s)),$$

due to the linearity of the trace. From this we can immediately conclude that the total variation of f over any interval $[a, b]$ with $a, b \in \mathbb{R}^+$, $a \leq b$ calculated in the trace norm is given by $\mathrm{tr}(f(b)) - \mathrm{tr}(f(a))$, which is finite. Hence, f is of finite variation on compacts. □

The trace norm has also been used in Barndorff-Nielsen and Pérez-Abreu (2007) and Pérez-Abreu and Rocha-Arteaga (2005) and thus seems to be very well adapted to the structure of matrix subordinators. The lemma could alternatively be easily established using the theory for general cones developed in Duda (2007) and the properties of the trace functional/norm.

Moreover, we need to consider an appropriate matrix extension of the inequality $\sqrt{a+b} - \sqrt{a} \leq \sqrt{b}$ for all $a, b \in \mathbb{R}^+$. Actually, the question whether $\sqrt{A+B} - \sqrt{A} \leq \sqrt{B}$ for $A, B \in \mathbb{S}_d^+$ seems not to have been discussed in the literature yet. However, the following norm version suffices for our purposes.

Definition 4.5.22. *Let $A, B \in M_d(\mathbb{R})$ then $|A| = (A^*A)^{1/2}$ is called the* modulus (absolute value) *of A.*

A norm $\|\cdot\|$ on $M_d(\mathbb{R})$ is said to be unitarily invariant*, if $\|UAV\| = \|A\|$ for all unitary matrices $U, V \in M_d(\mathbb{R})$.*

For more information see e.g. Bhatia (1997) and for unitarily invariant norms also Horn and Johnson (1985).

Lemma 4.5.23 (Ando (1988, Corollary 2))**.** *Let $A, B \in \mathbb{S}_d^+$ and $\|\cdot\|$ be any unitarily invariant norm. Then*

$$\|\sqrt{A} - \sqrt{B}\| \leq \|\sqrt{|A - B|}\|.$$

This result has originally been obtained in Birman, Koplienko and Solomjak (1975). We can simplify the result somewhat by using the operator norm associated to the usual Euclidean norm on $\mathbb{R}^d$.

Corollary 4.5.24 (cf. Bhatia (1997, Section X.1))**.** *Let $A, B \in \mathbb{S}_d^+$ and let $\|\cdot\|_2$ denote the operator norm associated with the Euclidean norm. Then*

$$\|\sqrt{A} - \sqrt{B}\|_2 \leq \sqrt{\|\,(|A - B|)\,\|_2}.$$

In particular, $\|\sqrt{A+B} - \sqrt{A}\|_2 \leq \sqrt{\|B\|_2}$.

Armed with these prerequisites we can now state criteria for the existence of the integrals in Theorem 4.5.19.

Lemma 4.5.25. *The integral $\int_0^t \int_{\mathbb{S}_d^+ \setminus \{0\}} w(s-, x)\mu(ds, dx)$ exists a.s. for all $t \in \mathbb{R}^+$ in the usual sense if the integrals*

$$\int_0^t \int_{\mathbb{S}_d^+ \setminus \{0\}} \sqrt{\|g(s-, x)\|_2}\mu(ds, dx) \quad \text{or} \quad \int_0^t \int_{\mathbb{S}_d^+ \setminus \{0\}} \sqrt{g(s-, x)}\mu(ds, dx)$$

exist a.s. for all $t \in \mathbb{R}^+$ or there is some $\mathbb{S}_d^{++}$-valued random variable C such that $X_t \geq C$ for all $t \in \mathbb{R}^+$.

Due to the equivalence of all norms one can actually use any other norm instead of $\|\cdot\|_2$. Moreover, the second case corresponds to Theorem 4.5.15.

Proof. First of all we note that $\int_0^t \int_{\mathbb{S}_d^+\setminus\{0\}} \sqrt{\|g(s-,x)\|_2}\mu(ds,dx)$ exists if and only if the integral $\int_0^t \int_{\mathbb{S}_d^+\setminus\{0\}} \sqrt{g(s-,x)}\mu(ds,dx)$ exists. This follows immediately, since according to the definition of integration with respect to Poisson random measures the integral $\int_0^t \int_{\mathbb{S}_d^+\setminus\{0\}} \sqrt{g(s-,x)}\mu(ds,dx)$ exists if and only if

$$\int_0^t \int_{\mathbb{S}_d^+\setminus\{0\}} \|\sqrt{g(s-,x)}\|\mu(ds,dx)$$

exists for one and hence all norms $\|\cdot\|$, and $\|\sqrt{x}\|_2 = \sqrt{\|x\|_2}$ for all $x \in \mathbb{S}_d^+$.

Noting that Corollary 4.5.24 gives $\|w(s-,x)\|_2 \leq \sqrt{\|g(s-,x)\|_2}$, a simple majorization argument establishes the existence of $\int_0^t \int_{\mathbb{S}_d^+\setminus\{0\}} w(s-,x)\mu(ds,dx)$ in the first case.

Assume now that $X_t \geq C$ for all $t \in \mathbb{R}^+$ holds with some $C \in \mathbb{S}_d^+$. Then we once again argue ω-wise. The square root function is Lipschitz on any set $\mathcal{A} \subset \mathbb{S}_d^{++}$ for which there is some $C_0 \in \mathbb{S}_d^{++}$ such that $C \geq C_0$ for all $C \in \mathcal{A}$ (see, for instance, Bhatia (1997, p. 305)). Thus there exists a constant K (possibly depending on C) such that $\big\|\sqrt{X_{s-}+g(s-,x)} - \sqrt{X_{s-}}\big\| \leq K\big\|g(s-,x)\big\|$. This implies the existence of the integral, as $\int_0^t \int_{\mathbb{S}_d^+\setminus\{0\}} g(s-,x)\mu(ds,dx)$ exists due to our assumptions on the process X. □

Lemma 4.5.26. *The integral $\int_0^t \int_{\mathbb{S}_d^+\setminus\{0\}} w(s-,x)\mu(ds,dx)$ exists a.s. for all $t \in \mathbb{R}^+$ in the usual sense provided $c_t \in \mathbb{S}_d^+$ for all $t \in \mathbb{R}^+$, i.e. the process X is $\mathbb{S}_d^+$-increasing.*

Proof. The $\mathbb{S}_d^+$-increasingness of X is clear. Since the square root preserves the ordering on $\mathbb{S}_d^+$, the process $\sqrt{X}$ is $\mathbb{S}_d^+$-increasing, as well. Thus, Lemma 4.5.21 ensures that $\sqrt{X}$ is of finite variation.

Now, we first assume $c_t = 0$ for all $t \in \mathbb{R}^+$. Denoting the total variation (in the trace norm) of a function f over a time interval $[t_1,t_2]$ with $0 \leq t_1 \leq t_2$ by $var(f;t_1,t_2)$, one deducts that $var(\sqrt{X},t_1,t_2) = \mathrm{tr}(\sqrt{X_{t_2}}) - \mathrm{tr}(\sqrt{X_{t_1}}) = \sum_{t_1<s\leq t_2} \|\Delta\left(\sqrt{X_s}\right)\|_{\mathrm{tr}}$. But,

$$\sum_{t_1<s\leq t_2} \big\|\Delta\big(\sqrt{X_s}\big)\big\|_{\mathrm{tr}} = \int_{t_1}^{t_2} \int_{\mathbb{S}_d^+\setminus\{0\}} \big\|\sqrt{X_{s-}+g(s-,x)} - \sqrt{X_{s-}}\big\|_{\mathrm{tr}}\mu(ds,dx)$$

obviously and hence the finite variation of $\sqrt{X}$ implies the existence of the integral.

If c_t does not vanish, we obtain $\mathrm{tr}(\sqrt{X_{t_2}}) - \mathrm{tr}(\sqrt{X_{t_1}}) \geq \sum_{t_1<s\leq t_2} \|\Delta\left(\sqrt{X_s}\right)\|_{\mathrm{tr}}$ and can argue as before. □

For the following recall that we refer to $\mathbb{S}_d^+$-increasing Lévy processes as matrix subordinators.

Lemma 4.5.27. *Suppose the function $g(s,x) = g(x)$ is deterministic and independent of s and the extended Poisson random measure μ is the jump measure of a matrix subordinator with Lévy measure ν. Then the integral*

$$\int_0^t \int_{\mathbb{S}_d^+\setminus\{0\}} \left(\sqrt{X_{s-}+g(x)} - \sqrt{X_{s-}}\right)\mu(ds,dx)$$

is indeed a.s. defined for all $t \in \mathbb{R}^+$ provided $\int_{0\leq\|x\|_2\leq 1,\, x\in\mathbb{S}_d^+\setminus\{0\}} \sqrt{\|g(x)\|_2}\nu(dx)$ is finite.

Again we can use any other norm instead of $\|\cdot\|_2$.

Proof. Recall that $E(\mu(ds,dx)) = ds \times \nu(dx)$ in the given set-up. The existence of the integral follows immediately by combining Lemma 4.5.25 and the fact that

$$\int_{\|x\|_2\leq 1} \sqrt{\|g(x)\|_2}\nu(dx) = \int_{\|x\|_2\leq 1} \|\sqrt{g(x)}\|_2\nu(dx) < \infty$$

implies the existence of

$$\int_0^t \int_{\mathbb{S}_d^+\setminus\{0\}} \sqrt{g(x)}\mu(ds,dx)$$

for all $t \in \mathbb{R}^+$ (cf. Marcus and Rosinski (2005, p. 113)). Here we note that

$$\int_0^t \int_{\mathbb{S}_d^+\setminus\{0\}} \min\left(\|g(x)\|_2, 1\right)\nu(dx)ds \leq t\left(\nu\left(\{x\in\mathbb{S}_d^+ : \|x\|_2 > 1\}\right) + \int_{\|x\|_2\leq 1} \sqrt{\|g(x)\|_2}\nu(dx)\right)$$

is finite. □

Regarding the existence of the integral with respect to the Lesbesgue measure, we only restate the criterion of Lemma 4.5.9 for the multivariate case.

Lemma 4.5.28. *Assume that there exists a (possibly random) function $f : \mathbb{R}^+ \to \mathbb{R}^+$ with $\int_0^t f(t)dt < \infty$ a.s. such that $\|\mathbf{X}_{t-}^{-1}c_t\| \leq f(t)$ for all $t \in \mathbb{R}^+$. Then the integral $\int_0^t \mathbf{X}_{t-}^{-1}c_t dt$ exists in the Lebesgue sense. The latter is in particular the case, if there are (possibly random) constants $C \geq 0$ and $\alpha > -1$ such that $\|\mathbf{X}_{t-}^{-1}c_t\| \leq Ct^\alpha$.*

After these general considerations we shall now turn to studying the roots of matrix subordinators.

Corollary 4.5.29. *Let $(L_t)_{t\in\mathbb{R}^+}$ be a matrix subordinator with initial value $L_0 \in \mathbb{S}_d^+$, associated drift γ and jump measure μ. Then the unique positive semi-definite process $\sqrt{L}$ is of finite variation and, provided that either $L_0 \in \mathbb{S}_d^{++}$ or $\gamma \in \mathbb{S}_d^{++}\cup\{0\}$,*

$$d\sqrt{L_t} = \mathbf{L}_{t-}^{-1}\gamma dt + \int_{\mathbb{S}_d^+\setminus\{0\}} \left(\sqrt{L_{t-}+x} - \sqrt{L_{t-}}\right)\mu(dt,dx),$$

where $\mathbf{L}_{t-}$ is the linear operator on $M_d(\mathbb{R})$ with $Z \mapsto \sqrt{L_{t-}}Z + Z\sqrt{L_{t-}}$. The drift $\mathbf{L}_{t-}^{-1}\gamma$ is predictable, and additionally locally bounded provided $L_0 \in \mathbb{S}_d^{++}$ or $\gamma = 0$.

Proof. As the square root preserves the ordering on $\mathbb{S}_d^+$, $\sqrt{L}$ is $\mathbb{S}_d^+$-increasing and thus of finite variation by Lemma 4.5.21.

In the case $L_0 \in \mathbb{S}_d^{++}$ the Corollary follows from Theorem 4.5.15.

Else we know from Lemma 4.5.26 that the integral $\int_0^t \int_{\mathbb{S}_d^+\setminus\{0\}} \left(\sqrt{L_{s-}+x} - \sqrt{L_{s-}}\right)\mu(ds,dx)$ exists a.s. for all $t \in \mathbb{R}^+$. Next we show that the integral $\int_0^t \mathbf{L}_{s-}^{-1}\gamma ds$ exists for all $t \in \mathbb{R}^+$. For $\gamma = 0$ this is trivial. For $\gamma \in \mathbb{S}_d^{++}$, we have that $L_s \geq \gamma s \in \mathbb{S}_d^{++}$. Using the variational characteristics of the eigenvalues as in the proof of Theorem 4.5.15 we get

$$\min_{\|x\|_{\tilde{2}}\neq 0}\left(\frac{\|\mathbf{L}_{s-}x\|_{\tilde{2}}}{\|x\|_{\tilde{2}}}\right) = \min\left(\sigma(\mathbf{L}_{s-})\right) = 2\min\left(\sigma(\sqrt{L_{s-}})\right) \geq 2\sqrt{s}\sqrt{\min\left(\sigma(\gamma)\right)}.$$

Therefore $\|\mathbf{L}_{s-}^{-1}\|_{\tilde{2}} \leq \big(\min\big(\sigma(\mathbf{L}_{s-})\big)\big)^{-1} \leq \big(2\sqrt{\min\big(\sigma(\gamma)\big)}\big)^{-1} s^{-1/2}$. Hence, $\|\mathbf{L}_{s-}^{-1}\gamma\| \leq Cs^{-1/2}$ for all $s \in \mathbb{R}^+$ with some constant $C \in \mathbb{R}^+$ and so Lemma 4.5.28 establishes the existence of $\int_0^t \mathbf{L}_{s-} c_s ds$ for all $t \in \mathbb{R}^+$ in the Lebesgue sense. Therefore Proposition 4.5.19 concludes the proof. □

Remark 4.5.30. *If the Lévy process is supposed to have initial value in $\partial\mathbb{S}_d^+$ (e.g. zero, as is usual) and non-zero drift $\gamma \in \partial\mathbb{S}_d^+$, then there appears to be basically no hope to obtain a representation of the above type.*

4.6. Roots of Ornstein-Uhlenbeck processes

Now we turn to studying the behaviour of the roots of positive Ornstein-Uhlenbeck processes as defined in Section 4.4. Recall in particular that the driving Lévy process L is assumed to be a (matrix) subordinator.

Straightforward calculations based on Theorems 4.5.1 and 4.5.4 establish the following result for a univariate OU process $dX_t = -\lambda X_{t-} dt + dL_t$.

Proposition 4.6.1. *Let $(X_t)_{t\in\mathbb{R}^+}$ be a positive univariate process of Ornstein-Uhlenbeck type driven by a Lévy subordinator L with drift γ and associated Poisson random measure μ. Then for $0 < r < 1$ the unique positive r-th power $Y = X^r$ is of finite variation and has the following representation:*

$$\begin{aligned} dY_t &= \left(-\lambda r X_{t-}^r + \gamma r X_{t-}^{r-1}\right) dt + \int_{\mathbb{R}^+\setminus\{0\}} \left((X_{t-}+x)^r - (X_{t-})^r\right) \mu(dt,dx) \\ &= \left(-\lambda r Y_{t-} + \gamma r Y_{t-}^{1-1/r}\right) dt + \int_{\mathbb{R}^+\setminus\{0\}} \left((Y_{t-}^{1/r}+x)^r - Y_{t-}\right) \mu(dt,dx), \end{aligned}$$

provided that the process X is locally bounded away from zero or the integrals $\int_0^t \gamma r X_{s-}^{r-1} ds$ and $\int_0^t \int_{\mathbb{R}^+\setminus\{0\}} \left((X_{s-}+x)^r - X_{s-}^r\right) \mu(ds,dx)$ exist a.s. for all $t \in \mathbb{R}$.

Before showing that the conditions are actually satisfied for all positive OU processes, we show this for stationary ones, as this case is of particular interest and the proof is very straightforward. Recall in particular that a stationary OU process can be represented as $\int_{-\infty}^t e^{-\lambda(t-s)} dL_s$, where the driving Lévy process has a finite logarithmic moment.

Proposition 4.6.2. *Let X be a stationary positive process of OU type with driving Lévy process L (having drift γ and non-zero Lévy measure ν). Then it is locally bounded away from zero.*

The same holds for any positive Ornstein-Uhlenbeck process X with $X_0 > 0$ a.s.

Proof. Let us first consider the stationary case. If $\gamma > 0$, we see from Proposition 4.4.6 that $X_t \geq \gamma/\lambda > 0$ for all t, which implies that X is locally bounded away from 0. Otherwise note first that $X_t \geq e^{-\lambda t} X_0$ for all $t \geq 0$ and that the stationary distribution is self-decomposable (cf. Sato (1999, Theorem 17.5)). As the driving Lévy process has a non-zero Lévy measure the stationary distribution must be non-trivial and thus by Sato (1999, Example 27.8) absolutely continuous with respect to the Lebesgue measure. Therefore we have $X_0 > 0$ a.s. Hence, there is a.s. a sequence of stopping times $(T_n)_{n\in\mathbb{N}}$ increasing to infinity such

that $X_t \geq 1/n$ for all $t \in [0, T_n)$ (actually we can set $T_n = \frac{\ln(X_0 n)}{\lambda}$), which gives that X is locally bounded away from the origin.

Obviously, the same arguments apply in the non-stationary case. □

Proposition 4.6.3. *Let $(X_t)_{t\in\mathbb{R}^+}$ be a positive univariate process of Ornstein-Uhlenbeck type driven by a Lévy subordinator L with drift γ and associated Poisson random measure μ. Then the integrals $\int_0^t \gamma r X_{s-}^{r-1} ds$ and $\int_0^t \int_{\mathbb{R}^+\setminus\{0\}} \big((X_{s-}+x)^r - X_{s-}^r\big)\mu(ds,dx)$ exist for all $t \in \mathbb{R}$ and $0 < r < 1$.*

Proof. To show this we introduce the auxiliary process $Z_t = X_0 + \int_0^t e^{\lambda s} dL_s$ for $t \in \mathbb{R}^+$. It holds that $Z_t = e^{\lambda t} X_t$ for all $t \in \mathbb{R}^+$, the process is monotonically increasing and $dZ_t = e^{\lambda t}\gamma dt + \int_{\mathbb{R}^+\setminus\{0\}} e^{\lambda t} x \mu(dt, dx)$.

The increasingness implies the existence of the integral

$$\begin{aligned}\int_0^t \int_{\mathbb{R}^+\setminus\{0\}} \left((Z_{s-} + e^{\lambda s}x)^r - Z_{s-}^r\right) \mu(ds,dx) \\ = \int_0^t \int_{\mathbb{R}^+\setminus\{0\}} e^{\lambda r s} \left((X_{s-}+x)^r - X_{s-}^r\right) \mu(ds,dx).\end{aligned}$$

Since $0 < \min\{1, e^{\lambda r t}\} \leq e^{\lambda r s} \leq \max\{1, e^{\lambda r t}\}$ for all $s \in [0,t]$, this shows that the integral $\int_0^t \int_{\mathbb{R}^+\setminus\{0\}} \left((X_{s-}+x)^r - X_{s-}^r\right)\mu(ds,dx)$ exists for all $t \in \mathbb{R}$.

Obviously, $Z_t \geq \int_0^t e^{\lambda s}\gamma ds = \frac{\gamma}{\lambda}\left(e^{\lambda t} - 1\right)$. Assuming first $\lambda \geq 0$, this gives

$$\begin{aligned}\int_0^t \gamma r X_{s-}^{r-1} ds &= \int_0^t r\gamma e^{-\lambda(r-1)s} Z_{s-}^{r-1} ds \leq \int_0^t r\gamma^r \lambda^{1-r} e^{-\lambda(r-1)s} \left(e^{\lambda s} - 1\right)^{r-1} ds \\ &= r\gamma^r \lambda^{1-r} e^{-\lambda(r-1)t} \int_0^t \left(e^{\lambda s} - 1\right)^{r-1} ds.\end{aligned}$$

Noting that $e^{\lambda s} - 1 \geq s$ for all $s \in \mathbb{R}^+$, this implies the existence of $\int_0^t \gamma r X_{s-}^{r-1} ds$ for all $t \in \mathbb{R}^+$ immediately. In the case $\lambda < 0$ one calculates

$$\int_0^t \gamma r X_{s-}^{r-1} ds \leq r\gamma^r |\lambda|^{1-r} \int_0^t (e^{-\lambda s} - 1)^{r-1} ds,$$

which likewise implies the existence of the integral for all $t \in \mathbb{R}^+$. □

Remark 4.6.4. *For a driftless driving Lévy process we see from*

$$dY_t = -\lambda r Y_{t-} dt + \int_{\mathbb{R}^+\setminus\{0\}} \left((Y_{t-}^{1/r} + x)^r - Y_{t-}\right) \mu(dt, dx) \tag{4.6.1}$$

for $0 < r < 1$ that the drift part is again that of an Ornstein-Uhlenbeck process.

Moreover, observe that (4.6.1) *gives a stochastic differential equation (cf. Applebaum (2004) for information on this type of SDEs) for the r-th power, $0 < r < 1$, of the OU process. Since the derivative of $y \mapsto (y^{1/r} + x)^r$ is given by $y \mapsto \left(y^{1/r}/(y^{1/r} + x)\right)^{1-r}$ and is thus obviously bounded by one for all $x \in \mathbb{R}^+$, the function $y \mapsto (y^{1/r} + x)^r$ is (globally) Lipschitz. This implies that for any initial value Y_0 the SDE* (4.6.1) *has a unique solution.*

If $\gamma > 0$ one likewise has the SDE

$$dY_t = \left(-\lambda r Y_{t-} + \gamma r Y_{t-}^{1-1/r}\right) dt + \int_{\mathbb{R}^+\setminus\{0\}} \left((Y_{t-}^{1/r} + x)^r - Y_{t-}\right) \mu(dt, dx)$$

for the r-th power of the OU process. In this case one has only local Lipschitz continuity in $\mathbb{R}^+$ for $y \mapsto \gamma r y^{1-1/r}$. In such a set-up results on the existence of unique solutions are still obtainable, but as these would require a rather lengthy discussion, we refrain from giving any details.

From the following proposition we see that the r-th power of a positive OU process X with $\gamma = 0$ has a representation quite similar to the one for the OU process given by $X_t = e^{-\lambda t} X_0 + \int_0^t \int_{\mathbb{R}^+\setminus\{0\}} e^{-\lambda(t-s)} x \mu(ds, dx)$:

Proposition 4.6.5. *Assume that $\gamma = 0$, $0 < r < 1$ and $X_0 \geq 0$ a.s. Then the process $Y = X^r$ can be represented as*

$$\begin{aligned} Y_t &= e^{-\lambda r t} X_0^r + \int_0^t \int_{\mathbb{R}^+\setminus\{0\}} \left((e^{-\lambda(t-s)} X_{s-} + e^{-\lambda(t-s)} x)^r - (e^{-\lambda(t-s)} X_{s-})^r\right) \mu(ds, dx) \\ &= e^{-\lambda r t} X_0^r + \int_0^t \int_{\mathbb{R}^+\setminus\{0\}} e^{-\lambda r(t-s)} \left((X_{s-} + x)^r - X_{s-}^r\right) \mu(ds, dx). \end{aligned}$$

Proof. As in the proof of Proposition 4.6.3 we use the auxiliary process $Z_t = X_0 + \int_0^t \int_{\mathbb{R}^+\setminus\{0\}} e^{\lambda s} x \mu(ds, dx)$. For the process Z_t^r we obtain from Proposition 4.5.4

$$\begin{aligned} dZ_t^r &= \int_{\mathbb{R}^+\setminus\{0\}} \left((Z_{s-} + e^{\lambda s} x)^r - Z_{s-}^r\right) \mu(ds, dx) \\ &= \int_{\mathbb{R}^+\setminus\{0\}} \left((e^{\lambda s} X_{s-} + e^{\lambda s} x)^r - (e^{\lambda s} X_{s-})^r\right) \mu(ds, dx). \end{aligned}$$

Thus,

$$Z_t^r = X_0^r + \int_0^t \int_{\mathbb{R}^+\setminus\{0\}} \left((e^{\lambda s} X_{s-} + e^{\lambda s} x)^r - (e^{\lambda s} X_{s-})^r\right) \mu(ds, dx).$$

This implies the assertion via $Y_t = X_t^r = e^{-\lambda r t} Z_t^r$. □

Finally let us improve the representation of Proposition 4.6.5 for a stationary Ornstein-Uhlenbeck process.

Proposition 4.6.6. *Let X be a stationary process of OU type with driving Lévy subordinator L (having non-zero Lévy measure) with a vanishing drift γ. Then for $0 < r < 1$ the stationary process $Y = X^r$ can be represented as*

$$\begin{aligned} Y_t &= \int_{-\infty}^t \int_{\mathbb{R}^+\setminus\{0\}} \left((e^{-\lambda(t-s)} X_{s-} + e^{-\lambda(t-s)} x)^r - (e^{-\lambda(t-s)} X_{s-})^r\right) \mu(ds, dx) \\ &= \int_{-\infty}^t \int_{\mathbb{R}^+\setminus\{0\}} e^{-\lambda r(t-s)} \left((X_{s-} + x)^r - X_{s-}^r\right) \mu(ds, dx). \end{aligned}$$

Proof. Note that as in Proposition 4.6.5 we have that

$$Y_t = e^{-\lambda r(t-\tau)}\sqrt{X_\tau} + \int_\tau^t \int_{\mathbb{R}^+\setminus\{0\}} e^{-\lambda r(t-s)} \left((X_{s-}+x)^r - X_{s-}^r\right) \mu(ds,dx)$$

holds for all $\tau \in (-\infty, 0]$. Letting τ go to $-\infty$ we see that $e^{-\lambda r(t-\tau)} X_\tau^r$ goes to zero, since for any stationary OU process $e^{-\lambda(t-\tau)} X_\tau$ converges to zero. As, moreover, the left hand side is independent of τ, the integral

$$\int_\tau^t \int_{\mathbb{R}^+\setminus\{0\}} e^{-\lambda r(t-s)} \left((X_{s-}+x)^r - X_{s-}^r\right) \mu(ds,dx)$$

exists for all $\tau \in (-\infty, 0]$ and is increasing for decreasing τ, the limit of the integrals for $\tau \to -\infty$ exists. This implies the result immediately. □

Having analysed the univariate positive Ornstein-Uhlenbeck processes in depth, let us now turn to multivariate positive definite ones and see which results can be extended. Here we state all results again only for the square root, but extensions to more general powers are immediate. The general result on the representation of the square root follows immediately from the results of Section 5.2.

Proposition 4.6.7. *Let $(X_t)_{t\in\mathbb{R}^+}$ be an $\mathbb{S}_d^+$-valued process of Ornstein-Uhlenbeck type driven by a matrix subordinator L with drift $\gamma \in \mathbb{S}_d^+$ and associated Poisson random measure μ. Then the unique positive square root $Y = \sqrt{X}$ is of finite variation and has the following representation:*

$$\begin{aligned} dY_t &= \mathbf{X}_{t-}^{-1}\left(AX_{t-} + X_{t-}A^* + \gamma\right) dt + \int_{\mathbb{S}_d^+\setminus\{0\}} \left(\sqrt{X_{t-}+x} - \sqrt{X_{t-}}\right) \mu(dt,dx) \\ &= \mathbf{Y}_{t-}^{-1}\left(AY_{t-}^2 + Y_{t-}^2A^* + \gamma\right) dt + \int_{\mathbb{S}_d^+\setminus\{0\}} \left(\sqrt{Y_{t-}^2+x} - Y_{t-}\right) \mu(dt,dx), \end{aligned}$$

provided that the process X is locally bounded within $\mathbb{S}_d^{++}$ or the integrals

$$\int_0^t \mathbf{X}_{s-}^{-1}\left(AX_{s-} + X_{s-}A^* + \gamma\right) ds \text{ and } \int_0^t \int_{\mathbb{S}_d^+\setminus\{0\}} \left(\sqrt{X_{s-}+x} - \sqrt{X_{s-}}\right) \mu(ds,dx)$$

exist a.s. for all $t \in \mathbb{R}$. Here, $\mathbf{X}_{t-}$ is the linear operator $Z \mapsto \sqrt{X_{t-}}Z + Z\sqrt{X_{t-}}$ and $\mathbf{Y}_{t-}$ the map $Z \mapsto Y_{t-}Z + ZY_{t-}$

For stationary OU processes one can again establish local boundedness, provided the driving Lévy process is non-degenerate.

Proposition 4.6.8. *Let X be a stationary positive semi-definite OU process and assume that the driving Lévy process L has drift $\gamma \in \mathbb{S}_d^{++}$ or Lévy measure ν such that $\nu(\mathbb{S}_d^{++}) > 0$. Then the process X is locally bounded within $\mathbb{S}_d^{++}$.*

The same holds for any positive definite OU process with initial value $X_0 \in \mathbb{S}_d^{++}$ a.s.

Proof. In the stationary case Theorem 4.4.9 implies $X_0 \in \mathbb{S}_d^{++}$ a.s. From (4.4.6) we thus always obtain that $X_t \geq e^{At}X_0e^{A^*t} \in \mathbb{S}_d^{++}$ for all $t \in \mathbb{R}$. As $\min\left(\sigma\left(e^{At}X_0e^{A^*t}\right)\right)$ is continuous in t and strictly positive, $\min\sigma\left(e^{At}X_0e^{A^*t}\right)$ is locally bounded away from 0; in particular, $T_n := \inf\{t \in \mathbb{R}^+ : e^{At}X_0e^{A^*t} < \frac{1}{n}I_d\}$ defines a sequence of stopping times that a.s. increases to infinity. But this implies $X_t \geq \frac{1}{n}I_d$ for all $t \in [0, T_n)$. Together with the local boundedness of X and the fact that sets of the form $\{x \in \mathbb{S}_d^+ : x \geq \epsilon I_d,\ \|x\| \leq K\}$ with $\epsilon, K > 0$ are convex and compact, this establishes the local boundedness of X within $\mathbb{S}_d^{++}$. □

In general we cannot obtain the existence of the relevant integrals for all positive definite OU processes, but the following proposition covers many cases of interest.

Proposition 4.6.9. *Let X be a positive definite OU process driven by a matrix subordinator L with drift γ and Lévy measure ν. Then the integral*

$$\int_0^t \mathbf{X}_{s-}^{-1}\left(AX_{s-} + X_{s-}A^* + \gamma\right) ds$$

exists a.s. for all $t \in \mathbb{R}$ provided $\gamma \in \mathbb{S}_d^{++}$ or $\gamma = 0$, $X_0 = 0$ and L is a compound Poisson process with $\nu(\mathbb{S}_d^+ \backslash \mathbb{S}_d^{++}) = 0$. Furthermore, the integral $\int_0^t \int_{\mathbb{S}_d^+\backslash\{0\}} \left(\sqrt{X_{s-} + x} - \sqrt{X_{s-}}\right)\mu(ds, dx)$ exist a.s. for all $t \in \mathbb{R}$, provided L is compound Poisson (with drift) or $\int_{0<\|x\|_2\leq 1} \sqrt{\|x\|_2}\nu(dx)$ is finite.

Proof. Let us first consider the second integral. Then $\int_{0<\|x\|_2\leq 1} \sqrt{\|x\|_2}\nu(dx) < \infty$ is trivially satisfied for any compound Poisson process and so Lemma 4.5.27 gives the result.

If $\gamma = 0$, $X_0 = 0$ and L is a compound Poisson process, $X_t = 0$ for all $t \in [0, T)$ where T denotes the first jump time of L. So the integral

$$\int_0^t \mathbf{X}_{s-}^{-1}\left(AX_{s-} + X_{s-}A^* + \gamma\right) ds$$

exists a.s. for all $t \in [0, T)$. The condition $\nu(\mathbb{S}_d^+ \backslash \mathbb{S}_d^{++}) = 0$ ensures that the first jump ΔL_T is a.s. strictly positive definite and hence $X_T \in \mathbb{S}_d^{++}$ a.s. Using basically the same arguments as in Proposition 4.6.8 this shows that the integral $\int_0^t \mathbf{X}_{s-}^{-1}\left(AX_{s-} + X_{s-}A^* + \gamma\right) ds$ exists also a.s. for all $t \in [T, \infty)$, which concludes the proof of this case.

Assume now that $\gamma \in \mathbb{S}_d^{++}$. We have

$$X_t \geq \int_0^t e^{A(t-s)}\gamma e^{A^*(t-s)} ds \geq \int_0^t \min\left(\sigma\left(e^{A(t-s)}\gamma e^{A^*(t-s)}\right)\right) I_d ds.$$

But $e^{A(t-s)}\gamma e^{A^*(t-s)} \in \mathbb{S}_d^{++}$ for all $t, s \in \mathbb{R}^+$ and so for any $M \in \mathbb{R}^+$ continuity and compactness ensures the existence of a constant $k_M > 0$ such that

$$\min\left(\sigma\left(e^{A(t-s)}\gamma e^{A^*(t-s)}\right)\right) \geq k_M$$

for all $t, s \in [0, M]$. Hence, $X_t \geq k_M t$ for all $t \in [0, M]$. Using the same matrix analytical arguments as in the proof of Corollary 4.5.29, this implies $\|\mathbf{X}_{t-}^{-1}\|_{\hat{2}} \leq \frac{1}{2\sqrt{k_M}} t^{-1/2}$ for all $t \in [0, M]$. Moreover, as X is locally bounded there is a.s. a constant K_M such that

$\|X_t\|_{\tilde{2}} \leq K_M$ for all $t \in [0, M]$. (Here we have fixed $\omega \in \Omega$, but recall that we can argue path-wise.) Since $\int_0^t \frac{\|\mathbf{A}\|_{\tilde{2}} K_M + \|\gamma\|_{\tilde{2}}}{2\sqrt{k_M}} s^{-1/2} ds$ is finite for all $t \in [0, M]$, where $\mathbf{A}$ is the linear operator $M_d(\mathbb{R}) \to M_d(\mathbb{R})$, $Z \mapsto AZ + ZA^*$, majorized convergence gives that $\int_0^t \mathbf{X}_{s-}^{-1} (AX_{s-} + X_{s-}A^* + \gamma)\, ds$ exists a.s. for all $t \in [0, M]$. As $M \in \mathbb{R}^+$ was arbitrary, this concludes the proof. □

However, one can again show that the square root of a positive definite OU process X with $\gamma = 0$ has a representation similar to the one for the OU process given by $X_t = e^{At} X_0 e^{A^* t} + \int_0^t \int_{\mathbb{S}_d^+ \setminus \{0\}} e^{A(t-s)} x e^{A^*(t-s)} \mu(ds, dx)$:

Proposition 4.6.10. *Assume that $\gamma = 0$ and $X_0 \geq 0$ a.s. Then the process $Y = \sqrt{X}$ can be represented as*

$$Y_t = \sqrt{e^{At} X_0 e^{A^* t}} + \int_0^t \int_{\mathbb{S}_d^+ \setminus \{0\}} \left(\sqrt{e^{A(t-s)}(X_{s-} + x) e^{A^*(t-s)}} - \sqrt{e^{A(t-s)} X_{s-} e^{A^*(t-s)}} \right) \mu(ds, dx).$$

Proof. Let $(Z_u)_{u \in \mathbb{R}^+}$ be the auxiliary process given by $Z_u = e^{A(t-u)} X_u e^{A^*(t-u)}$ where $t \in \mathbb{R}^+$ is fixed. Then $Z_u = e^{At} X_0 e^{A^* t} + \int_0^u \int_{\mathbb{S}_d^+ \setminus \{0\}} e^{A(t-s)} x e^{A^*(t-s)} \mu(ds, dx)$ is $\mathbb{S}_d^+$-increasing. Using Theorem 4.5.19 and Lemma 4.5.26 this implies that

$$\sqrt{Z_u} = \sqrt{e^{At} X_0 e^{A^* t}} + \int_0^u \int_{\mathbb{S}_d^+ \setminus \{0\}} \left(\sqrt{Z_{s-} + e^{A(t-s)} x e^{A^*(t-s)}} - \sqrt{Z_{s-}} \right) \mu(ds, dx).$$

Since $X_t = Z_t$ and $Z_{s-} = e^{A(t-s)} X_{s-} e^{A^*(t-s)}$, this immediately concludes the proof. □

Finally let us improve the above representation for a stationary positive definite Ornstein-Uhlenbeck process.

Proposition 4.6.11. *Let X be a stationary process of OU type with driving matrix subordinator L with a vanishing drift γ. Then the stationary process $Y = \sqrt{X}$ can be represented as*

$$Y_t = \int_{-\infty}^t \int_{\mathbb{S}_d^+ \setminus \{0\}} \left(\sqrt{e^{A(t-s)}(X_{s-} + x) e^{A^*(t-s)}} - \sqrt{e^{A(t-s)} X_{s-} e^{A^*(t-s)}} \right) \mu(dx, ds).$$

Proof. Follows from Proposition 4.6.10 using the same arguments as in the proof of Proposition 4.6.6. □

5. A Multivariate Generalization of the Ornstein-Uhlenbeck Stochastic Volatility Model

5.1. Introduction

A wide range of different univariate continuous-time stochastic volatility models has been developed in the financial literature aiming at capturing the most distinct features of the price process of a single financial asset. The adequacy of such models is essential for applications such as risk management and derivative pricing for which closed form expressions of the integrated variance process are desirable.

In a multivariate context modelling becomes even more challenging. Next to capturing the individual dynamics the model also needs to reproduce the comovements and spill-over effects across different assets. In particular, knowing the correlation structure is crucial for financial decision-making, such as portfolio risk management, asset allocation or the pricing of multi-asset options. In addition to those requirements, there also arise some technical issues in the multivariate setting. One is given by the necessity of a positive semi-definite covariance matrix. For stochastic volatility models this implies that the instantaneous covariance should be specified as a positive semi-definite process. Moreover, if the dimension of the return vector increases the number of parameters in the model should not explode. Hence, a parsimonious but at the same time accurate and flexible specification is needed.

Given these challenges the theoretical literature on multivariate stochastic volatility models has developed over the last few years, where the main focus was on discrete-time models as an alternative to the multivariate GARCH models, see e.g. Chib, Nardari and Shephard (2006) and Harvey, Ruiz and Shephard (1994). In contrast to the discrete time models, however, the continuous-time specification allows to infer the implied dynamics and properties of the estimated model at various frequencies differing from the one used in the estimation. This is important, inter alia, for forecasting the covariance over short term intervals, where the estimates can be based on lower frequent data. Moreover, the continuous time specification is very important from the asset pricing perspective. Despite these advantages, however, we are aware of only a few papers considering continuous time multivariate stochastic volatility models. In particular, Hubalek and Nicolato (2005) and Lindberg (2005) adopt a factor approach in which the volatility factors are independent and follow univariate positive non-Gaussian Ornstein-Uhlenbeck type (OU type hereafter) processes. But the flexibility of these models is accompanied by the difficulty to achieve identification, which is complicating the empirical application of these models. In Gourieroux (2006) the stochastic volatility is not driven by univariate factors, but the full covariance matrix is specified as the sum of outer products of Gaussian vector OU processes, which is referred to as a Wishart autoregressive process. Although this model provides closed form expressions for many applications, it also lacks a closed form representation for the

integrated covariance process.

In this chapter we therefore introduce a new continuous-time multivariate stochastic volatility model that is shown to meet the above mentioned requirements while providing a closed form and very simple structure for the integrated covariance process. We assume the general d-dimensional stochastic volatility model to be given by

$$dY_t = (\mu + \Sigma_t \beta)dt + \Sigma_t^{1/2} dW_t,\ Y_0 = 0, \tag{5.1.1}$$

where Y denotes the d-dimensional logarithmic stock price process, $\mu, \beta \in \mathbb{R}^d$ are the instantaneous drift and risk premium parameters, respectively, $(W_t)_{t\in\mathbb{R}^+}$ denotes a d-dimensional standard Brownian motion and $(\Sigma_t)_{t\in\mathbb{R}^+}$ is an adapted, stationary and square-integrable stochastic volatility process with values in the positive semi-definite matrices $\mathbb{S}_d^+$ being independent of $(W_t)_{t\in\mathbb{R}^+}$. Throughout the remainder of this chapter we refer to this model when we write "stochastic volatility model". Observe that this is a more restrictive use of the term "stochastic volatility (model)" than in all other parts of this thesis. The same general model has also been stated in e.g. Barndorff-Nielsen and Shephard (2001b), Barndorff-Nielsen, Nicolato and Shephard (2002) and Lindberg (2005). In the following the stochastic volatility process $(\Sigma_t)_{t\in\mathbb{R}^+}$ is usually given by a Lévy-driven positive semi-definite OU type process which was introduced in Chapter 4 and which is a multivariate extension of the positive non-Gaussian OU type process used in the context of univariate stochastic volatility models in Barndorff-Nielsen and Shephard (2001b, 2002). We therefore refer to this model as the "multivariate Ornstein-Uhlenbeck stochastic volatility model" in the sequel. Whereas the existing models mentioned above are factor models of some kind, the stochastic volatility process of our model is directly specified as a process in the positive semi-definite matrices.

It turns out that our model possesses many attractive features which are mainly a result of our stochastic volatility specification. First, let us note that similar to Barndorff-Nielsen and Shephard (2001b) our general multivariate specification as given in equation (5.1.1) implies that returns are scaled mixtures of multivariate normals with the scaling given by the integrated covariance matrix, so that the observed (semi-)heavy-tailedness can be reproduced. In addition, specifying the stochastic volatility by Lévy-driven positive semi-definite OU type processes provides a flexible dependence structure for the volatility, which can be made even more flexible by considering superpositions of these processes. Furthermore, we show that the vectorized outer product of the returns (the "squared returns" in a multivariate setting) of the multivariate OU type stochastic volatility model follows an ARMA(1,1) process. Furthermore, the first and second order moments of this variable are available in closed form – a result that facilitates the estimation of our model.

The remainder of the chapter is structured as follows. The next section establishes our notation. As our model builds on the positive semi-definite matrix process of OU type, Section 5.3 presents some important properties of these processes recalling and considerably extending the results of Section 4.4. The multivariate OU type stochastic volatility model is introduced in Section 5.4 and its properties are analysed in detail. Section 5.5 presents an empirical illustration and Section 5.6 finally concludes.

5.2. Notation

Throughout this chapter we write $\mathbb{R}^+$ for the positive real numbers including zero, $\mathbb{R}^{++}$ when zero is excluded and we denote the set of real $m \times n$ matrices by $M_{m,n}(\mathbb{R})$. If $m = n$

we simply write $M_n(\mathbb{R})$ and denote the group of invertible $n \times n$ matrices by $GL_n(\mathbb{R})$, the linear subspace of symmetric matrices by $\mathbb{S}_n$, the (closed) positive semi-definite cone by $\mathbb{S}_n^+$ and the open (in $\mathbb{S}_n$) positive definite cone by $\mathbb{S}_n^{++}$. I_n stands for the $n \times n$ identity matrix, $\sigma(A)$ for the spectrum (the set of all eigenvalues) of a matrix $A \in M_n(\mathbb{R})$ and $\rho(A)$ for its spectral radius. The natural ordering on the symmetric $n \times n$ matrices is denoted by $\leq$, i.e. for $A, B \in \mathbb{S}_n$ we have that $A \leq B$, if and only if $B - A \in \mathbb{S}_n^+$. The tensor (Kronecker) product of two matrices A, B is written as $A \otimes B$. vec denotes the well-known vectorisation operator that maps the $n \times n$ matrices to $\mathbb{R}^{n^2}$ by stacking the columns of the matrices below one another. For more information regarding the tensor product and vec operator we refer to Horn and Johnson (1991, Chapter 4). Likewise vech : $\mathbb{S}_d \to \mathbb{R}^{d(d+1)/2}$ denotes the "vector-half" operator that stacks the columns of the lower triangular part of a symmetric matrix below another. Finally, A^T is the transpose of a matrix $A \in M_n(\mathbb{R})$. For a matrix A we denote by A_{ij} the element in the i-th row and j-th column and this notation is extended to processes in a natural way.

Regarding all random variables and processes we assume that they are defined on a given appropriate filtered probability space $(\Omega, \mathcal{F}, P, (\mathcal{F}_t))$ satisfying the usual hypotheses. For random functions and measures we usually do not state the dependence on $\omega \in \Omega$ explicitly.

Norms of vectors or matrices are denoted by $\|\cdot\|$. If the norm is not specified, then it is irrelevant which particular norm is used.

Furthermore, we employ an intuitive notation with respect to the (stochastic) integration with matrix-valued integrators referring to any of the standard texts (e.g. Protter (2004) or Øksendal (1998) regarding Brownian motion) for a comprehensive treatment of the theory of stochastic integration. Let $(L_t)_{t\in\mathbb{R}^+}$ in $M_{n,r}(\mathbb{R})$ be a semi-martingale and $(A_t)_{t\in\mathbb{R}^+}$ in $M_{m,n}(\mathbb{R})$, $(B_t)_{t\in\mathbb{R}^+}$ in $M_{r,s}(\mathbb{R})$ be adapted integrable (w.r.t. L) processes. Then we denote by $\int_0^t A_s dL_s B_s$ the matrix C_t in $M_{m,s}(\mathbb{R})$ which has ij-th element $C_{ij,t} = \sum_{k=1}^n \sum_{l=1}^r \int_0^t A_{ik,s} B_{lj,s} dL_{kl,s}$. Equivalently such an integral can be understood in the sense of Métivier and Pellaumail (1980b), resp. Métivier (1982), by identifying it with the integral $\int_0^t \mathbf{A}_s dL_s$ with $\mathbf{A}_t$ being for each fixed t the linear operator $M_{n,r}(\mathbb{R}) \to M_{m,s}(\mathbb{R})$, $X \mapsto A_t X B_t$. Moreover, we always denote by $\int_a^b$ with $a \in \mathbb{R} \cup \{-\infty\}, b \in \mathbb{R}$ the integral over the half-open interval $(a, b]$ for notational convenience. If $b = \infty$ the integral is understood to be over (a, b).

For a set A the indicator function is denoted by I_A and the function $\log^+$ is defined as $\max(\log(x), 0)$.

5.3. Positive semi-definite matrix processes of Ornstein-Uhlenbeck type

In this section we briefly review the positive semi-definite OU type processes introduced in Chapter 4 where detailed proofs have been given. Additionally we present some new results for this class of processes which are particularly relevant in the context of our OU type stochastic volatility model, viz. properties of the stationary distribution and the marginal dynamics of the individual components.

5.3.1. Definition and probabilistic properties

In the following we first provide a definition of the positive semi-definite OU type processes and characterize their stationary distribution.

The construction of these processes is based on a special type of matrix-valued Lévy processes studied in detail in Barndorff-Nielsen and Pérez-Abreu (2007). For the relevant background on Lévy processes we refer to any of the standard references, for instance, Sato (1999).

Definition 5.3.1. *An $\mathbb{S}_d$-valued Lévy process $L = (L_t)_{t\in\mathbb{R}^+}$ is said to be a* matrix subordinator*, if $L_t - L_s \in \mathbb{S}_d^+$ for all $s, t \in \mathbb{R}^+$ with $t > s$.*

Matrix subordinators are a generalization of the concept of univariate Lévy subordinators to the matrix case, in particular, they are simply the same as Lévy subordinators for $d = 1$. As in the univariate case there are a lot of very different concrete examples of matrix subordinators. Barndorff-Nielsen and Pérez-Abreu (2007), for example, discuss matrix subordinators which are generalizations of stable, tempered stable and Gamma subordinators. These are examples having infinite activity. Of course, compound Poisson, i.e. finite activity, matrix subordinators can easily be constructed using any probability distribution on $\mathbb{S}_d^+$ (see e.g. Gupta and Nagar (2000) for some examples) for the jumps. In this context it should be noted that the outer product XX^T of any vector random variable X is positive semi-definite. Specifying the diagonal elements of the matrix process as (possibly dependent) univariate subordinators forming together a d-dimensional Lévy process and setting the off-diagonal elements to zero, leads to another simple example of a matrix subordinator (referred to as a diagonal matrix subordinator).

Importantly, it can easily be shown that the paths of a matrix subordinator are $\mathbb{S}_d^+$-increasing and of finite variation. Moreover, the trace $\mathrm{tr}(L)$ is a one-dimensional (Lévy) subordinator.

Based on matrix subordinators the existence of OU type processes assuming values in the positive semi-definite matrices is ensured by the following theorem, where the Lévy process L is extended to a Lévy process $(L_t)_{t\in\mathbb{R}}$ starting in the infinite past in the usual way.

Theorem 5.3.2 (Theorem 4.4.5 of Chapter 4)**.** *Let L be a matrix subordinator with $E(\log^+ \|L_1\|) < \infty$ and $A \in M_d(\mathbb{R})$ such that $\sigma(A) \subset (-\infty, 0) + i\mathbb{R}$. Then the stochastic differential equation of OU type*

$$d\Sigma_t = (A\Sigma_{t-} + \Sigma_{t-}A^T)dt + dL_t$$

has a unique stationary solution

$$\Sigma_t = \int_{-\infty}^{t} e^{A(t-s)} dL_s e^{A^T(t-s)}$$

or, in vectorial representation,

$$\mathrm{vec}(\Sigma_t) = \int_{-\infty}^{t} e^{(I_d\otimes A + A\otimes I_d)(t-s)} d\mathrm{vec}(L_s).$$

Moreover, $\Sigma_t \in \mathbb{S}_d^+$ for all $t \in \mathbb{R}$.

Remark 5.3.3. *It is shown in Appendix A that the linear operator $\mathbf{A} : \mathbb{S}_d \to \mathbb{S}_d$ given by $X \mapsto AX + XA^T$ with $A \in M_d(\mathbb{R})$ uniquely identifies A. Actually, to identify A it is already sufficient to know the values of $\mathbf{A}E_{ii}$ for $i = 1, \ldots, d$ where E_{ii} are the $d \times d$ matrices with only zero entries except for one entry of one at the i-th diagonal element.*

Moreover, the above mentioned appendix establishes that all linear operators $\mathbf{A}$ *which satisfy* $\exp(\mathbf{A}t)(\mathbb{S}_d) = \mathbb{S}_d^+$ $(*)$ *for all* $t \in \mathbb{R}$ *are necessarily of the form* $X \mapsto AX + XA^T$ *for some* $A \in M_d(\mathbb{R})$*. The property* $(*)$ *is important for the well-definedness of positive semi-definite OU type processes. As already noted in Chapter 4 it seems to be rather problematic to characterize linear operators which satisfy* $\exp(\mathbf{A}t)(\mathbb{S}_d) \subset \mathbb{S}_d^+$ $(*)$*. Obviously the equality also immediately implies that the distribution of the OU type process* Σ *is not concentrated on any proper linear subspace of* $\mathbb{S}_d$ *when the distribution of the driving Lévy process* L_t *is not concentrated on any proper linear subspace.*

Recall from Barndorff-Nielsen and Pérez-Abreu (2007) that any matrix subordinator $(L_t)_{t\in\mathbb{R}}$ can be represented as

$$L_t = \gamma_L t + \int_0^t \int_{\mathbb{S}_d^+\setminus\{0\}} x\mu(ds, dx) \tag{5.3.1}$$

where $\gamma_L \in \mathbb{S}_d^+$ is a deterministic drift and $\mu(ds, dx)$ an extended Poisson random measure on $\mathbb{R}^+ \times \mathbb{S}_d^+$ (regarding the definitions of random measures and the integration theory with respect to them we refer to Jacod and Shiryaev (2003, Section II.1)). Note also, that the integral exists without compensating. Moreover, the expectation of μ factorises, i.e. $E(\mu(ds, dx)) = Leb(ds)\nu_L(dx)$, Leb denoting the Lebesgue measure and ν_L the Lévy measure of L. The above equation (5.3.1) can be restated in a differential form as

$$dL_t = \gamma_L dt + \int_{\mathbb{S}_d^+\setminus\{0\}} x\mu(dt, dx). \tag{5.3.2}$$

The obvious extension of this to a Lévy process $(L_t)_{t\in\mathbb{R}}$ having been started in the infinite past gives another representation of the above stationary OU process.

Proposition 5.3.4 (Theorem 4.4.6 of Chapter 4)**.** *The positive semi-definite Ornstein-Uhlenbeck process* Σ *given in Theorem 5.3.2 can equivalently be represented as*

$$\begin{aligned}\Sigma_t &= \int_{-\infty}^t \int_{\mathbb{S}_d^+\setminus\{0\}} e^{A(t-s)} x e^{A^T(t-s)} \mu(ds, dx) + \int_{-\infty}^t e^{A(t-s)} \gamma_L e^{A^T(t-s)} ds \\ &= \int_{-\infty}^t \int_{\mathbb{S}_d^+\setminus\{0\}} e^{A(t-s)} x e^{A^T(t-s)} \mu(ds, dx) - \mathbf{A}^{-1}\gamma_L\end{aligned}$$

where $\mathbf{A}^{-1}$ *is the inverse of the linear operator* $\mathbf{A} : M_d(\mathbb{R}) \to M_d(\mathbb{R})$, $X \mapsto AX + XA^T$ *which can be represented as* $\mathrm{vec}^{-1} \circ ((I_d \otimes A) + (A \otimes I_d))^{-1} \circ \mathrm{vec}$.

Next we provide a characterization of the stationary distribution of the matrix OU process. To this end observe that $\mathrm{tr}(X^T Y)$ (with $X, Y \in M_d(\mathbb{R})$ and tr denoting the usual trace functional) defines a scalar product on $M_d(\mathbb{R})$ (respectively, $\mathbb{S}_d$). Moreover, note that the vec operator links the scalar product on $M_d(\mathbb{R})$ ($\mathbb{S}_d$) with the scalar product on $\mathbb{R}^{d^2}$ via $\mathrm{tr}(X^T Y) = \mathrm{vec}(X)^T \mathrm{vec}(Y)$ and that the norm on $M_d(\mathbb{R})$ induced by this scalar product is the Froebenius norm.

This, in particular, implies that the driving matrix subordinator L at time $t \in \mathbb{R}^+$ has characteristic function (cf. also Barndorff-Nielsen and Pérez-Abreu (2007))

$$\mu_{L_t}(Z) = \exp\Big(t\psi_L(Z)\Big), \quad Z \in \mathbb{S}_d, \quad \text{where} \tag{5.3.3}$$

$$\psi_L(Z) := i\mathrm{tr}(\gamma_L Z) + \int_{\mathbb{S}_d^+\setminus\{0\}} (e^{i\mathrm{tr}(XZ)} - 1)\nu_L(dX). \tag{5.3.4}$$

Proposition 5.3.5 (Proposition 4.4.7 of Chapter 4)**.** *The stationary distribution of the positive semi-definite Ornstein-Uhlenbeck process Σ is infinitely divisible with characteristic function*

$$\begin{aligned}\hat{\mu}_\Sigma(Z) &= \exp\left(\int_0^\infty \psi_L\left(e^{A^T s} Z e^{As}\right) ds\right) \\ &= \exp\left(i\mathrm{tr}(\gamma_\Sigma Z) + \int_{\mathbb{S}_d^+\setminus\{0\}} (e^{i\mathrm{tr}(XZ)} - 1)\nu_\Sigma(dX)\right),\ Z \in \mathbb{S}_d,\end{aligned} \tag{5.3.5}$$

where

$$\gamma_\Sigma = -\mathbf{A}^{-1}\gamma_L \tag{5.3.6}$$

with $\mathbf{A}$ *defined as in Proposition 5.3.4 and*

$$\nu_\Sigma(E) = \int_0^\infty \int_{\mathbb{S}_d^+\setminus\{0\}} I_E(e^{As} x e^{A^T s})\nu_L(dx)ds \tag{5.3.7}$$

for all Borel sets E in $\mathbb{S}_d^+\setminus\{0\}$.

Assume that the driving Lévy process is square-integrable. Then the second order moment structure is given by

$$E(\Sigma_t) = \gamma_\Sigma - \mathbf{A}^{-1}\int_{\mathbb{S}_d^+\setminus\{0\}} y\nu_L(dy) = -\mathbf{A}^{-1}E(L_1) \tag{5.3.8}$$

$$\begin{aligned}\mathrm{var}(\mathrm{vec}(\Sigma_t)) &= \int_0^\infty e^{(A\otimes I_d + I_d\otimes A)t}\mathrm{var}(\mathrm{vec}(L_1))e^{(A^T\otimes I_d + I_d\otimes A^T)t}dt \\ &= -\mathcal{A}^{-1}\mathrm{var}(\mathrm{vec}(L_1))\end{aligned} \tag{5.3.9}$$

$$\mathrm{cov}(\mathrm{vec}(\Sigma_{t+h}), \mathrm{vec}(\Sigma_t)) = e^{(A\otimes I_d + I_d\otimes A)h}\mathrm{var}(\mathrm{vec}(\Sigma_t)), \tag{5.3.10}$$

where $t \in \mathbb{R}$ and $h \in \mathbb{R}^+$ and $\mathcal{A} : M_{d^2}(\mathbb{R}) \to M_{d^2}(\mathbb{R})$, $X \mapsto (A\otimes I_d + I_d\otimes A)X + X(A^T\otimes I_d + I_d\otimes A^T)$. The linear operator $\mathcal{A}$ can be represented as

$$\mathrm{vec}^{-1}\circ((I_{d^2}\otimes(A\otimes I_d + I_d\otimes A)) + ((A\otimes I_d + I_d\otimes A)\otimes I_{d^2}))\circ\mathrm{vec}.$$

Remark 5.3.6. *(i) Equation* (5.3.9) *can alternatively be restated as*

$$-\mathrm{var}(\mathrm{vec}(L_1)) = (A\otimes I_d + I_d\otimes A)\mathrm{var}(\mathrm{vec}(\Sigma_t)) + \mathrm{var}(\mathrm{vec}(\Sigma_t))(A^T\otimes I_d + I_d\otimes A^T).$$

(ii) We used the vec *operator above, as this clarifies the order of the elements of the (co)variance matrix. One might wonder why one does not use the "vector half" operator* vech *that stacks the columns of the lower diagonal part (including the diagonal) of a symmetric matrix below one another. Although this would avoid the redundancies in the covariance matrices* $\mathrm{var}(\mathrm{vec}(L_1))$ *and* $\mathrm{var}(\mathrm{vec}(\Sigma_t))$ *caused by the symmetry of L_1 and Σ_t, it seems to be rather disadvantageous when seeking explicit expressions, since, to the best of our knowledge, there are far less results from linear algebra available for the* vech *operator than for the* vec*-operator. Hence, we will use the* vec *operator throughout most of this paper and merely note that one can, of course, switch to the* vech *operator by simply picking the relevant components.*

However, when it comes to the existence of a density or estimation it is sometimes easier or necessary to resort to the vech *operator,* $\mathrm{vech} : \mathbb{S}_d \to \mathbb{R}^{\frac{d(d+1)}{2}}$*. Defining* $\mathbf{A}_{\mathrm{vech}} := \mathrm{vech}\circ$

$\mathbf{A} \circ \mathrm{vech}^{-1}$, *where* $\mathbf{A}$ *is understood as linear operator on* $\mathbb{S}_d$, *we have that* $d\mathrm{vech}(\Sigma_t) = \mathbf{A}_{\mathrm{vech}}\mathrm{vech}(\Sigma_t) + d\mathrm{vech}(L_t)$ *and that the stationary solution is given by*

$$\mathrm{vech}(\Sigma_t) = \int_{-\infty}^{t} e^{\mathbf{A}_{\mathrm{vech}}(t-s)} d\mathrm{vech}(L_s).$$

In principle, it is now straightforward how to adjust all other results to the vech*-notation. Observe, moreover, that if one denotes the* $d^2 \times \frac{d(d+1)}{2}$ *duplication matrix by* $\mathbf{D}$ *and the* $\frac{d(d+1)}{2} \times d^2$ *elimination matrix by* $\mathbf{E}$ *(see Lütkepohl (2005, Appendix A.12), for instance) then* $\mathbf{A}_{\mathrm{vech}} = \mathbf{E}\left(A \otimes I_d + I_d \otimes A\right)\mathbf{D}$ *and* $e^{\mathbf{A}_{\mathrm{vech}}t} = \mathbf{E}\left(e^{At} \otimes e^{At}\right)\mathbf{D}$.

In general the finiteness of the moments of the stationary distribution of the OU type process is completely characterized by the driving Lévy process.

Proposition 5.3.7. *Let* $(\Sigma_t)_{t\in\mathbb{R}}$ *be a stationary OU type process in* $\mathbb{S}_d^+$ *with driving matrix subordinator* L *which has Lévy measure* ν_L *and be* $r \in \mathbb{R}^{++}$. *Then* $E(\|\Sigma_0\|^r) < \infty$, *if and only if* $E(\|L_1\|^r) < \infty$ *or equivalently* $\int_{\mathbb{S}_d^+, \|x\|\geq 1} \|x\|^r \nu_L(dx) < \infty$.

Proof. Follows by a straightforward adaptation of the proof of Proposition 2.3.30 in Chapter 2 to the matrix case. □

Noteworthy, the finiteness of some moment of the Lévy process also ensures that the stationary OU type process exhibits a very nice dependence structure. Let us thus introduce the notions of strong and β-mixing:

Definition 5.3.8 (cf. Davydov (1973))**.** *A continuous time stationary stochastic process* $X = \{X_t\}_{t\in\mathbb{R}}$ *is called* strongly (or α-) mixing, *if*

$$\alpha_l := \sup\left\{|P(A \cap B) - P(A)P(B)| : A \in \mathcal{F}_{-\infty}^0, B \in \mathcal{F}_l^\infty\right\} \to 0$$

as $l \to \infty$, *where* $\mathcal{F}_{-\infty}^0 := \boldsymbol{\sigma}\left(\{X_t\}_{t\leq 0}\right)$ *and* $\mathcal{F}_l^\infty = \boldsymbol{\sigma}\left(\{X_t\}_{t\geq l}\right)$ *(*$\boldsymbol{\sigma}(\cdot)$ *denoting the generated* σ*-algebra).*

It is said to be β- mixing (or completely regular), *if*

$$\beta_l := E\left(\sup\left\{\left|P(B|\mathcal{F}_{-\infty}^0) - P(B)\right| : B \in \mathcal{F}_l^\infty\right\}\right) \to 0 \;\; as \;\; l \to \infty.$$

Note that $\alpha_l \leq \beta_l$ and thus any β-mixing process is strongly mixing (see e.g. Doukhan (1994)), which implies that many results regarding statistics can be applied.

Proposition 5.3.9. *Let* Σ *be an OU type process in* $\mathbb{S}_d^+$. *Then* Σ *is a temporally homogeneous strong Markov process.*

If Σ *is stationary and the driving Lévy process* L *with Lévy measure* ν_L *satisfies additionally*

$$\int_{\mathbb{S}_d^+, \|x\|\geq 1} \|x\|^r \nu_L(dx) < \infty \tag{5.3.11}$$

for some $r > 0$, *then the stationary OU type process* Σ *is* β*-mixing with mixing coefficients* $\beta_l = O(e^{-al})$ *for some* $a > 0$. *In particular,* Σ *is strongly (or* α*-)mixing with exponential rate and ergodic.*

Proof. Follows from Protter (2004, Theorem V.32) and Masuda (2004, Theorem 4.3). □

Moreover, it is possible to derive conditions under which the stationary OU type process Σ is almost surely strictly positive definite.

Theorem 5.3.10 (Theorem 4.4.9 of Chapter 4)**.** *If $\gamma_L \in \mathbb{S}_d^{++}$ or $\nu_L(\mathbb{S}_d^{++}) > 0$, then the stationary distribution P_Σ of Σ is concentrated on $\mathbb{S}_d^{++}$, i.e. $P_\Sigma(\mathbb{S}_d^{++}) = 1$.*

The stationary distributions of multivariate Ornstein-Uhlenbeck type processes are operator self-decomposable. The operator self-decomposable distributions (cf. Jurek and Mason (1993) for a comprehensive treatment) form an important subclass of the infinitely divisible distributions and the definition adopted to our matrix case reads as follows:

Definition 5.3.11. *Let $Q : \mathbb{S}_d \to \mathbb{S}_d$ be a linear operator. A probability distribution μ on $\mathbb{S}_d$ is called* operator self-decomposable with respect to the operator Q*, if there exists a probability distribution ν_t on $\mathbb{S}_d$ such that $\mu = (e^{Qt}\mu) * \nu_t$ for all $t \in \mathbb{R}^+$, where $*$ denotes the usual convolution of probability measures.*

Remark 5.3.12. *(i) ν_t is for all $t \in \mathbb{R}^+$ infinitely divisible (Jurek and Mason (1993, Theorem 3.3.5).*

(ii) Rephrasing the definition on the level of random variables we call an $\mathbb{S}_d$-valued random variable X operator self-decomposable with respect to a linear operator $Q : \mathbb{S}_d \to \mathbb{S}_d$, if there exists a random variable Y_t independent of X such that $X \overset{\text{law}}{=} e^{Qt}X + Y_t$ for all $t \in \mathbb{R}^+$.

For our Ornstein-Uhlenbeck type processes in the positive semi-definite matrices we have immediately the following result.

Proposition 5.3.13. *The stationary distribution of the positive semi-definite OU type process Σ is operator self-decomposable with respect to the operator $\mathbf{A} : \mathbb{S}_d \to \mathbb{S}_d,\ X \mapsto AX + XA^T$.*

This observation allows us to use the well-established theory on operator self-decomposable distributions to study further properties of our processes. In particular, we can show that the stationary distribution always has a density (with respect to the Lebesgue measure) whenever the Lévy measure of the driving Lévy process has a non-degenerate support. Note that by the Lebesgue measure $\lambda_{\mathbb{S}_d}$ on $\mathbb{S}_d$ we mean the image of the Lebesgue measure $\lambda_{\mathbb{R}^{d(d+1)/2}}$ on $\mathbb{R}^{d(d+1)/2}$ under $\text{vech} : \mathbb{S}_d \to \mathbb{R}^{d(d+1)/2}$, i.e. $\lambda_{\mathbb{S}_d}(B) = \lambda_{\mathbb{R}^{d(d+1)/2}}(\text{vech}(B))$ for all Borel sets $B \subset \mathbb{S}_d$. (This definition has also been used in Barndorff-Nielsen and Pérez-Abreu (2007).)

Theorem 5.3.14. *Assume that the support of the Lévy measure $\nu_L \neq 0$ of the driving matrix subordinator L is contained in a linear subspace $V \subseteq \mathbb{S}_d$, that the smallest $\mathbf{A}$-invariant linear subspace of $\mathbb{S}_d$ containing V is $\mathbb{S}_d$ itself and that $\nu_L(W) = 0$ for every proper linear subspace $W \subset V$.*

Then the stationary distribution of the Ornstein-Uhlenbeck type process Σ is absolutely continuous with respect to the Lebesgue measure on $\mathbb{S}_d$. Moreover, the stationary distribution P_Σ of Σ is concentrated on $\mathbb{S}_d^{++}$, i.e. $P_\Sigma(\mathbb{S}_d^{++}) = 1$.

Proof. This follows by a straightforward adaptation of Jurek and Mason (1993, Proposition 3.8.6) or Yamazato (1983) to the $\mathbb{S}_d$-case and the fact that $\lambda_{\mathbb{S}_d}(\mathbb{S}_d^{+} \backslash \mathbb{S}_d^{++}) = 0$ shown in the next lemma. □

Observe that actually this result holds for OU type processes in $\mathbb{S}_d$ in general and not only for positive semi-definite ones and that the density can be characterised as the unique weak solution to a certain integro-differential equation (see Jurek and Mason (1993, Theorem 3.8.10)).

Lemma 5.3.15. *The Lebesgue measure on $\mathbb{S}_d$ satisfies $\lambda_{\mathbb{S}_d}(\mathbb{S}_d^+ \backslash \mathbb{S}_d^{++}) = 0$.*

Proof. It is clear that $\mathbb{S}_d^+ \backslash \mathbb{S}_d^{++} \subseteq \{X \in \mathbb{S}_d : \det(X) = 0\}$. The determinant is a polynomial in the entries of $X \in \mathbb{S}_d$. Hence, there exists a $d(d+1)/2$-variate polynomial $P : \mathbb{R}^{d(d+1)/2} \to \mathbb{R}$ such that $\det(X) = P(\mathrm{vech}(X))$ and P is not identically zero. As a polynomial P is a real analytic function and the set of zeros of real analytical functions not identically zero has Lebesgue measure zero (Jurek and Mason (1993, Lemma 3.8.4)), we have

$$\begin{aligned}\lambda_{\mathbb{S}_d}(\{X \in \mathbb{S}_d : \det(X) = 0\}) &= \lambda_{\mathbb{R}^{d(d+1)/2}}(\mathrm{vech}(\{X \in \mathbb{S}_d : \det(X) = 0\})) \\ &= \lambda_{\mathbb{R}^{d(d+1)/2}}(\{x \in \mathbb{R}^{d(d+1)/2} : P(x) = 0\}) = 0.\end{aligned}$$

□

The following is an immediate consequence of Theorem 5.3.14 which will be sufficient in most cases encountered in practice.

Corollary 5.3.16. *Assume that for every proper linear subspace V of $\mathbb{S}_d$ the Lévy measure ν_L of the driving matrix subordinator satisfies $\nu_L(V) = 0$ and that $\nu_L \neq 0$.*

Then the stationary distribution of the Ornstein-Uhlenbeck type process Σ is absolutely continuous with respect to the Lebesgue measure. Moreover, the stationary distribution P_Σ of Σ is concentrated on $\mathbb{S}_d^{++}$, i.e. $P_\Sigma(\mathbb{S}_d^{++}) = 1$.

A particular class of $d \times d$ matrix subordinators is given by the discrete part of the quadratic variation of a d-dimensional Lévy process. Let thus $\tilde{L}$ be a Lévy process in $\mathbb{R}^d$. Then the discrete part of the quadratic variation is given by

$$[\tilde{L}, \tilde{L}]_t^{\mathfrak{d}} := \sum_{0 \leq s \leq t} \Delta\tilde{L}_s (\Delta\tilde{L}_s)^T$$

and it is easy to see that $([\tilde{L}, \tilde{L}]_t^{\mathfrak{d}})_{t \in \mathbb{R}^+}$ is a matrix subordinator in $\mathbb{S}_d$ with the Lévy measure given by $\nu_{[\tilde{L},\tilde{L}]^{\mathfrak{d}}}(B) = \int_{\mathbb{R}^d} I_B(xx^T)\nu_{\tilde{L}}(dx) = \nu_{\tilde{L}}(\{x : xx^T \in B\})$ where $\nu_{\tilde{L}}$ denotes the Lévy measure of $\tilde{L}$. Note that $\nu_{[\tilde{L},\tilde{L}]}$ is obviously concentrated on the positive semi-definite rank one matrices and thus highly degenerate. Yet, under mild regularity conditions the stationary distribution of a positive semi-definite OU type process Σ driven by $[\tilde{L}, \tilde{L}]^{\mathfrak{d}}$ is absolutely continuous.

Proposition 5.3.17. *Let $\tilde{L}$ be a Lévy process in $\mathbb{R}^d$ with Lévy measure $\nu_{\tilde{L}} \neq 0$ and assume that $\nu_{\tilde{L}}(\{x : x^T Z x = 0\}) = 0$ for all $Z \in \mathbb{S}_d \backslash \{0\}$.*

Then the stationary distribution of the Ornstein-Uhlenbeck type process Σ driven by $[\tilde{L}, \tilde{L}]^{\mathfrak{d}}$ is absolutely continuous with respect to the Lebesgue measure. Moreover, the stationary distribution P_Σ of Σ is concentrated on $\mathbb{S}_d^{++}$, i.e. $P_\Sigma(\mathbb{S}_d^{++}) = 1$.

Proof. Recall that $\mathbb{S}_d$ with the scalar product $X, Y \mapsto \mathrm{tr}(X^T Y)$ is a Hilbert space and denote for a set $H \subset \mathbb{S}_d$ its orthogonal complement (with respect to this scalar product) by $H^\perp$.

Let now H be a proper linear subspace of $\mathbb{S}_d$ and $Z \in H^\perp \backslash \{0\}$. Then $H \subseteq \{Z\}^\perp$. Hence, $\nu_{[\tilde{L},\tilde{L}]^\mathfrak{d}}(H) \leq \nu_{[\tilde{L},\tilde{L}]^\mathfrak{d}}(\{Z\}^\perp) = \nu_{\tilde{L}}(\{x \in \mathbb{R}^d : xx^T \in \{Z\}^\perp\}) = \nu_{\tilde{L}}(\{x \in \mathbb{R}^d : \operatorname{tr}(xx^TZ) = x^TZx = 0\}) = 0$ by assumption.

Since $\nu_{[\tilde{L},\tilde{L}]^\mathfrak{d}} \neq 0$ is obviously true, the proposition is implied by Corollary 5.3.16. □

For applications the following corollary is usually sufficient:

Corollary 5.3.18. *Let $\tilde{L}$ be a Lévy process in $\mathbb{R}^d$ with Lévy measure $\nu_{\tilde{L}} \neq 0$ and assume that $\nu_{\tilde{L}}$ is absolutely continuous (with respect to the Lebesgue measure on $\mathbb{R}^d$).*

Then the stationary distribution of the Ornstein-Uhlenbeck type process Σ driven by $[\tilde{L},\tilde{L}]^\mathfrak{d}$ is absolutely continuous with respect to the Lebesgue measure. Moreover, the stationary distribution P_Σ of Σ is concentrated on $\mathbb{S}_d^{++}$, i.e. $P_\Sigma(\mathbb{S}_d^{++}) = 1$.

Proof. Let $Z \in \mathbb{S}_d\backslash\{0\}$. Then the function $\mathbb{R}^d \to \mathbb{R}$, $x = (x_1, x_2, \ldots, x_d)^T \mapsto x^TZx$ is a quadratic polynomial (in d variables) and thus a real analytic function not identically zero. The zeros of a real analytic function not identically zero form a set of Lebesgue measure zero (see Jurek and Mason (1993, Lemma 3.8.4), for instance). Hence, $\nu_{\tilde{L}}(\{x : x^TZx = 0\}) = 0$ for all $Z \in \mathbb{S}_d\backslash\{0\}$ and thus Proposition 5.3.17 concludes the proof. □

Furthermore, note that the result concerning the existence of an infinitely often differentiable density with all derivatives bounded given in Proposition 2.3.31 of Chapter 2 is also immediately applicable to our positive semi-definite OU type processes when using the vech notation.

It is also possible to construct positive semi-definite OU type processes with a prescribed marginal stationary distribution. The result is a version of Barndorff-Nielsen, Jensen and Sørensen (1998, Lemma 5.1) adapted to the matrix case considering the matrix subordinator case especially. "Differentiability" in the following means total differentiability (sometimes also referred to as Frechet differentiability) and "derivative" refers to the total derivative. Moreover, we need in the following that a probability distribution μ concentrated on $\mathbb{S}_d^+$ is infinitely divisible if and only if its characteristic function $\hat{\mu}$ is of the form

$$\hat{\mu}(\xi) = \exp\left(i\operatorname{tr}(\gamma_\mu\xi) + \int_{\mathbb{S}_d}\left(e^{i\operatorname{tr}(X\xi)} - 1\right)\nu_\mu(dX)\right),\ \xi \in \mathbb{S}_d, \tag{5.3.12}$$

where $\gamma_\mu \in \mathbb{S}_d^+$ and ν_μ is a Lévy measure on $\mathbb{S}_d$ satisfying $\nu_\mu(\mathbb{S}_d\backslash\mathbb{S}_d^+) = 0$ and also $\int_{\|X\|\leq 1}\|X\|\nu_\mu(dX) < \infty$. This follows from Skorohod (1991, p. 156) combined with arguments analogous to the ones in Sato (1999, Remark 24.9).

Theorem 5.3.19. *Let $\mathbf{A} : \mathbb{S}_d \to \mathbb{S}_d$ be a linear operator such that there is an $A \in M_d(\mathbb{R})$ with $\sigma(A) \subset (-\infty, 0)+i\mathbb{R}$ satisfying $\mathbf{A}X = AX+XA^T$ for all $X \in \mathbb{S}_d$. Furthermore, let μ be an operator self-decomposable (with respect to $\mathbf{A}$) distribution on $\mathbb{S}_d^+$ with associated γ_μ and ν_μ such that (5.3.12) holds and let $\kappa(\xi) = \log\left(\int_{\mathbb{S}_d^+} e^{i\operatorname{tr}(x\xi)}\mu(dx)\right)$, $\xi \in \mathbb{S}_d$, be its cumulant transform (logarithm of the characteristic function $\hat{\mu}$). Assume that $-A\gamma_\mu - \gamma_\mu A^T \in \mathbb{S}_d^+$, that $\kappa(\xi)$ is differentiable for all $\xi \neq 0$ with derivative $D\kappa(\xi)$ and that*

$$\kappa_L :\ \xi \mapsto \begin{cases} -D\kappa(\xi)(A^T\xi + \xi A) & \text{for } \xi \in \mathbb{S}_d\backslash\{0\} \\ 0 & \text{for } \xi = 0 \end{cases}$$

is continuous at zero. Then $\hat{\mu}_L : \xi \mapsto \exp(\kappa_L(\xi))$ *is the characteristic function of an infinitely divisible distribution* μ_L *on* $\mathbb{S}_d^+$.

Let L *be a matrix subordinator with characteristic function* $\hat{\mu}_L$ *at time one. Then the positive semi-definite Ornstein-Uhlenbeck type process* $d\Sigma_t = \mathbf{A}\Sigma_t dt + dL_t$ *driven by* L *has stationary distribution* μ.

Proof. Since the characteristic functions of infinitely divisible distributions have no zeros (Sato (1999, Lemma 7.5)), κ is well-defined. The definition of operator self-decomposability implies together with Remark 5.3.12 (i) that $\exp(\kappa(\xi))/\exp\left(\kappa\left(e^{A^T t}\xi e^{At}\right)\right)$ is for all $t \in \mathbb{R}^+$ the characteristic function of an infinitely divisible distribution. Hence the function $\xi \mapsto \left(\exp(\kappa(\xi))/\exp\left(\kappa\left(e^{A^T t}\xi e^{At}\right)\right)\right)^{1/t}$ is for all $t \in \mathbb{R}^+$ the characteristic function of an infinitely divisible distribution. Since obviously

$$\frac{\kappa(\xi) - \kappa\left(e^{A^T t}\xi e^{At}\right)}{t} \to -D\kappa(\xi)(A^T\xi + \xi A) \text{ as } t \to 0$$

for all $\xi \in \mathbb{S}_d$, we have that $\left(\exp(\kappa(\xi))/\exp\left(\kappa\left(e^{A^T t}\xi e^{At}\right)\right)\right)^{1/t} \to \exp(\kappa_L(\xi))$ point-wise for all $\xi \in \mathbb{S}_d$ as $t \to 0$. The continuity of κ_L at zero shows that $\hat{\mu}_L$ is a characteristic function. $\hat{\mu}_L$ belongs to an infinitely divisible distribution on $\mathbb{S}_d$, because infinite divisibility is preserved under weak convergence (Sato (1999, Lemma 7.8)).

Let $(L_t)_{t\in\mathbb{R}}$ now be a Lévy process with characteristic function $\hat{\mu}_L$ at time one and assume that $d\Sigma_t = \mathbf{A}\Sigma_t dt + dL_t$ is started at time zero with Σ_0 being independent of $(L_t)_{t\in\mathbb{R}}$ and having distribution μ. Thus $\Sigma_t = e^{At}\Sigma_0 e^{A^T t} + \int_0^t e^{A(t-s)} dL_s e^{A^T(t-s)}$ and we have from Rajput and Rosinski (1989) (see also Sato (2006), Jurek and Mason (1993, Lemma 3.6.4) or the review in Section 2.2.2 of Chapter 2) that

$$\begin{aligned}
E(\exp(i\mathrm{tr}(\Sigma_t\xi))) &= \hat{\mu}\left(e^{A^T t}\xi e^{At}\right)\exp\left(\int_0^t \kappa_L\left(e^{A^T s}\xi e^{As}\right) ds\right) \\
&= \hat{\mu}\left(e^{A^T t}\xi e^{At}\right)\exp\left(\int_0^t -D\kappa\left(e^{A^T s}\xi e^{As}\right)\left(A^T e^{A^T s}\xi e^{As} + e^{A^T s}\xi e^{As}A\right) ds\right) \\
&= \hat{\mu}\left(e^{A^T t}\xi e^{At}\right)\exp\left(\kappa(\xi) - \kappa\left(e^{A^T t}\xi e^{At}\right)\right) = \hat{\mu}(\xi).
\end{aligned}$$

Hence, Σ is stationary with stationary distribution μ. By Sato and Yamazato (1984, Theorem 4.2) it follows that $E(\log^+\|L_t\|) < \infty$. As the stationary distribution μ is concentrated on $\mathbb{S}_d^+$, μ_L has to be an infinitely divisible distribution on $\mathbb{S}_d^+$ and hence the Lévy process L a matrix subordinator. This can be seen in three steps as follows:

(i) Assume that the Lévy measure ν_L was not concentrated on $\mathbb{S}_d^+$, i.e. $\nu_L(\mathbb{S}_d\backslash\mathbb{S}_d^+) > 0$. Then, using the results of Appendix A, which imply for all $s \in \mathbb{R}^+$ that $e^{As}xe^{As} \in \mathbb{S}_d^+$ if and only if $x \in \mathbb{S}_d^+$, and (5.3.7) (which holds also in this general case, see e.g. Section 2.2.2), we have that $\int_{\mathbb{S}_d} I_{\mathbb{S}_d\backslash\mathbb{S}_d^+}(e^{As}xe^{A^T s})\nu_L(dx) > 0$ for all $s \in \mathbb{R}^+$ and hence $\nu_\mu(\mathbb{S}_d\backslash\mathbb{S}_d^+) = \int_0^\infty \int_{\mathbb{S}_d} I_{\mathbb{S}_d\backslash\mathbb{S}_d^+}(e^{As}xe^{A^T s})\nu_L(dx)ds > 0$, a contradiction.

(ii) We have for any $t^* \in \mathbb{R}^{++}$ by (5.3.7) and (5.3.12) that

$$\int_0^{t^*}\int_{\mathbb{S}_d^+}\left(\|e^{As}xe^{A^T s}\| \wedge 1\right)\nu_L(dx) \leq \int_{\mathbb{S}_d}(\|x\| \wedge 1)\nu_\Sigma(dx) < \infty.$$

Choosing t^* such that $\|e^{As}xe^{A^Ts}\| \geq \|x\|/2$ for all $s \in [0,t^*]$ and $x \in \mathbb{S}_d$ implies that

$$\int_0^{t^*} \int_{\mathbb{S}_d^+} \left(\|e^{As}xe^{A^Ts}\| \wedge 1\right) \nu_L(dx) \geq \int_0^{t^*} \int_{\mathbb{S}_d^+} ((\|x\|/2) \wedge 1)\, \nu_L(dx)$$
$$= 2t^* \int_{\mathbb{S}_d^+, \|x\| \leq 2} \|x\| \nu_L(dx).$$

Hence, $\int_{\mathbb{S}_d^+, \|x\| \leq 1} \|x\| \nu_L(dx) < \infty$ and therefore L is of finite variation.

(iii) Combining (i) and (ii) implies that $\hat{\mu}_L$ is of the form (5.3.12) with $\gamma_L \in \mathbb{S}_d$. Yet, (5.3.6) implies that $\gamma_L = -A\gamma_\mu - \gamma_\mu A^T$ and thus $\gamma_L \in \mathbb{S}_d^+$ by assumption. □

Remark 5.3.20. *(a) $D\kappa(\xi)$ is for each $\xi \in \mathbb{S}_d$ a linear operator from $\mathbb{S}_d$ to $\mathbb{R}$. Since $\mathbb{S}_d$ is a Hilbert space, the Riesz-Frechet theorem ensures that there exists a function $\hat{D}_\kappa : \mathbb{S}_d \to \mathbb{S}_d$ such that $D\kappa(\xi)X = \mathrm{tr}(\hat{D}_\kappa(\xi)X)$ for all $X \in \mathbb{S}_d$ which gives an alternative representation of κ_L above.*

(b) The condition $-A\gamma_\mu - \gamma_\mu A^T \in \mathbb{S}_d^+$ may appear superfluous at a first sight, since it obviously is always satisfied for $d = 1$. However, in general dimensions it need not hold. For example, for

$$A = \begin{pmatrix} -1/10 & -1/3 \\ -1/3 & -2 \end{pmatrix} \quad \textit{and} \quad \gamma_\mu = \begin{pmatrix} 2 & -2/3 \\ -2/3 & 2 \end{pmatrix}$$

we have that

$$\sigma(A) = \left\{ -\frac{21}{20} \pm \frac{1}{60}\sqrt{3649} \right\} \approx \{-0.043, -2.06\}, \ \sigma(\gamma_\mu) = \{8/3, 4/3\},$$
$$\textit{but} \quad \sigma(-A\gamma_\mu - \gamma_\mu A^T) = \left\{ \frac{169}{45} \pm \frac{1}{3}\sqrt{130} \right\} \approx \{7.56, -.045\}.$$

If the condition is not satisfied, all assertions of the theorem remain valid except that then μ_L is not concentrated on $\mathbb{S}_d^+$ and that L is a finite variation Lévy process in $\mathbb{S}_d$ with all jumps in $\mathbb{S}_d^+$ but a drift in $\mathbb{S}_d \backslash \mathbb{S}_d^+$.

Having introduced the positive semi-definite OU processes and given some of its probabilistic properties, we now focus on its integrated version and marginal dynamics.

5.3.2. The integrated process

From a financial point of view the integrated process is of major importance, as it corresponds to the integrated volatility, which is a main variable of interest in financial applications. Thus, a simple representation as the one for the positive semi-definite OU type processes below is desirable.

Proposition 5.3.21 (Proposition 4.4.10 of Chapter 4)**.** *Let Σ be a positive semi-definite OU process with initial value $\Sigma_0 \in \mathbb{S}_d^+$ and driven by the Lévy process L. Then the integrated OU process Σ^+ is given by*

$$\Sigma_t^+ := \int_0^t \Sigma_t dt = \mathbf{A}^{-1}(\Sigma_t - \Sigma_0 - L_t)$$

for $t \in \mathbb{R}^+$, where $\mathbf{A}$ is the linear operator defined in Proposition 5.3.5.

5.3.3. Marginal dynamics

Deriving the marginal dynamics, i.e. the behaviour of the individual components $\Sigma_{ij} = (\Sigma_{ij,t})_{t\in\mathbb{R}^+}$ of a positive semi-definite OU type process Σ, facilitates the comparison with the univariate OU type processes. To this end, we assume that A is real diagonalizable and $\sigma(A) = \{\lambda_1, \ldots, \lambda_d\}$. Let $U \in GL_d(\mathbb{R})$ be such that

$$UAU^{-1} = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_d \end{pmatrix} := D. \tag{5.3.13}$$

Denoting $(U^T)^{-1} = (U^{-1})^T$ by U^{-T}, it follows that

$$\begin{aligned} \Sigma_t &= U^{-1}\left(\int_{-\infty}^t e^{D(t-s)} U dL_s U^T e^{D^T(t-s)}\right) U^{-T} \\ &= U^{-1}\left(\int_{-\infty}^t e^{D(t-s)} d(UL_sU^T) e^{D^T(t-s)}\right) U^{-T}. \end{aligned} \tag{5.3.14}$$

Defining $\widetilde{L}_t = UL_tU^T$ for $t \in \mathbb{R}$ we have that $\widetilde{L}$ is again a Lévy process in $M_d(\mathbb{R})$, or more specifically it is even a matrix subordinator. Moreover, one obtains that

$$\left(\int_{-\infty}^t e^{D(t-s)} d\widetilde{L}_s e^{D^T(t-s)}\right)_{ij} = \int_{-\infty}^t e^{(\lambda_i+\lambda_j)(t-s)} d\widetilde{L}_{ij,s} \text{ for } i,j = 1,\ldots,d,$$

which obviously shows that the individual components of

$$U\Sigma_t U^T = \int_{-\infty}^t e^{D(t-s)} d\widetilde{L}_s e^{D^T(t-s)} =: \widetilde{\Sigma}_t$$

are stationary one-dimensional Ornstein-Uhlenbeck type processes with associated stochastic differential equations

$$d\widetilde{\Sigma}_{ij,t} = (\lambda_i + \lambda_j)\widetilde{\Sigma}_{ij,t}dt + d\widetilde{L}_{ij,t}. \tag{5.3.15}$$

Note further that $\widetilde{L}_{ii}$ for $1 \le i \le d$ is necessarily a subordinator and $\widetilde{\Sigma}_{ii}$ has to be a positive OU type process. These assertions do, however, fail in general for $\widetilde{L}_{ij}$ and $\widetilde{\Sigma}_{ij}$ with $i \neq j$.

Together with (5.3.14) the above considerations show that the individual components Σ_{ij} of Σ are superpositions of (at most d^2) univariate OU type processes. However, unlike in the univariate superposition model, see Barndorff-Nielsen and Shephard (2001b), the individual OU-processes superimposed are in general not independent. Actually, they can only be independent when the Lévy measure of $\widetilde{L}$ is concentrated on the diagonal matrices.

With the obvious modifications the above results hold also true for general diagonalizable $A \in M_d(\mathbb{R})$. Then X^T has to be replaced by the Hermitian of a matrix $X \in M_d(\mathbb{C})$ and $\widetilde{\Sigma}$ is an OU type processes in the positive semi-definite complex matrices. Note that $\widetilde{\Sigma}_{ii}$ still have to be real (even positive) and $\widetilde{L}_{ii}$ a real matrix subordinator. Furthermore, (5.3.15) becomes

$$d\widetilde{\Sigma}_{ij,t} = (\lambda_i + \overline{\lambda}_j)\widetilde{\Sigma}_{ij,t}dt + d\widetilde{L}_{ij,t}.$$

This result adds important insight regarding the behaviour of the autocovariance functions of the volatility of the individual assets. In particular, in order to obtain a more realistic decay (compared to using a single univariate OU type process in a univariate model) of these functions it is no longer necessary to consider superpositions of different OU type processes. So, although it is also possible to build superpositions of positive semi-definite OU type processes (see Section 5.4.6), we expect them to be less important for financial applications as in the univariate case where it has been shown that sufficiently realistic patterns of the autocorrelation functions can only be obtained by superpositions of OU type processes. As such, the multivariate specification obviously introduces more flexibility.

5.4. The multivariate Ornstein-Uhlenbeck stochastic volatility model

Based on the above results for the positive semi-definite OU type process we can now introduce our multivariate stochastic volatility model. As already stated in the introductory section of this chapter, the general d-dimensional stochastic volatility stock price model is given by

$$dY_t = (\mu + \Sigma_t \beta)dt + \Sigma_t^{1/2} dW_t,\ Y_0 = 0, \tag{5.4.1}$$

where Y denotes the d-dimensional logarithmic stock price process, $\mu, \beta \in \mathbb{R}^d$ are the drift and so-called risk premium parameters, respectively, $(W_t)_{t\in\mathbb{R}^+}$ denotes a d-dimensional standard Brownian motion and $(\Sigma_t)_{t\in\mathbb{R}^+}$ is an adapted, stationary and square integrable stochastic process with values in $\mathbb{S}_d^+$ and independent of $(W_t)_{t\in\mathbb{R}^+}$. As common in the finance literature, $(\Sigma_t)_{t\in\mathbb{R}^+}$ represents the stochastic volatility or instantaneous covariance process.

In this chapter instead we mainly focus on a specification in which the volatility process is given by a Lévy-driven positive semi-definite OU type process where the driving Lévy process $(L_t)_{t\in\mathbb{R}^+}$ and the Brownian Motion of the price process are independent. We refer to this model as the "multivariate Ornstein-Uhlenbeck stochastic volatility model". However, whenever possible we state our results for the general model given in (5.4.1) and only assume that Σ is an adapted, stationary and square-integrable process. Note that for the multivariate OU type stochastic volatility model the formulae turn out to be very explicit.

Following Barndorff-Nielsen and Shephard (2001b) we presume $Y_0 = 0$, which is no real constraint as it just corresponds to a normalization of the prices at time zero. In the OU type stochastic volatility model we extend the driving Lévy process to one defined on the whole real line and write

$$\Sigma_t = \int_{-\infty}^{t} e^{A(t-s)} dL_s e^{A^T(t-s)}. \tag{5.4.2}$$

Note that this corresponds to starting the OU type process at time zero with Σ_0 having the stationary distribution and being independent of $(L_t)_{t\in\mathbb{R}^+}$.

The subsequent returns over time intervals of length $\Delta \in \mathbb{R}^{++}$ are denoted by $\mathbf{Y} = (\mathbf{Y}_n)_{n\in\mathbb{N}}$. In many financial applications this time interval, i.e. $[(n-1)\Delta, n\Delta]$ with $n \in \mathbb{N}$, will represent a trading day, for example. So, the logarithmic price increments are defined

by

$$\mathbf{Y}_n := Y_{n\Delta} - Y_{(n-1)\Delta} = \int_{(n-1)\Delta}^{n\Delta} (\mu + \Sigma_t \beta)dt + \int_{(n-1)\Delta}^{n\Delta} \Sigma_t^{1/2} dW_t, \;\; n \in \mathbb{N}.$$

As already stated in Barndorff-Nielsen and Shephard (2001b) it is easy to see that

$$\mathbf{Y}_n \,|\boldsymbol{\Sigma}_n \sim N_d\left(\mu\Delta + \boldsymbol{\Sigma}_n\beta, \boldsymbol{\Sigma}_n\right) \tag{5.4.3}$$

where

$$\boldsymbol{\Sigma}_n := \int_{(n-1)\Delta}^{n\Delta} \Sigma_t dt = \Sigma_{n\Delta}^{+} - \Sigma_{(n-1)\Delta}^{+}$$

is the integrated volatility over the unit time interval and $N_d(m, s)$ denotes the d-dimensional normal distribution with mean m and covariance matrix s.

Note that the multivariate OU type stochastic volatility model can easily be extended to account for the leverage effect by simply specifying

$$dY_t = (\mu + \Sigma_t\beta)dt + \Sigma_t^{1/2} dW_t + \psi dL_t,$$

with ψ being a linear operator from $\mathbb{S}_d$ to $\mathbb{R}^d$. This is a straightforward generalization of the univariate OU type models with leverage effect. However, as the derivation of the properties of the OU type models is markedly complicated by the inclusion of a leverage effect – even in the univariate case – we solely focus here on the model without leverage effect.

5.4.1. Second order structure

In this section we study the second order moments of the multivariate stochastic volatility model. In doing so we consider the general multivariate model whenever possible, but the most explicit results are obtained for the OU type stochastic volatility model. This analysis provides a basis for estimation and forecasting in our model.

Henceforth we make the following assumption:

Assumption 5.1. *The stationary stochastic volatility process $(\Sigma_t)_{t\in\mathbb{R}^+}$ has a finite second moment.*

Recall that in the OU-case this means that the driving Lévy process has to be square-integrable.

Before moving on to the second order properties of our model, we first introduce some notation regarding the autocovariance function.

Definition 5.4.1. *Let $(X_t)_{t\in\mathbb{T}}$ (with $\mathbb{T}$ being either $\mathbb{N}_0$ or $\mathbb{R}^+$) be a second order stationary process with values in $\mathbb{R}^d$. Then the autocovariance function* $\operatorname{acov}_X : \mathbb{T} \cup (-\mathbb{T}) \mapsto M_d(\mathbb{R})$ *of X is given by* $\operatorname{acov}_X(h) = \operatorname{cov}(X_h, X_0) = E(X_h X_0^T) - E(X_0)E(X_0)^T$ *for $h \geq 0$ and by* $\operatorname{acov}_X(h) = \operatorname{acov}_X(-h)^T$ *for $h < 0$.*

If $(X_t)_{t\in\mathbb{T}}$ is a second order stationary process with values in $M_d(\mathbb{R})$ (or $\mathbb{S}_d$) then we set $\operatorname{acov}_X := \operatorname{acov}_{\operatorname{vec}(X)}$.

As the twice integrated autocovariance function of the stationary volatility process Σ will be of particular importance, we define

$$r^+(t) := \int_0^t \operatorname{acov}_\Sigma(u)du \quad \text{and} \quad r^{++}(t) := \int_0^t r^+(u)du. \tag{5.4.4}$$

Theorem 5.4.2. *For the* general stochastic volatility model *with* $(\Sigma_t)_{t\in\mathbb{R}^+}$ *being stationary and square-integrable it holds that:*

$$\begin{aligned} E(\Sigma_t^+) &= tE(\Sigma_0) && (5.4.5)\\ \mathrm{var}(\mathrm{vec}(\Sigma_t^+)) &= r^{++}(t) + r^{++}(t)^T && (5.4.6)\\ E(Y_t) &= (\mu + E(\Sigma_0)\beta)t && (5.4.7)\\ \mathrm{var}(Y_t) &= E(\Sigma_0)t + (\beta^T \otimes I_d)\mathrm{var}(\mathrm{vec}(\Sigma_t^+))(\beta \otimes I_d). && (5.4.8) \end{aligned}$$

Furthermore, the increments of the integrated volatility $(\mathbf{\Sigma}_n)_{n\in\mathbb{N}}$ *are stationary and square-integrable. We have:*

$$\begin{aligned} E(\mathbf{\Sigma}_n) &= \Delta E(\Sigma_0) && (5.4.9)\\ \mathrm{var}(\mathrm{vec}(\mathbf{\Sigma}_n)) &= r^{++}(\Delta) + r^{++}(\Delta)^T && (5.4.10)\\ \mathrm{acov}_{\mathbf{\Sigma}}(h) &= r^{++}(h\Delta + \Delta) - 2r^{++}(h\Delta) + r^{++}(h\Delta - \Delta),\ h \in \mathbb{N}. && (5.4.11) \end{aligned}$$

Likewise we have that the discretely observed log-price increments $(\mathbf{Y}_n)_{n\in\mathbb{N}}$ *are stationary and square-integrable. It holds that:*

$$\begin{aligned} E(\mathbf{Y}_n) &= (\mu + E(\Sigma_0)\beta)\Delta && (5.4.12)\\ \mathrm{var}(\mathbf{Y}_n) &= E(\Sigma_0)\Delta + (\beta^T \otimes I_d)\mathrm{var}(\mathrm{vec}(\Sigma_\Delta^+))(\beta \otimes I_d) && (5.4.13)\\ \mathrm{acov}_{\mathbf{Y}}(h) &= (\beta^T \otimes I_d)\mathrm{acov}_{\mathbf{\Sigma}}(h)(\beta \otimes I_d),\ h \in \mathbb{N}. && (5.4.14) \end{aligned}$$

Let Σ *now be a* positive semi-definite OU type process *with driving matrix subordinator* L *then*

$$\begin{aligned} E(\Sigma_0) &= -\mathbf{A}^{-1}E(L_1) && (5.4.15)\\ r^{++}(t) &= \left(\mathscr{A}^{-2}\left(e^{\mathscr{A}t} - I_{d^2}\right) - \mathscr{A}^{-1}t\right)\mathrm{var}(\mathrm{vec}(\Sigma_0))\\ &= -\left(\mathscr{A}^{-2}\left(e^{\mathscr{A}t} - I_{d^2}\right) - \mathscr{A}^{-1}t\right))\mathcal{A}^{-1}\mathrm{var}(\mathrm{vec}(L_1)) && (5.4.16)\\ \mathrm{acov}_{\mathbf{\Sigma}}(h) &= e^{\mathscr{A}\Delta(h-1)}\mathscr{A}^{-2}\left(I_{d^2} - e^{\mathscr{A}\Delta}\right)^2\mathrm{var}(\mathrm{vec}(\Sigma_0))\\ &= -e^{\mathscr{A}\Delta(h-1)}\mathscr{A}^{-2}\left(I_{d^2} - e^{\mathscr{A}\Delta}\right)^2\mathcal{A}^{-1}\mathrm{var}(\mathrm{vec}(L_1)),\ h \in \mathbb{N}, && (5.4.17) \end{aligned}$$

where $\mathbf{A}$ *and* $\mathcal{A}$ *are defined in Proposition 5.3.5 and* $\mathscr{A} := A \otimes I_d + I_d \otimes A$. *Observe that* $\mathscr{A}$ *and* $\mathcal{A}$ *commute (as linear operators over* $M_{d^2}(\mathbb{R})$*).*

Proof. It is immediate from the definitions that the stationarity of $(\Sigma_t)_{t\in\mathbb{R}^+}$ implies the stationarity of $(\mathbf{\Sigma}_n)_{n\in\mathbb{N}}$ and $(\mathbf{Y}_n)_{n\in\mathbb{N}}$. So it remains to verify the above stated formulae.

Equations (5.4.5) and (5.4.9) follow immediately from the integral representation $\Sigma_t^+ =$

$\int_0^t \Sigma_t dt$ and (5.4.6) can be obtained by using a straightforward Fubini argument:

$$\begin{aligned}
\operatorname{var}(\operatorname{vec}(\Sigma_t^+)) &= E\left(\int_0^t \operatorname{vec}(\Sigma_s)ds \int_0^t (\operatorname{vec}(\Sigma_u))^T du\right) \\
&\quad -E\left(\int_0^t \operatorname{vec}(\Sigma_s)ds\right) E\left(\int_0^t (\operatorname{vec}(\Sigma_u))^T du\right) \\
&= \int_0^t \int_0^t \left(E(\operatorname{vec}(\Sigma_s)\operatorname{vec}(\Sigma_u)^T) - E(\operatorname{vec}(\Sigma_s))E(\operatorname{vec}(\Sigma_u)^T)\right) duds \\
&= \int_0^t \int_0^t \operatorname{acov}_\Sigma(s-u)duds = \int_0^t \int_0^s \operatorname{acov}_\Sigma(s-u)duds \\
&\quad + \int_0^t \int_s^t \operatorname{acov}_\Sigma(u-s)^T duds = r^{++}(t) + r^{++}(t)^T.
\end{aligned}$$

As $\mathbf{\Sigma}_n$ is equal to Σ_Δ^+ in distribution this implies (5.4.10) immediately.

Equation (5.4.7) is an immediate consequence of $Y_t = \int_0^t (\mu + \Sigma_s\beta)ds + \int_0^t \Sigma_s^{1/2} dW_s$ and $E\left(\int_0^t \Sigma_s^{1/2} dW_s\right) = 0$. This also establishes (5.4.12).

Regarding (5.4.8) elementary arguments imply

$$\begin{aligned}
\operatorname{var}(Y_t) &= \operatorname{var}\left(\int_0^t (\mu + \Sigma_s\beta)ds + \int_0^t \Sigma_s^{1/2} dW_s\right) = \operatorname{var}\left(\int_0^t \Sigma_s\beta ds\right) \\
&\quad + \operatorname{var}\left(\int_0^t \Sigma_s^{1/2} dW_s\right) + \operatorname{cov}\left(\int_0^t \Sigma_s\beta ds, \int_0^t \Sigma_s^{1/2} dW_s\right) \\
&\quad + \operatorname{cov}\left(\int_0^t \Sigma_s\beta ds, \int_0^t \Sigma_s^{1/2} dW_s\right)^T.
\end{aligned}$$

The standard Itô isometry implies

$$\operatorname{var}\left(\int_0^t \Sigma_s^{1/2} dW_s\right) = E(\Sigma_t^+) = tE(\Sigma_0).$$

For all matrices $A \in M_{m,n}(\mathbb{R})$, $B \in M_{n,p}(\mathbb{R})$ and $C \in M_{p,q}(\mathbb{R})$ with arbitrary $m, n, p, q \in \mathbb{N}$ it holds that $\operatorname{vec}(ABC) = (C^T \otimes A)\operatorname{vec}(B)$ (see Horn and Johnson (1991, Lemma 4.3.1)). Thus

$$\begin{aligned}
\operatorname{var}\left(\int_0^t \Sigma_s\beta ds\right) &= \operatorname{var}\left(\int_0^t \operatorname{vec}(\Sigma_s\beta)ds\right) = \operatorname{var}\left(\int_0^t (\beta^T \otimes I_d)\operatorname{vec}(\Sigma_s)ds\right) \\
&= (\beta^T \otimes I_d)\operatorname{var}(\operatorname{vec}(\Sigma_t^+))(\beta \otimes I_d).
\end{aligned}$$

Moreover, the independence of $(\Sigma_t)_{t\in\mathbb{R}^+}$ and $(W_t)_{t\in\mathbb{R}^+}$ gives

$$\begin{aligned}
\operatorname{cov}\left(\int_0^t \Sigma_s\beta ds, \int_0^t \Sigma_s^{1/2} dW_s\right) &= E\left(\int_0^t \Sigma_s\beta ds \int_0^t dW_s^T \Sigma_s^{1/2}\right) \\
&= E\left(\int_0^t \Sigma_s\beta ds E\left(\int_0^t dW_s^T \Sigma_s^{1/2} \,\middle|\, (\Sigma_s)_{s\in[0,t]}\right)\right) \\
&= E\left(\int_0^t \Sigma_s\beta ds \cdot 0\right) = 0.
\end{aligned}$$

Combining the above formulae gives (5.4.8) which in turn implies (5.4.13).

Formula (5.4.11): For $h \in \mathbb{N}$ we have:

$$\begin{aligned}
\operatorname{acov}_{\mathbf{\Sigma}}(h) &= E\left(\int_{h\Delta}^{(h+1)\Delta} \operatorname{vec}(\Sigma_s)ds \int_0^{\Delta} \operatorname{vec}(\Sigma_u)^T du\right) \\
&\quad -E\left(\int_0^{\Delta} \operatorname{vec}(\Sigma_s)ds\right) E\left(\int_0^{\Delta} \operatorname{vec}(\Sigma_u)^T du\right) \\
&= \int_{h\Delta}^{(h+1)\Delta} \int_0^{\Delta} \left(E\left(\operatorname{vec}(\Sigma_s)\operatorname{vec}(\Sigma_u)^T\right) - E\left(\operatorname{vec}(\Sigma_s)\right) E\left(\operatorname{vec}(\Sigma_u)^T\right)\right) duds \\
&= \int_{h\Delta}^{(h+1)\Delta} \int_0^{\Delta} \operatorname{acov}_{\Sigma}(s-u)duds \\
&= r^{++}(h\Delta+\Delta) - 2r^{++}(h\Delta) + r^{++}(h\Delta-\Delta).
\end{aligned}$$

Formula (5.4.14): Arguments analogous to the ones given for (5.4.13) imply

$$\operatorname{acov}_{\mathbf{Y}}(h) = \operatorname{cov}\left(\int_{h\Delta}^{(h+1)\Delta} \Sigma_t \beta dt, \int_0^{\Delta} \Sigma_t \beta dt\right) = (\beta^T \otimes I_d)\operatorname{acov}_{\mathbf{\Sigma}}(h)(\beta \otimes I_d)$$

which is (5.4.14).

Turning to the positive semi-definite OU type stochastic volatility model (5.4.15) has already been given in (5.3.8). For the twice integrated covariance $r^{++}(t)$ formula (5.3.10) and elementary integration shows (5.4.16):

$$\begin{aligned}
r^{+}(t) &= \int_0^t e^{\mathscr{A}s}\operatorname{var}(\operatorname{vec}(\Sigma_0))ds = \mathscr{A}^{-1}\left(e^{\mathscr{A}t} - I_{d^2}\right)\operatorname{var}(\operatorname{vec}(\Sigma_0)) \\
r^{++}(t) &= \mathscr{A}^{-1}\int_0^t \left(e^{\mathscr{A}s} - I_{d^2}\right)\operatorname{var}(\operatorname{vec}(\Sigma_0))ds \\
&= \left(\mathscr{A}^{-2}\left(e^{\mathscr{A}t} - I_{d^2}\right) - \mathscr{A}^{-1}t\right)\operatorname{var}(\operatorname{vec}(\Sigma_0)).
\end{aligned}$$

Together with (5.3.9) this shows (5.4.16).

Regarding equation (5.4.17) a combination of (5.4.11) and (5.4.16) implies:

$$\begin{aligned}
\operatorname{acov}_{\mathbf{\Sigma}}(h) &= \left(\mathscr{A}^{-2}\left(e^{\mathscr{A}(\Delta h+\Delta)} - 2e^{\mathscr{A}\Delta h} + e^{\mathscr{A}(\Delta h-\Delta)}\right)\right. \\
&\quad \left. -\mathscr{A}^{-1}\left(\Delta h + \Delta - 2\Delta h + \Delta h - \Delta\right)\right)\operatorname{var}(\operatorname{vec}(\Sigma_0)) \\
&= \mathscr{A}^{-2}e^{\mathscr{A}\Delta(h-1)}\left(e^{2\mathscr{A}\Delta} - 2e^{\mathscr{A}\Delta} + I_{d^2}\right)\operatorname{var}(\operatorname{vec}(\Sigma_0)) \\
&= e^{\mathscr{A}\Delta(h-1)}\mathscr{A}^{-2}\left(I_{d^2} - e^{\mathscr{A}\Delta}\right)^2 \operatorname{var}(\operatorname{vec}(\Sigma_0)).
\end{aligned}$$

This shows the first equality in (5.4.17) and the second follows immediately by (5.3.9).

Finally observe that

$$\begin{aligned}
\mathscr{A}\mathcal{A}X &= \left((A \otimes I_d) + (I_d \otimes A)\right)^2 X \\
&\quad + \left((A \otimes I_d) + (I_d \otimes A)\right) X \left((A^T \otimes I_d) + (I_d \otimes A^T)\right) = \mathcal{A}\mathscr{A}X
\end{aligned}$$

for all $X \in M_{d^2}(\mathbb{R})$ and thus $\mathcal{A}$ and $\mathscr{A}$ commute. □

Remark 5.4.3. *(a) The above formulae imply that for $\beta = 0$ the log-price increments $(\mathbf{Y}_n)_{n\in\mathbb{N}}$ form an uncorrelated sequence and are thus white noise (in the second order sense).*

(b) Note that for a second order stationary causal m-dimensional ARMA(1,1) process $(X_t)_{t\in\mathbb{Z}}$ given by $X_t - \Phi X_{t-1} = Z_t + \Theta Z_{t-1}$ with $\Phi, \Theta \in M_m(\mathbb{R})$ and $(Z_t)_{t\in\mathbb{Z}}$ being m-dimensional white noise with covariance matrix Σ_Z the following autocovariance function can be obtained using the general formulae of Brockwell and Davis (1991, p. 420) (see also Wei (1990, Section 14.3.5)):

$$\begin{aligned}\operatorname{acov}_X(0) &= \Sigma_Z + (I_m - B)^{-1}(\Phi+\Theta)\Sigma_Z(\Phi+\Theta)^T\\ \operatorname{acov}_X(1) &= (\Phi+\Theta)\Sigma_Z + \Phi(I_m - B)^{-1}(\Phi+\Theta)\Sigma_Z(\Phi+\Theta)^T\\ \operatorname{acov}_X(h) &= \Phi^{h-1}\operatorname{acov}_X(1),\ h \geq 1\end{aligned}$$

where $B : M_m(\mathbb{R}) \to M_m(\mathbb{R})$, $X \mapsto \Phi X \Phi^T$. Since we consider a stationary causal ARMA process, $\rho(\Phi) < 1$. Hence, as $\operatorname{vec}(BX) = (\Phi \otimes \Phi)\operatorname{vec}(X)$, it is obvious that $\rho(B) < 1$ and thus $I_m - B$ is invertible.

Comparing equation (5.4.17) *with the general autocovariance function of an ARMA(1,1) process, immediately reveals that in the positive semi-definite OU type stochastic volatility model the process $(\operatorname{vec}(\mathbf{\Sigma}_n))_{n\in\mathbb{N}}$ is an ARMA(1,1) process with autoregressive parameter $e^{\mathscr{A}\Delta}$.*

Moreover, since we assume $\sigma(A) \subset (-\infty, 0) + i\mathbb{R}$, we have from Horn and Johnson (1991, Theorem 4.4.5) that $\sigma(\mathscr{A}) \subset (-\infty, 0) + i\mathbb{R}$ and thus all elements of $\sigma(e^{\mathscr{A}\Delta})$ are less than one in modulus, which implies that this ARMA(1,1) process is causal.

The ARMA(1,1) structure of $\mathbf{\Sigma}$ might seem to provide a natural starting point for making inference on the OU type stochastic volatility model. However, usually $\mathbf{\Sigma}$ is unobservable and so inference can only be based on the observed returns $\mathbf{Y}$. But the second order structure of the returns obviously does not allow for an in-depth analysis of the latent stochastic volatility model. Yet, the squared log-price increments $\mathbf{Y}\mathbf{Y}^T := (\mathbf{Y}_n\mathbf{Y}_n^T)_{n\in\mathbb{N}}$ are not only observable, but also exhibit a useful second order structure, which is no surprise, since their sum converges in probability to the integrated volatility when letting the time increment Δ go to zero.

Theorem 5.4.4. *In the* general stochastic volatility model *with $\mu = \beta = 0$ the second order structure of the squared log price process is given by*

$$E(Y_tY_t^T) = \operatorname{var}(Y_t) + E(Y_t)E(Y_t^T) = E(\Sigma_0)t \tag{5.4.18}$$

$$\begin{aligned}\operatorname{var}(\operatorname{vec}(Y_tY_t^T)) &= (I_{d^2} + \mathbf{Q} + \mathbf{PQ})\left(r^{++}(t) + (r^{++}(t))^T\right)\\ &\quad + (I_{d^2} + \mathbf{P})\left(E(\Sigma_0) \otimes E(\Sigma_0)\right)t^2\end{aligned} \tag{5.4.19}$$

and the one of the squared log return increments $(\mathbf{Y}_n\mathbf{Y}_n^T)_{n\in\mathbb{N}}$ is given by

$$E(\mathbf{Y}_n\mathbf{Y}_n^T) = \operatorname{var}(\mathbf{Y}_n) + E(\mathbf{Y}_n)E(\mathbf{Y}_n^T) = E(\Sigma_0)\Delta \tag{5.4.20}$$

$$\begin{aligned}\operatorname{var}(\operatorname{vec}(\mathbf{Y}_n\mathbf{Y}_n^T)) &= (I_{d^2} + \mathbf{Q} + \mathbf{PQ})\left(r^{++}(\Delta) + r^{++}(\Delta)^T\right)\\ &\quad + (I_{d^2} + \mathbf{P})\left(E(\Sigma_0) \otimes E(\Sigma_0)\right)\Delta^2\end{aligned} \tag{5.4.21}$$

$$\operatorname{acov}_{\mathbf{YY}^T}(h) = \operatorname{acov}_{\mathbf{\Sigma}}(h) \ \text{ for } \ h \in \mathbb{N} \tag{5.4.22}$$

where

$$\begin{aligned}&\mathbf{P} : M_{d^2}(\mathbb{R}) \to M_{d^2}(\mathbb{R}),\\ &(\mathbf{P}X)_{i,(p-1)d+q} = X_{i,(q-1)d+p} \text{ for all } i = \{1, 2, \ldots, d^2\},\ p, q = \{1, 2, \ldots d\},\end{aligned} \tag{5.4.23}$$

$$\mathbf{Q} : M_{d^2}(\mathbb{R}) \to M_{d^2}(\mathbb{R}), \tag{5.4.24}$$
$$(\mathbf{Q}X)_{(k-1)d+l,(p-1)d+q} = X_{(k-1)d+p,(l-1)d+q} \text{ for all } k,l,p,q = \{1,2,\dots d\}$$

are linear operators. Obviously $\mathbf{P}^{-1} = \mathbf{P}$, $\mathbf{Q}^{-1} = \mathbf{Q}$ *and* $\mathbf{P}$ *is representable as* $X \mapsto XP$ *with* $P \in M_{d^2}(\mathbb{R})$ *being a permutation matrix. Moreover,* $\mathbf{Q}(\mathrm{vec}(X)\mathrm{vec}(Z)^T) = X \otimes Z$ *for all* $X, Z \in \mathbb{S}_d$.

Component-wise we have for the variance

$$\begin{aligned}\mathrm{cov}\left(\mathbf{Y}_{i,n}\mathbf{Y}_{j,n}, \mathbf{Y}_{k,n}\mathbf{Y}_{l,n}\right) &= \mathrm{var}(\mathrm{vec}(\mathbf{Y}_n\mathbf{Y}_n^T))_{(j-1)d+i,(l-1)d+k} \\ &= \int_0^t \int_0^z \left(\mathrm{cov}(\Sigma_{ij,z}, \Sigma_{kl,u}) + \mathrm{cov}(\Sigma_{ij,u}, \Sigma_{kl,z})\right) dudz \\ &+ \int_0^t \int_0^z \left(E(\Sigma_{jl,z}\Sigma_{ik,u}) + E(\Sigma_{jl,u}\Sigma_{ik,z}) + E(\Sigma_{jk,z}\Sigma_{il,u}) + E(\Sigma_{jk,u}\Sigma_{il,z})\right) dudz\end{aligned} \tag{5.4.25}$$

In the OU stochastic volatility model *the process* $(\mathrm{vec}(\mathbf{Y}_n\mathbf{Y}_n^T))_{n\in\mathbb{N}}$ *is thus a causal ARMA(1,1) process with autoregressive parameter* $e^{\mathscr{A}\Delta}$.

Proof. The first equation in (5.4.18) is standard and the second then follows immediately from (5.4.7) and (5.4.8). This in turn implies (5.4.20).

Observe that (5.4.19) implies immediately (5.4.21), since $\mathbf{Y}_n\mathbf{Y}_n^T$ and $Y_\Delta Y_\Delta^T$ are equal in law. Turning to the proof of (5.4.19) we have from Lemma 4.5.11 in Chapter 4 that

$$\begin{aligned}Y_tY_t^T &= \int_0^t \Sigma_s^{1/2}dW_s \int_0^t dW_s^T\Sigma_s^{1/2} = \int_0^t \left(\int_0^s \Sigma_u^{1/2}dW_u\right) dW_s^T\Sigma_s^{1/2} \\ &+ \int_0^t \Sigma_s^{1/2}dW_s \left(\int_0^s dW_u^T\Sigma_u^{1/2}\right) + \left[\int_0^t \Sigma_s^{1/2}dW_s, \int_0^t dW_s^T\Sigma_s^{1/2}\right]_t^M \\ &= \int_0^t \left(\int_0^s \Sigma_u^{1/2}dW_u\right) dW_s^T\Sigma_s^{1/2} + \int_0^t \Sigma_s^{1/2}dW_s \left(\int_0^s dW_u^T\Sigma_u^{1/2}\right) + \Sigma_t^+\end{aligned} \tag{5.4.26}$$

referring to Lemma 4.5.11 in Chapter 4 for the definition of the matrix covariation $[\cdot,\cdot]^M$ and observing that we do not have to take left limits as integrals with respect to Brownian motion are necessarily continuous.

This stochastic integral representation implies that

$$\begin{aligned}\mathrm{var}(\mathrm{vec}(Y_tY_t^T)) &= \mathrm{var}(\mathrm{vec}(\Sigma_t^+)) + \mathrm{var}\left(\mathrm{vec}\left(\int_0^t \left(\int_0^s \Sigma_u^{1/2}dW_u\right) dW_s^T\Sigma_s^{1/2}\right)\right) \\ &+ \mathrm{var}\left(\mathrm{vec}\left(\int_0^t \Sigma_s^{1/2}dW_s \left(\int_0^s dW_u^T\Sigma_u^{1/2}\right)\right)\right) \\ &+ \mathscr{D}\mathrm{cov}\left(\mathrm{vec}(\Sigma_t^+), \mathrm{vec}\left(\int_0^t \left(\int_0^s \Sigma_u^{1/2}dW_u\right) dW_s^T\Sigma_s^{1/2}\right)\right) \\ &+ \mathscr{D}\mathrm{cov}\left(\mathrm{vec}(\Sigma_t^+), \mathrm{vec}\left(\int_0^t \Sigma_s^{1/2}dW_s \left(\int_0^s dW_u^T\Sigma_u^{1/2}\right)\right)\right) \\ &+ \mathscr{D}\mathrm{cov}\left(\mathrm{vec}\left(\int_0^t \left(\int_0^s \Sigma_u^{1/2}dW_u\right) dW_s^T\Sigma_s^{1/2}\right), \mathrm{vec}\left(\int_0^t \Sigma_s^{1/2}dW_s \left(\int_0^s dW_u^T\Sigma_u^{1/2}\right)\right)\right)\end{aligned} \tag{5.4.27}$$

setting $\mathscr{D} : M_{d^2} \to M_{d^2},\ X \mapsto X + X^T$

In the following we will now calculate the individual summands above in order to obtain an explicit expression for the variance of $Y_t Y_t^T$. To this end we first of all note that

$$E\left(\int_0^t \left(\int_0^s \Sigma_u^{1/2} dW_u\right) dW_s^T \Sigma_s^{1/2}\right) = E\left(\int_0^t \Sigma_s^{1/2} dW_s \left(\int_0^s dW_u^T \Sigma_u^{1/2}\right)\right) = 0.$$

Moreover, we already know from (5.4.6) that

$$\operatorname{var}(\operatorname{vec}(\Sigma_t^+)) = r^{++}(t) + r^{++}(t)^T.$$

Next we use the independence of $(\Sigma_t)_{t\in\mathbb{R}^+}$ and $(W_t)_{t\in\mathbb{R}^+}$ to obtain

$$\begin{aligned}
&E\left(\operatorname{vec}(\Sigma_t^+)\left(\int_0^t \left(\int_0^s \Sigma_u^{1/2} dW_u\right) dW_s^T \Sigma_s^{1/2}\right)^T\right) \\
&= E\left(\operatorname{vec}(\Sigma_t^+) E\left(\int_0^t \left(\int_0^s \Sigma_u^{1/2} dW_u\right) dW_s^T \Sigma_s^{1/2} \,\middle|\, (\Sigma_s)_{s\in[0,t]}\right)^T\right) \\
&= E\left(\operatorname{vec}(\Sigma_t^+)\cdot 0\right) = 0. \qquad (5.4.28)
\end{aligned}$$

Thus

$$\operatorname{cov}\left(\operatorname{vec}(\Sigma_t^+), \operatorname{vec}\left(\int_0^t \left(\int_0^s \Sigma_u^{1/2} dW_u\right) dW_s^T \Sigma_s^{1/2}\right)\right) = 0$$

and likewise

$$\operatorname{cov}\left(\operatorname{vec}(\Sigma_t^+), \operatorname{vec}\left(\int_0^t \Sigma_s^{1/2} dW_s \left(\int_0^s dW_u^T \Sigma_u^{1/2}\right)\right)\right) = 0.$$

In order to calculate the remaining covariances we have to study the individual entries. In the following let $k,l,m,n \in \{1,2,\cdots d\}$, $g := (k-1)d + l$, $g := (m-1)d + n$ and we write moreover $\Sigma_{ij,s}^{1/2}$ for $(\Sigma_s^{1/2})_{ij}$.

$$\begin{aligned}
&E\left(\operatorname{vec}\left(\int_0^t \left(\int_0^s \Sigma_u^{1/2} dW_u\right) dW_s^T \Sigma_s^{1/2}\right) \operatorname{vec}\left(\int_0^t \left(\int_0^s \Sigma_u^{1/2} dW_u\right) dW_s^T \Sigma_s^{1/2}\right)^T\right)_{g,h} \\
&= E\left(\int_0^t \int_0^s \sum_{p=1}^d \sum_{q=1}^d \Sigma_{pk,s}^{1/2} \Sigma_{lq,u}^{1/2} dW_{q,u} dW_{p,s} \int_0^t \int_0^s \sum_{a=1}^d \sum_{b=1}^d \Sigma_{am,s}^{1/2} \Sigma_{nb,u}^{1/2} dW_{b,u} dW_{a,s}\right) \\
&\overset{(*)}{=} \int_0^t \sum_{a=1}^d E\left(\int_0^s \sum_{q=1}^d \Sigma_{ak,s}^{1/2} \Sigma_{lq,u}^{1/2} dW_{q,u} \int_0^s \sum_{b=1}^d \Sigma_{am,s}^{1/2} \Sigma_{nb,u}^{1/2} dW_{b,u}\right) ds \\
&\overset{(*)}{=} \int_0^t \int_0^s \sum_{a=1}^d \sum_{b=1}^d E\left(\Sigma_{ak,s}^{1/2} \Sigma_{lb,u}^{1/2} \Sigma_{am,s}^{1/2} \Sigma_{nb,u}^{1/2}\right) du ds \\
&= \int_0^t \int_0^s \sum_{a=1}^d \sum_{b=1}^d E\left(\Sigma_{ka,s}^{1/2} \Sigma_{lb,u}^{1/2} \Sigma_{am,s}^{1/2} \Sigma_{bn,u}^{1/2}\right) du ds \\
&= \int_0^t \int_0^s E\left((\Sigma_s^{1/2} \otimes \Sigma_u^{1/2})(\Sigma_s^{1/2} \otimes \Sigma_u^{1/2})\right)_{(k-1)d+l,(m-1)d+n} du ds \\
&= \int_0^t \int_0^s E\left(\Sigma_s \otimes \Sigma_u\right)_{(k-1)d+l,(m-1)d+n} du ds.
\end{aligned}$$

The $(*)$ above indicates that we have used the Itô isometry and the fact that stochastic integrals with respect to two independent Brownian motions are uncorrelated.

Thus we have established that

$$\operatorname{var}\left(\operatorname{vec}\left(\int_0^t\left(\int_0^s \Sigma_u^{1/2} dW_u\right) dW_s^T \Sigma_s^{1/2}\right)\right) = \int_0^t \int_0^s E\left(\Sigma_s \otimes \Sigma_u\right) duds$$

and calculations analogous to the above ones give

$$\operatorname{var}\left(\operatorname{vec}\left(\int_0^t \Sigma_s^{1/2} dW_s \left(\int_0^s dW_u^T \Sigma_u^{1/2}\right)\right)\right) = \int_0^t \int_0^s E\left(\Sigma_u \otimes \Sigma_s\right) duds.$$

Using again the Itô isometry twice and the uncorrelatedness of integrals with respect to independent Brownian motions we have with $g := (k-1)d + l$, $h := (m-1)d + n$:

$$\begin{aligned}
&\operatorname{cov}\left(\operatorname{vec}\left(\int_0^t\left(\int_0^s \Sigma_u^{1/2} dW_u\right) dW_s^T \Sigma_s^{1/2}\right), \operatorname{vec}\left(\int_0^t \Sigma_s^{1/2} dW_s \left(\int_0^s dW_u^T \Sigma_u^{1/2}\right)\right)\right)_{g,h} \\
&= E\left(\operatorname{vec}\left(\int_0^t\left(\int_0^s \Sigma_u^{1/2} dW_u\right) dW_s^T \Sigma_s^{1/2}\right) \operatorname{vec}\left(\int_0^t \Sigma_s^{1/2} dW_s \left(\int_0^s dW_u^T \Sigma_u^{1/2}\right)\right)^T\right)_{g,h} \\
&= E\left(\int_0^t \int_0^s \sum_{p=1}^d \sum_{q=1}^d \Sigma_{pk,s}^{1/2} \Sigma_{lq,u}^{1/2} dW_{q,u} dW_{p,s} \int_0^t \int_0^s \sum_{a=1}^d \sum_{b=1}^d \Sigma_{na,s}^{1/2} \Sigma_{bm,u}^{1/2} dW_{b,u} dW_{a,s}\right) \\
&= \int_0^t \sum_{a=1}^d E\left(\int_0^s \sum_{q=1}^d \Sigma_{ak,s}^{1/2} \Sigma_{lq,u}^{1/2} dW_{q,u} \int_0^s \sum_{b=1}^d \Sigma_{na,s}^{1/2} \Sigma_{bm,u}^{1/2} dW_{b,u}\right) ds \\
&= \int_0^t \int_0^s \sum_{a=1}^d \sum_{b=1}^d E\left(\Sigma_{ak,s}^{1/2} \Sigma_{lb,u}^{1/2} \Sigma_{na,s}^{1/2} \Sigma_{bm,u}^{1/2}\right) duds \\
&= \int_0^t \int_0^s \sum_{a=1}^d \sum_{b=1}^d E\left(\Sigma_{ka,s}^{1/2} \Sigma_{lb,u}^{1/2} \Sigma_{an,s}^{1/2} \Sigma_{bm,u}^{1/2}\right) duds \\
&= \int_0^t \int_0^s E\left((\Sigma_s^{1/2} \otimes \Sigma_u^{1/2})(\Sigma_s^{1/2} \otimes \Sigma_u^{1/2})\right)_{(k-1)d+l,(n-1)d+m} duds \\
&= \int_0^t \int_0^s \left(\mathbf{P} E\left(\Sigma_s \otimes \Sigma_u\right)\right)_{(k-1)d+l,(m-1)d+n} duds.
\end{aligned}$$

Thus

$$\begin{aligned}
&\operatorname{cov}\left(\operatorname{vec}\left(\int_0^t\left(\int_0^s \Sigma_u^{1/2} dW_u\right) dW_s^T \Sigma_s^{1/2}\right), \operatorname{vec}\left(\int_0^t \Sigma_s^{1/2} dW_s \left(\int_0^s dW_u^T \Sigma_u^{1/2}\right)\right)\right) \\
&= \mathbf{P} \int_0^t \int_0^s E\left(\Sigma_s \otimes \Sigma_u\right) duds.
\end{aligned}$$

Observing that

$$\left(\mathbf{P}(A \otimes B)\right)_{(k-1)d+l,(m-1)d+n}^T = a_{ml} b_{nk} = a_{lm} b_{kn} = \left(\mathbf{P}(B \otimes A)\right)_{(k-1)d+l,(m-1)d+n},$$

i.e. $(\mathbf{P}(A\otimes B))^T = \mathbf{P}(B\otimes A)$, for all $A=(a_{ij}), B=(b_{ij}) \in \mathbb{S}_d$, we finally obtain

$$\left(\operatorname{cov}\left(\operatorname{vec}\left(\int_0^t \left(\int_0^s \Sigma_u^{1/2} dW_u\right) dW_s^T \Sigma_s^{1/2}\right), \operatorname{vec}\left(\int_0^t \Sigma_s^{1/2} dW_s \left(\int_0^s dW_u^T \Sigma_u^{1/2}\right)\right)\right)\right)^T$$
$$= \left(\mathbf{P}\int_0^t\int_0^s E\left(\Sigma_s\otimes\Sigma_u\right) duds\right)^T = \mathbf{P}\int_0^t\int_0^s E\left(\Sigma_u\otimes\Sigma_s\right) duds.$$

Inserting all the obtained results into (5.4.27) leads to

$$\begin{aligned}\operatorname{var}\left(\operatorname{vec}\left(Y_t Y_t^T\right)\right) &= r^{++}(t) + r^{++}(t)^T \\ &+ \left(I_{d^2}+\mathbf{P}\right)\left(\int_0^t\int_0^s E\left(\Sigma_s\otimes\Sigma_u\right) duds + \int_0^t\int_0^s E\left(\Sigma_u\otimes\Sigma_s\right) duds\right).\end{aligned} \tag{5.4.29}$$

Component-wise this gives (5.4.25). As we have that for any $X=(x_{ij}), Z=(z_{ij}) \in \mathbb{S}_d$

$$\begin{aligned}&\left(\mathbf{Q}(\operatorname{vec}(X)\operatorname{vec}(Z)^T)\right)_{(k-1)d+l,(m-1)d+n} = \left(\operatorname{vec}(X)\operatorname{vec}(Z)^T\right)_{(k-1)d+m,(l-1)d+n} \\ &= x_{mk}y_{nl} = x_{km}y_{ln} = (X\otimes Y)_{(k-1)d+l,(m-1)d+n},\end{aligned}$$

it is immediate that $\mathbf{Q}(\operatorname{vec}(X)\operatorname{vec}(Y)^T) = X\otimes Y$ for all $X, Y \in \mathbb{S}_d$. Using this property we get from (5.4.29):

$$\begin{aligned}&\operatorname{var}\left(\operatorname{vec}\left(Y_t Y_t^T\right)\right) = r^{++}(t) + r^{++}(t)^T \\ &+ \left(I_{d^2}+\mathbf{P}\right)\left(\int_0^t\int_0^s \left(E\left(\Sigma_s\otimes\Sigma_u\right) - E(\Sigma_0)\otimes E(\Sigma_0)\right) duds\right. \\ &\left.+ \int_0^t\int_0^s \left(E\left(\Sigma_u\otimes\Sigma_s\right) - E(\Sigma_0)\otimes E(\Sigma_0)\right) duds\right) \\ &+ 2\left(I_{d^2}+\mathbf{P}\right)\int_0^t\int_0^s E(\Sigma_0)\otimes E(\Sigma_0) duds \\ &= r^{++}(t) + r^{++}(t)^T + \\ &\left(\mathbf{Q}+\mathbf{PQ}\right)\left(\int_0^t\int_0^s \left(E\left(\operatorname{vec}(\Sigma_s)(\operatorname{vec}(\Sigma_u))^T\right) - E(\operatorname{vec}(\Sigma_0))E(\operatorname{vec}(\Sigma_0))^T\right) duds\right. \\ &\left.+ \int_0^t\int_0^s \left(E\left(\operatorname{vec}(\Sigma_u)(\operatorname{vec}(\Sigma_s))^T\right) - E(\operatorname{vec}(\Sigma_0))E(\operatorname{vec}(\Sigma_0))^T\right) duds\right) \\ &+ \left(I_{d^2}+\mathbf{P}\right)\left(E(\Sigma_0)\otimes E(\Sigma_0)\right)t^2 \\ &= r^{++}(t) + r^{++}(t)^T + \left(\mathbf{Q}+\mathbf{PQ}\right)\left(\int_0^t\int_0^s \left(\operatorname{acov}_\Sigma(s-u) + \operatorname{acov}_\Sigma(u-s)\right) duds\right) \\ &+ \left(I_{d^2}+\mathbf{P}\right)\left(E(\Sigma_0)\otimes E(\Sigma_0)\right)t^2\end{aligned}$$

Together with (5.4.4) this finally shows (5.4.19).

It remains to show (5.4.22). Applying (5.4.26) we have for $h \in \mathbb{N}$:

$$
\begin{aligned}
&\operatorname{acov}_{\mathbf{Y}\mathbf{Y}^T}(h) = \operatorname{cov}\left(\operatorname{vec}\left(\mathbf{Y}_{h+1}\mathbf{Y}_{h+1}^T\right), \operatorname{vec}\left(\mathbf{Y}_1\mathbf{Y}_1^T\right)\right) \\
&= \operatorname{cov}\left(\operatorname{vec}\left(\int_{h\Delta}^{(h+1)\Delta} \Sigma_s^{1/2} dW_s \left(\int_{h\Delta}^{s} dW_u^T \Sigma_u^{1/2}\right)\right), \operatorname{vec}\left(\int_0^{\Delta} \Sigma_s^{1/2} dW_s \left(\int_0^{s} dW_u^T \Sigma_u^{1/2}\right)\right)\right) \\
&+ \operatorname{cov}\left(\operatorname{vec}\left(\int_{h\Delta}^{(h+1)\Delta} \Sigma_s^{1/2} dW_s \left(\int_{h\Delta}^{s} dW_u^T \Sigma_u^{1/2}\right)\right), \operatorname{vec}\left(\int_0^{\Delta} \left(\int_0^{s} \Sigma_u^{1/2} dW_u\right) dW_s^T \Sigma_s^{1/2}\right)\right) \\
&+ \operatorname{cov}\left(\operatorname{vec}\left(\int_{h\Delta}^{(h+1)\Delta} \Sigma_s^{1/2} dW_s \left(\int_{h\Delta}^{s} dW_u^T \Sigma_u^{1/2}\right)\right), \operatorname{vec}\left(\mathbf{\Sigma}_1\right)\right) \\
&+ \operatorname{cov}\left(\operatorname{vec}\left(\int_{h\Delta}^{(h+1)\Delta} \left(\int_{h\Delta}^{s} \Sigma_u^{1/2} dW_u\right) dW_s^T \Sigma_s^{1/2}\right), \operatorname{vec}\left(\int_0^{\Delta} \Sigma_s^{1/2} dW_s \left(\int_0^{s} dW_u^T \Sigma_u^{1/2}\right)\right)\right) \\
&+ \operatorname{cov}\left(\operatorname{vec}\left(\int_{h\Delta}^{(h+1)\Delta} \left(\int_{h\Delta}^{s} \Sigma_u^{1/2} dW_u\right) dW_s^T \Sigma_s^{1/2}\right), \operatorname{vec}\left(\int_0^{\Delta} \left(\int_0^{s} \Sigma_u^{1/2} dW_u\right) dW_s^T \Sigma_s^{1/2}\right)\right) \\
&+ \operatorname{cov}\left(\operatorname{vec}\left(\int_{h\Delta}^{(h+1)\Delta} \left(\int_{h\Delta}^{s} \Sigma_u^{1/2} dW_u\right) dW_s^T \Sigma_s^{1/2}\right), \operatorname{vec}\left(\mathbf{\Sigma}_1\right)\right) \\
&+ \operatorname{cov}\left(\operatorname{vec}\left(\mathbf{\Sigma}_{h+1}\right), \operatorname{vec}\left(\int_0^{\Delta} \Sigma_s^{1/2} dW_s \left(\int_0^{s} dW_u^T \Sigma_u^{1/2}\right)\right)\right) \\
&+ \operatorname{cov}\left(\operatorname{vec}\left(\mathbf{\Sigma}_{h+1}\right), \operatorname{vec}\left(\int_0^{\Delta} \left(\int_0^{s} \Sigma_u^{1/2} dW_u\right) dW_s^T \Sigma_s^{1/2}\right)\right) + \operatorname{cov}\left(\operatorname{vec}\left(\mathbf{\Sigma}_{h+1}\right), \operatorname{vec}\left(\mathbf{\Sigma}_1\right)\right).
\end{aligned}
$$

The independence of the increments of Brownian motion over distinct time intervals imply that the first, second, fourth and fifth covariance terms above vanish. Likewise conditioning on $(\Sigma_t)_{t\in\mathbb{R}^+}$ and using the independence of $(\Sigma_t)_{t\in\mathbb{R}^+}$ and $(W_t)_{t\in\mathbb{R}^+}$ (i.e. arguing basically as in (5.4.28)) show that the third, sixth, seventh and eighth covariance term are actually zero. Thus only the last term remains which gives

$$\operatorname{acov}_{\mathbf{Y}\mathbf{Y}^T}(h) = \operatorname{cov}\left(\operatorname{vec}\left(\mathbf{\Sigma}_{h+1}\right), \operatorname{vec}\left(\mathbf{\Sigma}_1\right)\right) = \operatorname{acov}_{\mathbf{\Sigma}}(h).$$

Combining this with the ARMA(1,1) property of the process $(\operatorname{vec}(\mathbf{\Sigma}_n))_{n\in\mathbb{N}}$ we see that in the OU type stochastic volatility model the process $(\operatorname{vec}(\mathbf{Y}_n\mathbf{Y}_n^T))_{n\in\mathbb{N}}$ has a causal ARMA(1,1) structure with autoregressive coefficient $e^{\mathscr{A}\Delta}$. □

Note that the ARMA(1,1) structure of $\operatorname{vec}(\mathbf{Y}\mathbf{Y}^T)$, of course, means that $\mathbf{Y}\mathbf{Y}^T$ itself is an ARMA(1,1) process. Its autoregressive coefficient is given by the linear operator $\mathbb{S}_d \to \mathbb{S}_d$, $X \mapsto e^{A\Delta} X e^{A\Delta}$.

For the positive semi-definite OU type stochastic volatility model the closed form expressions for the first and second moments can be obtained by applying Theorem 5.4.4 and by using the expression for the autocovariance, integrated autocovariance and for the mean as given in Theorem 5.4.2.

Moreover, recall that a d-dimensional sequence $(\epsilon_n)_{n\in\mathbb{N}}$ is said to be generated by a *weak GARCH(p, q) process*, if the best linear predictor of ϵ_n with respect to $\epsilon_{n-1}, \epsilon_{n-2}, \ldots, \epsilon_{n-1}\epsilon_{n-1}^T, \epsilon_{n-2}\epsilon_{n-2}^T, \ldots$ is zero and the best linear predictor of $\epsilon_n\epsilon_n^T$ with respect to $\epsilon_{n-1}, \epsilon_{n-2}, \ldots, \epsilon_{n-1}\epsilon_{n-1}^T, \epsilon_{n-2}\epsilon_{n-2}^T, \ldots$ is a (strong) GARCH(p, q) process. In the univariate case this definition goes back to Drost and Nijman (1993) and in the multivariate case it

has been given in Hafner (2007). As it is easy to see that $\text{cov}(\text{vec}(\mathbf{Y}_n\mathbf{Y}_n^T), \mathbf{Y}_m) = 0$ for $n, m \in \mathbb{N}$ and $n \neq m$ (actually this holds also true for $n = m$ as shown in the proof of Proposition 5.4.8), combining the last theorem with Brockwell and Davis (1991, Example 11.4.1) gives the following result.

Proposition 5.4.5. *In the OU-type multivariate stochastic volatility model with $\mu = \beta = 0$ the log-return sequence $(\mathbf{Y}_n)_{n\in\mathbb{N}}$ is generated by a weak GARCH(1,1) process.*

In order to obtain consistency results and central limit theorems for the estimation of the multivariate OU type model based on the moments of $\mathbf{Y}$ and $\mathbf{Y}\mathbf{Y}^T$, we need to show that the discretely observed stationary log-returns $\mathbf{Y}$ form a strongly mixing and, thus, ergodic sequence. For $\mu = \beta = 0$ strong mixing in the univariate OU type stochastic volatility model has been obtained in Sørensen (2000) and Genon-Catalot, Jeantheau and Larédo (2000). For details on mixing we refer again to Doukhan (1994) and regarding ergodicity to Ash and Gardner (1975) or Krengel (1985). In our set-up the most important implication of ergodicity is that the usual empirical moments converge almost surely (and in L^1) to the true moments (provided they are finite) as the number of observations goes to infinity.

Proposition 5.4.6. *(i) Assume that in the* general stochastic volatility model *the stationary and square integrable process Σ is strongly mixing with the mixing coefficients being $(\alpha_k(\Sigma))_{k\in\mathbb{N}}$. Then the process $\mathbf{Y}$ is strongly mixing with mixing coefficients $\alpha_k(\mathbf{Y}) \leq \alpha_k(\Sigma)$ for all $k \in \mathbb{N}$. Thus $\mathbf{Y}$ is ergodic.*

(ii) In the positive semi-definite OU type stochastic volatility model *the process $\mathbf{Y}$ is always strongly mixing with mixing coefficients $(\alpha_k)_{k\in\mathbb{N}}$ decaying at least at an exponential rate. Thus $\mathbf{Y}$ is ergodic.*

Proof. Part (i) follows from an immediate adaptation of the proof of Sørensen (2000, Lemma 6.3) to the multivariate case and the case $\mu, \beta \neq 0$. (ii) results from combining (i) with Proposition 5.3.9. □

Based on these results and the closed form expression for the second moments a simple moment matching estimator within the framework of the generalized method of moment (GMM) estimation of the model can be implemented, which is consistent and asymptotically normal. However, this involves the determination of an optimal number of lags to be included, as well as the specification of an optimal weighting scheme of the different lags. Moreover, the estimation procedure does not allow to filter the current volatility states. To overcome these problems we derive a state-space representation for the joint series of the returns and squared returns, which allows us to use the Kalman recursions for estimation and filtering.

5.4.2. State-space representation

The aim of this section is to establish a state-space representation for the joint process $(\mathbf{Y}_n, \mathbf{Y}_n\mathbf{Y}_n^T)_{n\in\mathbb{N}}$. Throughout we assume $\beta = 0$. As before, we first analyse the general stochastic volatility model and then focus on the OU type specification.

Recall that

$$\mathbf{Y}_n = \Delta\mu + \int_{(n-1)\Delta}^{n\Delta} \Sigma_s^{1/2} dW_s,$$

which immediately implies

$$\begin{aligned}\mathbf{Y}_n\mathbf{Y}_n^T &= \Delta^2\mu\mu^T + \int_{(n-1)\Delta}^{n\Delta}\Sigma_s^{1/2}dW_s\int_{(n-1)\Delta}^{n\Delta}dW_s^T\Sigma_s^{1/2} \\ &\quad +\Delta\int_{(n-1)\Delta}^{n\Delta}\Sigma_s^{1/2}dW_s\mu^T + \Delta\mu\int_{(n-1)\Delta}^{n\Delta}dW_s^T\Sigma_s^{1/2}. \qquad (5.4.30)\end{aligned}$$

Setting

$$u_n = \begin{pmatrix}u_{1,n}\\ u_{2,n}\end{pmatrix} \qquad (5.4.31)$$

$$u_{1,n} = \int_{(n-1)\Delta}^{n\Delta}\Sigma_s^{1/2}dW_s \qquad (5.4.32)$$

$$\begin{aligned}u_{2,n} &= \int_{(n-1)\Delta}^{n\Delta}\Sigma_s^{1/2}dW_s\int_{(n-1)\Delta}^{n\Delta}dW_s^T\Sigma_s^{1/2} + \Delta\int_{(n-1)\Delta}^{n\Delta}\Sigma_s^{1/2}dW_s\mu^T \\ &\quad +\Delta\mu\int_{(n-1)\Delta}^{n\Delta}dW_s^T\Sigma_s^{1/2} - \mathbf{\Sigma}_n, \qquad (5.4.33)\end{aligned}$$

it follows that

$$\mathbf{Y}_n = \Delta\mu + u_{1,n} \qquad (5.4.34)$$

$$\mathbf{Y}_n\mathbf{Y}_n^T = \Delta^2\mu\mu^T + \mathbf{\Sigma}_n + u_{2,n}. \qquad (5.4.35)$$

The martingale property in the following theorem is, of course, understood w.r.t. the filtration $(\mathcal{F}_t)$ which we assume to be given. Recall that all processes (in particular, L_t, Σ_t, W_t, Y_t, $\mathbf{Y}_n, \mathbf{\Sigma}_n$) are adapted with respect to this filtration. Moreover, $(W_s - W_t)_{s\geq t}$ is independent of $(\Sigma_s)_{s\in\mathbb{R}^+}$ as well as of $\mathcal{F}_t$ for all $t \in \mathbb{R}^+$. For technical reasons this is, however, not fully sufficient. Thus, we henceforth assume:

Assumption 5.2. *$(W_s - W_t)_{s\geq t}$ is independent of $\boldsymbol{\sigma}(\mathcal{F}_t, (\Sigma_s)_{s\in\mathbb{R}^+})$ for all $t \in \mathbb{R}^+$ ($\boldsymbol{\sigma}(\cdot)$ denoting the generated σ-algebra).*

Remark 5.4.7. *In the OU type stochastic volatility model this assumption is satisfied if the σ-algebras $\mathcal{F}_t$, $\boldsymbol{\sigma}((L_s - L_t)_{s\geq t})$ and $\boldsymbol{\sigma}((W_s - W_t)_{s\geq t})$ are independent for all $t \in \mathbb{R}^+$ and not only pairwise independent.*

Clearly the last condition will usually be satisfied. In particular, it is satisfied when the pair (L, W) of the driving Lévy process and Wiener process forms a Lévy process in $\mathbb{S}_d^+ \times \mathbb{R}^d$.

Recall further that a sequence $(X_n)_{n\in\mathbb{N}}$ is said to be a martingale difference sequence with respect to a filtration $\mathcal{G} = (\mathcal{G}_n)_{n\in\mathbb{N}}$ if it is adapted to $\mathcal{G}$ and $E(X_n|\mathcal{G}_{n-1}) = 0$. It is straightforward to see that a martingale difference sequence is uncorrelated provided the second moments are finite.

Proposition 5.4.8. *The sequence $(u_n)_{n\in\mathbb{N}}$ is a (second order) stationary zero-mean martingale difference sequence w.r.t. the filtration $(\mathcal{G}_n)_{n\in\mathbb{N}} := (\mathcal{F}_{n\Delta})_{n\in\mathbb{N}}$ and thus in particular*

white noise. It holds that

$$\begin{aligned}
\operatorname{var}(u_{1,n}) &= E(\mathbf{\Sigma}_n) = E(\Sigma_0)\Delta && (5.4.36)\\
\operatorname{var}(\operatorname{vec}(u_{2,n})) &= \Delta^2\big(E(\mathbf{\Sigma}_n)\otimes(\mu\mu^T) + (\mu\mu^T)\otimes E(\mathbf{\Sigma}_n) + \mu^T\otimes E(\mathbf{\Sigma}_n)\otimes\mu \\
&\quad +\mu\otimes E(\mathbf{\Sigma}_n)\otimes\mu^T\big) + (I_{d^2}+\mathbf{P})\left(E(\mathbf{\Sigma}_n)\otimes E(\mathbf{\Sigma}_n)\right) \\
&\quad +(\mathbf{Q}+\mathbf{PQ})\operatorname{var}(\operatorname{vec}(\mathbf{\Sigma}_n)) && (5.4.37)\\
&= \Delta^3\big(E(\Sigma_0)\otimes(\mu\mu^T) + (\mu\mu^T)\otimes E(\Sigma_0) + \mu^T\otimes E(\Sigma_0)\otimes\mu \\
&\quad +\mu\otimes E(\Sigma_0)\otimes\mu^T\big) + \Delta^2(I_{d^2}+\mathbf{P})\left(E(\Sigma_0)\otimes E(\Sigma_0)\right) \\
&\quad +(\mathbf{Q}+\mathbf{PQ})\left(r^{++}(\Delta) + (r^{++}(\Delta))^T\right) \\
\operatorname{cov}(u_{1,n},\operatorname{vec}(u_{2,n})) &= \Delta\left(E(\mathbf{\Sigma}_n)\otimes\mu^T + \mu^T\otimes E(\mathbf{\Sigma}_n)\right) \\
&= \Delta^2\left(E(\Sigma_0)\otimes\mu^T + \mu^T\otimes E(\Sigma_0)\right) && (5.4.38)
\end{aligned}$$

Proof. The stationarity follows immediately from the stationarity of the processes involved in the definition. For the martingale difference sequence property observe

$$\begin{aligned}
E(u_{1,n}|\mathcal{G}_{n-1}) &= E\left(E\left(\int_{(n-1)\Delta}^{n\Delta}\Sigma_s^{1/2}dW_s\,\middle|\,\boldsymbol{\sigma}\left(\mathcal{G}_{n-1},(\Sigma_s)_{s\in[(n-1)\Delta,n\Delta]}\right)\right)\middle|\,\mathcal{G}_{n-1}\right)\\
&= E(0|\mathcal{G}_{n-1}) = 0, && (5.4.39)
\end{aligned}$$

since obviously $\boldsymbol{\sigma}\left(\mathcal{G}_{n-1},(\Sigma_s)_{s\in[(n-1)\Delta,n\Delta]}\right)\subseteq\boldsymbol{\sigma}\left(\mathcal{G}_{n-1},(\Sigma_s)_{s\in\mathbb{R}^+}\right)$ and thus the Brownian increments $(W_s-W_{(n-1)\Delta})_{s\in[(n-1)\Delta,n\Delta]}$ are independent of the σ-algebra we are conditioning upon, and likewise

$$E(u_{2,n}|\mathcal{G}_{n-1}) = E\left(\int_{(n-1)\Delta}^{n\Delta}\Sigma_s ds + 0 + 0 - \mathbf{\Sigma}_n\,\middle|\,\mathcal{G}_{n-1}\right) = 0. \tag{5.4.40}$$

Taking unconditional expectations above gives that u_n has mean zero. The Itô isometry immediately implies (5.4.36).

Turning to (5.4.38) we observe that

$$\begin{aligned}
&\operatorname{cov}(u_{1,n},\operatorname{vec}(u_{2,n})) && (5.4.41)\\
&= E\left(\int_{(n-1)\Delta}^{n\Delta}\Sigma_s^{1/2}dW_s\operatorname{vec}\left(\int_{(n-1)\Delta}^{n\Delta}\Sigma_s^{1/2}dW_s\int_{(n-1)\Delta}^{n\Delta}dW_s^T\Sigma_s^{1/2}\right)^T\right)\\
&\quad + E\left(\int_{(n-1)\Delta}^{n\Delta}\Sigma_s^{1/2}dW_s\operatorname{vec}\left(\Delta\int_{(n-1)\Delta}^{n\Delta}\Sigma_s^{1/2}dW_s\mu^T\right)^T\right)\\
&\quad + E\left(\int_{(n-1)\Delta}^{n\Delta}\Sigma_s^{1/2}dW_s\operatorname{vec}\left(\Delta\mu\int_{(n-1)\Delta}^{n\Delta}dW_s^T\Sigma_s^{1/2}\right)^T\right)\\
&\quad - E\left(\int_{(n-1)\Delta}^{n\Delta}\Sigma_s^{1/2}dW_s\operatorname{vec}\left(\mathbf{\Sigma}_n\right)^T\right).
\end{aligned}$$

Conditioning upon $(\Sigma_s)_{s\in[(n-1)\Delta,n\Delta]}$, it is once again easy to see that the last expectation in (5.4.41) vanishes. Set now for a moment $X=(x_i)_{1\le i\le d}:=\int_{(n-1)\Delta}^{n\Delta}\Sigma_s^{1/2}dW_s$. Then X is

conditionally upon $(\Sigma_s)_{s\in[(n-1)\Delta,n\Delta]}$ a d-dimensional normal random variable and assume for a moment that the components of $X|(\Sigma_s)_{s\in[(n-1)\Delta,n\Delta]}$ are moreover independent. For any $i,k,l,\in\{1,2,\dots,d\}$ we have that $\left(X\mathrm{vec}(XX^T)^T\right)_{i,(k-1)d+l} = x_ix_kx_l$ thus clearly has zero expectation conditional upon $(\Sigma_s)_{s\in[(n-1)\Delta,n\Delta]}$ due to the conditional independence and Gaussianity. Hence $E(X\mathrm{vec}(XX^T)^T|(\Sigma_s)_{s\in[(n-1)\Delta,n\Delta]}) = 0$. If the components of $X|(\Sigma_s)_{s\in[(n-1)\Delta,n\Delta]}$ are not independent, then due the conditional Gaussianity there is a matrix $C\in M_d(\mathbb{R})$ such that $CX|(\Sigma_s)_{s\in[(n-1)\Delta,n\Delta]}$ has independent components. Therefore

$$\begin{aligned}&E(X\mathrm{vec}(XX^T)^T|(\Sigma_s)_{s\in[(n-1)\Delta,n\Delta]})\\&= C^{-1}E(CX\mathrm{vec}(CX(CX)^T)^T|(\Sigma_s)_{s\in[(n-1)\Delta,n\Delta]})(C^{-T}\otimes C^{-T}) = C^{-1}0(C^{-T}\otimes C^{-T}).\end{aligned}$$

Thus also the first term in (5.4.41) above vanishes.

Let us now calculate the remaining expectations:

$$\begin{aligned}&E\left(\int_{(n-1)\Delta}^{n\Delta}\Sigma_s^{1/2}dW_s\mathrm{vec}\left(\Delta\int_{(n-1)\Delta}^{n\Delta}\Sigma_s^{1/2}dW_s\mu^T\right)^T\right)\\&= \Delta E\left(\int_{(n-1)\Delta}^{n\Delta}\Sigma_s^{1/2}dW_s\left((\mu\otimes I_d)\int_{(n-1)\Delta}^{n\Delta}\Sigma_s^{1/2}dW_s\right)^T\right)\\&= \Delta E\left(\int_{(n-1)\Delta}^{n\Delta}\Sigma_s^{1/2}dW_s\left(\int_{(n-1)\Delta}^{n\Delta}\Sigma_s^{1/2}dW_s\right)^T\right)(\mu^T\otimes I_d)\\&= \Delta E\left(\mathbf{\Sigma}_n\right)(\mu^T\otimes I_d) = \Delta\left(1\otimes E\left(\mathbf{\Sigma}_n\right)\right)(\mu^T\otimes I_d) = \Delta\left(\mu^T\otimes E\left(\mathbf{\Sigma}_n\right)\right)\qquad\text{and}\\&E\left(\int_{(n-1)\Delta}^{n\Delta}\Sigma_s^{1/2}dW_s\mathrm{vec}\left(\Delta\mu\int_{(n-1)\Delta}^{n\Delta}dW_s^T\Sigma_s^{1/2}\right)^T\right)\\&= \Delta E\left(\int_{(n-1)\Delta}^{n\Delta}\Sigma_s^{1/2}dW_s\left((I_d\otimes\mu)\mathrm{vec}\left(\int_{(n-1)\Delta}^{n\Delta}dW_s^T\Sigma_s^{1/2}\right)\right)^T\right)\\&= \Delta E\left(\int_{(n-1)\Delta}^{n\Delta}\Sigma_s^{1/2}dW_s\left(\int_{(n-1)\Delta}^{n\Delta}\Sigma_s^{1/2}dW_s\right)^T\right)(I_d\otimes\mu^T)\\&= \Delta\left(E\left(\mathbf{\Sigma}_n\right)\otimes 1\right)(I_d\otimes\mu^T) = \Delta\left(E\left(\mathbf{\Sigma}_n\right)\otimes\mu^T\right).\end{aligned}$$

Combining these formulae with (5.4.41) and (5.4.9) establishes (5.4.38).

It remains to show (5.4.37). Let $\mathscr{D}$ be defined as in the proof of Proposition 5.4.4

$$\begin{aligned}\mathrm{var}(\mathrm{vec}(u_{2,n})) &= \mathrm{var}\left(\mathrm{vec}\left(\int_{(n-1)\Delta}^{n\Delta}\Sigma_s^{1/2}dW_s\int_{(n-1)\Delta}^{n\Delta}dW_s^T\Sigma_s^{1/2}\right)\right) + \mathrm{var}\left(\mathrm{vec}\left(\mathbf{\Sigma}_n\right)\right)\\&+\mathrm{var}\left(\mathrm{vec}\left(\Delta\int_{(n-1)\Delta}^{n\Delta}\Sigma_s^{1/2}dW_s\mu^T\right)\right) + \mathrm{var}\left(\mathrm{vec}\left(\Delta\mu\int_{(n-1)\Delta}^{n\Delta}dW_s^T\Sigma_s^{1/2}\right)\right)\\&+\mathscr{D}\mathrm{cov}\left(\mathrm{vec}\left(\int_{(n-1)\Delta}^{n\Delta}\Sigma_s^{1/2}dW_s\int_{(n-1)\Delta}^{n\Delta}dW_s^T\Sigma_s^{1/2}\right),\mathrm{vec}\left(\Delta\int_{(n-1)\Delta}^{n\Delta}\Sigma_s^{1/2}dW_s\mu^T\right)\right)\end{aligned}$$

$$
\begin{aligned}
&+ \mathscr{D}\mathrm{cov}\left(\mathrm{vec}\left(\int_{(n-1)\Delta}^{n\Delta}\Sigma_s^{1/2}dW_s\int_{(n-1)\Delta}^{n\Delta}dW_s^T\Sigma_s^{1/2}\right),\mathrm{vec}\left(\Delta\mu\int_{(n-1)\Delta}^{n\Delta}dW_s^T\Sigma_s^{1/2}\right)\right)\\
&- \mathscr{D}\mathrm{cov}\left(\mathrm{vec}\left(\int_{(n-1)\Delta}^{n\Delta}\Sigma_s^{1/2}dW_s\int_{(n-1)\Delta}^{n\Delta}dW_s^T\Sigma_s^{1/2}\right),\mathrm{vec}\left(\mathbf{\Sigma}_n\right)\right)\\
&+ \mathscr{D}\mathrm{cov}\left(\mathrm{vec}\left(\Delta\int_{(n-1)\Delta}^{n\Delta}\Sigma_s^{1/2}dW_s\mu^T\right),\mathrm{vec}\left(\Delta\mu\int_{(n-1)\Delta}^{n\Delta}dW_s^T\Sigma_s^{1/2}\right)\right)\\
&- \mathscr{D}\mathrm{cov}\left(\mathrm{vec}\left(\Delta\int_{(n-1)\Delta}^{n\Delta}\Sigma_s^{1/2}dW_s\mu^T\right),\mathrm{vec}\left(\mathbf{\Sigma}_n\right)\right)\\
&- \mathscr{D}\mathrm{cov}\left(\mathrm{vec}\left(\Delta\mu\int_{(n-1)\Delta}^{n\Delta}dW_s^T\Sigma_s^{1/2}\right),\mathrm{vec}\left(\mathbf{\Sigma}_n\right)\right). \qquad (5.4.42)
\end{aligned}
$$

The only term we already know above is $\mathrm{var}(\mathrm{vec}(\mathbf{\Sigma}_n)) = r^{++}(\Delta) + r^{++}(\Delta)^T$. Therefore we will now calculate the remaining covariances. As we assumed $\mu = 0$ in Theorem 5.4.4, we can use (5.4.21) and obtain

$$
\begin{aligned}
\mathrm{var}\left(\mathrm{vec}\left(\int_{(n-1)\Delta}^{n\Delta}\Sigma_s^{1/2}dW_s\int_{(n-1)\Delta}^{n\Delta}dW_s^T\Sigma_s^{1/2}\right)\right) =& (I_{d^2}+\mathbf{Q}+\mathbf{PQ})\mathrm{var}(\mathrm{vec}(\mathbf{\Sigma}_n))\\
&+ (I_{d^2}+\mathbf{P})\left(E(\mathbf{\Sigma}_n)\otimes E(\mathbf{\Sigma}_n)\right).
\end{aligned}
$$

Using the Itô isometry once again gives:

$$
\begin{aligned}
&\mathrm{var}\left(\mathrm{vec}\left(\Delta\int_{(n-1)\Delta}^{n\Delta}\Sigma_s^{1/2}dW_s\mu^T\right)\right) = \Delta^2\mathrm{var}\left((\mu\otimes I_d)\mathrm{vec}\left(\int_{(n-1)\Delta}^{n\Delta}\Sigma_s^{1/2}dW_s\right)\right)\\
&= \Delta^2(\mu\otimes I_d)E(\mathbf{\Sigma}_n)(\mu^T\otimes I_d) = \Delta^2(\mu\otimes I_d)(1\otimes E(\mathbf{\Sigma}_n))(\mu^T\otimes I_d)\\
&= \Delta^2(\mu\mu^T)\otimes E(\mathbf{\Sigma}_n)
\end{aligned}
$$

and

$$
\begin{aligned}
&\mathrm{var}\left(\mathrm{vec}\left(\Delta\mu\int_{(n-1)\Delta}^{n\Delta}dW_s^T\Sigma_s^{1/2}\right)\right) = \Delta^2\mathrm{var}\left((I_d\otimes\mu)\mathrm{vec}\left(\int_{(n-1)\Delta}^{n\Delta}dW_s^T\Sigma_s^{1/2}\right)\right)\\
&= \Delta^2(I_d\otimes\mu)E(\mathbf{\Sigma}_n)(I_d\otimes\mu^T) = \Delta^2E(\mathbf{\Sigma}_n)\otimes(\mu\mu^T).
\end{aligned}
$$

Defining X as in the discussion of the first summand in (5.4.41) and applying the result obtained there we get

$$
\begin{aligned}
&\mathrm{cov}\left(\mathrm{vec}\left(\int_{(n-1)\Delta}^{n\Delta}\Sigma_s^{1/2}dW_s\int_{(n-1)\Delta}^{n\Delta}dW_s^T\Sigma_s^{1/2}\right),\mathrm{vec}\left(\Delta\int_{(n-1)\Delta}^{n\Delta}\Sigma_s^{1/2}dW_s\mu^T\right)\right)\\
&= \Delta E(\mathrm{vec}(XX^T)((\mu\otimes I_d)X)^T) = \Delta E(\mathrm{vec}(XX^T)X^T)(\mu^T\otimes I_d) = 0
\end{aligned}
$$

and likewise

$$
\mathrm{cov}\left(\mathrm{vec}\left(\int_{(n-1)\Delta}^{n\Delta}\Sigma_s^{1/2}dW_s\int_{(n-1)\Delta}^{n\Delta}dW_s^T\Sigma_s^{1/2}\right),\mathrm{vec}\left(\Delta\mu\int_{(n-1)\Delta}^{n\Delta}dW_s^T\Sigma_s^{1/2}\right)\right) = 0.
$$

Furthermore, the standard conditioning argument and the independence of $(\Sigma_t)_{t\in\mathbb{R}^+}$ and $(W_t)_{t\in\mathbb{R}^+}$ imply

$$\begin{aligned}
&\operatorname{cov}\left(\operatorname{vec}\left(\int_{(n-1)\Delta}^{n\Delta}\Sigma_s^{1/2}dW_s\int_{(n-1)\Delta}^{n\Delta}dW_s^T\Sigma_s^{1/2}\right),\operatorname{vec}\left(\mathbf{\Sigma}_n\right)\right)\\
&=E\left(\operatorname{vec}\left(E\left(\int_{(n-1)\Delta}^{n\Delta}\Sigma_s^{1/2}dW_s\int_{(n-1)\Delta}^{n\Delta}dW_s^T\Sigma_s^{1/2}\middle|(\Sigma_s)_{s\in[(n-1)\Delta,n\Delta]}\right)\right)(\operatorname{vec}(\mathbf{\Sigma}_n))^T\right)\\
&\quad-\operatorname{vec}\left(E\left(\int_{(n-1)\Delta}^{n\Delta}\Sigma_s^{1/2}dW_s\int_{(n-1)\Delta}^{n\Delta}dW_s^T\Sigma_s^{1/2}\right)\right)E(\operatorname{vec}(\mathbf{\Sigma}_n))^T\\
&=E(\operatorname{vec}(\mathbf{\Sigma}_n)(\operatorname{vec}(\mathbf{\Sigma}_n))^T)-E(\operatorname{vec}(\mathbf{\Sigma}_n))E(\operatorname{vec}(\mathbf{\Sigma}_n))^T=\operatorname{var}(\operatorname{vec}(\mathbf{\Sigma}_n)).
\end{aligned}$$

Next we observe that

$$\begin{aligned}
&\operatorname{cov}\left(\operatorname{vec}\left(\Delta\int_{(n-1)\Delta}^{n\Delta}\Sigma_s^{1/2}dW_s\mu^T\right),\operatorname{vec}\left(\Delta\mu\int_{(n-1)\Delta}^{n\Delta}dW_s^T\Sigma_s^{1/2}\right)\right)\\
&=\Delta^2(\mu\otimes I_d)\operatorname{var}\left(\int_{(n-1)\Delta}^{n\Delta}\Sigma_s^{1/2}dW_s\right)(I_d\otimes\mu^T)\\
&=\Delta^2(\mu\otimes I_d)E(\mathbf{\Sigma}_n)(I_d\otimes\mu^T)=\Delta^2\mu\otimes E(\mathbf{\Sigma}_n)\otimes\mu^T,
\end{aligned}$$

since

$$\begin{aligned}
&\left((\mu\otimes I_d)E(\mathbf{\Sigma}_n)(I_d\otimes\mu^T)\right)_{(i-1)d+j,(k-1)d+l}=\mu_iE(\mathbf{\Sigma}_{jk,n})\mu_l\\
&\quad=\left(\mu\otimes E(\mathbf{\Sigma}_n)\otimes\mu^T\right)_{(i-1)d+j,(k-1)d+l}
\end{aligned}$$

for all $i,j,k,l\in\{1,2,\dots,d\}$.

Finally using again the conditioning and independence it is immediate that

$$\begin{aligned}
&\operatorname{cov}\left(\operatorname{vec}\left(\Delta\int_{(n-1)\Delta}^{n\Delta}\Sigma_s^{1/2}dW_s\mu^T\right),\operatorname{vec}\left(\mathbf{\Sigma}_n\right)\right)=0\quad\text{and}\\
&\operatorname{cov}\left(\operatorname{vec}\left(\Delta\mu\int_{(n-1)\Delta}^{n\Delta}dW_s^T\Sigma_s^{1/2}\right),\operatorname{vec}\left(\mathbf{\Sigma}_n\right)\right)=0.
\end{aligned}$$

Inserting all these results into (5.4.42) gives

$$\begin{aligned}
\operatorname{var}(\operatorname{vec}(u_{2,n}))&=(I_{d^2}+\mathbf{Q}+\mathbf{PQ})\operatorname{var}(\operatorname{vec}(\mathbf{\Sigma}_n))+(I_{d^2}+\mathbf{P})E(\mathbf{\Sigma}_n)\otimes E(\mathbf{\Sigma}_n)\\
&\quad+\Delta^2\left(E(\mathbf{\Sigma}_n)\otimes(\mu\mu^T)+(\mu\mu^T)\otimes E(\mathbf{\Sigma}_n)+\mu\otimes E(\mathbf{\Sigma}_n)\otimes\mu^T+\left(\mu\otimes E(\mathbf{\Sigma}_n)\otimes\mu^T\right)^T\right)\\
&\quad+\operatorname{var}(\operatorname{vec}(\mathbf{\Sigma}_n))-\operatorname{var}(\operatorname{vec}(\mathbf{\Sigma}_n))-\operatorname{var}(\operatorname{vec}(\mathbf{\Sigma}_n))^T\\
&=(\mathbf{Q}+\mathbf{PQ})\operatorname{var}(\operatorname{vec}(\mathbf{\Sigma}_n))+(I_{d^2}+\mathbf{P})E(\mathbf{\Sigma}_n)\otimes E(\mathbf{\Sigma}_n)\\
&\quad+\Delta^2\left(E(\mathbf{\Sigma}_n)\otimes(\mu\mu^T)+(\mu\mu^T)\otimes E(\mathbf{\Sigma}_n)+\mu\otimes E(\mathbf{\Sigma}_n)\otimes\mu^T+\mu^T\otimes E(\mathbf{\Sigma}_n)\otimes\mu\right),
\end{aligned}$$

which is (5.4.37). □

Remark 5.4.9. *For the commutation matrix K_d as defined in Magnus and Neudecker (1979), for instance, it can be shown that $K_d(X \otimes Z) = \mathbf{P}(X \otimes Z)$ for all $X, Z \in \mathbb{S}_d$. Thus, the operator $\mathbf{P}$ can be replaced by the commutation matrix K_d in Theorem 5.4.4 and Proposition 5.4.8. Note, however, that for $X \in M_{d^2}(\mathbb{R})$ multiplication by K_d is in general not the same as applying $\mathbf{P}$.*

Although the noise u_n is independent of $\mathbf{\Sigma}_n$, we are still confronted with the problem that inference is infeasible, as the latent process $\mathbf{\Sigma}_n$ still appears in the equations for our observables (5.4.34) and (5.4.35). Clearly, it would be desirable for the process $(\mathbf{\Sigma}_n)_{n\in\mathbb{N}}$ to be also representable as a linear process, preferably with a noise sequence that is uncorrelated with $(u_n)_{n\in\mathbb{N}}$, because then the equations (5.4.34) and (5.4.35) could be extended to a state-space model (for a detailed treatment see e.g. Brockwell and Davis (1991, Chapter 12)) and all the tools developed for these models would be available.

In the following we show that at least for the OU type stochastic volatility model such a state-space representation is indeed available.

To this end, define

$$\eta_{1,n} := \int_{(n-1)\Delta}^{n\Delta} e^{A(n\Delta-s)} dL_s e^{A^T(n\Delta-s)}, \qquad \eta_{2,n} := \int_{(n-1)\Delta}^{n\Delta} dL_s = L_{n\Delta} - L_{(n-1)\Delta}$$

and $\eta_n := (\eta_{1,n}, \eta_{2,n})$. Then for all $n \in \mathbb{N}$ it is obvious that

$$\Sigma_{n\Delta} = e^{A\Delta}\Sigma_{(n-1)\Delta}e^{A^T\Delta} + \eta_{1,n} \text{ and } L_{n\Delta} = L_{(n-1)\Delta} + \eta_{2,n}.$$

Before showing that this leads to a helpful state-space representation we first study the properties of the noise sequence $(\eta_n)_{n\in\mathbb{N}}$.

Proposition 5.4.10. *The sequence of random variables $(\eta_n)_{n\in\mathbb{N}}$ is i.i.d. and uncorrelated with $(u_n)_{n\in\mathbb{N}}$. Moreover, it has finite second moments and*

$$\begin{aligned}
E(\eta_{1,n}) &= -\mathbf{A}^{-1}\left(E(L_1) - e^{A\Delta}E(L_1)e^{A^T\Delta}\right) \\
&= E(\Sigma_0) - e^{A\Delta}E(\Sigma_0)e^{A^T\Delta} && (5.4.43)\\
E(\eta_{2,n}) &= \Delta E(L_1) = -\Delta\mathbf{A}E(\Sigma_0) && (5.4.44)\\
\operatorname{var}(\operatorname{vec}(\eta_{1,n})) &= -\mathcal{A}^{-1}\left(\operatorname{var}(\operatorname{vec}(L_1)) - e^{\mathscr{A}\Delta}\operatorname{var}(\operatorname{vec}(L_1))e^{\mathscr{A}^T\Delta}\right) \\
&= \operatorname{var}(\operatorname{vec}(\Sigma_0)) - e^{\mathscr{A}\Delta}\operatorname{var}(\operatorname{vec}(\Sigma_0))e^{\mathscr{A}^T\Delta} && (5.4.45)\\
\operatorname{var}(\operatorname{vec}(\eta_{2,n})) &= \Delta\operatorname{var}(\operatorname{vec}(L_1)) = -\Delta\mathcal{A}\operatorname{var}(\operatorname{vec}(\Sigma_0)) && (5.4.46)\\
\operatorname{cov}\left(\operatorname{vec}(\eta_{1,n}), \operatorname{vec}(\eta_{2,n})\right) &= -\mathscr{A}^{-1}\left(\operatorname{var}(\operatorname{vec}(L_1)) - e^{\mathscr{A}\Delta}\operatorname{var}(\operatorname{vec}(L_1))\right) \\
&= \mathscr{A}^{-1}\mathcal{A}(I_{d^2} - e^{\mathscr{A}\Delta})\operatorname{var}(\operatorname{vec}(\Sigma_0)). && (5.4.47)
\end{aligned}$$

Proof. The i.i.d. property immediately follows from the i.i.d. property of the increments of a Lévy process over disjoint intervals of common length. To see that $(\eta_n)_{n\in\mathbb{N}}$ and $(u_n)_{n\in\mathbb{N}}$ are uncorrelated, it suffices to note that

$$E(u_n|(L_s)_{s\in\mathbb{R}^+}) = 0 \text{ for all } n \in \mathbb{N}$$

and $(\eta_n)_{n\in\mathbb{N}}$ is measurable with respect to the σ-algebra generated by $(L_s)_{s\in\mathbb{R}^+}$.

The second equalities in all the above formulae are immediate consequences of Proposition 5.3.5 and the commutativity of $\mathcal{A}$ and $\mathscr{A}$ (see Theorem 5.4.2). Let us thus proof the first equalities now. Regarding (5.4.43) we have

$$E(\eta_{1,n}) = \int_{(n-1)\Delta}^{n\Delta} e^{A(n\Delta-s)} E(L_1) e^{A^T(n\Delta-s)} ds = -\mathbf{A}^{-1}\left(E(L_1) - e^{A\Delta}E(L_1)e^{A^T\Delta}\right)$$

and (5.4.44) follows analogously. Turning to (5.4.45) we obtain

$$\begin{aligned}\mathrm{var}(\mathrm{vec}(\eta_1)) &= \mathrm{var}\left(\int_{(n-1)\Delta}^{n\Delta} e^{\mathscr{A}(n\Delta-s)} d\mathrm{vec}(L_s)\right)\\ &= \int_{(n-1)\Delta}^{n\Delta} e^{\mathscr{A}(n\Delta-s)} \mathrm{var}(\mathrm{vec}(L_1)) e^{\mathscr{A}^T(n\Delta-s)}\\ &= -\mathcal{A}^{-1}\left(\mathrm{var}(\mathrm{vec}(L_1)) - e^{\mathscr{A}\Delta}\mathrm{var}(\mathrm{vec}(L_1))e^{\mathscr{A}^T\Delta}\right)\end{aligned}$$

and (5.4.46) is again shown along the same lines. Finally, (5.4.47) follows from

$$\begin{aligned}\mathrm{cov}(\mathrm{vec}(\eta_{1,n}), \mathrm{vec}(\eta_{2,n})) &= \mathrm{cov}\left(\int_{(n-1)\Delta}^{n\Delta} e^{\mathscr{A}(n\Delta-s)} d\mathrm{vec}(L_s), \int_{(n-1)\Delta}^{n\Delta} d\mathrm{vec}(L_s)\right)\\ &= \int_{(n-1)\Delta}^{n\Delta} e^{\mathscr{A}(n\Delta-s)} \mathrm{var}(\mathrm{vec}(L_1)) ds\\ &= -\mathscr{A}^{-1}\left(\mathrm{var}(\mathrm{vec}(L_1)) - e^{\mathscr{A}\Delta}\mathrm{var}(\mathrm{vec}(L_1))\right).\end{aligned}$$

□

Proposition 5.3.21 implies

$$\mathbf{\Sigma}_n = \Sigma^+_{n\Delta} - \Sigma^+_{(n-1)\Delta} = \mathbf{A}^{-1}\left(\Sigma_{n\Delta} - \Sigma_{(n-1)\Delta} - L_{n\Delta} + L_{(n-1)\Delta}\right)$$

for all $n \in \mathbb{N}$. Recalling the definition of η_n one thus obtains

$$\mathbf{A}\mathbf{\Sigma}_n = e^{A\Delta}\Sigma_{(n-1)\Delta}e^{A^T\Delta} - \Sigma_{(n-1)\Delta} + \eta_{1,n} - \eta_{2,n}.$$

Combining this with the representation (5.4.34) and (5.4.35) of the observable log price $\mathbf{Y}_n$ and its "square" $\mathbf{Y}_n\mathbf{Y}_n^T$ and setting $\alpha_{1,n} = \mathbf{A}\mathbf{\Sigma}_n$ and $\alpha_{2,n} = \Sigma_{n\Delta}$ yields the desired state-space representation:

$$\mathbf{Y}_n = \Delta\mu + u_{1,n} \tag{5.4.48}$$

$$\mathbf{Y}_n\mathbf{Y}_n^T = \Delta^2\mu\mu^T + \mathbf{A}^{-1}\alpha_{1,n} + u_{2,n}. \tag{5.4.49}$$

where

$$\alpha_{1,n} = e^{A\Delta}\alpha_{2,n-1}e^{A^T\Delta} - \alpha_{2,n-1} + \eta_{1,n} - \eta_{2,n} \tag{5.4.50}$$

$$\alpha_{2,n} = e^{A\Delta}\alpha_{2,n-1}e^{A^T\Delta} + \eta_{1,n} \tag{5.4.51}$$

or in pure vector notation with $\alpha_n := \begin{pmatrix} \mathrm{vec}(\alpha_{1,n}) \\ \mathrm{vec}(\alpha_{2,n}) \end{pmatrix}$

$$\begin{pmatrix} \mathbf{Y}_n \\ \mathrm{vec}(\mathbf{Y}_n \mathbf{Y}_n^T) \end{pmatrix} = \begin{pmatrix} \Delta\mu \\ \Delta^2(\mu \otimes \mu) \end{pmatrix} + \begin{pmatrix} 0_{M_{d,d^2}(\mathbb{R})} & 0_{M_{d,d^2}(\mathbb{R})} \\ \mathscr{A}^{-1} & 0_{M_{d^2,d^2}(\mathbb{R})} \end{pmatrix} \alpha_n + \begin{pmatrix} u_{1,n} \\ \mathrm{vec}(u_{2,n}) \end{pmatrix} \tag{5.4.52}$$

where

$$\alpha_n = \begin{pmatrix} 0_{M_{d^2}(\mathbb{R})} & e^{A\Delta} \otimes e^{A\Delta} - I_{d^2} \\ 0_{M_{d^2}(\mathbb{R})} & e^{A\Delta} \otimes e^{A\Delta} \end{pmatrix} \alpha_{n-1} + \begin{pmatrix} \mathrm{vec}(\eta_{1,n} - \eta_{2,n}) \\ \mathrm{vec}(\eta_{1,n}) \end{pmatrix}. \tag{5.4.53}$$

Observe that $0_{M_{d,d^2}(\mathbb{R})}$ denotes the zero matrix in $M_{d,d^2}(\mathbb{R})$.

As regards the noise terms we have that $(u_n)_{n\in\mathbb{N}}$ and $(\eta_n)_{n\in\mathbb{N}}$ are uncorrelated. Furthermore $(\eta_n)_{n\in\mathbb{N}}$ is an i.i.d. sequence and $(u_n)_{n\in\mathbb{N}}$ a martingale difference sequence. This state-space representation can be used to conduct model inference. Moreover, the volatility states can be inferred straightforwardly using the Kalman filter.

5.4.3. Realised quadratic variation

As in Barndorff-Nielsen and Shephard (2003) we will now show that we can determine the second-order structure of the realised quadratic variation as well as a state space representation for it. Note that a detailed asymptotic distribution theory for the realised quadratic variation and its generalizations has been provided in Barndorff-Nielsen and Shephard (2004), Barndorff-Nielsen, Graversen, Jacod and Shephard (2006) and Barndorff-Nielsen, Graversen, Jacod, Podolskij and Shephard (2006). In the following we start again with the general stochastic volatility model and focus on the OU type model later on.

Let us thus assume that we do not only observe the price increments $\mathbf{Y}_n$ over increments of length Δ, but also for each $n \in \mathbb{N}$ the price increments $\mathbf{Y}_{j,n} := Y_{((n-1)+\frac{j}{M})\Delta} - Y_{((n-1)+\frac{j-1}{M})\Delta}$ with $j = 1, \ldots, M$ over subintervals of length Δ/M for some $M \in \mathbb{N}$ (for example, Δ might be one day and Δ/M ten minutes). Moreover, we aggregate the $M \in \mathbb{N}$ consecutive squared increments over a period of length Δ to realised variation $[Y]_n^{(M)}$ and set $\mathbf{\Sigma}_{j,n} = \Sigma^+_{((n-1)+\frac{j}{M})\Delta} - \Sigma^+_{((n-1)+\frac{j-1}{M})\Delta}$ with $n \in \mathbb{N}$, $j = 1, \ldots, M$. Precisely, we have for $n \in \mathbb{N}$:

$$\mathbf{Y}_n = Y_{n\Delta} - Y_{(n-1)\Delta} = \sum_{j=1}^{M} \mathbf{Y}_{j,n} \tag{5.4.54}$$

$$[\mathbf{Y}]_n^{(M)} := \sum_{j=1}^{M} \mathbf{Y}_{j,n} \mathbf{Y}_{j,n}^T \tag{5.4.55}$$

$$\mathbf{\Sigma}_n = \Sigma^+_{n\Delta} - \Sigma^+_{(n-1)\Delta} = \sum_{j=1}^{M} \mathbf{\Sigma}_{j,n}. \tag{5.4.56}$$

Note that in law $\mathbf{Y}_{j,n}$ and $\mathbf{\Sigma}_{j,n}$ are just the same as $\mathbf{Y}_n$ and $\mathbf{\Sigma}_n$ calculated not over intervals of length Δ but over intervals of length Δ/M, whereas the realised variation $[\mathbf{Y}]_n^{(M)}$ is distinct from the squared increments over intervals of length Δ.

Theorem 5.4.11. *Consider the general stochastic volatility model with $\mu = \beta = 0$. Then the sequence $([\mathbf{Y}]_n^{(M)})_{n\in\mathbb{N}}$ is (second order) stationary and*

$$E\left([\mathbf{Y}]_n^{(M)}|\mathbf{\Sigma}_n\right) = \mathbf{\Sigma}_n \tag{5.4.57}$$

$$E\left([\mathbf{Y}]_n^{(M)}\right) = E(\mathbf{\Sigma}_n) = \Delta E(\Sigma_0) \tag{5.4.58}$$

$$\begin{aligned}\operatorname{var}\left(\operatorname{vec}\left([\mathbf{Y}]_n^{(M)}\right)\right) &= \operatorname{var}(\operatorname{vec}(\mathbf{\Sigma}_n)) + M(\mathbf{Q}+\mathbf{QP})\left(r^{++}\left(\frac{\Delta}{M}\right) + \left(r^{++}\left(\frac{\Delta}{M}\right)\right)^T\right)\\ &\quad + \frac{\Delta^2}{M}(I_{d^2}+\mathbf{P})(E(\Sigma_0)\otimes E(\Sigma_0)) \qquad (5.4.59)\\ &= r^{++}(\Delta) + (r^{++}(\Delta))^T\\ &\quad + M(\mathbf{Q}+\mathbf{QP})\left(r^{++}\left(\frac{\Delta}{M}\right) + \left(r^{++}\left(\frac{\Delta}{M}\right)\right)^T\right)\\ &\quad + \frac{\Delta^2}{M}(I_{d^2}+\mathbf{P})(E(\Sigma_0)\otimes E(\Sigma_0))\end{aligned}$$

$$\begin{aligned}\operatorname{acov}_{[\mathbf{Y}]^{(M)}}(h) &= \operatorname{acov}_{\mathbf{\Sigma}}(h) \qquad (5.4.60)\\ &= r^{++}(h\Delta+\Delta) - 2r^{++}(h\Delta) + r^{++}(h\Delta-\Delta),\ h\in\mathbb{N}.\end{aligned}$$

In the Ornstein-Uhlenbeck case the realised quadratic variation $(\operatorname{vec}([\mathbf{Y}]_n^{(M)}))_{n\in\mathbb{N}}$ is thus a causal ARMA(1,1) process with autoregressive parameter $e^{\mathscr{A}\Delta}$.

Proof. Formulae (5.4.57) and (5.4.58) are immediate as is the stated (second order) stationarity.

Using the operator $\mathscr{D}$ as defined before we have:

$$\begin{aligned}&\operatorname{var}\left(\operatorname{vec}\left([\mathbf{Y}]_1^{(M)}\right)\right)\\ &\quad= \sum_{j=1}^M \operatorname{var}(\operatorname{vec}(\mathbf{Y}_{j,1}\mathbf{Y}_{j,1}^T)) + \sum_{j=1}^M\sum_{k=j+1}^M \mathscr{D}\operatorname{cov}(\operatorname{vec}(\mathbf{Y}_{j,1}\mathbf{Y}_{j,1}^T),\operatorname{vec}(\mathbf{Y}_{k,1}\mathbf{Y}_{k,1}^T))\\ &\quad= M\operatorname{var}(\operatorname{vec}(\mathbf{Y}_{1,1}\mathbf{Y}_{1,1}^T)) + \sum_{j=1}^M\sum_{k=j+1}^M \mathscr{D}\operatorname{cov}(\operatorname{vec}(\mathbf{\Sigma}_{j,1}),\operatorname{vec}(\mathbf{\Sigma}_{k,1}))\\ &\quad= M\operatorname{var}(\operatorname{vec}(\mathbf{Y}_{1,1}\mathbf{Y}_{1,1}^T)) + \operatorname{var}(\operatorname{vec}(\mathbf{\Sigma}_1)) - M\operatorname{var}(\operatorname{vec}(\mathbf{\Sigma}_{1,1}))\\ &\quad= \operatorname{var}(\operatorname{vec}(\mathbf{\Sigma}_1)) + M(\mathbf{Q}+\mathbf{QP})\left(r^{++}(\Delta/M) + \left(r^{++}(\Delta/M)\right)^T\right)\\ &\qquad+ \frac{\Delta^2}{M}(I_{d^2}+\mathbf{P})(E(\Sigma_0)\otimes E(\Sigma_0))\\ &\quad= r^{++}(\Delta) + (r^{++}(\Delta))^T + M(\mathbf{Q}+\mathbf{QP})\left(r^{++}(\Delta/M) + (r^{++}(\Delta/M))^T\right)\\ &\qquad+ \frac{\Delta^2}{M}(I_{d^2}+\mathbf{P})(E(\Sigma_0)\otimes E(\Sigma_0))\end{aligned}$$

and

$$\operatorname{acov}_{[\mathbf{Y}]^{(M)}}(h) = \operatorname{cov}\left(\operatorname{vec}\left([\mathbf{Y}]_{h+1}^{(M)}\right), \operatorname{vec}\left([\mathbf{Y}]_1^{(M)}\right)\right)$$

$$
\begin{aligned}
&= \sum_{j=1}^{M}\sum_{k=1}^{M} \operatorname{cov}\left(\operatorname{vec}(\mathbf{Y}_{j,h+1}\mathbf{Y}_{j,h+1}^T), \operatorname{vec}(\mathbf{Y}_{k,1}\mathbf{Y}_{k,1}^T)\right) \\
&= \sum_{j=1}^{M}\sum_{k=1}^{M} \operatorname{cov}\left(\operatorname{vec}(\mathbf{\Sigma}_{j,h+1}), \operatorname{vec}(\mathbf{\Sigma}_{k,1})\right) \\
&= \operatorname{cov}(\operatorname{vec}(\mathbf{\Sigma}_{h+1}), \operatorname{vec}(\mathbf{\Sigma}_1)) = \operatorname{acov}_{\mathbf{\Sigma}}(h).
\end{aligned}
$$

In the OU case the structure of the autocovariance function now implies immediately that the realised quadratic variation follows an ARMA(1,1) process with AR parameter $e^{\mathscr{A}\Delta}$. □

Moreover, we can also generalize the state space representation of the last section by simple addition.

Assuming again given the general stochastic volatility model with $\beta = 0$ we define for $n \in \mathbb{N}$ and $j = 1, 2, \ldots, M$

$$
\begin{aligned}
u_n^{(M)} &= \begin{pmatrix} u_{1,n}^{(M)} \\ u_{2,n}^{(M)} \end{pmatrix} && (5.4.61)\\
u_{1,j,n}^{(M)} &= \int_{((n-1)+\frac{j-1}{M})\Delta}^{(n+\frac{j}{M})\Delta} \Sigma_s^{1/2} dW_s && (5.4.62)\\
u_{2,j,n}^{(M)} &= \int_{((n-1)+\frac{j-1}{M})\Delta}^{(n+\frac{j}{M})\Delta} \Sigma_s^{1/2} dW_s \int_{((n-1)+\frac{j-1}{M})\Delta}^{(n+\frac{j}{M})\Delta} dW_s^T \Sigma_s^{1/2} + \Delta \int_{((n-1)+\frac{j-1}{M})\Delta}^{(n+\frac{j}{M})\Delta} \Sigma_s^{1/2} dW_s \mu^T \\
&\quad + \Delta\mu \int_{((n-1)+\frac{j-1}{M})\Delta}^{(n+\frac{j}{M})\Delta} dW_s^T \Sigma_s^{1/2} - \mathbf{\Sigma}_{j,n} && (5.4.63)\\
u_{1,n}^{(\mathcal{M})} &= \sum_{j=1}^{M} u_{1,j,n}^{(M)} && (5.4.64)\\
u_{2,n}^{(\mathcal{M})} &= \sum_{j=1}^{M} u_{2,j,n}^{(M)}. && (5.4.65)
\end{aligned}
$$

Then the sequence $\left(u_{j,n}^{(M)}\right)_{n\in\mathbb{N}} := \left(u_{1,j,n}^{(M)}, u_{2,j,n}^{(M)}\right)_{n\in\mathbb{N}}$ is in law equal to the sequence $(u_n)_{n\in\mathbb{N}}$ with Δ replaced by Δ/M. It is straightforward to see that

$$
\mathbf{Y}_n = \Delta\mu + u_{1,n}^{(\mathcal{M})} \tag{5.4.66}
$$

$$
[\mathbf{Y}]_n^{(M)} = \frac{\Delta^2}{M}\mu\mu^T + \mathbf{\Sigma}_n + u_{2,n}^{(\mathcal{M})} \tag{5.4.67}
$$

and $(u_n^{(\mathcal{M})})_{n\in\mathbb{N}} := (u_{1,n}^{(\mathcal{M})}, u_{2,n}^{(\mathcal{M})})_{n\in\mathbb{N}}$ is a (second order) stationary zero-mean martingale difference sequence with respect to the filtration $\mathcal{G}_n := \mathcal{F}_{n\Delta}$, $n \in \mathbb{N}$. Moreover,

$$
\operatorname{var}(u_{1,n}^{(\mathcal{M})}) = E(\mathbf{\Sigma}_n) = \Delta E(\Sigma_0) \tag{5.4.68}
$$

$$\begin{aligned}\operatorname{var}(\operatorname{vec}(u_{2,n}^{(\mathcal{M})})) &= M\operatorname{var}(\operatorname{vec}(u_{2,1,n})) = \frac{\Delta^3}{M^2}\big(E(\Sigma_0)\otimes(\mu\mu^T) + (\mu\mu^T)\otimes E(\Sigma_0) \\ &\quad + \mu^T\otimes E(\Sigma_0)\otimes\mu + \mu\otimes E(\Sigma_0)\otimes\mu^T\big) \\ &\quad + \frac{\Delta^2}{M}(I_{d^2}+\mathbf{P})\left(E(\Sigma_0)\otimes E(\Sigma_0)\right) \\ &\quad + M(\mathbf{Q}+\mathbf{PQ})\left(r^{++}(\Delta/M) + (r^{++}(\Delta/M))^T\right) \qquad (5.4.69)\\ \operatorname{cov}(u_{1,n}^{(\mathcal{M})}, \operatorname{vec}(u_{2,n}^{(\mathcal{M})})) &= \frac{\Delta}{M}\left(E(\mathbf{\Sigma}_n)\otimes\mu^T + \mu^T\otimes E(\mathbf{\Sigma}_n)\right) \\ &= \frac{\Delta^2}{M}\left(E(\Sigma_0)\otimes\mu^T + \mu^T\otimes E(\Sigma_0)\right). \qquad (5.4.70)\end{aligned}$$

In the Ornstein-Uhlenbeck type stochastic volatility model we can use equations (5.4.50) and (5.4.51) to obtain a recursion for $(\mathbf{A\Sigma}_n, \Sigma_{n\Delta})$. Hence, we obtain the following state space representation

$$\mathbf{Y}_n = \Delta\mu + u_{1,n}^{(\mathcal{M})} \qquad (5.4.71)$$

$$[\mathbf{Y}]_n^{(M)} = \frac{\Delta^2}{M}\mu\mu^T + \alpha_{1,n} + u_{2,n}^{(\mathcal{M})} \qquad (5.4.72)$$

where

$$\alpha_{1,n} = e^{A\Delta}\alpha_{2,n-1}e^{A^T\Delta} - \alpha_{2,n-1} + \eta_{1,n} - \eta_{2,n} \qquad (5.4.73)$$

$$\alpha_{2,n} = e^{A\Delta}\alpha_{2,n-1}e^{A^T\Delta} + \eta_{1,n}. \qquad (5.4.74)$$

Alternatively in pure vector notation with $\alpha_n := \begin{pmatrix}\operatorname{vec}(\alpha_{1,n})\\ \operatorname{vec}(\alpha_{2,n})\end{pmatrix}$:

$$\begin{pmatrix}\mathbf{Y}_n\\ \operatorname{vec}([\mathbf{Y}]_n^{(M)})\end{pmatrix} = \begin{pmatrix}\Delta\mu\\ \frac{\Delta^2}{M}(\mu\otimes\mu)\end{pmatrix} + \begin{pmatrix}0_{M_{d,d^2}(\mathbb{R})} & 0_{M_{d,d^2}(\mathbb{R})}\\ \mathscr{A}^{-1} & 0_{M_{d^2,d^2}(\mathbb{R})}\end{pmatrix}\alpha_n + \begin{pmatrix}u_{1,n}^{(\mathcal{M})}\\ \operatorname{vec}(u_{2,n}^{(\mathcal{M})})\end{pmatrix} \qquad (5.4.75)$$

where

$$\alpha_n = \begin{pmatrix}0_{M_{d^2}(\mathbb{R})} & e^{A\Delta}\otimes e^{A\Delta} - I_{d^2}\\ 0_{M_{d^2}(\mathbb{R})} & e^{A\Delta}\otimes e^{A\Delta}\end{pmatrix}\alpha_{n-1} + \begin{pmatrix}\operatorname{vec}(\eta_{1,n}-\eta_{2,n})\\ \operatorname{vec}(\eta_{1,n})\end{pmatrix}. \qquad (5.4.76)$$

Furthermore, recall that the sequence $(\eta_n)_{n\in\mathbb{N}} := (\eta_{1,n}, \eta_{2,n})_{n\in\mathbb{N}}$ is i.i.d. and note that it is uncorrelated with $(u_n^{(\mathcal{M})})_{n\in\mathbb{N}}$. The further second order properties are given by equations (5.4.43) to (5.4.47).

5.4.4. Identifiability

In this section we discuss conditions ensuring that the second order structure of $\mathbf{YY}^T$ uniquely identifies $A, E(L_1)$ and $\operatorname{var}(\operatorname{vec}(L_1))$ in the OU type stochastic volatility model with $\mu = \beta = 0$. Note that the identification can already be obtained by using only $E(\mathbf{Y}_1\mathbf{Y}_1^T), \operatorname{acov}_{\mathbf{YY}^T}(1)$ and $\operatorname{acov}_{\mathbf{YY}^T}(2)$. Intuitively, one may think that one can improve upon our results below by using also $\operatorname{var}(\operatorname{vec}(\mathbf{Y}_1\mathbf{Y}_1^T))$. However, we have not been able do this so far due to the rather complicated expression for $\operatorname{var}(\operatorname{vec}(\mathbf{Y}_1\mathbf{Y}_1^T))$.

In the following we consider $\text{vech}(\Sigma)$ and $\text{vech}(\mathbf{Y}\mathbf{Y}^T)$ rather than $\text{vec}(\Sigma)$ and $\text{vec}(\mathbf{Y}\mathbf{Y}^T)$, because otherwise we would necessarily be dealing with singular covariance matrices as the symmetric $d \times d$ matrices are a proper subspace of $M_d(\mathbb{R})$.

Thus we need to use the set-up of Remark 5.3.6 (ii) which we are recalling now. Defining $\mathbf{A}_{\text{vech}} := \text{vech} \circ \mathbf{A} \circ \text{vech}^{-1}$ with $\mathbf{A} : \mathbb{S}_d \to \mathbb{S}_d$, $X \mapsto AX + XA^T$ it is easy to see that

$$d\text{vech}(\Sigma_t) = \mathbf{A}_{\text{vech}}\text{vech}(\Sigma_t)dt + d\text{vech}L_t, \tag{5.4.77}$$

with the stationary solution being given by $\text{vech}(\Sigma_t) = \int_{-\infty}^t e^{\mathbf{A}_{\text{vech}}(t-s)}d\text{vech}L_s$. Observe, moreover, that, if $\mathbf{D}$ denotes the $d^2 \times d(d+1)/2$ duplication matrix and $\mathbf{E}$ the $d(d+1)/2 \times d^2$ elimination matrix by $\mathbf{E}$ (see Lütkepohl (2005), for instance), then $\mathbf{A}_{\text{vech}} = \mathbf{E}\left(A \otimes I_d + I_d \otimes A\right)\mathbf{D}$ and $e^{\mathbf{A}_{\text{vech}}t} = \mathbf{E}\left(e^{At} \otimes e^{At}\right)\mathbf{D}$. Furthermore, in Appendix A it is shown that the linear operator $\mathbf{A}$ on $\mathbb{S}_d$ uniquely identifies A. Actually, to identify A it is already sufficient to know the values of $\mathbf{A}E_{ii}$ for $i = 1, \ldots, d$ where E_{ii} are the $d \times d$ matrices with zero entries except for one entry of one at the i-th diagonal element. Thus, there is a one-to-one correspondence between A and $\mathbf{A}_{\text{vech}}$.

Note that basically all formulae obtained so far in this chapter can be rewritten using vech instead of vec in a straightforward manner. For example, the second order structure of $\mathbf{\Sigma}$ can also be expressed using vech:

$$\begin{aligned}
\text{var}(\text{vech}(\Sigma_0)) &= -\mathcal{A}_{\text{vech}}^{-1}\text{var}(\text{vech}(L_1)) && (5.4.78)\\
&\quad \text{where } \mathcal{A}_{\text{vech}} : M_{d(d+1)/2}(\mathbb{R}) \to M_{d(d+1)/2}(\mathbb{R}),\ X \mapsto \mathbf{A}_{\text{vech}}X + X\mathbf{A}_{\text{vech}}^T \\
\text{acov}_{\text{vech}(\mathbf{\Sigma})}(h) &= \text{acov}_{\text{vech}(\mathbf{Y}\mathbf{Y}^T)}(h) && (5.4.79)\\
&= e^{\mathbf{A}_{\text{vech}}\Delta(h-1)}\mathbf{A}_{\text{vech}}^{-2}\left(I_{d(d+1)/2} - e^{\mathbf{A}_{\text{vech}}\Delta}\right)^2\text{var}(\text{vech}(\Sigma_0)),\ h \in \mathbb{N}.
\end{aligned}$$

Proposition 5.4.12. *Assume that the OU type stochastic volatility model with $\mu = \beta = 0$ and $\Delta \in \mathbb{R}^{++}$ are given and that the possible $A \in M_d(\mathbb{R})$ and matrix subordinators L are restricted such that:*

(i) $\sigma(A) \subset (-\infty, 0) + i\mathbb{R}$.

(ii) $e^{\mathbf{A}_{\text{vech}}\Delta}$ *uniquely identifies* $\mathbf{A}_{\text{vech}}$.

(iii) $\text{var}(\text{vech}(\Sigma_0)) = -\mathcal{A}_{\text{vech}}^{-1}\text{var}(\text{vech}(L_1)) \in GL_{d(d+1)/2}(\mathbb{R})$.

Then $A, E(L_1)$ and $\text{var}(\text{vech}(L_1))$ are uniquely identified by $E(\mathbf{Y}_1\mathbf{Y}_1^T), \text{acov}_{\mathbf{Y}\mathbf{Y}^T}(1)$ and $\text{acov}_{\mathbf{Y}\mathbf{Y}^T}(2)$.

Proof. By construction $\sigma(\mathbf{A}_{\text{vech}}) \subseteq \sigma(\mathbf{A}) = \sigma(A) + \sigma(A)$ and thus (i) ensures that the matrix $\mathbf{A}_{\text{vech}}^{-2}\left(I_{d(d+1)/2} - e^{\mathbf{A}_{\text{vech}}\Delta}\right)^2$ is invertible. Using also (iii) this gives that $\text{acov}_{\text{vech}(\mathbf{Y}\mathbf{Y}^T)}(1)$ is invertible and that $e^{\mathbf{A}_{\text{vech}}\Delta} = \text{acov}_{\text{vech}(\mathbf{Y}\mathbf{Y})}(2)(\text{acov}_{\text{vech}(\mathbf{Y}\mathbf{Y})}(1))^{-1}$. Using (ii) $\mathbf{A}_{\text{vech}}$ is therefore identified from $\text{acov}_{\text{vech}(\mathbf{Y}\mathbf{Y})}(1)$ and $\text{acov}_{\text{vech}(\mathbf{Y}\mathbf{Y})}(2)$. Hence, $\mathbf{A}_{\text{vech}}$ and A can be treated as known and so

$$\text{var}(\text{vech}(L_1)) = -\mathcal{A}_{\text{vech}}\left(I_{d(d+1)/2} - e^{\mathbf{A}_{\text{vech}}\Delta}\right)^{-2}\mathbf{A}_{\text{vech}}^2\text{acov}_{\text{vech}(\mathbf{Y}\mathbf{Y}^T)}(1)$$

and $E(L_1) = -\mathbf{A}\Delta^{-1}E(\mathbf{Y}_1\mathbf{Y}_1^T)$ conclude. □

The Assumption (*ii*) from above is crucial for the identifiability of the OU type stochastic volatility model. It requests that $\exp(\mathbf{A}_{\text{vech}}\Delta)$ has a unique real logarithm. Other statistical papers where it was necessary to assume the existence of a unique real logarithm of a matrix in order to obtain identifiability are Kessler and Rahbek (2004) and Bladt and Sørensen (2005), for instance.

The following results deal with criteria ensuring the existence of a unique real logarithm of $\exp(\mathbf{A}_{\text{vech}}\Delta)$.

Lemma 5.4.13. *Assume that A is required to satisfy $\sigma(\mathbf{A}_{\text{vech}}\Delta) \subseteq (-\infty, 0) + i(-\pi, \pi)$. Then $e^{\mathbf{A}_{\text{vech}}\Delta}$ uniquely identifies $\mathbf{A}_{\text{vech}}$.*

Proof. This follows immediately from Horn and Johnson (1991, Section 6.4). □

Lemma 5.4.14. *Assume that A is required to satisfy $\sigma(\mathbf{A}_{\text{vech}}) \subseteq (-\infty, 0)$ and that all Jordan blocks belonging to the same eigenvalue of $\mathbf{A}_{\text{vech}}$ have to be of a different size. Then $e^{\mathbf{A}_{\text{vech}}\Delta}$ uniquely identifies $\mathbf{A}_{\text{vech}}$ for all $\Delta \in \mathbb{R}^{++}$.*

Proof. Combine Culver (1966, Theorem 2), $\sigma(e^{\mathbf{A}_{\text{vech}}\Delta}) = e^{\sigma(\mathbf{A}_{\text{vech}})\Delta}$ and the fact that the matrix exponential preserves the Jordan block structure (Horn and Johnson (1991, Theorem 6.2.25)). □

The last lemma is comparable to the identifiability restrictions of Kessler and Rahbek (2004) and Bladt and Sørensen (2005). However, it appears to be preferable to have a condition involving only restrictions on A, such that $\mathbf{A}_{\text{vech}}$ does not have to be computed first. To this end we give the following purely linear algebraic lemma which we state for real diagonalizable matrices. Its generalization to diagonalizable matrices is straightforward. Below we denote by $\mathbb{SK}_d$ the $d \times d$ skew-symmetric matrices (i.e. the matrices $X \in M_d(\mathbb{R})$ with $X^T = -X$).

Lemma 5.4.15. *Let $A \in M_d(\mathbb{R})$ be real diagonalizable with (not necessarily distinct) eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_d$.*

Then the linear operator $\mathbf{A} : M_d(\mathbb{R}) \to M_d(\mathbb{R})$, $X \mapsto AX + XA^T$ satisfies $\mathbf{A}(\mathbb{S}_d) \subseteq \mathbb{S}_d$ and $\mathbf{A}(\mathbb{SK}_d) \subseteq \mathbb{SK}_d$. Moreover, $\mathbf{A}$ has $d(d+1)/2$ linearly independent eigenvectors in $\mathbb{S}_d$ with associated eigenvalues $\{\lambda_i + \lambda_j : i = 1, \ldots, d;\ j = 1, \ldots, i\}$ and $d(d-1)/2$ linearly independent eigenvectors in $\mathbb{SK}_d$ with associated eigenvalues $\{\lambda_i + \lambda_j : i = 1, \ldots, d;\ j = 1, \ldots, i-1\}$, which are also linearly independent of the eigenvectors in $\mathbb{S}_d$. Hence, every eigenvalue has an eigenvector in $\mathbb{S}_d \cup \mathbb{SK}_d$.

Proof. That $\mathbf{A}$ preserves (skew-)symmetry is trivial. Assume that $U \in GL_d(\mathbb{R})$ is such that $U^{-1}AU =: D$ is diagonal. Then

$$AX + XA^T = U\left(DU^{-1}XU^{-T} + U^{-1}XU^{-T}D^T\right)U^T$$

and $M_d(\mathbb{R}) \to M_d(\mathbb{R})$, $X \mapsto U^{-1}XU^{-T}$ is an invertible linear map on $M_d(\mathbb{R})$ preserving both $\mathbb{S}_d$ and $\mathbb{SK}_d$ and having inverse $M_d(\mathbb{R}) \to M_d(\mathbb{R})$, $X \mapsto UXU^T$. This implies that we can without loss of generality take A to be diagonal, i.e.

$$A = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_d \end{pmatrix}.$$

Let $\{E_{ij}\}_{i,j=1,\dots,d}$ be the standard basis of $M_d(\mathbb{R})$, i.e. E_{ij} is a matrix having only zero entries except for one entry of one in the i-th row and j-th column. We then have $\mathbf{A}E_{ii} = 2\lambda_i$ for $i = 1, \dots, d$ and $\mathbf{A}(E_{ij} + E_{ji}) = (\lambda_i + \lambda_j)(E_{ij} + E_{ji})$ for $i = 1, \dots, d$ and $j = 1, \dots, i-1$. These are $d(d+1)/2$ linearly independent eigenvectors in $\mathbb{S}_d$. Likewise $\mathbf{A}(E_{ij} - E_{ji}) = (\lambda_i + \lambda_j)(E_{ij} - E_{ji})$ for $i = 1, \dots, d$ and $j = 1, \dots, i-1$ gives $d(d-1)/2$ linearly independent eigenvectors in $\mathbb{SK}_d$.

Since $\mathbf{A}$ has d^2 eigenvalues, the fact that $\mathbb{S}_d \cap \mathbb{SK}_d = \{0\}$ and both are linear subspaces of $M_d(\mathbb{R})$ implies that every eigenvalue has an eigenvector in $\mathbb{S}_d \cup \mathbb{SK}_d$. □

Lemma 5.4.16. *Assume that A is required to be real diagonalizable with eigenvalues $\lambda_1, \dots, \lambda_d$ and that the set $\{\lambda_i + \lambda_j : i = 1, \dots, d, j = 1, \dots, i\}$ has to consist of $d(d+1)/2$ pairwise distinct elements. Then $e^{\mathbf{A}_{\text{vech}}\Delta}$ uniquely identifies $\mathbf{A}_{\text{vech}}$ for all $\Delta \in \mathbb{R}^{++}$.*

Proof. $\mathbf{A}_{\text{vech}}$ corresponds to the restriction of $\mathbf{A} : M_d(\mathbb{R}) \to M_d(\mathbb{R})$, $X \mapsto AX + XA^T$ to $\mathbb{S}_d$. Hence, Lemma 5.4.15 implies that the $d(d+1)/2$ real eigenvalues of $\mathbf{A}_{\text{vech}}$ are pairwise distinct. Thus Culver (1966, Theorem 2) gives that $e^{\mathbf{A}_{\text{vech}}\Delta}$ uniquely identifies $\mathbf{A}_{\text{vech}}$ for all $\Delta \in \mathbb{R}^{++}$. □

Thus, if Lemma 5.4.16 holds, Assumption (*ii*) for the identifiability of our model is fulfilled.

5.4.5. The covariance structure of subordinators

A topic related to identifiability is the characterization of the second order structure of matrix subordinators, i.e. which positive semi-definite $d^2 \times d^2$ matrices appear as covariance matrices of $d \times d$ matrix subordinators. Until now we have not been able to find a useful characterization of the covariance matrices of matrix subordinators in general. Thus we discuss only an important tractable subclass which we shall also use in our empirical examples later on.

Let $L_1, L_2, \dots, L_d$ be d univariate Lévy subordinators forming together a d-dimensional Lévy process. Then the process

$$L = \begin{pmatrix} L_1 & 0 & \cdots & 0 \\ 0 & L_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & L_d \end{pmatrix}$$

is a matrix subordinator with non-zero elements only on the diagonal. Such matrix subordinators are referred to as *diagonal matrix subordinators* in the following. Obviously, all information about L, in particular the second order structure, is already contained in the diagonal $\left((L_{1,t}, L_{2,t}, \dots, L_{d,t})^T\right)_{t \in \mathbb{R}^+}$. It should be noted that $\left((L_{1,t}, L_{2,t}, \dots, L_{d,t})^T\right)_{t \in \mathbb{R}^+}$ is increasing in the usual order of the cone $(\mathbb{R}^+)^d$ and thus the d-dimensional Lévy process $\left((L_{1,t}, L_{2,t}, \dots, L_{d,t})^T\right)_{t \in \mathbb{R}^+}$ is another generalization of a subordinator. In the following we call a d-dimensional Lévy processes with all components being univariate subordinators a *multivariate subordinator*.

For the analysis of the second order structure of multivariate subordinators (and thus of diagonal matrix subordinators) we need the following notions from linear algebra referring to Berman (1988), Berman and Shaked-Monderer (2003) or Xu (2004) and the references therein for further details.

Definition 5.4.17. *A matrix $A \in M_d(\mathbb{R})$ is called* completely positive *if there exist a $k \in \mathbb{N}$ and a $B \in M_{d,k}(\mathbb{R})$ with all entries being non-negative such that $A = BB^T$.*

A matrix $A \in M_d(\mathbb{R})$ is called doubly non-negative *if A is positive semi-definite and all entries are non-negative.*

Proposition 5.4.18. *(a) If L is a d-dimensional multivariate subordinator with finite second moments,* $\operatorname{var}(L_1)$ *is completely positive.*

(b) If $C \in M_d(\mathbb{R})$ is completely positive, there exists a d-dimensional multivariate subordinator L such that $C = \operatorname{var}(L_1)$.

(c) For $C \in M_d(\mathbb{R})$ there exists a d-dimensional multivariate subordinator L such that $\operatorname{var}(L_1) = C$ if and only if C is completely positive.

(d) Provided $d \leq 4$, there exists a d-dimensional multivariate subordinator L such that $\operatorname{var}(L_1) = C$ if and only if $C \in M_d(\mathbb{R})$ is doubly non-negative.

Proof. Regarding Part (a) observe that $\operatorname{var}(L_1) = \int_{\mathbb{R}^d} xx^T \nu_L(dx)$ (cf. Sato (1999, p. 163)) and that the Lévy measure ν_L of L is necessarily concentrated on $(\mathbb{R}^+)^d$. Thus approximating $\int_{(\mathbb{R}^+)^d} xx^T \nu_L(dx)$ by a sequence of integrals of simple functions $(f_n)_{n\in\mathbb{N}}$ of the form $f_n(x) = \sum_{i=1}^k x_{i,n} x_{i,n}^T I_{A_{i,n}}(x)$ with appropriate $x_{i,n} \in (\mathbb{R}^+)^d$ and $A_{i,n} \subset (\mathbb{R}^+)^d$ converging to $x \mapsto xx^T$ gives that $\operatorname{var}(L_1)$ is the limit of a sequence of completely positive matrices using Berman and Shaked-Monderer (2003, Proposition 2.2). Since the set of completely positive matrices is closed (cf. Berman and Shaked-Monderer (2003, Theorem 2.2)), this implies that $\operatorname{var}(L_1)$ is completely positive.

For Part (b) let $k \in \mathbb{N}$ and $B \in M_{d,k}(\mathbb{R})$ with only non-negative entries be such that $C = BB^T$. Let further $(L_{i,t})_{t\in\mathbb{R}^+}$ with $i = 1, 2, \ldots, k$ be k independent univariate subordinators with finite second moments and $\operatorname{var}(\tilde{L}_1) = I_k$ where $\tilde{L}_t = (L_{1,t}, L_{2,t}, \ldots, L_{k,t})^T$ for $t \in \mathbb{R}^+$. Then $L := (B\tilde{L}_t)_{t\in\mathbb{R}^+}$ is a d-dimensional multivariate subordinator with $\operatorname{var}(L_1) = BB^T = C$.

Combing (a) and (b) gives (c) and, finally, (d) follows by from (c), because for $d \leq 4$ a matrix is doubly non-negative if and only if it is completely positive (cf. Maxfield and Minc (1962) or Berman and Shaked-Monderer (2003, Theorem 2.4)). □

Part (d) provides a very nice complete characterization of the covariance matrices of multivariate subordinators for $d \leq 4$. However, it does not extend to higher dimensions, since for dimensions five and greater there exist examples of doubly non-negative matrices which are not completely positive (see e.g. Berman (1988)). In general dimensions it should thus be noted that there are easy to check sufficient conditions for complete positivity, but the necessary and sufficient conditions known until now are more involved (cf. Xu (2004)).

To conclude this discussion of the second order properties of multivariate subordinators, we strengthen Part (b) of the last theorem by considering also the expected value.

Proposition 5.4.19. *Let $C \in M_d(\mathbb{R})$ be completely positive and $\mu \in \mathbb{R}^d$ have only strictly positive entries. Then there exists a d-dimensional multivariate subordinator L such that $E(L_1) = \mu$ and $\operatorname{var}(L_1) = C$.*

Proof. W.l.o.g. assume $C \neq 0$. (In the case $C = 0$ simply take L as the deterministic Lévy process with drift μ.) Let $k \in \mathbb{N}$ and $B \in M_{d,k}(\mathbb{R})$ with only non-negative entries be such that $C = BB^T$. Denote by $\mathbf{e} \in \mathbb{R}^k$ the vector $(1, 1, \ldots, 1)^T$ and set

$$\lambda = \min_{i=1,2,\ldots,d} \left\{ \frac{\mu_i}{(B\mathbf{e})_i} \right\} \quad \text{and} \quad r = \frac{\lambda^3}{2}$$

where the well-definedness of the minimum follows from $C \neq 0$. Defining now $\gamma_L = \mu - 0.5\lambda B\mathbf{e}$ the choice of λ ensures that γ_L has only strictly positive entries.

Let further $(L_{i,t})_{t\in\mathbb{R}^+}$ with $i = 1, 2, \ldots, k$ be k independent univariate compound Poisson processes with rate r and the jump distribution being the exponential distribution with parameter λ. Define a k-dimensional Lévy process $\tilde{L}$ by $\tilde{L}_t = (L_{1,t}, L_{2,t}, \ldots, L_{k,t})^T$ for $t \in \mathbb{R}^+$. Then elementary calculations imply $E(\tilde{L}_1) = 0.5\lambda\mathbf{e}$ and $\mathrm{var}(\tilde{L}_1) = I_k$.

Hence, defining L by $L_t = \gamma_L t + B\tilde{L}_t$ for $t \in \mathbb{R}^+$ concludes the proof. □

5.4.6. Superpositions of Ornstein-Uhlenbeck type processes

In this section we modify our multivariate OU stochastic volatility model by using a superposition of independent positive semi-definite OU type processes for the volatility process. As in the univariate case this makes the model more flexible without loosing much of its tractability.

Let K be a natural number, $(L^{(1)})_{t\in\mathbb{R}}, (L^{(2)})_{t\in\mathbb{R}}, \ldots, (L^{(K)})_{t\in\mathbb{R}}$ be independent matrix subordinators which are jointly independent of the Brownian motion $(W_t)_{t\in\mathbb{R}^+}$ of our general stochastic volatility model and let $A^{(1)}, A^{(2)}, \ldots, A^{(K)} \in M_d(\mathbb{R})$ be matrices with all eigenvalues in $(-\infty, 0)+i\mathbb{R}$. Then we define K independent stationary positive semi-definite OU type processes by $\Sigma_t^{(i)} = \int_{-\infty}^t e^{A^{(i)}(t-s)} dL_s^{(i)} e^{A^{(i)T}(t-s)}$ with $i = 1, 2, \ldots, K$ and the stationary stochastic volatility process $\Sigma_t = \sum_{i=1}^K \Sigma_t^{(i)}$. Due to the independence it is clear that the expected value, variance, autocovariance function and integrated volatility of Σ_t are simply the sum over the respective quantities of the individual processes $\Sigma_t^{(i)}$. Thus closed form formulae for these quantities follow immediately from the results of Section 5.3.

Moreover, also the results on the second order structure of the increments of the integrated volatility generalize.

Proposition 5.4.20. *Define* $r^{(i)+}(t) = \int_0^t \mathrm{acov}_{\Sigma^{(i)}}(u)du$, $r^{(i)++}(t) = \int_0^t r^{(i)+}(u)du$, $\mathbf{A}^{(i)} : M_d(\mathbb{R}) \to M_d(\mathbb{R}), X \mapsto A^{(i)}X + XA^{(i)T}$, $\mathscr{A}^{(i)} = (A^{(i)} \otimes I_d) + (I_d \otimes A^{(i)})$ *and* $\mathcal{A}^{(i)} : M_{d^2}(\mathbb{R}) \to M_{d^2}(\mathbb{R}), X \mapsto \mathscr{A}^{(i)}X + X\mathscr{A}^{(i)T}$ *for* $i = 1, 2, \ldots, K$. *Then we have for the stochastic volatility model with a superposition of positive semi-definite OU processes as volatility process:*

$$
\begin{aligned}
E(\Sigma_0) &= -\sum_{i=1}^K (\mathbf{A}^{(i)})^{-1} E(L_1^{(i)}) \\
\mathrm{var}(\mathrm{vec}(\Sigma_0)) &= -\sum_{i=1}^K (\mathcal{A}^{(i)})^{-1} \mathrm{var}(\mathrm{vec}(L_1^{(i)})) \\
r^{++}(t) &= \sum_{i=1}^K \left((\mathscr{A}^{(i)})^{-2} \left(e^{\mathscr{A}^{(i)}t} - I_{d^2} \right) - (\mathscr{A}^{(i)})^{-1} t \right) \mathrm{var}(\mathrm{vec}(\Sigma_0^{(i)})) \\
&= -\sum_{i=1}^K \left((\mathscr{A}^{(i)})^{-2} \left(e^{\mathscr{A}^{(i)}t} - I_{d^2} \right) - (\mathscr{A}^{(i)})^{-1} t \right) (\mathcal{A}^{(i)})^{-1} \mathrm{var}(\mathrm{vec}(L_1^{(i)})) \\
\mathrm{acov}_{\mathbf{\Sigma}}(h) &= \sum_{i=1}^K e^{\mathscr{A}^{(i)}\Delta(h-1)} (\mathscr{A}^{(i)})^{-2} \left(I_{d^2} - e^{\mathscr{A}^{(i)}\Delta} \right)^2 \mathrm{var}(\mathrm{vec}(\Sigma_0^{(i)})) \\
&= -\sum_{i=1}^K e^{\mathscr{A}^{(i)}\Delta(h-1)} (\mathscr{A}^{(i)})^{-2} \left(I_{d^2} - e^{\mathscr{A}^{(i)}\Delta} \right)^2 (\mathcal{A}^{(i)})^{-1} \mathrm{var}(\mathrm{vec}(L_1^{(i)})),\ h \in \mathbb{N}.
\end{aligned}
$$

Thus we obtain very explicit formulae whenever they are also available in the simple multivariate OU type stochastic volatility model. However, the integrated volatilities $\mathbf{\Sigma}$ and the logarithmic price increments $\mathbf{Y}\mathbf{Y}^T$ are no longer multivariate ARMA(1,1) processes.

But it is easy to see that the state space representations presented in Sections 5.4.2 and 5.4.3 generalize to the case of superpositions. We illustrate this for the realised quadratic variation. We have

$$\mathbf{Y}_n = \Delta\mu + u_{1,n}^{(\mathcal{M})} \tag{5.4.80}$$

$$[\mathbf{Y}]_n^{(M)} = \frac{\Delta^2}{M}\mu\mu^* + \mathbf{\Sigma}_n + u_{2,n}^{(\mathcal{M})} \tag{5.4.81}$$

$$= \frac{\Delta^2}{M}\mu\mu^* + \sum_{i=1}^{K}\mathbf{\Sigma}_n^{(i)} + u_{2,n}^{(\mathcal{M})} \tag{5.4.82}$$

This immediately gives rise to a state space representation of

$$\left(\mathbf{Y}_n, [\mathbf{Y}]_n^{(M)}\right), \left(\mathbf{A}^{(1)}\mathbf{\Sigma}_n^{(1)}, \Sigma_{n\Delta}^{(1)}\right), \left(\mathbf{A}^{(2)}\mathbf{\Sigma}_n^{(2)}, \Sigma_{n\Delta}^{(2)}\right), \ldots, , \left(\mathbf{A}^{(K)}\mathbf{\Sigma}_n^{(K)}, \Sigma_{n\Delta}^{(K)}\right)$$

using equations (5.4.50) and (5.4.51) to obtain independent recursions for $(\mathbf{A}^{(i)}\mathbf{\Sigma}_n^{(i)}, \Sigma_{n\Delta}^{(i)})$.

As in the univariate case (cf. Barndorff-Nielsen (2001b), Barndorff-Nielsen and Shephard (2001b)), one can also model long-range dependence by superimposing infinitely (but countably) many appropriate positive semi-definite OU type processes. This follows from a straightforward generalization of the arguments in Barndorff-Nielsen (1998b, Section 4).

5.5. Empirical illustration

In this section we provide a small empirical application of the multivariate OU type stochastic volatility model using two bivariate data sets in order to illustrate the features of the model developed further.

5.5.1. Data

Our empirical illustration is based on daily prices of four major US stocks, viz. Applied Materials Inc. (AMAT), Amgen Inc. (AMGN), Pfizer Inc. (PFE) and Wal-Mart Stores Inc. (WMT). The sample covers the period from January 2nd, 1985 to December 29th, 2006. Note that splits and dividends are incorporated into the prices. Moreover, the continuously compounded returns are mean-adjusted. Figure 5.1 presents the time evolvement of the corresponding return series exhibiting the usual empirical characteristics. Moreover, apparently the stocks tend to move similarly.

5.5.2. Estimation methods

The estimation of continuous-time stochastic volatility models is complicated by the unavailability of the likelihood function. However, based on the theoretical results derived in the foregoing Section 5.4, the multivariate OU type stochastic volatility model can be estimated either by using the second order dependence structure of the squared returns or by exploiting its state-space representation.

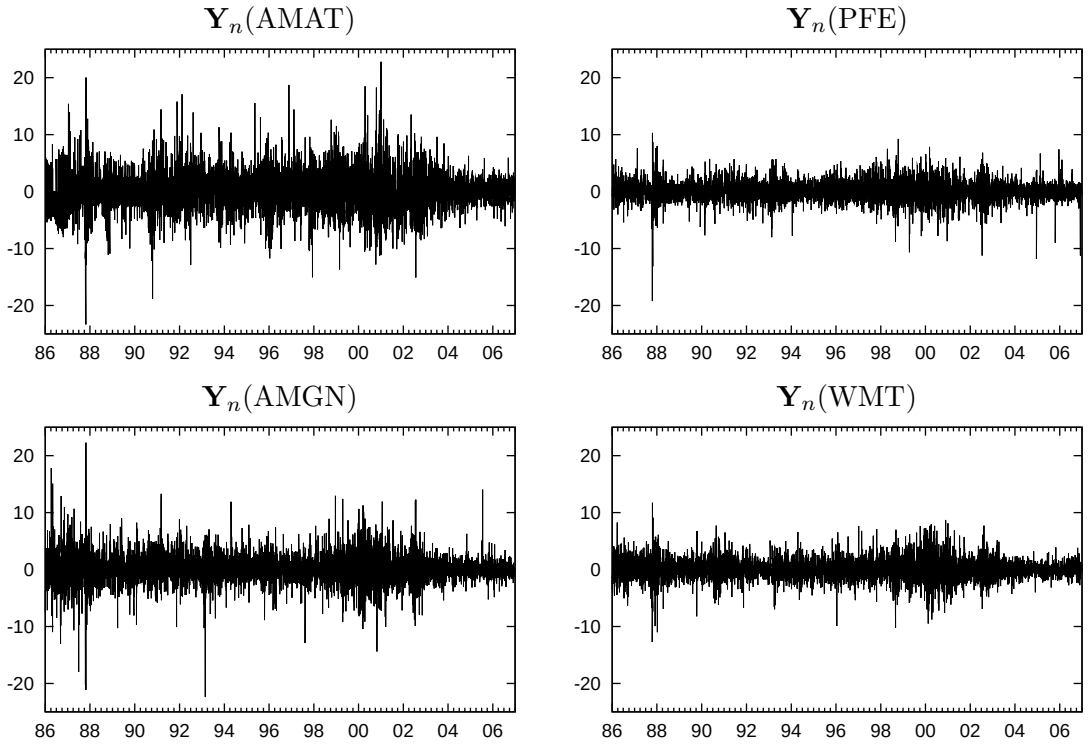


Figure 5.1.: **Daily returns.** Time series plots of the daily returns (January 2nd, 1985 until December 29th, 2006) of the four stocks under consideration.

5.5.2.1. Estimation via the second order dependence structure

Based on the ergodicity of the discretely sampled returns in our model a simple estimator can be obtained by matching a set of empirical moments with their model implied counterparts given in Theorem 5.4.4. Using the identification results natural candidates are $\text{vech}(E\left(\mathbf{Y}\mathbf{Y}^T\right)$, $\text{acov}_{\text{vech}(\mathbf{Y}\mathbf{Y}^T)}(1)$ and $\text{acov}_{\text{vech}(\mathbf{Y}\mathbf{Y}^T)}(2)$. However to gain efficiency we include lags of the autocovariance function of different order such that our objective function is finally given by

$$SSR(\mathcal{L}) = \left\| m - E\left(\text{vech}(\mathbf{Y}\mathbf{Y}^T)\right)\right\|_F + \sum_{l\in\mathcal{L}} \left\| a_l - \text{acov}_{\text{vech}(\mathbf{Y}\mathbf{Y}^T)}(l)\right\|_F \tag{5.5.1}$$

with $\mathcal{L}$ denoting the selected lags of the autocovariance function and m and a_l for $l \in \mathcal{L}$ the empirical mean and autocovariance function of $\text{vech}(\mathbf{Y}\mathbf{Y}^T)$, respectively. Although several matrix norms can be considered we use here the Frobenius norm, i.e.

$$\|C\|_F = \sum_{i=1}^{d}\sum_{j=1}^{d} c_{ij}^2 = \text{tr}(CC^T),$$

such that the objective function (5.5.1) leads to a non-linear least squares problem.

As noted in Barndorff-Nielsen and Shephard (2001b, Section 5.3), who apply this procedure for the estimation of a univariate OU type stochastic volatility model, the estimator

is independent of the assumption of a particular OU type process. More precisely, based on the results discussed in Section 5.4.4, we know that this estimator generally identifies the mean and the variance of the Lévy process L driving the OU type process (referred to as the background-driving Lévy process or BDLP, for short, in the following), as well as the matrix A. So, rather than assuming a specific parametric model for the BDLP, we optimize over the first and second moments of L_1. Often the parameters of a specific BDLP can be identified solely from the mean and the variance of the BDLP. In this case, the autocovariance fit also identifies these parameters.

Although this estimator is computationally very fast it lacks from the perspective of optimal weighting, especially the autocovariance terms have usually a higher variance as the empirical mean of the time series, which is not incorporated in the estimation by minimizing the objective function (5.5.1). We account for these effects by applying the GMM estimation proposed by Hansen (1982) using the previous estimates based on (5.5.1) as starting values. For our weighting matrix we use a HAC estimator of the long run covariance matrix with Parzen kernel and a lag of 25, which is continuously updated.

Although a large number of moment conditions would improve the asymptotic efficiency of the estimator it lacks from the fact that the weighting matrix is estimated with less precision. Given this trade-off in a finite sample set-up our moment conditions are given by the mean and the autocovariances of the time series $\text{vech}(\mathbf{Y}\mathbf{Y}^T)$ at lags $(1, 3, 5, 7, 9, 15, 20, 30, 40, 60, 80, 100)$.

5.5.2.2. Estimation via the state-space representation

Based on the state-space representation for the squared returns (see Section 5.4.2) the Kalman recursions can be used to obtain the quasi-likelihood function of the model. This estimation approach has also been considered in Barndorff-Nielsen and Shephard (2001b) for the univariate OU type stochastic volatility model. In that paper they also show that the Kalman filter is suboptimal but provides consistent and asymptotically normal estimators. The state-space representation also illustrates once more that the multivariate model can again be estimated without the need of a prespecified parametric BDLP. So, estimates of the mean and variance of the BDLP as well as of the elements of the matrix A can straightforwardly be obtained by the Kalman filter. Moreover, in contrast to the above GMM approach, the latent volatility process can be inferred.

5.5.3. Estimation results

In the following we estimate the multivariate OU type stochastic volatility model as given in equations (5.4.1) and (5.4.2), where we follow Barndorff-Nielsen and Shephard (2001b) by assuming that $\mu = \beta = 0$, i.e. the means of the returns are set to zero. Recall in this context that we use mean-adjusted returns and note that for $\mu \neq 0$ and $\beta = 0$ one can still use our state space representation of Section 5.4.2. Moreover, we assume that the off-diagonal elements of the matrix subordinator are zero, i.e. we restrict the BDLP to be a diagonal matrix subordinator as introduced in Section 5.4.5. The vector of the diagonal elements $(L_1, L_2, \ldots, L_d)^T$ of L is to be denoted by $\text{diag}(L)$. The d elements of $\text{diag}(L)$ may not exhibit negative correlations due to Proposition 5.4.18 which is incorporated as a restriction into our estimation procedure. Hence, in the model estimated the correlation between the variances of the different assets is determined by both the correlation structure

of diag(L) as well as by the entries of A. Although this specification seems rather restrictive it turns out that even this model can quite adequately describe the joint dynamics.

In the following we mainly present the estimation results obtained by the GMM method. From the results obtained using the Kalman filter we only show the estimates of the latent stochastic volatility process.

Before presenting the estimation results of the multivariate OU type stochastic volatility model, we first consider its univariate counterpart. This allows for a more detailed analysis of the multivariate model. We therefore start with a discussion of the univariate estimation results.

The upper two panels in Table 5.1 present the univariate estimation results for the four assets. The first panel gives the parameter estimates of the univariate models with one OU type process as the stochastic volatility process, whereas the middle panel shows the estimates of a univariate model with a superposition of two OU type processes where the parameters of each individual OU type process are given in a separate line and where w_i is the weight of the individual OU type processes in the superposition. As we can see for the models with a single OU type process the magnitudes are comparable to the ones obtained in the existing literature, see e.g. Barndorff-Nielsen and Shephard (2001b, 2002).

Figure 5.2 depicts the empirical and estimated autocovariance functions. In particular, the upper four panels show the estimated model-implied autocovariance functions of the squared daily returns along with the empirical one (given by dots) for AMAT, AMGN, PFE and WMT, respectively. In particular, the solid line refers to the daily autocovariance function implied by the estimated univariate model and the dotted lines depict the autocovariance function implied by the superposition model with two processes superimposed. Obviously the superposition model provides a better fit than the simple model. More precisely, in contrast to the simple models its autocovariance function decreases faster for short lags (up to the 10th or 20th lag, depending on the asset) and slower for longer lags, because it is the weighted sum of two exponentials decreasing at different rates. These results are in line with the findings of Barndorff-Nielsen and Shephard (2001b, 2002), who encourage the use of superposition models in the univariate case.

We now turn to the estimation of the bivariate OU type stochastic volatility model for the two systems (AMAT, AMGN) and (PFE, WMT) where we ensure identifiability using Lemma 5.4.13 noting that $\Delta = 1$ equals one day. Table 5.1 presents the bivariate GMM estimation results in its last panel. Obviously, the parameter estimates are somewhat similar (in magnitude) to those of the univariate models. Note in this connection that the eigenvalues of A are $-0.0148 \pm 0.1036i$ for (AMAT, AMGN) and -0.0482, -0.0107 for (PFE,WMT).

The estimated autocovariance functions of the daily squared returns are also depicted in the upper four panels of Figure 5.2 (dashed lines). The bottom panels show the empirical and estimated autocovariance function of the cross product of the returns of the two stocks used in the respective bivariate models. Two effects of our multivariate model stand out. The results for the first system (AMAT and AMGN), see left panels, illustrate that our model does not necessarily impose monotonically decreasing autocovariance functions. This behaviour is mainly due to the exponentially damped sinusoidal behaviour of the matrix exponential induced by eigenvalues with non-vanishing complex part. The results for the second system (PFE and WMT), see right panels, nicely illustrate the property derived in Section 5.3.3 that the multivariate model is able to mimic the behaviour of univariate

Table 5.1.: Estimation Results

	asset	w_i	A	$E(\mathrm{diag}(L_1))$	$\mathrm{var}(\mathrm{diag}(L_1))$
univariate	AMAT		-0.0109	0.3301	2.0453
	AMGN		-0.0118	0.2843	1.8324
	PFE		-0.0158	0.1658	0.7718
	WMT		-0.0066	0.0564	0.4999
superposition	AMAT	0.9222	-0.0837	0.9101	5.9075
		0.0773	-0.0060	0.4685	4.2608
	AMGN	0.8324	-0.0719	0.5452	6.2873
		0.1676	-0.0008	0.0243	0.6591
	PFE	0.6231	-0.0998	0.4499	2.9555
		0.3768	-0.0061	0.0717	0.5223
	WMT	0.9961	-0.0059	0.0265	0.4561
		0.0039	-0.1806	16.2901	29.2314
bivariate	AMAT & AMGN		$\begin{pmatrix} -0.0527 & 0.1320 \\ -0.0921 & 0.0230 \end{pmatrix}$	$\begin{pmatrix} 0.7838 \\ 0.0002 \end{pmatrix}$	$\begin{pmatrix} 2.5501 & 0.7026 \\ 2.3626 & 4.4340 \end{pmatrix}$
	PFE & WMT		$\begin{pmatrix} -0.0396 & 0.0183 \\ 0.0135 & -0.0193 \end{pmatrix}$	$\begin{pmatrix} 0.2864 \\ 0.0719 \end{pmatrix}$	$\begin{pmatrix} 1.9126 & 0.7241 \\ 0.8522 & 0.7240 \end{pmatrix}$

In the upper two panels the estimates for a univariate model with one (univariate) and a superposition of two Ornstein-Uhlenbeck type processes as the stochastic volatility process are shown. The lower panel shows the estimation results for the bivariate data sets.

The first column states the company of the stock(s) considered, the second the weight of the OU type processes superimposed (only relevant for the superposition of univariate OU type processes), the third the estimate of the matrix A and the third and fourth the expectation and variance of the driving Lévy process L at time one. In the bivariate case the lower left element of $\mathrm{var}(\mathrm{diag}(L_1))$ gives the covariance and the upper right the corresponding correlation.

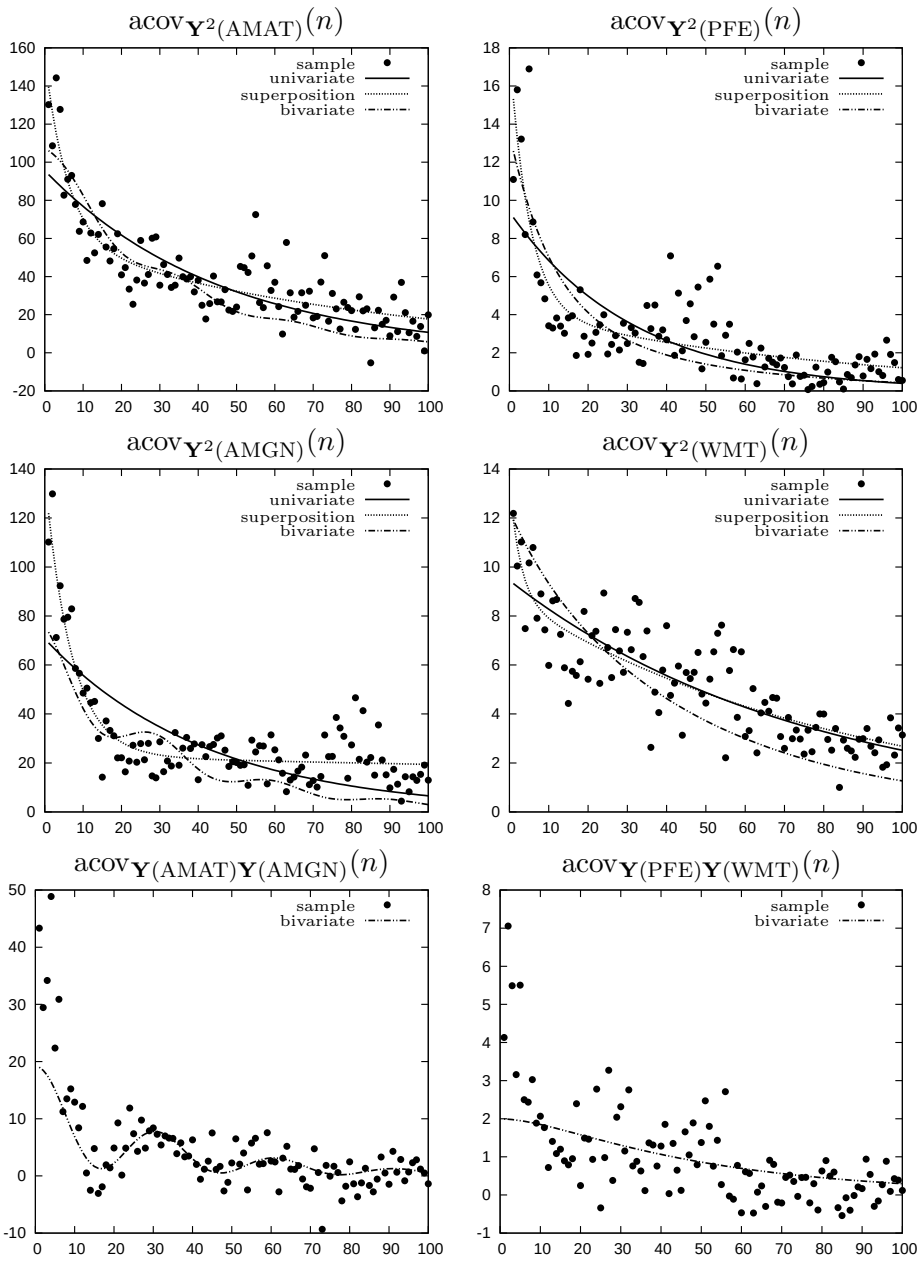


Figure 5.2.: **Autocovariances.** Estimated and empirical (black dots) daily autocovariance functions of the squared returns (upper four panels) and of the cross products of the returns of the two bivariate systems (bottom panels). The left panels show the result of the first system (AMAT, AMGN) and the right panels those of the second system (PFE, WMT). The autocovariance functions based on the different models are characterized by different line styles: the continuous line refers to the fit of the univariate model, the dotted line corresponds to the univariate superposition model with two OU type processes, and the dashed line corresponds to the autocovariances based on the bivariate system.

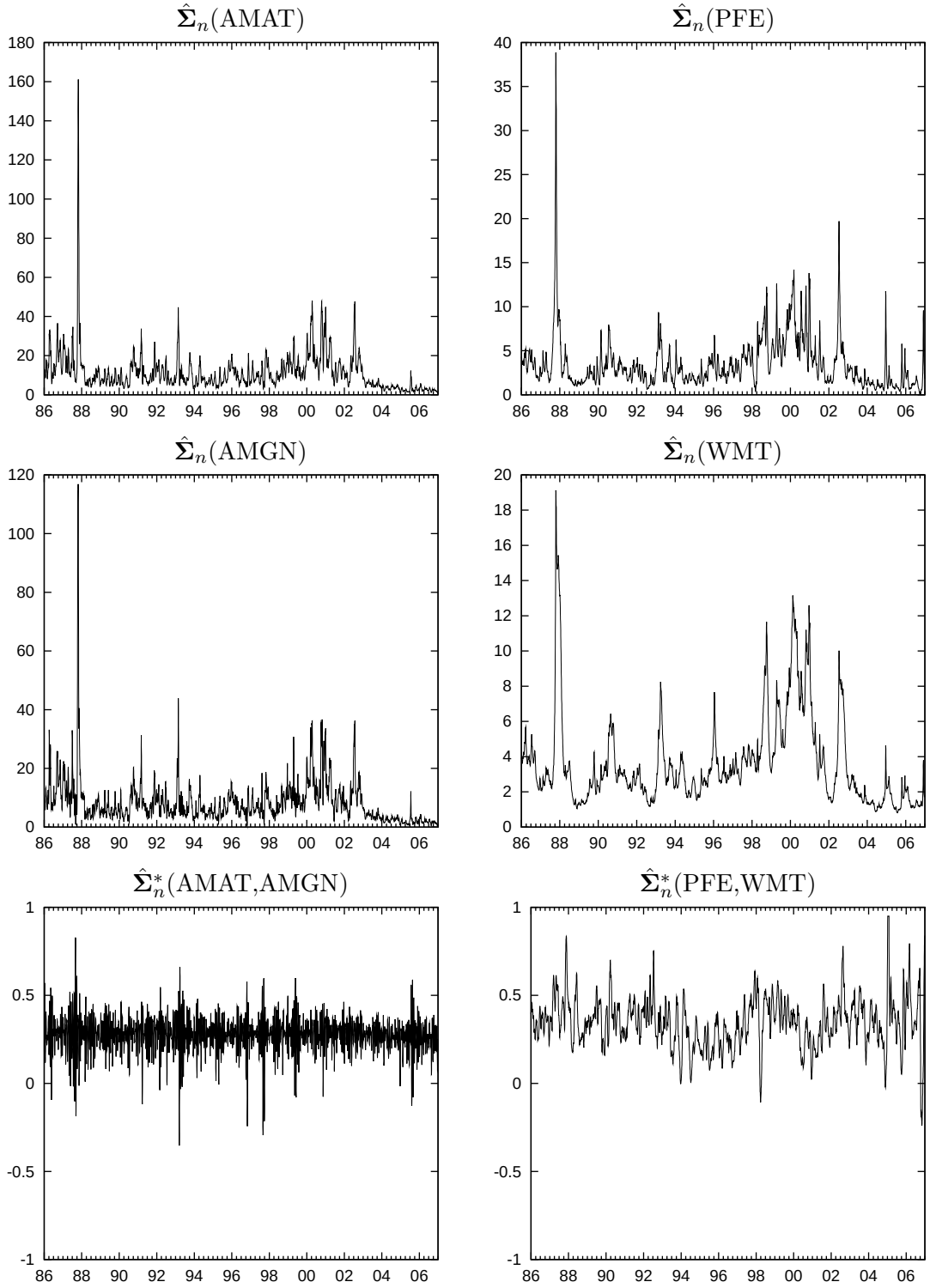


Figure 5.3.: **Filtered integrated variances.** Note that $\hat{\mathbf{\Sigma}}^*_n$(A,B) denotes the filtered correlation instead of the filtered covariance between asset A and B.

superpositions. Especially the autocovariance function of PFE in the bivariate model is remarkably close to the autocovariance function of the univariate superposition model.

Note that these properties of our model already emerge empirically within our rather simple specification of the matrix subordinator, viz. a diagonal subordinator. But we conjecture that the flexibility of our model can even be further improved by considering more sophisticated specifications of the background-driving Lévy process or by allowing for superpositions.

As the knowledge of the latent stochastic volatility process $(\mathbf{\Sigma}_n)_{n\in\mathbb{N}}$ is important for many financial applications, we also present here the model-implied volatility states $(\hat{\mathbf{\Sigma}}_n)_{n\in\mathbb{N}}$ that can straightforwardly be obtained using the Kalman filter. Figure 5.3 depicts the filtered volatility and correlation states of the two bivariate models. Obviously, in both systems the volatilities tend to move together. Moreover, the correlations, especially the one between AMAT and AMGN, show a relatively constant mean. Note that the volatilities and correlation of AMAT and AMGN are themselves quite volatile.

5.6. Conclusion

Given the relevance of a joint modelling of the dynamics of multiple assets for portfolio and risk management decisions, we have generalized the non-Gaussian OU type stochastic volatility model proposed by Barndorff-Nielsen and Shephard (2001b) to the multivariate case. It turns out that our model possesses many attractive features which are mainly a result of our stochastic volatility specification.

Specifying the stochastic volatility by Lévy-driven positive semi-definite OU type processes provides a flexible dependence structure for the volatility. In particular, we show that the increments of the integrated covariance and the outer product of the returns ("squared returns") of a stochastic volatility model based on a single positive semi-definite OU type process follow ARMA(1,1) processes. Furthermore, closed form expressions are given for the first and second moments of these variables. These results facilitate the implementation of financial decisions, such as the choice of e.g. a minimum-variance portfolio or other types of risk assessment, and the estimation of our model. Moreover, we derived a state-space representation for the joint process of the returns and the outer product of the returns, which provides an additional approach for the estimation of the model as well as for the estimation and forecasting of the volatility states using the Kalman recursions.

Since our model is defined in terms of a matrix subordinator its particular specification may depend on the application at hand. In the empirical part of this chapter we focused on models with a simple diagonal matrix subordinator, which already exhibit some nice properties, see Section 5.5.3. However, studying the empirical relevance of alternative classes of matrix subordinators deserves more attention in future research.

Further improvements in the estimation of the model may be obtained by incorporating the high-frequency based and thus more informative realized covariation measure using the results of Section 5.4.3. We are currently working on this extension.

6. Multivariate Continuous Time Lévy-driven GARCH Processes

6.1. Introduction

Since the introduction of the ARCH process in Engle (1982) and the extension to GARCH by Bollerslev (1986), these processes have been widely used to successfully model discretely observed time series. In particular, they have often been used in financial econometrics to model the returns of stocks, currencies and other assets, since they are capable of reproducing several of the "stylized facts" of financial data, viz. stochastic volatility (=variance), volatility clusters, tail-heaviness and dependence without autocorrelation. In order to be able to treat irregularly spaced data and to use continuous time financial theory, already Nelson (1990) started to study diffusions which appear as the limit of rescaled GARCH processes. Thereafter, such diffusion limits have been further studied by various authors (see Corradi (2000) and references therein). However, many of the desirable features of GARCH are lost in these limits. Thus, Klüppelberg et al. (2004) introduced a continuous time analogue of GARCH processes – the COGARCH(1,1) process – by using a Lévy process and generalizing the GARCH recursions. This process inherits many of the appealing features from the GARCH process.

As can be seen from Klüppelberg et al. (2004, Equation 3.8), the COGARCH(1,1) process can be defined as the solution of

$$dG_t = \sqrt{v_{t-}}dL_t \tag{6.1.1}$$

$$dv_t = -\beta(v_{t-} - c)dt + \alpha v_{t-}d[L,L]^{\mathfrak{d}}_t \tag{6.1.2}$$

using the discrete quadratic variation $[L,L]^{\mathfrak{d}}$ of a univariate Lévy process L, parameters $\alpha, \beta, c > 0$ and initial values $G_0 = 0, v_0 > 0$. The process G is referred to as the COGARCH(1,1) process and v as the COGARCH(1,1) volatility process. To be in line with what follows we have defined v to be càdlàg instead of càglàd and changed the parametrization compared to Klüppelberg et al. (2004). In a financial context G is understood to resemble a log-price and v its stochastic volatility.

By now the COGARCH process has been investigated in several papers. Klüppelberg, Lindner and Maller (2006) studied the COGARCH(1,1) process further in comparison to the popular stochastic volatility model of Barndorff-Nielsen and Shephard (2001b) and Fasen et al. (2006) presented a detailed analysis of the extremal behaviour. Moreover, Brockwell et al. (2006) introduced univariate COGARCH processes of arbitrary orders (p,q) with $q > p$ by using insight from the theory of CARMA processes. Different estimation procedures for the COGARCH(1,1) process were discussed and successfully applied to data in Haug, Klüppelberg, Lindner and Zapp (2007), Müller (2006) and Maller et al. (2006). The latter paper also constructed sequences of discrete time GARCH(1,1) processes on a finite time interval converging to the COGARCH(1,1) process in the Skorokhod metric in

probability. A similar but weaker result is to be found in Kallsen and Vesenmayer (2006) who also obtained the infinitesimal generator of (G, v). Furthermore, the volatility process in the COGARCH(1,1) model is a particular example of a generalized Ornstein-Uhlenbeck process (cf. Lindner and Maller (2005) and references therein). The asymptotics of the sample autocovariance functions of such processes are discussed in Fasen (2007) with a special emphasis on the COGARCH case.

In this paper we introduce a multivariate version of the COGARCH(1,1) process, to be dubbed "MUCOGARCH(1,1)", and analyse its properties (Markov properties, existence of moments, stationarity and second order structure) in detail. In analogy to Brockwell et al. (2006) we use a d-dimensional Lévy process and define the volatility process using a "self-exciting" autoregressive structure of order one and the discrete quadratic variation of the Lévy process. Clearly multivariate models are needed in order to model and understand the joint behaviour of several time series (e.g. the prices of assets) which exhibit non-trivial interdependencies. But in a financial context they are also necessary to obtain a reasonable basis for portfolio optimization or the pricing of multi-asset options. So far only few multivariate stochastic volatility models have been discussed in the literature. Lindberg (2005) and Hubalek and Nicolato (2005) introduced different factor models where the volatility is driven by factors which are univariate positive Ornstein-Uhlenbeck (OU) type processes as used in the works of Barndorff-Nielsen and Shephard. In Chapter 4 we introduced positive semi-definite OU type processes which were used in Chapter 5 to define a multivariate extension of the popular model of Barndorff-Nielsen and Shephard (2001b).

In the discrete time GARCH world the general multivariate model (known as vec-model) seems to have been first stated in Bollerslev, Engle and Wooldridge (1988). The first comprehensive study of multivariate GARCH models is Engle and Kroner (1995) where the so-called BEKK model was introduced which provides an easy parametrization ensuring positive semi-definiteness of the variance process. Until now various variants of multivariate GARCH processes have been proposed, analysed and fitted to data. For a comprehensive review paper we refer to Bauwens, Laurent and Rombouts (2006).

Generally the analysis of multivariate GARCH models is considerably more involved than the univariate case due to the non-linearity of the variance equation and the "noise" being concentrated on the rank one-matrices, for example, explicit results for the moments are only obtainable under certain restrictions. Similar problems are appearing in our analysis of the MUCOGARCH processes which will be defined via a stochastic differential equation (SDE) which is only defined on the positive semi-definite matrices and has only locally Lipschitz coefficients. However, we are able to show that the MUCOGARCH process inherits many nice features from the univariate COGARCH processes. Yet, to do this we often need to establish technical results first and the formulae obtained are more complicated.

The remainder of this paper is structured as follows. First we summarize important notation to be used throughout this paper below.

In Section 6.2 we review discrete time multivariate GARCH processes, multivariate Lévy processes and positive semi-definite processes of Ornstein-Uhlenbeck processes, because they are intrinsically related to the MUCOGARCH process to be defined.

Thereafter we introduce and study the MUCOGARCH(1,1) process in Section 6.3 starting with the definition and establishing well-definedness in Part 6.3.1. In the next part of this section we turn to presenting a univariate COGARCH(1,1) process that bounds the volatility process in a norm intrinsically related to the autoregressive parameter and use this

bound to give sufficient conditions for the finiteness of moments. This is followed by showing that the volatility process alone and the MUCOGARCH process together with its volatility are strong Markov processes. Moreover, we establish conditions for the existence of a stationary distribution of the volatility in Subsection 6.3.3. In the last part we calculate the second order structure of the volatility process explicitly under certain assumptions on the moments of the driving Lévy process and establish (asymptotic) second order stationarity.

In the following brief Section 6.4 we turn our focus on the increments of the MUCOGARCH process itself showing that it has stationary increments provided the volatility is stationary. Thereafter we calculate the second order moment structure of the increments (i.e. the returns in a financial context) observed on a regularly spaced discrete grid. Here we obtain in particular that the returns have zero autocorrelation but their "squares" have non-vanishing autocorrelation.

In Section 6.5 we study several illustrative examples and show simulations of them.

Finally, we present in the last section an idea how to define MUCOGARCH processes of general orders higher than $(1,1)$. As it is based on multivariate CARMA processes, this is preceded by a brief review of those processes.

In the appendix of this chapter we present results on stochastic differential equations which are only defined on an open subset of $\mathbb{R}^d$ and which have only locally Lipschitz coefficients. After establishing existence and uniqueness of solutions we turn to their Markovian properties where we also recall the necessary notions from the theory of Markov processes needed. Additionally, several other auxiliary results needed are to be found in the appendix.

6.1.1. Notation

We denote the set of real $m \times n$ matrices by $M_{m,n}(\mathbb{R})$. If $m = n$, we simply write $M_n(\mathbb{R})$ and denote the group of invertible $n \times n$ matrices by $GL_n(\mathbb{R})$, the linear subspace of symmetric matrices by $\mathbb{S}_n$, the (closed) positive semi-definite cone by $\mathbb{S}_n^+$ and the open positive definite cone by $\mathbb{S}_n^{++}$. I_n stands for the $n \times n$ identity matrix. The natural ordering on the symmetric $n \times n$ matrices shall be denoted by $\leq$, i.e. for $A, B \in M_n(\mathbb{R})$ we have that $A \leq B$ if and only if $B - A \in \mathbb{S}_n^+$. The tensor (Kronecker) product of two matrices A, B is written as $A \otimes B$. vec denotes the well-known vectorisation operator that maps the $n \times n$ matrices to $\mathbb{R}^{n^2}$ by stacking the columns of the matrices below another. For more information regarding the tensor product and vec operator we refer to Horn and Johnson (1991, Chapter 4). Likewise $\text{vech} : \mathbb{S}_d \to \mathbb{R}^{d(d+1)/2}$ denotes the "vector-half" operator that stacks the columns of the lower triangular part of a symmetric matrix including the diagonal below another. The spectrum of a matrix is denoted by $\sigma(\cdot)$ and the spectral radius by $\rho(\cdot)$. Finally, A^* is the transpose (adjoint) of a matrix $A \in M_{m,n}(\mathbb{R})$.

For a matrix A we denote by A_{ij} the element in the i-th row and j-th column and this notation is extended to processes in a natural way.

For some set B the indicator function is denoted by $I_B(\cdot)$ and operations on sets are defined element-wise as usual, for instance, if $A, B \subseteq \mathbb{C}$ we have $A + B = \{a + b : a \in A,\, b \in B\}$ and $\Re(A) = \{\Re(a) : a \in A\}$ where $\Re$ denotes the real part of a complex number.

Norms of vectors or matrices are denoted by $\|\cdot\|$. If the norm is not specified, then it is irrelevant which particular norm is used.

Throughout we assume that all random variables and processes are defined on a given filtered probability space $(\Omega, P, \mathcal{F}, (\mathcal{F}_t)_{t\in\mathcal{T}})$ with $\mathcal{T} = \mathbb{N}$ in the discrete time case and

$\mathcal{T} = \mathbb{R}^+$ in the continuous one. Moreover, in the continuous setting we assume the usual conditions to hold.

Furthermore, we employ an intuitive notation with respect to the (stochastic) integration with matrix-valued integrators referring to any of the standard texts (e.g. Protter (2004)) for a comprehensive treatment of the theory of stochastic integration. Let $(A_t)_{t\in\mathbb{R}^+}$ in $M_{m,n}(\mathbb{R})$, $(B_t)_{t\in\mathbb{R}^+}$ in $M_{r,s}(\mathbb{R})$ be càdlàg and adapted processes and $(L_t)_{t\in\mathbb{R}^+}$ in $M_{n,r}(\mathbb{R})$ be a semi-martingale. Then we denote by $\int_0^t A_{s-}dL_sB_{s-}$ the matrix C_t in $M_{m,s}(\mathbb{R})$ which has ij-th element $C_{ij,t} = \sum_{k=1}^n \sum_{l=1}^r \int_0^t A_{ik,s-}B_{lj,s-}dL_{kl,s}$. Equivalently such an integral can be understood in the sense of Métivier and Pellaumail (1980b), resp. Métivier (1982), by identifying it with the integral $\int_0^t \mathbf{A}_{s-}dL_s$ with $\mathbf{A}_t$ being for each fixed t the linear operator $M_{n,r}(\mathbb{R}) \to M_{m,s}(\mathbb{R})$, $X \mapsto A_tXB_t$. Moreover, we always denote by $\int_a^b$ with $a \in \mathbb{R}\cup\{-\infty\}, b \in \mathbb{R}$ the integral over the half-open interval $(a, b]$ for notational convenience. If $b = \infty$ the integral is understood to be over (a, b). If $(X_t)_{t\in\mathbb{R}^+}$ is a semi-martingale in $\mathbb{R}^m$ and $(Y_t)_{t\in\mathbb{R}^+}$ one in $\mathbb{R}^n$ then the quadratic variation $([X, Y]_t)_{t\in\mathbb{R}^+}$ is defined as the finite variation process in $M_{m,n}(\mathbb{R})$ with components $[X, Y]_{ij,t} = [X_i, Y_j]_t$ for $t \in \mathbb{R}^+$ and $i = 1, \ldots, m$, $j = 1, \ldots, n$.

6.2. A review of related processes

In this section we briefly review some processes which are related to the multivariate continuous time GARCH (MUCOGARCH) process to be introduced in this chapter. The insights gained from these processes will be of use for the definition of the MUCOGARCH processes in particular.

6.2.1. Discrete time multivariate GARCH

The well-known one dimensional GARCH(p, q) model with $p, q \in \mathbb{N}$ is defined via an i.i.d. sequence $(\epsilon_n)_{n\in\mathbb{Z}}$ and the equations

$$X_n = \sqrt{v_n}\epsilon_n \tag{6.2.1}$$

$$v_n = \alpha_0 + \sum_{i=1}^p \alpha_i X_{n-i}^2 + \sum_{j=1}^q \beta_j v_{n-j} \tag{6.2.2}$$

where the parameters satisfy $\alpha_0, \ldots, \alpha_p, \beta_1, \ldots, \beta_q \geq 0$.

When one moves from a scalar X to a multidimensional X, the variance process v becomes a covariance matrix process V. The most general d-dimensional discrete time GARCH(p, q) model based on an i.i.d. sequence $(\epsilon_n)_{n\in\mathbb{Z}}$ in $\mathbb{R}^d$ thus is the following:

$$X_n = V_n^{1/2}\epsilon_n \tag{6.2.3}$$

$$V_n = C + \sum_{i=1}^p A_i X_{n-i}X_{n-i}^* + \sum_{j=1}^q B_j V_{n-j}. \tag{6.2.4}$$

Here $V_n^{1/2}$ denotes the unique square root defined by functional calculus. Moreover, A_i and B_j are linear operators that map the positive semi-definite $d \times d$ matrices into themselves and $C \in \mathbb{S}_d^+$.

This model is implicit in most multivariate GARCH models studied so far. In particular, it is obviously equivalent to the vec-model (Engle and Kroner (1995)) which is given by:

$$X_n = V_n^{1/2}\epsilon_n \tag{6.2.5}$$

$$\mathrm{vec}(V_n) = \mathrm{vec}(C) + \sum_{i=1}^{p} \tilde{A}_i \mathrm{vec}(X_{n-i}X_{n-i}^*) + \sum_{j=1}^{q} \tilde{B}_j \mathrm{vec}(V_{n-j}). \tag{6.2.6}$$

$\tilde{A}_i$ and $\tilde{B}_j$ are now $d^2 \times d^2$ matrices mapping the vectorized positive semi-definite matrices into themselves.

Establishing conditions for the existence of stationary multivariate GARCH(p,q) processes is considerably more involved than in the univariate case (see e.g. Bougerol and Picard (1992)), since one cannot utilize the theory of linear random recurrence equations. We refrain from giving details and refer the interested reader to Boussama (1998 or 2006).

Moreover, the restrictions on the linear operators A_i and B_j necessary to ensure positive semi-definiteness led to the introduction of the so-called BEKK model (see again Engle and Kroner (1995)), which automatically ensures positive semi-definiteness:

$$X_n = V_n^{1/2}\epsilon_n \tag{6.2.7}$$

$$V_n = C + \sum_{i=1}^{p}\sum_{k=1}^{l_i} \bar{A}_{i,k} X_{n-i} X_{n-i}^* \bar{A}_{i,k}^* + \sum_{j=1}^{q}\sum_{r=1}^{s_j} \bar{B}_{j,r} V_{n-j} \bar{B}_{j,r}^*, \tag{6.2.8}$$

with $l_i, s_j \in \mathbb{N}_0$ for $i = 1,2,\ldots,p$, $j = 1,2,\ldots,q$ and where $\bar{A}_{i,k}, \bar{B}_{j,r}$ are now arbitrary elements of $M_d(\mathbb{R})$.

The BEKK model corresponds to the vec model with $\tilde{A}_i = \sum_{k=1}^{l_i} \bar{A}_{i,k} \otimes \bar{A}_{i,k}$ and $\tilde{B}_j = \sum_{r=1}^{s_j} \bar{B}_{j,r} \otimes \bar{B}_{j,r}$ with $\otimes$ denoting the tensor (Kronecker) product. Conversely, all vec-models with $\tilde{A}_i(\mathbb{S}_d^+) = \mathbb{S}_d^+$ and $\tilde{B}_j(\mathbb{S}_d^+) = \mathbb{S}_d^+$ for $i = 1,\ldots,p, j = 1,\ldots,q$ can be represented as BEKK models with $l_i = s_j = 1$ and in dimension two all vec-models have BEKK representations (see Appendix B for details). Thus, it is not too restrictive to look only at BEKK models with $l_i = s_j = 1$ as we shall do in the following.

6.2.2. Multivariate Lévy processes

Later on we shall use both Lévy processes in $\mathbb{R}^d$ and in the symmetric matrices $\mathbb{S}_d$. Thus we briefly recall the relevant basic facts on them now. For a comprehensive treatment of Lévy processes confer Applebaum (2004), Protter (2004) or Sato (1999), for instance.

We consider a Lévy process $L = (L_t)_{t\in\mathbb{R}^+}$ (where $L_0 = 0$ a.s.) in $\mathbb{R}^d$ determined by its characteristic function in the Lévy-Khintchine form $E\left[e^{i\langle u, L_t\rangle}\right] = \exp\{t\psi_L(u)\}$ for $t \in \mathbb{R}^+$ with

$$\psi_L(u) = i\langle \gamma_L, u\rangle - \frac{1}{2}\langle u, \tau_L u\rangle + \int_{\mathbb{R}^d} (e^{i\langle u,x\rangle} - 1 - i\langle u,x\rangle I_{[0,1]}(\{\|x\|\}))\, \nu_L(dx), \quad u \in \mathbb{R}^d, \tag{6.2.9}$$

where $\gamma_L \in \mathbb{R}^d$, $\tau_L \in \mathbb{S}_d^+$ and ν_L is a measure on $\mathbb{R}^d$ that satisfies $\nu_L(\{0\}) = 0$ and $\int_{\mathbb{R}^d}(\|x\|^2 \wedge 1)\, \nu_L(dx) < \infty$. The measure ν_L is referred to as the Lévy measure of L. Moreover, $\langle\cdot,\cdot\rangle$ denotes the usual Euclidean scalar product on $\mathbb{R}^d$ and $\|\cdot\|$ the associated norm.

We always assume L to be càdlàg. It is a well-known fact that to every càdlàg Lévy process L on $\mathbb{R}^d$ one can associate a random measure μ_L on $\mathbb{R}^+ \times \mathbb{R}^d \setminus \{0\}$ describing the jumps of L. For any measurable set $B \subset \mathbb{R}^+ \times \mathbb{R}^d \setminus \{0\}$,

$$\mu_L(B) = \sharp\{s \geq 0 : \ (s, L_s - L_{s-}) \in B\}.$$

The jump measure μ_L is an extended Poisson random measure (see Jacod and Shiryaev (2003, Section II.1) for a definition and the relevant integration theory) on $\mathbb{R}^+ \times \mathbb{R}^d \setminus \{0\}$ with intensity measure $E(\mu_L(ds, dx)) = ds\, \nu_L(dx)$. By the Lévy-Itô decomposition we can rewrite L almost surely as

$$L_t = \gamma_L t + B_t + \int\limits_{\|x\|>1, s\in[0,t]} x\, \mu_L(ds, dx) + \lim_{\varepsilon \downarrow 0} \int\limits_{\varepsilon \leq \|x\| \leq 1,\, s\in[0,t]} x \tilde{\mu}_L(ds, dx) \tag{6.2.10}$$

for all $t \in \mathbb{R}^+$. Here $\tilde{\mu}_L(ds, dx) = \mu_L(ds, dx) - ds\nu_L(dx)$ is the compensated jump measure and $(B_t)_{t\in\mathbb{R}^+}$ is a driftless Brownian motion in $\mathbb{R}^d$ with $\mathrm{var}(B_1) = \tau_L$. Furthermore, the terms in (6.2.10) are independent and the convergence in the last term is a.s. and locally uniform in $t \geq 0$.

In the sequel we will sometimes work with a two-sided Lévy process $L = (L_t)_{t\in\mathbb{R}}$, constructed by taking two independent copies $(L_{1,t})_{t\in\mathbb{R}^+}$, $(L_{2,t})_{t\in\mathbb{R}^+}$ of a one-sided Lévy process and setting

$$L_t = \begin{cases} L_{1,t} & \text{if} \quad t \geq 0 \\ -L_{2,-t-} & \text{if} \quad t < 0. \end{cases} \tag{6.2.11}$$

Assuming that ν satisfies additionally

$$\int\limits_{\|x\|>1} \|x\|^2\, \nu_L(dx) < \infty, \tag{6.2.12}$$

L has finite mean and covariance matrix given by

$$E(L_1) = \gamma + \int_{\|x\|>1} x\, \nu(dx), \tag{6.2.13}$$

$$\mathrm{var}(L_1) = \tau_L + \int\limits_{\mathbb{R}^d} xx^*\, \nu_L(dx). \tag{6.2.14}$$

Of particular importance in the following will be Lévy processes with paths of finite variation. A Lévy process L is of finite variation, if $\tau_L = 0$ and $\int_{\|x\|\leq 1} \|x\| \nu_L(dx) < \infty$. Then the Lévy process can be represented as

$$L_t = \tilde{\gamma}_L t + \int_0^t \int_{\mathbb{R}^d} x \mu_L(ds, dx)$$

with $\tilde{\gamma}_L = \gamma_L - \int_{\|x\|\leq 1} x\nu_L(dx)$. We have for the characteristic function $E\left[e^{i\langle u, L_t\rangle}\right] = \exp\{t\psi_L(u)\}$, $t \geq 0$, that

$$\psi_L(u) = i\langle \tilde{\gamma}_L, u\rangle + \int\limits_{\mathbb{R}^d} (e^{i\langle u,x\rangle} - 1)\, \nu_L(dx), \quad u \in \mathbb{R}^d. \tag{6.2.15}$$

An important Lévy process of finite variation in $\mathbb{S}_d$ associated to any Lévy process L in $\mathbb{R}^d$ is the quadratic variation

$$[L,L]_t = \tau_L t + \int_0^t \int_{\mathbb{R}^d} xx^* \mu_L(ds,dx) = \tau_L t + \sum_{0\le s\le t} (\Delta L_s)(\Delta L_s)^*.$$

The only stochastic part is the discrete part

$$[L,L]_t^{\mathfrak{d}} := \int_0^t \int_{\mathbb{R}^d} xx^* \mu_L(ds,dx) = \sum_{0\le s\le t} (\Delta L_s)(\Delta L_s)^*,$$

which we henceforth call "discrete quadratic variation". As for any $x \in \mathbb{R}^d$ one has that $xx^* \in \mathbb{S}_d^+$, it is easy to see that these processes are examples of matrix subordinators. Matrix subordinators are Lévy processes in the symmetric $d \times d$ matrices (which can be identified with $\mathbb{R}^{d(d+1)/2}$) whose paths remain in $\mathbb{S}_d^+$ at all times. This class of Lévy processes has been studied in detail by Barndorff-Nielsen and Pérez-Abreu (2007). For the characteristic function observe that $\mathrm{tr}(X^*Y)$ (with $X, Y \in M_d(\mathbb{R})$ and tr denoting the usual trace functional) defines a scalar product on $M_d(\mathbb{R})$ (respectively, $\mathbb{S}_d$) and note that the vec operator links the scalar product on $M_d(\mathbb{R})$ (respectively, $\mathbb{S}_d$) with the Euclidean scalar product on $\mathbb{R}^{d^2}$ via $\mathrm{tr}(X^*Y) = \mathrm{vec}(X)^*\mathrm{vec}(Y) = \langle \mathrm{vec}(Y), \mathrm{vec}(X)\rangle$ and that the norm on $M_d(\mathbb{R})$ induced by this scalar product is the Froebenius norm. Thus a matrix subordinator L has characteristic function (cf. also Barndorff-Nielsen and Pérez-Abreu (2007))

$$E\left(e^{i\mathrm{tr}(L_t^* Z)}\right) = \exp\left(t\psi_L(Z)\right), \quad Z \in \mathbb{S}_d, \quad \text{where} \tag{6.2.16}$$

$$\psi_L(Z) := i\mathrm{tr}(\gamma_L Z) + \int_{\mathbb{S}_d^+} (e^{i\mathrm{tr}(XZ)} - 1)\nu_L(dX) \tag{6.2.17}$$

with drift $\gamma_L \in \mathbb{S}_d^+$ and Lévy measure ν_L.

Matrix subordinators are necessarily of finite variation and the paths are increasing with respect to the natural ordering "$\ge$" on $\mathbb{S}_d$ (i.e. $A \ge B$, if and only if $A - B \in \mathbb{S}_d^+$). Observe that the Lévy measure is concentrated on $\mathbb{S}_d^+$ and thus all jumps are positive semi-definite. From this it is clear that matrix subordinators are a generalization of the univariate concept of Lévy subordinators.

Straightforward calculations show that the characteristic function of the discrete quadratic variation of a d-dimensional Lévy process L is given by

$$E\left(e^{i\mathrm{tr}(([L,L]_t^{\mathfrak{d}})^* Z)}\right) = \exp\left(t\psi_{[L,L]^{\mathfrak{d}}}(Z)\right), \quad Z \in \mathbb{S}_d, \quad \text{where} \tag{6.2.18}$$

$$\psi_{[L,L]^{\mathfrak{d}}}(Z) := \int_{\mathbb{R}^d} (e^{i\mathrm{tr}(xx^* Z)} - 1)\nu_L(dx). \tag{6.2.19}$$

In particular the Lévy measure of the discrete quadratic variation is given by

$$\nu_{[L,L]^{\mathfrak{d}}}(B) = \int_{\mathbb{R}^d} I_B(xx^*)\nu_L(dx) \tag{6.2.20}$$

for all Borel sets $B \subseteq \mathbb{S}_d$. It is obvious that the Lévy measure of $[L,L]^{\mathfrak{d}}$ is concentrated on the rank one matrices, thus all jumps of $[L,L]^{\mathfrak{d}}$ are rank one positive semi-definite matrices. Yet, the degeneracy of the Lévy measure does not necessarily imply degeneracy of the distribution of $[L,L]_t^{\mathfrak{d}}$ at any fixed time $t \in \mathbb{R}^+$, since elementary linear algebra implies that any matrix $\mathbb{S}_d^+$ can be written as $\sum_{i=1}^k x_i x_i^*$ with $k \in \mathbb{N}, k \le d$ and $x_i \in \mathbb{R}^d \backslash \{0\}$ such that $x_i^* x_j = 0$ for $i \ne j$.

6.2.3. Positive semi-definite processes of Ornstein-Uhlenbeck type

Before turning to a continuous time GARCH setting, let us briefly recall processes of Ornstein-Uhlenbeck type assuming values in the positive semi-definite matrices which have been introduced in Chapter 4. For more details and a statistical analysis of a stochastic volatility model involving these processes see Chapter 5. As it turns out later on, the insight gained in this section will be very helpful when defining and analysing the multivariate COGARCH process.

In the univariate case modelling the variance $(v_t)_{t\in\mathbb{R}^+}$ by Ornstein-Uhlenbeck processes has become very popular in recent years (see in particular the works of Barndorff-Nielsen and Shephard). There one assumes given a real Lévy process $(L_t)_{t\in\mathbb{R}^+}$, a parameter $\lambda \in \mathbb{R}$ and considers the SDE

$$dv_t = -\lambda v_{t-}dt + dL_t. \tag{6.2.21}$$

The solution can be shown to be

$$v_t = e^{-\lambda t}v_0 + \int_0^t e^{-\lambda(t-s)}dL_s. \tag{6.2.22}$$

Provided that $\lambda > 0, v_0 \geq 0$ and the Lévy process L is a subordinator (a.s. non-decreasing Lévy process), the unique (càdlàg adapted) solution v is positive and thus can be taken as a variance process. After extending the Lévy process to $(L_t)_{t\in\mathbb{R}}$, living on the whole real line, by (6.2.11) one can show that (6.2.21) has a unique stationary solution given by

$$v_t = \int_{-\infty}^t e^{-\lambda(t-s)}dL_s,$$

if the Lévy process has a finite logarithmic moment, i.e. $E(\log^+(L_1)) < \infty$.

Let $(L_t)_{t\in\mathbb{R}}$ now be a Lévy process with values in $M_d(\mathbb{R})$, $\mathbf{A} : M_d(\mathbb{R}) \to M_d(\mathbb{R})$ a linear operator and consider the SDE

$$dV_t = \mathbf{A}V_{t-}dt + dL_t. \tag{6.2.23}$$

As in the univariate case one can show that the unique strong solution is given by

$$V_t = e^{\mathbf{A}t}V_0 + \int_0^t e^{\mathbf{A}(t-s)}dL_s.$$

Provided $E(\log^+ \|L_1\|) < \infty$ and $\sigma(\mathbf{A}) \in (-\infty, 0) + i\mathbb{R}$, there exists a unique stationary solution given by

$$V_t = \int_{-\infty}^t e^{\mathbf{A}(t-s)}dL_s.$$

Provided that the driving Lévy process L is a matrix subordinator and $\exp(\mathbf{A}(\mathbb{S}_d^+)) \subseteq \mathbb{S}_d^+$, the process V assumes values in the symmetric positive semi-definite matrices only (assuming that the starting value $V_0 \in \mathbb{S}_d^+$, of course).

The main question now is, which possible linear operators $\mathbf{A}$ one can actually take to obtain both a unique stationary solution and positive definiteness. One possible choice is to take an $\mathbf{A}$ representable as $X \mapsto AX + XA^*$ for some $A \in M_d(\mathbb{R})$. Note the close relation of this kind of operators to Kronecker sums (see Horn and Johnson (1991, Ch. 4)) and that

Appendix A shows that all linear operators $\mathbf{A}$ on $\mathbb{S}_d$ satisfying $e^{\mathbf{A}t}(\mathbb{S}_d^+) = \mathbb{S}_d^+$ are necessarily of this form. In this set-up (6.2.23) becomes

$$dV_t = (AV_{t-} + V_{t-}A^*)dt + dL_t \tag{6.2.24}$$

and observing that $e^{\mathbf{A}t}$ has the representation $X \mapsto e^{At}Xe^{A^*t}$ we see that the solution is

$$V_t = e^{At}V_0e^{A^*t} + \int_0^t e^{A(t-s)}dL_se^{A^*(t-s)}, \tag{6.2.25}$$

compare also Horn and Johnson (1991, p. 440) for a related deterministic differential equation. Obviously, $V_t \in \mathbb{S}_d^+$ if $V_0 \in \mathbb{S}_d^+$.

Using the vec transformation and Horn and Johnson (1991, Theorem 4.4.5) we see that $\sigma(\mathbf{A}) = \sigma(A) + \sigma(A)$. Thus

Theorem 6.2.1 (Chapter 4, Theorem 4.4.5)**.** *Let $(L_t)_{t\in\mathbb{R}}$ be a matrix subordinator with $E(\log^+ \|L_1\|) < \infty$ and $A \in M_d(\mathbb{R})$ such that $\sigma(A) \subset (-\infty, 0) + i\mathbb{R}$. Then the SDE of Ornstein-Uhlenbeck type*

$$dV_t = (AV_{t-} + V_{t-}A^*)dt + dL_t$$

has a unique stationary solution

$$V_t = \int_{-\infty}^t e^{A(t-s)}dL_se^{A^*(t-s)},$$

or in vectorial representation

$$\mathrm{vec}(V_t) = \int_{-\infty}^t e^{(I\otimes A + A\otimes I)(t-s)}d\mathrm{vec}(L_s).$$

Moreover, $V_t \in \mathbb{S}_d^+$ for all $t \in \mathbb{R}$.

We conclude the discussion of positive semi-definite OU type processes by observing that even if the driving matrix subordinator is the discrete quadratic variation $[L, L]^{\mathfrak{d}}$ of a Lévy process L in $\mathbb{R}^d$ and the Lévy measure of the driving process is thus highly degenerated (concentrated on the rank one positive semi-definite matrices), one can easily have a nice stationary distribution. For instance, Corollary 5.3.18 in Chapter 5 shows that the stationary distribution has a density whenever the Lévy measure of L is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^d$.

6.3. The multivariate COGARCH(1,1) process

6.3.1. Definition and elementary properties

In order to define a continuous time multivariate GARCH(1,1) process the main idea is to replace the noise ϵ of a multivariate GARCH(1,1) process by the jumps of a multivariate Lévy process L and the autoregressive structure of the covariance matrix process by a multivariate continuous time autoregressive structure (OU structure). So the idea is again basically the same as in Brockwell et al. (2006) for the univariate case.

In the simplest BEKK GARCH(1,1) model the volatility process is given by

$$\Sigma_n = C + A\Sigma_{n-1}^{1/2}\epsilon_{n-1}\epsilon_{n-1}^*\Sigma_{n-1}^{1/2}A^* + B\Sigma_{n-1}B^* \tag{6.3.1}$$

with $C \in \mathbb{S}_d^+$, $A, B \in M_d(\mathbb{R})$ and $(\epsilon_n)_{n\in\mathbb{N}_0}$ being an i.i.d. sequence in $\mathbb{R}^d$.

This shows that the dynamics of $(\Sigma_n)_{n\in\mathbb{N}_0}$ are those of a multivariate AR process, which is "self-exciting" in the sense that we have an AR structure with the noise given by

$$\left(\Sigma_{n-1}^{1/2}\epsilon_{n-1}\epsilon_{n-1}^*\Sigma_{n-1}^{1/2}\right)_{n\in\mathbb{N}}.$$

Replacing (like in the univariate COGARCH(p,q) case) the AR structure with a CAR (OU type) one, using $V_{t-}^{1/2}d[L,L]_t^{\mathfrak{d}}V_{t-}^{1/2}$ as "noise" where L is a d-dimensional Lévy process and using the same linear operators as for positive semi-definite processes of Ornstein-Uhlenbeck type, leads now to a multivariate continuous time GARCH(1,1) process G (referred to as MUCOGARCH(1,1) process in the following) with associated volatility process V given by:

Definition 6.3.1 (MUCOGARCH(1,1))**.** *Let L be a d-dimensional Lévy process and $A, B \in M_d(\mathbb{R})$, $C \in \mathbb{S}_d^+$. Then the process $G = (G_t)_{t\in\mathbb{R}^+}$ solving*

$$\begin{aligned} dG_t &= V_{t-}^{1/2}dL_t && (6.3.2)\\ V_t &= C + Y_t && (6.3.3)\\ dY_t &= (BY_{t-} + Y_{t-}B^*)dt + AV_{t-}^{1/2}d[L,L]_t^{\mathfrak{d}}V_{t-}^{1/2}A^* && (6.3.4)\end{aligned}$$

with initial values G_0 in $\mathbb{R}^d$ and Y_0 in $\mathbb{S}_d^+$ is called a MUCOGARCH(1,1) process*.*

The process $Y = (Y_t)_{t\in\mathbb{R}^+}$ with paths in $\mathbb{S}_d^+$ is referred to as a MUCOGARCH(1,1) volatility process*.*

As except in the very last section we are only dealing with MUCOGARCH processes of order (1,1), we often omit the orders (1,1) in the following and write only "MUCOGARCH" instead of "MUCOGARCH(1,1)".

For the MUCOGARCH process to be well-defined it is necessary that there exists a unique solution to the above system of stochastic differential equations and that $Y_t \in \mathbb{S}_d^+$ for all $t \in \mathbb{R}^+$. That this is indeed the case is shown in the following theorems.

We can also directly give a stochastic differential equation for the covariance matrix process V:

$$dV_t = (B(V_{t-} - C) + (V_{t-} - C)B^*)dt + AV_{t-}^{1/2}d[L,L]_t^{\mathfrak{d}}V_{t-}^{1/2}A^*. \tag{6.3.5}$$

Obviously this SDE has a "mean reverting structure", viz. V is returning to the level C at a rate determined by B. However, since all jumps are positive semi-definite, as we shall see, C is not a "mean" level but a lower bound.

Equivalently we can use the following representation using the vec-operator:

$$\begin{aligned} dG_t &= V_{t-}^{1/2}dL_t,\ V_t = C + Y_t\\ d\mathrm{vec}(Y_t) &= (B\otimes I + I\otimes B)\mathrm{vec}(Y_{t-})dt + (A\otimes A)(V_{t-}^{1/2}\otimes V_{t-}^{1/2})d\mathrm{vec}([L,L]_t^{\mathfrak{d}})\\ d\mathrm{vec}(V_t) &= (B\otimes I + I\otimes B)(\mathrm{vec}(V_{t-}) - \mathrm{vec}(C))dt\\ &\quad +(A\otimes A)(V_{t-}^{1/2}\otimes V_{t-}^{1/2})d\mathrm{vec}([L,L]_t^{\mathfrak{d}}).\end{aligned}$$

Remark 6.3.2. *(a) The introduction of the process Y may seem to be superfluous. However, for the following results and calculations it is often advantageous to work with Y rather than with V.*

(b) Likewise one might wonder why we do not use vech *instead of* vec*, as this was an isomorphism between $\mathbb{S}_d$ and $\mathbb{R}^{d(d+1)/2}$. However, there are much more useful results for the* vec*-operator available from linear algebra than for the* vech *operator. In particular, the interplay between the* vec *operator and the Kronecker product is very helpful at numerous places.*

To see the connection between our multivariate COGARCH(1,1) process for $d = 1$ and the COGARCH(1,1) process defined in Klüppelberg et al. (2004) set $\alpha = A^2$ and $\beta = -2B$. Then equations (6.3.2)-(6.3.4) become for $d = 1$

$$dG_t = V_{t-}^{1/2} dL_t, \;\; V_t = C + Y_t \;\text{ and }\; dY_t = -\beta Y_{t-}dt + \alpha V_{t-} d[L,L]_t^{\mathfrak{d}}. \tag{6.3.6}$$

Replacing Y_t by $\tilde{Y}_t := Y_t/\alpha$ we obtain finally

$$dG_t = V_{t-}^{1/2} dL_t, \;\; V_t = C + \alpha\tilde{Y}_t \;\text{ and }\; d\tilde{Y}_t = -\beta\tilde{Y}_{t-}dt + V_{t-} d[L,L]_t^{\mathfrak{d}}.$$

From the last set of equations it is immediate that for $d = 1$ our definition agrees with the case $p = q = 1$ of the general COGARCH(p, q) definition given in Brockwell et al. (2006), i.e. $\tilde{Y}_t$ agrees with their process $\mathbf{Y}_t$ and our V_t with their V_{t+} (note that this is simply due to the fact that for convenience we have defined all processes such that they are càdlàg, whereas the previous COGARCH papers choose V to be càglàd). Hence, Brockwell et al. (2006, Theorem 2.2) implies that our definition conforms with the original definition given in Klüppelberg et al. (2004) which can also be seen from $dV_t = -\beta(V_{t-} - C)dt + \alpha V_{t-} d[L,L]_t^{\mathfrak{d}}$ and Equations (6.1.1) and (6.1.2).

The first thing we need to show is that Y and V indeed assume only values in the positive semi-definite matrices, as long as they do not explode.

Theorem 6.3.3. *If the MUCOGARCH(1,1) process is started with a positive semi-definite initial value Y_0, then it assumes values in the positive semi-definite matrices for all $t \in [0, T)$ with $T := \sup_{n\in\mathbb{N}} \inf_{t\in\mathbb{R}^+} \{Y_t \geq n\}$.*

Moreover, $Y_t \geq e^{Bt} Y_0 e^{B^* t}$ and so $Y_t \in \mathbb{S}_d^{++}$ for all $t \in [0, T)$ provided $Y_0 \in \mathbb{S}_d^{++}$.

Proof. Assume first that L is the zero process. Then we have

$$Y_t = \exp(Bt) Y_0 \exp(B^* t),$$

From the last representation it is obvious that Y_t is positive (semi-)definite for all $t \in \mathbb{R}^+$ provided we start the process with a positive (semi-)definite initial value Y_0.

Assume now that L is a compound Poisson process. As there are only finitely many jumps, it suffices to consider the jump times of the Poisson process. Let Γ_1 be the first jump time, then $Y_t \in \mathbb{S}_d^+$ for all $t \in [0, \Gamma_1)$ in view of the above result for $L = 0$. However, as

$$Y_{\Gamma_1} = Y_{\Gamma_1 -} + A(C + Y_{\Gamma_1 -})\Delta[L,L]_{\Gamma_1}^{\mathfrak{d}}(C + Y_{\Gamma_1 -})A^*,$$

we immediately deduce from $\Delta[L,L]_t^{\mathfrak{d}} \in \mathbb{S}_d^+$ that $Y_{\Gamma_1} \in \mathbb{S}_d^+$ and $Y_{\Gamma_1} \geq e^{B\Gamma_1} Y_0 e^{B^*\Gamma_1}$. Iterating the argument we get $Y_t \in \mathbb{S}_d^+$ and $Y_t \geq e^{Bt} Y_0 e^{B^* t}$ for all $t \in [0, T)$. Finally, as $e^{Bt} \in GL_d(\mathbb{R})$, we have $Y_t \in \mathbb{S}_d^{++}$ for all $t \in [0, T)$, if $Y_0 \in \mathbb{S}_d^{++}$.

In the general case we note that the driving Lévy process $[L,L]^{\mathfrak{d}}$ is of finite variation and thus it is clear that the above path-wise approach carries through to the infinite activity case. □

Implicitly we often use the following property of $t \mapsto e^{Bt}Y_0e^{B^*t}$ in the analysis of the MUCOGARCH process.

Lemma 6.3.4. *Let $Y_0 \in \mathbb{S}_d^+$ and be $\lambda_{\min}$ the eigenvalue of $B \in M_d(\mathbb{R})$ satisfying $\Re(\lambda_{\min}) \leq \Re(\lambda)$ for all $\lambda \in \sigma(B)$. Then for any $c < \Re(\lambda_{\min})$ there exists a $K > 0$ such that $e^{Bt}Y_0e^{B^*t} \geq Ke^{ct}\min(\sigma(Y_0))I_d$. for all $t \in \mathbb{R}^+$.*

Proof. Denote the usual Euclidean norm on $\mathbb{R}^d$ and the associated operator norm by $\|\cdot\|_2$. Then for any $x \in \mathbb{R}^d$ we have $x^*e^{Bt}Y_0e^{B^*t}x \geq \min(\sigma(Y_0))\|e^{B^*t}x\|_2$. Since $\|e^{B^*t}x\|_2 \geq \|e^{-B^*t}\|_2^{-1}\|x\|$ and $\|e^{-B^*t}\|_2 \leq De^{dt}$ for any $d > \max(\Re(\sigma(-B^*)) = -\Re(\lambda_{\min})$ and an appropriate $D > 0$, it follows that

$$x^*e^{Bt}Y_0e^{B^*t}x \geq D^{-1}e^{-dt}\min(\sigma(Y_0))\|x\|_2$$

for all $t \in \mathbb{R}^+$, which concludes. □

Some more subtle issues appear when considering the differential equation (6.3.5).

Proposition 6.3.5. *Let V be a solution of (6.3.5) with initial value $V_0 \in \mathbb{S}_d^+$. If $V_0 \geq C$, we have that $V_t \geq C + e^{Bt}(V_0 - C)e^{B^*t} \geq C$ for all $t \in [0,T)$ with $T := \sup_{n\in\mathbb{N}} \inf_{t\in\mathbb{R}^+}\{Y_t \geq n\}$.*

If $0 \leq V_0 < C$, however, then $V_t \geq C + e^{Bt}(V_0 - C)e^{B^*t}$ still holds, but $C + e^{Bt}(V_0 - C)e^{B^*t} < C$ and V_t need not be in $\mathbb{S}_d^+$.

Proof. The case $V_0 \geq C$ follows immediately from Theorem 6.3.3 and the relations between the processes Y and V.

In the case $0 \leq V_0 < C$ it suffices to consider the case, when L is the zero process, since the generalization can be done analogously to the previous proof.

In this case elementary analysis gives

$$V_t = e^{Bt}V_0e^{B^*t} + C - e^{Bt}Ce^{B^*t}.$$

This shows the claimed inequality. Therefore it only remains to give a counterexample showing that V_t may leave $\mathbb{S}_d^+$. Let us take

$$C = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}, \qquad V_0 = \begin{pmatrix} 0.5 & 0 \\ 0 & 0.5 \end{pmatrix},$$
$$B = \begin{pmatrix} -0.5\ln(10/9) & 0 \\ 1 & -0.5\ln(10/9) \end{pmatrix} \text{ and } x = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

Then we obtain that

$$e^B = \sqrt{\frac{9}{10}}\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \quad \text{and} \quad x^*V_1x = -\frac{11}{4}.$$

So $V_1 \notin \mathbb{S}_d^{++}$, although $V_0 \in \mathbb{S}_d^+$. Note that this problem arises also with positive probability, if the driving Lévy process is compound Poisson, as it then may well happen that there is no jump until time 1. □

If neither $V_0 \geq C$ nor $V_0 < C$, then $V_t \geq C + e^{Bt}(V_0 - C)e^{B^*t}$ is valid as well, but this cannot be compared to C.

Remark 6.3.6. *The insight gained from positive semi-definite OU-processes suggests that all eigenvalues of B should have negative real part, if one wants a "stable" COGARCH volatility process as covariance matrix process. The arguments in the last proof show immediately, that in this case $V_t \to C$ as $t \to \infty$, if the Lévy process was zero. Thus, in general the process V tends to C, as long as the driving Lévy process does not jump. The given counterexample shows, however, that below C this does not occur in a "straight" manner. Observe that this behaviour is comparable to the one of the univariate COGARCH(1,1), but there the tendency to a constant goes "straight" (see Klüppelberg et al. (2006, Proposition 2)).*

Now it is time to address the central question, the existence of a unique solution to the stochastic differential equations (6.3.2), (6.3.4) defining the MUCOGARCH(1,1). Naturally we are only interested in positive semi-definite solutions and hence restrict ourselves to cases in which Theorem 6.3.3 or Proposition 6.3.5 ensure this property, as long as there are no explosions. Moreover, we switch frequently between (6.3.4), respectively (6.3.5), and their counterparts in vectorial representation choosing the one which it is best to work with.

In the following it is implicitly understood that our processes and stochastic differential equations are not living on the whole space $M_d(\mathbb{R})$ (respectively $\mathbb{R}^{d^2}$), but on the linear sub-space $\mathbb{S}_d$ of symmetric matrices (repectively $\mathrm{vec}(\mathbb{S}_d)$). The latter can, as usual, be identified with $\mathbb{R}^{d(d+1)/2}$, when appropriate. The importance of this lies in the fact that $\mathbb{S}_d^{++}$ is an open subset of $\mathbb{S}_d$ and thus the SDE theory developed in Appendix 6.7.1 of this chapter applies immediately using the Lipschitz properties of the SDE (6.3.4) studied in Appendix 6.7.2 of this chapter.

Theorem 6.3.7. *Let $A, B \in M_d(\mathbb{R})$, $C \in \mathbb{S}_d^{++}$ and L be a d-dimensional Lévy process. Then the SDE* (6.3.4) *with initial value $Y_0 \in \mathbb{S}_d^+$ has a unique positive semi-definite solution $(Y_t)_{t\in\mathbb{R}^+}$. The solution $(Y_t)_{t\in\mathbb{R}^+}$ is locally bounded and of finite variation.*

Proof. Define the maps F, G by $F(\mathrm{vec}(y)) = (I_d \otimes B + B \otimes I_d)\mathrm{vec}(y)$ and $G(y) = (A \otimes A)\left((C+y)^{1/2} \otimes (C+y)^{1/2}\right)$. Then the SDE (6.3.4) can be written as

$$d\mathrm{vec}(Y_t) = F(\mathrm{vec}(Y_{t-}))dt + G(Y_{t-})d\mathrm{vec}([L,L]_t^{\mathfrak{d}}).$$

Moreover, we define the set $U_{C,\epsilon} = \{x \in \mathbb{S}_d : x > -\epsilon I_d\}$ for some ϵ with $0 < \epsilon < \min\sigma(C)$. Then the set $U_{C,\epsilon}$ (and thus $\mathrm{vec}(U_{C,\epsilon})$) is open and for each $x \in U_{C,\epsilon}$ we have $x + C > (\min\sigma(C) - \epsilon)\, I_d \in \mathbb{S}_d^{++}$. Being a linear map F is trivially Lipschitz and Lemma 6.7.16 shows that G is Lipschitz on any set of the form $\mathrm{vec}(U_{C,\epsilon} \cap \{x \in \mathbb{S}_d^+ : \|x\| \leq c\})$ with $c > 0$, thus locally Lipschitz on $\mathrm{vec}(U_{C,\epsilon})$. Applying Theorem 6.7.3 with $U_n = \mathrm{vec}(\{x \in \mathbb{S}_d : x \geq (1/n - \epsilon)\, I_d\})$ implies the existence of a unique solution Y_t in $U_{C,\epsilon}$ to the SDE up to a random stopping time T. Theorem 6.3.3 ensures that any solution necessarily stays in $\mathbb{S}_d^+ \subset U_{C,\epsilon}$. Hence, on $T < \infty$ the only thing that can happen is an explosion, i.e. $\limsup_{t\to T} \|Y_t\| = \infty$. However, the linearity of F and Lemma 6.7.16 imply that both F and G grow at most linearly, i.e. (6.7.3) holds, and so Theorem 6.7.3 shows that T is infinite. This establishes the existence of a unique solution on $[0, \infty)$ and the local boundedness of Y.

Since the SDE implies that $(Y_t)_{t\in\mathbb{R}^+}$ is the sum of an integral with respect to time and one with respect to a finite variation Lévy process, it is straightforward to see that $(Y_t)_{t\in\mathbb{R}^+}$ is of finite variation on compacts. □

When considering only initial values in $\mathbb{S}_d^{++}$, we can relax the condition $C \in \mathbb{S}_d^{++}$. Moreover, we need a replacement for the concept of local boundedness.

Definition 6.3.8 (Ch. 4, Def. 4.3.1)**.** *Let $(W, \|\cdot\|_W)$ be either $\mathbb{R}^d, M_d(\mathbb{R})$ or $\mathbb{S}_d(\mathbb{R})$ with $d \in \mathbb{N}$ and equipped with the norm $\|\cdot\|_W$, $U \subseteq W$ open and let $(X_t)_{t\in\mathbb{R}^+}$ be a U-valued stochastic process. We say that the process X is* locally bounded within U *if there exists a sequence of stopping times $(T_n)_{n\in\mathbb{N}}$ increasing to infinity almost surely and a sequence of compact convex subsets $D_n \subset U$ with $D_n \subset D_{n+1}$ $\forall n \in \mathbb{N}$ such that $X_t \in D_n$ for all $0 \le t < T_n$.*

Theorem 6.3.9. *Let $A, B \in M_d(\mathbb{R})$, $C \in \mathbb{S}_d^+$ and L be a d-dimensional Lévy process. Then the SDE* (6.3.4) *with initial value $Y_0 \in \mathbb{S}_d^{++}$ has a unique positive definite solution $(Y_t)_{t\in\mathbb{R}^+}$. $(Y_t)_{t\in\mathbb{R}^+}$ is locally bounded within $\mathbb{S}_d^{++}$ and of finite variation.*

Proof. As $\mathbb{S}_d^{++}$ is an open set Theorem 6.7.3 (with $U_n = \{x \in \mathbb{S}_d : x > (1/n)I_d\}$, for instance) gives the existence of a unique solution Y_t in $\mathbb{S}_d^{++}$ up to a stopping time T, at which it either explodes, jumps out of $\mathbb{S}_d^+$ or reaches the boundary of $\mathbb{S}_d^{++}$. Yet, as Theorem 6.3.3 ensures that any solution Y_t satisfies $Y_t \ge e^{Bt}Y_0e^{B^*t}$, it cannot happen that it hits the boundary or jumps out of $\mathbb{S}_d^+$ and arguing as before there can be no explosions in finite time. Therefore, we have again $T = \infty$.

The proof of Theorem 6.7.3 shows that $(Y_t)_{t\in\mathbb{R}^+}$ is locally bounded within $\mathbb{S}_d^{++}$ and the finite variation follows as in the last theorem. □

Likewise we could have considered the SDE (6.3.5). Using the relationship between (6.3.5) and (6.3.4) we obtain the following.

Corollary 6.3.10. *Let $A, B \in M_d(\mathbb{R})$, $C \in \mathbb{S}_d^+$, and L be a d-dimensional Lévy process. Assume that the initial value satisfies $V_0 \ge C$ and either $C \in \mathbb{S}_d^{++}$ or $V_0 > C$ holds. Then the SDE* (6.3.5) *has a unique positive definite solution $(V_t)_{t\in\mathbb{R}^+}$.*

Remark 6.3.11. *It is immediate to see that all results stated for the stochastic differential equations (6.3.4) and (6.3.5) remain valid when replacing the differential $d[L,L]_t^{\mathfrak{d}}$ with the differential $d\tilde{L}_t$ of a matrix subordinator $\tilde{L}_t$ (see Barndorff-Nielsen and Pérez-Abreu (2007)).*

For the sake of simplicity we state the following two theorems only for $C \in \mathbb{S}_d^{++}$ noting that the analogues for $C \in \mathbb{S}_d^+$ and $Y_0 \in \mathbb{S}_d^{++}$ are obvious.

That positive semi-definite OU type processes and MUCOGARCH volatility processes are related is also reflected in the following representation of $(Y_t)_{t\in\mathbb{R}^+}$ and Equation (6.2.25).

Theorem 6.3.12. *Let $C \in \mathbb{S}_d^{++}$ and Y_0 in $\mathbb{S}_d^+$. Then the MUCOGARCH(1,1) volatility process Y satisfies*

$$Y_t = e^{Bt}Y_0e^{B^*t} + \int_0^t e^{B(t-s)}A(C + Y_{s-})^{1/2}d[L,L]_s^{\mathfrak{d}}(C + Y_{s-})^{1/2}A^*e^{B^*(t-s)} \tag{6.3.7}$$

for all $t \in \mathbb{R}^+$.

Proof. Define $M_t = \int_0^t A(C + Y_{s-})^{1/2} d[L, L]_s^{\mathfrak{d}} (C + Y_{s-})^{1/2} A^*$. Then M is $\mathbb{S}_d^+$-increasing and of finite variation and Y obviously solves the stochastic differential equation $dX_t = (BX_{t-} + X_{t-}B^*)dt + dM_t$ $(*)$. Standard theory implies that this differential equation has a unique solution and the same elementary calculations as for Ornstein-Uhlenbeck processes show that the solution of $(*)$ with initial value Y_0, which is necessarily equal to Y, is given by

$$e^{Bt} Y_0 e^{B^* t} + \int_0^t e^{B(t-s)} A dM_s A^* e^{B^*(t-s)} =$$
$$e^{Bt} Y_0 e^{B^* t} + \int_0^t e^{B(t-s)} A (C + Y_{s-})^{1/2} d[L, L]_s^{\mathfrak{d}} (C + Y_{s-})^{1/2} A^* e^{B^*(t-s)}.$$

□

Recently Reiß, Riedle and van Gaans (2007) studied univariate equations of the form $X(t) = J(t) + \int_0^t g(t-s) f(X_{s-}) dZ_s$ and their relation to certain SDEs. In particular, they obtained uniqueness of the solutions under uniform Lipschitz assumptions on f. Our equation (6.3.7) is basically a multivariate equation of this type with f being only locally Lipschitz. From the arguments given in Reiß et al. (2007) it is clear that their Theorem 5.2 has a straightforward multivariate extension. Using a localization procedure as in the proof of Theorem 6.7.3 this uniqueness result can then be easily extended to f being only defined on an open subset and locally Lipschitz. Hence, (6.3.7) provides an alternative characterization for the MUCOGARCH volatility process.

So far we have excluded the MUCOGARCH process G itself from the analysis. However, the following result should be obvious by now.

Theorem 6.3.13. *Let $C \in \mathbb{S}_d^{++}$, $B, A \in M_d(\mathbb{R})$ and L be a d-dimensional Lévy process. Then the system of SDEs* (6.3.2), (6.3.4) *has a unique solution $(G_t, Y_t)_{t \in \mathbb{R}^+}$ with paths in $\mathbb{R}^d \times \mathbb{S}_d^+$ for any initial value (G_0, Y_0) in $\mathbb{R}^d \times \mathbb{S}_d^+$.*

The solution $(G_t, Y_t)_{t \in \mathbb{R}^+}$ is a semi-martingale and locally bounded.

6.3.2. Univariate COGARCH(1,1) bounds and finiteness of moments

In this section we show that similar to the COGARCH(p, q) case (cf. Brockwell et al. (2006, Lemma 9.1)) the norm of a MUCOGARCH(1,1) volatility process can be bounded by a univariate COGARCH(1,1) volatility process. This will first be shown for processes driven by compound Poisson processes and then for the general case using an approximation by compound Poisson processes which is of interest in its own. It should be noted that all processes, i.e. the majorizing and approximating ones, can all be defined on the original probability space.

The majorization will then immediately imply sufficient conditions for the existence of moments and in the next section it will be utilized to obtain criteria for the existence of stationary distributions.

To see that the processes defined in the following are indeed univariate COGARCH processes, we need the following general lemma saying that any driftless Lévy subordinator is the discrete quadratic variation of a Lévy process.

Lemma 6.3.14. *Let $(L_t)_{t \in \mathbb{R}^+}$ be a driftless Lévy subordinator. Then there exists a Lévy process $(\overline{L}_t)_{t \in \mathbb{R}^+}$ in $\mathbb{R}$ such that $L_t = [\overline{L}, \overline{L}]_t^{\mathfrak{d}}$ for all $t \in \mathbb{R}^+$.*

Proof. Denote the jump measure associated to L by μ_L, i.e. $L_t = \int_0^t \int_{\mathbb{R}^+} x\mu_L(ds, dx)$, and its Lévy measure by ν_L. Set

$$\overline{L}_t = \int_0^t \int_{0<x\leq 1} \sqrt{x}(\mu_L(ds, dx) - ds\nu_L(dx)) + \int_0^t \int_{x>1} \sqrt{x}\mu_L(ds, dx)$$

noting that the existence of the first integral is ensured by the fact that the integral $\int_{0<x\leq 1} \sqrt{x}^2 \nu_L(dx) = \int_{0<x\leq 1} x\nu_L(dx)$ is finite, since L is of finite variation. Obviously $(\overline{L}_t)_{t\in\mathbb{R}^+}$ is a Lévy process and $L_t = [\overline{L}, \overline{L}]_t^{\mathfrak{d}}$. □

It should be noted that this result cannot be generalized to matrix subordinators and d-dimensional Lévy processes.

In the following we shall consider a special norm that fits particularly well to our model. Again, $\|\cdot\|_2$ denotes the operator norm on $M_{d^2}(\mathbb{R})$ associated to the usual Euclidean norm. Assume now that B is diagonalizable and let $S \in GL_d(\mathbb{C})$ be such that $S^{-1}BS$ is diagonal. Then we define the norm $\|\cdot\|_{B,S}$ on $M_{d^2}(\mathbb{R})$ by $\|X\|_{B,S} := \|(S^{-1} \otimes S^{-1})X(S \otimes S)\|_2$ for $X \in M_{d^2}(\mathbb{R})$. It should be noted that $\|\cdot\|_{B,S}$ depends both on B and on the choice of the matrix S diagonalizing B. Actually, $\|\cdot\|_{B,S}$ is again an operator norm, namely the one associated to the norm $\|x\|_{B,S} := \|(S^{-1} \otimes S^{-1})x\|_2$ on $\mathbb{R}^{d^2}$. Besides, $\|\cdot\|_{B,S}$ actually is simply the norm $\|\cdot\|_2$ provided S is a unitary matrix, since the norm $\|\cdot\|_2$ is unitarily invariant (cf. Horn and Johnson (1985, p. 308)). As is well-known from basic linear algebra, S can be chosen to be unitary if and only if B is normal.

A very similar norm has also been used in Brockwell et al. (2006). There, a more general class of norms, including the Euclidean norm, was used leading to great flexibility. In principle our results below can be adapted to this more general class of norms. However, it is very natural to consider only the Euclidean norm for the multivariate COGARCH, since we need a unitarily invariant norm to deal with the matrix square root and this immediately restricts the possible operator norms to the one induced by the Euclidean norm (see Horn and Johnson (1985, Corollary 5.6.35)).

The following elementary results are straightforward to obtain.

Lemma 6.3.15. *It holds that* $\|S\otimes S\|_{B,S} = \|S\|_2^2$ *and* $\|S^{-1}\otimes S^{-1}\|_{B,S} = \|S^{-1}\|_2^2$. *Moreover,*

$$\|x\|_{B,S} \leq \|S^{-1}\|_2^2\|x\|_2 \text{ and } \|x\|_2 \leq \|S\|_2^2\|x\|_{B,S} \text{ for all } x \in \mathbb{R}^{d^2},$$
$$\|X\|_{B,S} \leq \|S\|_2^2\|S^{-1}\|_2^2\|X\|_2 \text{ and } \|X\|_2 \leq \|S\|_2^2\|S^{-1}\|_2^2\|X\|_{B,S} \text{ for all } X \in M_{d^2}(\mathbb{R}).$$

Now we can analyse the norm of compound Poisson driven MUCOGARCH volatility processes. Recall that in the univariate case the MUCOGARCH volatility process Y is just a (deterministically) scaled version of the COGARCH volatility process $\mathbf{Y}$ defined in Brockwell et al. (2006).

Theorem 6.3.16. *Let* $C \in \mathbb{S}_d^+$ *and* Y *be a MUCOGARCH volatility process with initial value* $Y_0 \in \mathbb{S}_d^+$ *and driven by a compound Poisson process* L *in* $\mathbb{R}^d$. *Assume further that* $B \in M_d(\mathbb{R})$ *is diagonalizable and let* $S \in GL_d(\mathbb{C})$ *be such that* $S^{-1}BS$ *is diagonal. The process solving the SDE*

$$dy_t = 2\lambda y_{t-}dt + \|S\|_2^2\|S^{-1}\|_2^2 K_{2,B}\|A\otimes A\|_{B,S}\left(\frac{\|C\|_2}{K_{2,B}} + y_{t-}\right)d\tilde{L}_t, y_0 = \|\mathrm{vec}(Y_0)\|_{B,S} \quad (6.3.8)$$

with

$$\tilde{L}_t := \int_0^t \int_{\mathbb{R}^d} \|\mathrm{vec}(xx^*)\|_{B,S} \mu_L(ds, dx), \;\; \lambda := \max(\Re(\sigma(B)))$$
$$\textit{and} \;\; K_{2,B} := \max_{X \in \mathbb{S}_d^+, \|X\|_2 = 1} \left(\frac{\|X\|_2}{\|\mathrm{vec}(X)\|_{B,S}} \right)$$

is the volatility process of a univariate MUCOGARCH(1,1) process and y satisfies

$$\|\mathrm{vec}(Y_t)\|_{B,S} \leq y_t \textit{ for all } t \in \mathbb{R}^+ \textit{ a.s.} \tag{6.3.9}$$

Moreover,

$$K_{2,B} \leq \|S\|_2^2 \max_{X \in \mathbb{S}_d^+, \|X\|_2 = 1} \left(\frac{\|X\|_2}{\|\mathrm{vec}(X)\|_2} \right) \leq \|S\|_2^2.$$

Proof. Combining Lemma 6.3.14 with (6.3.6) it is clear that the above defined process $(y_t)_{t \in \mathbb{R}}$ is a univariate MUCOGARCH(1,1) process.

Let Γ_1 be the time of the first jump of L and $t \in [0, \Gamma_1)$. Since $\|e^{(I_d \otimes B + B \otimes I_d)t}\|_{B,S} = e^{2\lambda t}$, it holds that

$$\begin{aligned} \|\mathrm{vec}(Y_t)\|_{B,S} = \|e^{(I_d \otimes B + B \otimes I_d)t} \mathrm{vec}(Y_0)\|_{B,S} &\leq \|e^{(I_d \otimes B + B \otimes I_d)t}\|_{B,S} \|\mathrm{vec}(Y_0)\|_{B,S} \\ &= e^{2\lambda t} y_0 = y_t. \end{aligned}$$

Thus, (6.3.9) is shown for all $t \in [0, \Gamma_1)$. At time Γ_1 we have

$$\begin{aligned} &\|\mathrm{vec}(Y_{\Gamma_1})\|_{B,S} \\ &\quad = \|\mathrm{vec}(Y_{\Gamma_1 -}) + (A \otimes A)((C + Y_{\Gamma_1 -})^{1/2} \otimes (C + Y_{\Gamma_1 -})^{1/2}) \mathrm{vec}(\Delta L_{\Gamma_1} (\Delta L_{\Gamma_1})^*)\|_{B,S} \\ &\quad \leq y_{\Gamma_1 -} + \|A \otimes A\|_{B,S} \|(C + Y_{\Gamma_1 -})^{1/2} \otimes (C + Y_{\Gamma_1 -})^{1/2}\|_{B,S} \|\mathrm{vec}(\Delta L_{\Gamma_1} (\Delta L_{\Gamma_1})^*)\|_{B,S} \\ &\quad \leq y_{\Gamma_1 -} + \|A \otimes A\|_{B,S} \|S\|_2^2 \|S^{-1}\|_2^2 \|(C + Y_{\Gamma_1 -})^{1/2} \otimes (C + Y_{\Gamma_1 -})^{1/2}\|_2 \Delta \tilde{L}_{\Gamma_1} \\ &\quad \leq y_{\Gamma_1 -} + \|A \otimes A\|_{B,S} \|S\|_2^2 \|S^{-1}\|_2^2 \left(\|C\|_2 + \|Y_{\Gamma_1 -}\|_2 \right) \Delta \tilde{L}_{\Gamma_1} \\ &\quad \leq y_{\Gamma_1 -} + \|A \otimes A\|_{B,S} \|S\|_2^2 \|S^{-1}\|_2^2 K_{2,B} \left(K_{2,B}^{-1} \|C\|_2 + \|\mathrm{vec}(Y_{\Gamma_1 -})\|_{B,S} \right) \Delta \tilde{L}_{\Gamma_1} = y_{\Gamma_1}, \end{aligned}$$

which establishes (6.3.9) for $t = \Gamma_1$. Iterating these arguments shows (6.3.9) for all $t \in \mathbb{R}^+$.

The first inequality for $K_{2,B}$ follows immediately from Lemma 6.3.15 and the second one by Horn and Johnson (1985, p. 314). □

Remark 6.3.17. *If S is unitary, $K_{2,B} = 1$ is easily seen.*

Otherwise, an inspection of the proof shows that the inequality (6.3.9) *also holds if $K_{2,B}$ is replaced by $\|S\|_2$ in* (6.3.8) *which saves one from calculating the value of $K_{2,B}$ in practice. Likewise, $\|A \otimes A\|_{B,S}$ can be replaced by $\|A \otimes A\|_2 = \|A\|_2^2$, since $\|(A \otimes A)((C + Y_{\Gamma_1 -})^{1/2} \otimes (C + Y_{\Gamma_1 -})^{1/2})\|_{B,S} \leq \|S\|_2^2 \|S^{-1}\|_2^2 \|(A \otimes A)((C + Y_{\Gamma_1 -})^{1/2} \otimes (C + Y_{\Gamma_1 -})^{1/2})\|_2$.*

This can be done in all upcoming results involving $K_{2,B}$ or $\|A \otimes A\|_{B,S}$ as well.

In order to extend this result to MUCOGARCH processes driven by general Lévy processes, we need to show that we can approximate the MUCOGARCH volatility processes by approximating the driving Lévy process. The following result is very similar to Brockwell et al. (2006, Lemma 8.2). Yet, we need to give a detailed proof, since the standard results

cannot be applied due to the fact that we have only locally Lipschitz coefficients. For the definition of prelocal convergence in $\underline{\underline{H}}^2$ we refer to Protter (2004, Section V.4), as we will use this technical device only here and later on employ only the ucp-convergence. Observe that several results we use in the following are stated for one-dimensional SDEs in Protter (2004), but that they extend immediately to the multidimensional case as remarked on p. 257 of that book (cf. also Protter (1978, p. 346) in this context and for a definition of the multidimensional $\underline{\underline{H}}^2$-norm).

Proposition 6.3.18. *Let Y be a MUCOGARCH volatility process with $C \in \mathbb{S}_d^{++}$ and $Y_0 \in \mathbb{S}_d^+$ driven by a Lévy process L in $\mathbb{R}^d$ and $(\epsilon_n)_{n\in\mathbb{N}}$ a monotonically decreasing sequence in $\mathbb{R}^{++}$ with $\lim_{n\to\infty}\epsilon_n = 0$. Assume that B is diagonalizable, define for $n \in \mathbb{N}$ compound Poisson Lévy processes L_n by $L_{n,t} = \int_0^t \int_{\mathbb{R}^d,\|x\|\geq\epsilon_n} x\mu_L(ds,dx)$ and associated MUCOGARCH volatility processes Y_n by*

$$dY_{n,t} = (BY_{n,t-} + Y_{n,t-}B^*)dt + A(C + Y_{n,t-})^{1/2}d[L_n, L_n]_t^{\mathfrak{d}}(C + Y_{n,t-})^{1/2}A^*, \;\; Y_{n,0} = Y_0.$$

Then $Y_n \to Y$ as $n \to \infty$ in ucp and prelocally in $\underline{\underline{H}}^2$.

Proof. Using Theorems V.2 and V.12 of Protter (2004) it suffices to show prelocal convergence in $\underline{\underline{H}}^2$.

Noting that the existence is ensured by Protter (2004, Theorem V.4), choose a sequence of finite stopping times $(T_{L,k})_{k\in\mathbb{N}}$ increasing to infinity a.s. such that $([L,L]^{\mathfrak{d}})^{T_{L,k}-} \in \underline{\underline{H}}^2$ for all $k \in \mathbb{N}$. ($X_t^{T-} := X_t I_{[0,T)}(t) + X_{T-}I_{[T,\infty)}(t)$ for any process X and stopping time T.) The construction of L_n implies that $([L_n, L_n]^{\mathfrak{d}})^{T_{L,k}-} \in \underline{\underline{H}}^2$ for all $n, k \in \mathbb{N}$ and the definition of the $\underline{\underline{H}}^2$ norm gives

$$\|([L,L]^{\mathfrak{d}} - [L_n,L_n]^{\mathfrak{d}})^{T_{L,k}-}\|_{\underline{\underline{H}}^2} = \left\| \int_0^{T_{L,k}-} \int_{\mathbb{R}^d,\|x\|<\epsilon_n} xx^*\mu_L(ds,dx) \right\|_{L^2}$$

which obviously tends to zero as $n \to \infty$.

Let $S \in GL_d(\mathbb{C})$ be such that $S^{-1}BS$ is diagonal and denote for $n \in \mathbb{N}$ by y_n the univariate MUCOGARCH(1,1) processes constructed in Theorem 6.3.16 satisfying $\|\mathrm{vec}(Y_{n,t})\|_{B,S} \leq y_{n,t}$ for all $t \in \mathbb{R}^+$. $y_{n,t} + K_{2,B}^{-1}\|C\|$ is a univariate COGARCH(1,1) volatility process as defined in Klüppelberg et al. (2004) where it is denoted by σ_{t+}^2. Since we only add more jumps in $[L_n, L_n]^{\mathfrak{d}}$ when we increase n, it is straightforward to see from equations (3.3) and (3.4) in Klüppelberg et al. (2004) that $y_{n+l,t} \geq y_{n,t}$ for all $n, l \in \mathbb{N}$ and $t \in \mathbb{R}^+$. Moreover, defining the process y by

$$\begin{aligned} dy_t &= 2\lambda y_{t-}dt + \|S\|_2^2\|S^{-1}\|_2^2 K_{2,B}\|A\otimes A\|_{B,S}\left(\frac{\|C\|_2}{K_{2,B}} + y_{t-}\right)d\tilde{L}_t, \qquad (6.3.10)\\ y_0 &= \|\mathrm{vec}(Y_0)\|_{B,S} \end{aligned}$$

with $\tilde{L}_t := \int_0^t \int_{\mathbb{R}^d} \|\mathrm{vec}(xx^*)\|_{B,S}\mu_L(ds,dx)$, the same argument implies $y_{n,t} \leq y_t$ for all $n \in \mathbb{N}$ and $t \in \mathbb{R}^+$. Note $(\tilde{L}_t)_{t\in\mathbb{R}^+}$ is a well-defined Lévy process, because there is a $K > 0$ such that

$$\int_{\mathbb{R}^d}(\|\mathrm{vec}(xx^*)\|_{B,S} \wedge 1)\nu_L(dx) \leq K\int_{\mathbb{R}^d}(\|xx^*\|_2 \wedge 1)\nu_L(dx) = K\int_{\mathbb{R}^d}(\|x\|_2^2 \wedge 1)\nu_L(dx) < \infty.$$

Define now for $k \in \mathbb{N}$ stopping times $T_{y,k} := \inf\{t \in \mathbb{R}^+ : y_t > k\}$, $T_{Y,k} := \inf\{t \in \mathbb{R}^+ : \|\mathrm{vec}(Y_t)\|_{B,S} > k\}$ and sets $W_k := \{x \in \mathbb{S}_d^+ : \|\mathrm{vec}(x)\|_{B,S} \leq k\}$ which are compact and convex. From the arguments in the proof of Theorem 6.3.7, it follows that $T_{y,k}$, $T_{Y,k}$ and, hence, $T_k := \min(T_{y,k}, T_{Y,k}, T_{L,k})$ are sequences of stopping times increasing to infinity. Moreover, we have $Y_{n,t}, Y_t \in W_k$ for all $t \in [0, T_k)$ and $n \in \mathbb{N}$.

Using implicitly the usual scalar product, we denote the orthogonal projection on W_k by Π_{W_k}. Then it follows that $Y_n^{T_k-}$ satisfies

$$\begin{aligned} dX_{n,t} =&(B\Pi_{W_k}(X_{n,t-}) + \Pi_{W_k}(X_{n,t-})B^*)dt^{T_k-} \\ &+ A(C + \Pi_{W_k}(X_{n,t-}))^{1/2} d[L_n^{T_k-}, L_n^{T_k-}]_t^{\mathfrak{d}} (C + \Pi_{W_k}(X_{n,t-}))^{1/2} A^* \end{aligned}$$

with $X_{n,0} = Y_0$ and Y^{T_k-} satisfies

$$\begin{aligned} dX_t =&(B\Pi_{W_k}(X_{t-}) + \Pi_{W_k}(X_{t-})B^*)dt^{T_k-} \\ &+ A(C + \Pi_{W_k}(X_{t-}))^{1/2} d[L_n^{T_k-}, L_n^{T_k-}]_t^{\mathfrak{d}} (C + \Pi_{W_k}(X_{t-}))^{1/2} A \end{aligned}$$

with $X_0 = Y_0$. Clearly for $k \in \mathbb{N}$ fixed the functions appearing in the coefficients of the above SDEs are uniformly bounded and uniformly Lipschitz.

Assume without loss of generality that the semi-martingale with the constant value Y_0 and t^{T_k-} are in $\underline{\underline{H}}^2$ for all $k \in \mathbb{N}$ and that the Lévy process $L \in \mathcal{S}(1/(2\sqrt{8}a)$ where a is some finite uniform bound on the Lipschitz coefficients. For the definition of $\mathcal{S}(\cdot)$ see Protter (2004, p. 248). (Otherwise one would just have to take the minimum with three more sequences of stopping times increasing to infinity whose existence is ensured by Protter (2004, Theorems V.4 and V.5).)

Combining the above results, Protter (2004, Theorem V.9) implies that

$$\lim_{n\to\infty} \|(Y_n - Y)^{T_k-}\|_{\underline{\underline{H}}^2} = 0$$

for all $k \in \mathbb{N}$. This shows the prelocal convergence of Y_n to Y in $\underline{\underline{H}}^2$. □

Theorem 6.3.19. *For $C \in \mathbb{S}_d^{++}$ Theorem 6.3.16 holds with any driving Lévy process L in $\mathbb{R}^d$.*

Proof. Let $(Y_n)_{n\in\mathbb{N}}$ be the sequence of MUCOGARCH(1,1) processes converging in ucp to Y constructed in the proof of the last proposition. In that proof it has already been shown that $\|\mathrm{vec}(Y_{n,t})\|_{B,S} \leq y_t$ for all $t \in \mathbb{R}^+$. Picking a subsequence converging a.s. on compacts immediately concludes. □

Since the finiteness of moments of univariate COGARCH(1,1) processes is well-known from Klüppelberg et al. (2004, Section 4), we can now give sufficient conditions for the MUCOGARCH volatility process to have some finite moments.

Proposition 6.3.20. *Let $k \in \mathbb{N}$, $C \in \mathbb{S}_d^{++}$, $Y_0 \in \mathbb{S}_d^+$ such that $E(\|Y_0\|^k) < \infty$ and B be diagonalizable. Assume further that the MUCOGARCH volatility process Y is driven by a Lévy process L satisfying $E(\|L_1\|^{2k}) < \infty$ and that $\lambda = \max(\Re(\sigma(B))) < 0$.*

Then $E(\|Y_t\|^k) < \infty$ for all $t \in \mathbb{R}^+$ and $t \mapsto E(\|Y_t\|^k)$ is locally bounded.

Proof. Let y be the process constructed in Theorems 6.3.16 and 6.3.19. Then it suffices to show that $E(y_t^k) < \infty$ and that this is locally bounded in t. By construction $E(y_0)$ is finite. Moreover, let $\bar{L}$ be the Lévy process constructed in Lemma 6.3.14 such that $\tilde{L}_t = [\bar{L}, \bar{L}]_t^{\mathfrak{d}}$. The finiteness of $E(\|L_1\|^{2k})$ implies that $\int_{\mathbb{R}^d} \|x\|_2^{2k} \nu_L(dx) = \int_{\mathbb{R}^d} \|xx^*\|_2^k \nu_L(dx) < \infty$ (with $\|\cdot\|$ denoting the Euclidean norm in the first integral and the associated operator norm in the second one). Since the finiteness of the integrals is independent of the particular norm used, it follows that $\int_{\mathbb{R}^d} \|\mathrm{vec}(xx^*)\|_{B,S}^k \nu_L(dx) = \int_{\mathbb{R}} |x|^k \nu_{\tilde{L}}(dx) = \int_{\mathbb{R}} |x|^{2k} \nu_{\bar{L}}(dx) < \infty$. Hence, $E(|\bar{L}_1|^{2k})$ is finite and using the results of Klüppelberg et al. (2004, Section 4) as in the proof of Brockwell et al. (2006, Proposition 4.1) concludes. □

Some explicit expressions for the first and second moment are presented in the upcoming Section 6.3.4.

6.3.3. Markovian properties and stationarity

Turning to the study of the Markovian properties of a MUCOGARCH process we refer to Appendix 6.7.1.2 of this chapter for definitions and necessary general results. Moreover, we implicitly assume that our given filtered probability space is enlarged as there to allow for arbitrary initial conditions.

From the general results of Appendix 6.7.1.2 we can immediately deduce important Markovian properties of the MUCOGARCH, provided we only want to study the volatility process on the open set $\mathbb{S}_d^{++}$, i.e. we take only initial values in $\mathbb{S}_d^{++}$.

Theorem 6.3.21. *Let $C \in \mathbb{S}_d^+$. The MUCOGARCH process (G, Y) as well as its volatility process Y alone are temporally homogeneous weakly Fellerian strong Markov processes on $\mathbb{R}^d \times \mathbb{S}_d^{++}$ and $\mathbb{S}_d^{++}$, respectively.*

It is clear that $(G_t)_{t\in\mathbb{R}^+}$ alone cannot be Markovian. The above result is, unfortunately, not sufficient. In order to apply a tightness argument, we do need Y to be a Markov process on $\mathbb{S}_d^+$, i.e. we also need to allow initial values that are only positive semi-definite and not strictly positive definite. To achieve this we need to restrict ourselves to C being positive definite.

Theorem 6.3.22. *Let $C \in \mathbb{S}_d^{++}$. The MUCOGARCH process (G, Y) as well as its volatility process Y alone are temporally homogeneous weakly Fellerian strong Markov processes on $\mathbb{R}^d \times \mathbb{S}_d^+$ and $\mathbb{S}_d^+$, respectively.*

Proof. Inspecting the arguments of Appendix 6.7.1.2 of this chapter it is clear that to show the above result it suffices to generalize Lemmata 6.7.9 and 6.7.10 to the closed set $\mathbb{S}_d^+$ instead of the open set U for the volatility process Y.

Reconsider thus the set-up of the proof of Theorem 6.3.7.

Since the approximating processes corresponding to the sets U_n as constructed in the proof of Theorem 6.7.3 can be assumed without loss of generality to be $\boldsymbol{\sigma}(\mathcal{B}(\mathbb{S}_d) \times \mathcal{B}(\mathbb{R}^+) \times \mathcal{F})$-measurable, they can be assumed to be $\boldsymbol{\sigma}(\mathcal{B}(\mathbb{S}_d^+) \times \mathcal{B}(\mathbb{R}^+) \times \mathcal{F})$-measurable when considering only initial values in $\mathbb{S}_d^+$. The ucp-convergence of the approximating processes to Y for any initial value in $\mathbb{S}_d^+$ and Protter (2004, Theorem IV.62) thus imply that there exists a version of Y that is $\boldsymbol{\sigma}(\mathcal{B}(\mathbb{S}_d^+) \times \mathcal{B}(\mathbb{R}^+) \times \mathcal{F})$-measurable. Hence, the assertions of Lemma 6.7.10 still hold.

Likewise Lemma 6.7.9 extends, since the adaptation of the arguments of Protter (2004, Theorem V.38) as in the proof of Lemma 6.7.9 actually gives that the flow of the volatility process is continuous on $[0, T(x))$ ($T(x)$ being the exit time of the volatility process from the set $U_{C,\epsilon}$ when started at x) and $T(x)$ is infinite for all $x \in \mathbb{S}_d^+$, as has already been shown. □

In order to show the existence of a stationary distribution of the MUCOGARCH volatility process Y we need to recall some notions and results from the theory of weak convergence. For more details we refer to any of the standard texts (e.g. Billingsley (1999), Jacod and Shiryaev (2003) or Pollard (1984)). Below we denote by $\mathcal{M}_1(E)$ the set of all probability measures on the Borel σ-algebra of a Polish space E .

Definition 6.3.23 (Tightness)**.** *Let E be a Polish space. A subset $\mathcal{I} \subseteq \mathcal{M}_1(E)$ is said to be* tight *if for every $\epsilon > 0$ there is a compact subset $K \subseteq E$ such that $\mu(K) \geq 1 - \epsilon$ for all $\mu \in \mathcal{I}$.*

The following theorem on the existence of a stationary distribution for a Markov process is called the "Krylov-Bogoliubov existence theorem" in the literature. For a proof see da Prato and Zabczyk (1996, Section 3.1) or Reiß, Riedle and van Gaans (2006, Theorem 4.6).

Theorem 6.3.24. *Let E be a Polish space and $(P_s)_{s\in\mathbb{R}^+}$ the transition semi-group of an E-valued weak Feller Markov process. Assume that there is an $\eta \in \mathcal{M}_1(E)$ such that the set $\{P_t^*\eta : t \in \mathbb{R}^+\}$ is tight. Then there exists a $\mu \in \mathcal{M}_1(E)$ such that $P_t^*\mu = \mu$ for all $t \in \mathbb{R}^+$, i.e. μ is an invariant measure for $(P_s)_{s\in\mathbb{R}^+}$, and μ is in the closed (w.r.t. to weak convergence) convex hull of $\{P_t^*\eta : t \in \mathbb{R}^+\}$.*

Using this result we obtain the following sufficient criterion for the existence of a stationary MUCOGARCH volatility process Y. Of course, this immediately translates to stationarity of V. Moreover, for $d = 1$ it recovers the necessary and sufficient stationarity condition of Klüppelberg et al. (2004).

Theorem 6.3.25. *Let $C \in \mathbb{S}_d^{++}$ and $B \in M_d(\mathbb{R})$ be diagonalizable with $S \in GL_d(\mathbb{C})$ such that $S^{-1}BS$ is diagonal. Furthermore, let L be a d-dimensional Lévy process with non-zero Lévy measure, $\alpha_1 := \|S\|_2^2\|S^{-1}\|_2^2 K_{2,B}\|A \otimes A\|_{B,S}$ and λ be defined as in Theorem 6.3.16. Assume that*

$$\int_{\mathbb{R}^d} \log\left(1 + \alpha_1\|\mathrm{vec}(yy^*)\|_{B,S}\right)\nu_L(dy) < -2\lambda. \tag{6.3.11}$$

Then there exists a stationary distribution $\mu \in \mathcal{M}_1(\mathbb{S}_d^+)$ for the MUCOGARCH(1,1) volatility process Y such that

$$\int_{\mathbb{R}^d} \left(\left(1 + \alpha_1\|\mathrm{vec}(yy^*)\|_{B,S}\right)^k - 1\right)\nu_L(dy) < -2\lambda k. \tag{6.3.12}$$

for some $k \in \mathbb{N}$ implies that $\int_{\mathbb{S}_d^+} \|x\|^k \mu(dx) < \infty$, i.e. that the k-th moment of the stationary distribution is finite.

Proof. Let $\lambda, \tilde{L}$ be defined as in Theorem 6.3.16 and $\bar{L}$ be the Lévy process constructed in Lemma 6.3.14 such that $\tilde{L}_t = [\bar{L}, \bar{L}]_t^{\mathfrak{d}}$. Then $\int_{\mathbb{R}^d} \log\left(1 + \alpha_1\|\mathrm{vec}(yy^*)\|_{B,S}\right)\nu_L(dy) =$

$\int_{\mathbb{R}^d} \log\left(1+\alpha_1 y^2\right) \nu_{\bar{L}}(dy)$ and thus Brockwell et al. (2006, Theorem 3.1) (see also Klüppelberg et al. (2004, Theorem 3.1)) show that the process y satisfying (6.3.8) converges in distribution to a distribution concentrated on $\mathbb{R}^+$. Assume now that y_0 has this stationary probability distribution and is independent of $(L_s)_{s\in\mathbb{R}^+}$. Setting $Y_0 = \frac{y_0}{\|\mathrm{vec}(I_d)\|_{B,S}} I_d$ gives an initial value for the MUCOGARCH volatility process that is independent of L and, moreover, $\|\mathrm{vec}(Y_0)\|_{B,S} = y_0$. Thus the process y satisfying $\|\mathrm{vec}(Y_t)\|_{B,S} \leq y_t$ for all $t \in \mathbb{R}^+$ (cf. Theorems 6.3.16, 6.3.19) is stationary. Since for every $K > 0$ the set $\{x \in \mathbb{S}_d : \|x\| \leq K\}$ is compact in $\mathbb{S}_d^+$, $P(\|Y_t\|_{B,S} \leq K) \geq P(y_t \leq K)$ and y is stationary with a stationary distribution concentrated on $\mathbb{R}^+$, it follows that the set $\{\mathscr{L}(Y_t) : t \in \mathbb{R}^+\}$ of laws $\mathscr{L}(Y_t)$ of Y_t forms a tight subset of $\mathcal{M}_1(\mathbb{S}_d^+)$. Therefore Theorem 6.3.25 combined with Theorem 6.3.22 implies that there exists a stationary distribution $\mu \in \mathcal{M}_1(\mathbb{S}_d^+)$ for the MUCOGARCH volatility process Y such that μ is in the closed convex hull of $\{\mathscr{L}(Y_t) : t \in \mathbb{R}^+\}$.

If (6.3.12) holds for some $k \in \mathbb{N}$, Brockwell et al. (2006, Proposition 4.1) (cf. also Klüppelberg et al. (2004, Section 4)) shows that the stationary distribution of y has a finite k-th moment. This in turn implies that $E(\|Y_t\|^k) \leq c$ for some finite $c \in \mathbb{R}^+$ and all $t \in \mathbb{R}^+$. Hence, $\int_{\mathbb{S}_d^+} \|x\|^k \mu(dx) < \infty$, because μ is in the closed convex hull of $\{\mathscr{L}(Y_t) : t \in \mathbb{R}^+\}$. □

Remark 6.3.26. *(a) From Klüppelberg et al. (2004, Lemma 4.1 (d)) it follows that, if* (6.3.12) *is satisfied for $k \in \mathbb{N}$, then it is satisfied for all $\tilde{k} \in \mathbb{N}$, $\tilde{k} \leq k$ as well.*

(b) Combining results of Section 6.3.2 shows that $\alpha_1 = \|A\|_2^2$ and $\|\cdot\|_{B,S} = \|\cdot\|_2$ if B is normal.

The following result implies that when $C = cI_d$ for some $c \in \mathbb{R}^{++}$ then a change in c affects the stationary distribution of V in a simple multiplicative way. This result is also new in the univariate case, although the proof is straightforward and thus left out.

Proposition 6.3.27. *Let $c \in \mathbb{R}^{++}$, $A, B \in M_d(\mathbb{R})$ and L be a d-dimensional Lévy process. If V satisfies*

$$dV_t = (B(V_{t-} - cI_d) + (V_{t-} - cI_d)B^*)\,dt + AV_{t-}^{1/2} d[L,L]_t^{\mathfrak{d}} V_{t-}^{1/2} A^* \tag{6.3.13}$$

then Z defined by $Z_t = V_t/c$ satisfies

$$dZ_t = (B(Z_{t-} - I_d) + (Z_{t-} - I_d)B^*)\,dt + AZ_{t-}^{1/2} d[L,L]_t^{\mathfrak{d}} Z_{t-}^{1/2} A^*, \tag{6.3.14}$$

which does not depend on c.

In particular, if $\mu \in \mathcal{M}_1(\mathbb{S}_d^+)$ is a stationary distribution for (6.3.14)*, then $\mu_c \in \mathcal{M}_1(\mathbb{S}_d^+)$ defined by $\mu_c(W) = \mu(W/c)$ for all Borel sets $W \subset \mathbb{S}_d^+$ is a stationary distribution for* (6.3.13)*.*

Establishing uniqueness of the stationary distribution and convergence to the stationary distribution for arbitrary starting values appears to be a rather intricate question due to the Lipschitz property holding only locally and the fact that $d[L,L]^{\mathfrak{d}}$ lives on the rank one matrices. However, in the next section we obtain asymptotic second order stationarity and that the stationary second order structure is unique under some technical conditions.

Regarding the relation to the stationarity of univariate COGARCH processes let us consider the following example. Assume that A, B, C and Y_0 are diagonal, the components of Y_0 are independent and that the components of the Lévy process are completely independent, i.e. whenever L has a jump then only one of the d components jumps. In this case it is

easy the see that Y or V, respectively, consists of d independent univariate COGARCH(1,1) processes. It is clear that if each of the d univariate COGARCH(1,1) processes converges in distribution to a stationary distribution then Y or V, respectively, converges in distribution to a stationary distribution. In this example Condition (6.3.11) can be shown to imply the necessary and sufficient stationarity condition of Klüppelberg et al. (2004, Theorem 3.1) for all components simultaneously. Actually, Condition (6.3.11) is easily seen to be stronger than requiring that the univariate stationarity condition be satisfied for all components.

However, it should be noted that the picture is very different when Y_0 is not diagonal, because then jumps in one component of L may affect all components of Y, as can easily be seen. Hence, it is not clear if one still has convergence to a stationary distribution and whether this has to be the same distribution as the limit distribution when Y_0 is diagonal. When we have that Y is asymptotically second order stationary (cf. the upcoming Definition 6.3.41) and the limiting distribution for a diagonal Y_0 has finite second moments, the off-diagonal (covariance) elements of Y or V, respectively, necessarily converge to zero in L^2 as $t \to \infty$.

Another degeneracy occurs in the following example. Take $d = 2$, $A = \alpha I_2$, $B = -\beta I_2$ with $\alpha, \beta \in \mathbb{R}^{++}$ and $C = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$. If one has that $Y_0 = yC$ with $y \in \mathbb{R}^+$ (possibly random) then Y is at all times a scalar multiple of C and when L jumps all components (variance and covariance ones) have a jump of the same height. However, one again has a completely different picture if Y_0 is chosen differently, e.g. Y is in $\mathbb{S}_d^{++}$ at all times, provided $Y_0 \in \mathbb{S}_d^{++}$.

6.3.4. Second order properties of the volatility process

Assuming stationarity and the existence of the relevant moments of the stationary solution we calculate explicit expressions for the moments of a stationary MUCOGARCH(1,1) volatility process in this section, treat also the non-stationary case along the way and present results regarding (asymptotic) second order stationarity. Due to the special structure of the stochastic differential equation (6.3.4), especially due to the presence of the matrix square root, it is only under certain assumptions on the Lévy process possible to obtain explicit formulae. Note, however that the assumptions made are natural and that similar ones are typically made in the discrete time GARCH literature as well (cf., in particular, Hafner (2003)).

Henceforth we often assume the following in this section.

Assumption 6.1. *$(Y_t)_{t\in\mathbb{R}^+}$ is a second order stationary MUCOGARCH(1,1) volatility process.*

Assumption 6.2. *The pure jump part of the driving Lévy process $(L_t)_{t\in\mathbb{R}^+}$ has finite variance which is a scalar multiple of the identity: let $L_t^{\mathfrak{d}} := \int_0^t \int_{\|x\|\leq 1} x(\mu_L(ds,dx) - ds\nu_L(dx)) + \int_0^t \int_{\|x\|>1} x\mu_L(ds,dx)$ denote the pure jump part of L, then this means that there exists a $\sigma_L \in \mathbb{R}^+$ such that $\operatorname{var}(L_1^{\mathfrak{d}}) = \int_{\mathbb{R}^d} xx^*\nu_L(dx) = \sigma_L I_d$.*

We start the calculation of the moments of Y with the expected value of the volatility noting that some elementary auxiliary results used are to be found in Appendix 6.7.3 of this chapter.

Theorem 6.3.28. *Assume that Assumption 6.2 holds.*

(i) If the MUCOGARCH volatility process Y has a finite first moment for all $t \in \mathbb{R}^+$, i.e. $E(\|Y_t\|) < \infty \, \forall \, t \in \mathbb{R}^+$, and $t \mapsto E(\|Y_t\|)$ is locally bounded, then

$$E(\mathrm{vec}(Y_t)) = e^{\mathscr{B}t} E(\mathrm{vec}(Y_0)) + \int_0^t e^{\mathscr{B}(t-s)} ds \, \sigma_L (A \otimes A) \mathrm{vec}(C)$$

with $\mathscr{B} := B \otimes I_d + I_d \otimes B + \sigma_L A \otimes A$. If $\mathscr{B}$ is invertible, then

$$\begin{aligned} E(\mathrm{vec}(Y_t)) =& e^{\mathscr{B}t} \left(E(\mathrm{vec}(Y_0)) + \sigma_L \mathscr{B}^{-1} (A \otimes A) \mathrm{vec}(C) \right) \\ & - \sigma_L \mathscr{B}^{-1} (A \otimes A) \mathrm{vec}(C) \, \forall \, t \in \mathbb{R}^+. \end{aligned} \tag{6.3.15}$$

(ii) Under Assumption 6.1 the stationary expected value $E(Y_0)$ of the MUCOGARCH volatility process satisfies

$$BE(Y_0) + E(Y_0)B^* + \sigma_L A E(Y_0) A^* = -\sigma_L A C A^*. \tag{6.3.16}$$

If $\mathscr{B}$ is invertible, the following formulae hold

$$\begin{aligned} & E(\mathrm{vec}(Y_0)) = -\sigma_L \mathscr{B}^{-1} (A \otimes A) \mathrm{vec}(C) \text{ and} \\ & E(\mathrm{vec}(V_0)) = E(\mathrm{vec}(Y_0)) + \mathrm{vec}(C) = \mathscr{B}^{-1} \left(B \otimes I_d + I_d \otimes B \right) \mathrm{vec}(C). \end{aligned} \tag{6.3.17}$$

Proof. From the defining stochastic differential equation (6.3.4) we have

$$Y_t = Y_0 + \int_0^t (BY_{s-} + Y_{s-}B^*) ds + \int_0^t A(Y_{s-} + C)^{1/2} d[L,L]_s^{\mathfrak{d}} (Y_{s-} + C)^{1/2} A^*.$$

Therefore

$$E(Y_t) = E(Y_0) + \int_0^t (BE(Y_s) + E(Y_s)B^*) ds + \sigma_L \int_0^t AE(Y_s + C) A^* ds \tag{6.3.18}$$

using a Fubini argument, Lemmata 6.7.20 and 6.7.21 and observing that

$$E([L,L]_1^{\mathfrak{d}}) = \int_{\mathbb{R}^d} xx^* \nu(dx) = \mathrm{var}(L_1^{\mathfrak{d}}) \tag{6.3.19}$$

is implied by equations (6.2.13) and (6.2.20). Thus

$$\begin{aligned} & \mathrm{vec} \left(E \left(\int_0^t A(Y_{s-} + C)^{1/2} d[L,L]_s^{\mathfrak{d}} (Y_{s-} + C)^{1/2} A^* \right) \right) \\ & \quad = E \left(\int_0^t (A \otimes A) \left((Y_{s-} + C)^{1/2} \otimes (Y_{s-} + C)^{1/2} \right) d\mathrm{vec}([L,L]_s^{\mathfrak{d}}) \right) \\ & \quad = \int_0^t (A \otimes A) E \left((Y_{s-} + C)^{1/2} \otimes (Y_{s-} + C)^{1/2} \right) \mathrm{vec}(E([L,L]_1^{\mathfrak{d}})) ds \\ & \quad = \sigma_L \int_0^t (A \otimes A) E \left((Y_{s-} + C)^{1/2} \otimes (Y_{s-} + C)^{1/2} \mathrm{vec}(I_d) \right) ds \\ & \quad = \sigma_L \mathrm{vec} \left(\int_0^t AE \left(Y_{s-} + C \right) A^* ds \right). \end{aligned}$$

Equation (6.3.18) therefore implies the following differential equation after vectorizing:

$$\frac{d}{dt}E(\mathrm{vec}(Y_t)) = \mathscr{B}E(\mathrm{vec}(Y_t)) + \sigma_L(A \otimes A)\mathrm{vec}(C).$$

Solving this ODE establishes (i).

Turning to (ii) the assumed second order stationarity and (6.3.18) immediately imply

$$BE(Y_0) + E(Y_0)B^* + \sigma_L A(E(Y_0) + C)A^* = 0.$$

The rest is just rewriting this linear equation. □

Remark 6.3.29. *(a) Observe that the stationary expectation is the limit of the expected value in (i) for $t \to \infty$ provided $\sigma(\mathscr{B}) \subset (-\infty, 0) + i\mathbb{R}$.*

(b) Of course, for $d = 1$ the formulae agree (after the relevant transformations and substitutions) with those of the univariate COGARCH(1,1) (see especially Brockwell et al. (2006)).

(c) The existence of a positive semi-definite solution of (6.3.16) *is ensured by our assumptions. Results from linear algebra (Ran and Reurings (2002)) imply that* (6.3.16) *has a positive definite solution, if there is an $X \in \mathbb{S}_d^{++}$ such that $-BX - XB^* - \sigma_L AXA^*$ is positive definite.*

Theorem 6.3.25 cannot only be used to show that Assumption 6.1 is satisfied, but also to ensure the invertibility of $\mathscr{B}$.

Lemma 6.3.30. *Assume that* (6.3.12) *is satisfied with $k = 1$ for the MUCOGARCH volatility process Y and that Assumption 6.2 holds. Then $\mathscr{B}$ as defined in Theorem 6.3.28 is invertible and $\sigma(\mathscr{B}) \subset (-\infty, 0) + i\mathbb{R}$.*

Proof. We have from Assumption 6.2 that

$$\sigma_L\|\mathrm{vec}(I_d)\|_{B,S} = \left\|\int_{\mathbb{R}^d} \mathrm{vec}(xx^*)\nu_L(dx)\right\|_{B,S} \leq \int_{\mathbb{R}^d} \|\mathrm{vec}(xx^*)\|_{B,S}\nu_L(dx). \tag{6.3.20}$$

For $k = 1$ condition (6.3.12) becomes

$$\|S\|_2^2\|S^{-1}\|_2^2 K_{2,B}\|A \otimes A\|_{B,S} \int_{\mathbb{R}^d} \|\mathrm{vec}(xx^*)\|_{B,S}\nu_L(dx) < -2\lambda.$$

Using (6.3.20), $\|S\|_2\|S^{-1}\|_2 \geq 1$ and that $K_{2,B}\|\mathrm{vec}(I_d)\|_{B,S} \geq \|I_d\|_2 = 1$ due the definition of $K_{2,B}$, one obtains

$$\sigma_L\|A \otimes A\|_{B,S} = \sigma_L\|(S^{-1} \otimes S^{-1})(A \otimes A)(S \otimes S)\|_2 < -2\lambda.$$

Let μ now be any eigenvalue of $\mathscr{B}$ and note that $(S^{-1} \otimes S^{-1})(B \otimes I_d + I_d \otimes B)(S \otimes S)$ is diagonal. Thus, the Bauer-Fike theorem (see Horn and Johnson (1985, Theorem 6.3.2 and its proof), for instance) gives that there exists a $\tilde{\mu} \in \sigma(B \otimes I_d + I_d \otimes B)$ such that

$$\begin{aligned}|\Re(\mu) - \Re(\tilde{\mu})| \leq |\mu - \tilde{\mu}| &\leq \|(S^{-1} \otimes S^{-1})(\mathscr{B} - B \otimes I_d - I_d \otimes B)(S \otimes S)\| \\ &= \sigma_L\|(S^{-1} \otimes S^{-1})(A \otimes A)(S \otimes S)\|_2 < -2\lambda.\end{aligned}$$

Hence, $\Re(\mu) < \max\{\Re(\tilde{\mu}) : \tilde{\mu} \in \sigma(B \otimes I_d + I_d \otimes B)\} - 2\lambda = 0$, because the maximum equals 2λ due to $\sigma(B \otimes I_d + I_d \otimes B) = \sigma(B) + \sigma(B)$ and the definition of $\lambda = \max(\Re(\sigma(B)))$. Therefore $\sigma(\mathscr{B}) \subset (-\infty, 0) + i\mathbb{R}$ and $\mathscr{B}$ is invertible. □

Before analysing the variance, let us study the autocovariance function.

Definition 6.3.31. *Let $(X_t)_{t\in\mathbb{T}}$ (with $\mathbb{T}$ being either $\mathbb{N}_0$ or $\mathbb{R}^+$) be a second order stationary process with values in $\mathbb{R}^d$. Then the autocovariance function $\operatorname{acov}_X : \mathbb{T}\cup(-\mathbb{T}) \mapsto M_d(\mathbb{R})$ of X is given by $\operatorname{acov}_X(h) = \operatorname{cov}(X_h, X_0) = E(X_h X_0^*) - E(X_0)E(X_0)^*$ for $h \geq 0$ and by $\operatorname{acov}_X(h) = (\operatorname{acov}_X(-h))^*$ for $h < 0$.*

If $(X_t)_{t\in\mathbb{T}}$ is a second order stationary process with values in $M_d(\mathbb{R})$ (or $\mathbb{S}_d$) then we set $\operatorname{acov}_X := \operatorname{acov}_{\operatorname{vec}(X)}$.

Theorem 6.3.32. *(i) Under Assumptions 6.1 and 6.2 the autocovariance function of the MUCOGARCH volatility process satisfies*

$$\frac{d}{dh}\operatorname{acov}_Y(h) = (B\otimes I_d + I_d\otimes B + \sigma_L A\otimes A)\operatorname{acov}_Y(h) \tag{6.3.21}$$

for $h \geq 0$. Hence,

$$\operatorname{acov}_Y(h) = \operatorname{acov}_V(h) = e^{(B\otimes I_d + I_d\otimes B + \sigma_L A\otimes A)h}\operatorname{var}(\operatorname{vec}(Y_0)),\ h \geq 0. \tag{6.3.22}$$

(ii) If Assumption 6.2 is satisfied and $E(\|Y_t\|^2)$ is finite for all $t \in \mathbb{R}^+$ and $t \mapsto E(\|Y_t\|^2)$ is locally bounded, it holds that

$$\operatorname{cov}(Y_{u+h}, Y_u) = \operatorname{cov}(V_{u+h}, V_u) = e^{(B\otimes I_d + I_d\otimes B + \sigma_L A\otimes A)h}\operatorname{var}(\operatorname{vec}(Y_u)) \tag{6.3.23}$$

for all $u, h \geq 0$.

Proof. We only proof (i), because the proof of (ii) proceeds along the same lines

The equality $\operatorname{acov}_Y(\cdot) = \operatorname{acov}_V(\cdot)$ is obvious. Due to the second order stationarity we have

$$\begin{aligned}
\operatorname{acov}_Y(h) =& \operatorname{cov}\left(\operatorname{vec}\left(Y_0 + \int_0^h (BY_{s-} + Y_{s-}B^*)ds\right.\right.\\
&\left.\left. + \int_0^h A(Y_{s-}+C)^{1/2} d[L,L]_s^{\mathfrak{d}} (Y_{s-}+C)^{1/2}A^*\right), \operatorname{vec}(Y_0)\right)\\
=& \operatorname{var}(\operatorname{vec}(Y_0)) + E\left(\int_0^h (B\otimes I_d + I_d\otimes B)\operatorname{vec}(Y_{s-})\operatorname{vec}(Y_0)^* ds\right)\\
&- E\left(\int_0^t (B\otimes I_d + I_d\otimes B)\operatorname{vec}(Y_{s-})ds\right) E(\operatorname{vec}(Y_0)^*)\\
&+ E\left(\int_0^h (A\otimes A)\left((Y_{s-}+C)^{1/2}\otimes(Y_{s-}+C)^{1/2}\right) d\operatorname{vec}([L,L]_s^{\mathfrak{d}})\operatorname{vec}(Y_0)^*\right)\\
&- E\left(\int_0^h (A\otimes A)\left((Y_{s-}+C)^{1/2}\otimes(Y_{s-}+C)^{1/2}\right) d\operatorname{vec}([L,L]_s^{\mathfrak{d}})\right) E(\operatorname{vec}(Y_0)^*)\\
=& \operatorname{var}(\operatorname{vec}(Y_0)) + \int_0^h (B\otimes I_d + I_d\otimes B)E\left(\operatorname{vec}(Y_s)\operatorname{vec}(Y_0)^*\right) ds\\
&- \int_0^t (B\otimes I_d + I_d\otimes B)E\left(\operatorname{vec}(Y_s)\right) ds E(\operatorname{vec}(Y_0)^*)\\
&+ \sigma_L\int_0^h (A\otimes A)E\left(\left((Y_{s-}+C)^{1/2}\otimes(Y_{s-}+C)^{1/2}\right)\operatorname{vec}(I_d)\operatorname{vec}(Y_0)^*\right) ds\\
&- \sigma_L\int_0^h (A\otimes A)E\left(\left((Y_{s-}+C)^{1/2}\otimes(Y_{s-}+C)^{1/2}\right)\operatorname{vec}(I_d)\right) ds\, E(\operatorname{vec}(Y_0)^*)
\end{aligned}$$

$$
\begin{aligned}
=&\mathrm{var}(\mathrm{vec}(Y_0)) + \int_0^h (B\otimes I_d + I_d \otimes B)\mathrm{acov}_Y(s)ds \\
&+ \sigma_L \int_0^h (A\otimes A) E\left(\mathrm{vec}(Y_s + C)\mathrm{vec}(Y_0)^*\right) ds \\
&- \sigma_L \int_0^h (A\otimes A) E\left(\mathrm{vec}(Y_s + C)\right) E(\mathrm{vec}(Y_0)^*) ds \\
=&\mathrm{var}(\mathrm{vec}(Y_0)) + \int_0^h (B\otimes I_d + I_d \otimes B + \sigma_L A\otimes A)\mathrm{acov}_Y(s)ds
\end{aligned}
$$

where we used a Fubini argument, Lemmata 6.7.19 and 6.7.21 and $E([L,L]^{\mathfrak{d}}_1) = \sigma_L I_d$. Regarding the use of Lemma 6.7.21 we observe that $\|(Y_{s-}+C)^{1/2}\otimes(Y_{s-}+C)^{1/2}\|_2 = \|Y_{s-}+C\|_2$ and hence the required local boundedness is ensured by the second order stationarity of Y.

The ordinary differential equation (6.3.21) is now immediate and to conclude the proof it suffices to note that $\mathrm{acov}_Y(0) = \mathrm{var}(\mathrm{vec}(Y_0))$ and thus solving the ODE gives $\mathrm{acov}_Y(h) = \mathrm{acov}_V(h) = e^{(B\otimes I_d + I_d\otimes B + \sigma_L A\otimes A)h}\mathrm{var}(\mathrm{vec}(Y_0))$, $h \geq 0$. □

The autocovariance function of the volatility process Y is thus exponentially decreasing and it is also easy to see that there is an OU type process having the same second order structure. This corresponds to the univariate COGARCH(p,q) volatility process having the same second order structure as a certain CARMA$(q,p-1)$ process (cf. Brockwell et al. (2006, Section 4)). However, we are so far lacking an explicit expression for $\mathrm{var}(\mathrm{vec}(Y_0))$. Unfortunately, our Assumption 6.2 on the second moment of the driving Lévy process $L^{\mathfrak{d}}$ made so far seems not to be sufficient to obtain an explicit expression for the variance.

As we shall see, the discrete quadratic variation of the vectorized discrete quadratic variation of the driving Lévy process

$$
\begin{aligned}
[\mathrm{vec}([L,L]^{\mathfrak{d}}), \mathrm{vec}([L,L]^{\mathfrak{d}})]^{\mathfrak{d}}_t &= \sum_{0\leq s\leq t} \mathrm{vec}(\Delta L_s(\Delta L_s)^*)\mathrm{vec}(\Delta L_s(\Delta L_s)^*)^* \\
&= \int_0^t \int_{\mathbb{R}^d} \mathrm{vec}(xx^*)\mathrm{vec}(xx^*)^* \mu_L(ds,dx)
\end{aligned}
$$

which is again a pure jump Lévy process of finite variation will appear in our calculations of the second moment and we need it to have finite expectation and even to make specific assumptions on its expectation

$$
E([\mathrm{vec}([L,L]^{\mathfrak{d}}), \mathrm{vec}([L,L]^{\mathfrak{d}})]^{\mathfrak{d}}_1) = \int_{\mathbb{R}^d} \mathrm{vec}(xx^*)\mathrm{vec}(xx^*)^* \nu_L(dx). \tag{6.3.24}
$$

To see what are reasonable assumptions, let us assume for a moment that L is a d-dimensional compound Poisson process with rate one and the jump distribution being the d-dimensional standard normal distribution. This implies that $[L,L]^{\mathfrak{d}}$ is a compound Poisson process with rate one and the jump distribution being a Wishart distribution. Then denoting the d-dimensional normal distribution by $N(dx)$ and noting that $\mathrm{vec}(xx^*)\mathrm{vec}(xx^*)^* = (x\otimes x)(x^*\otimes x^*) = (xx^*)\otimes(xx^*)$ we have

$$
\begin{aligned}
E([\mathrm{vec}([L,L]^{\mathfrak{d}}), \mathrm{vec}([L,L]^{\mathfrak{d}})]^{\mathfrak{d}}_1) &= \int_{\mathbb{R}^d} (xx^*)\otimes(xx^*) N(dx) \\
&= I_{d^2} + K_d + \mathrm{vec}(I_d)\mathrm{vec}(I_d)^*
\end{aligned} \tag{6.3.25}
$$

from Magnus and Neudecker (1979, Theorem 4.1). Here $K_d \in M_{d^2}(\mathbb{R})$ denotes the commutation matrix which can be characterized by $K_d\text{vec}(A) = \text{vec}(A^*)$ for all $A \in M_d(\mathbb{R})$ (see Magnus and Neudecker (1979) for more details). This can be easily generalized to the following result:

Lemma 6.3.33. *Let L be a d-dimensional compound Poisson process with rate c and the jumps being distributed like $\sqrt{\epsilon}X$ where X is a d-dimensional standard normal random variable and ϵ is a random variable in $\mathbb{R}^+$ with finite variance and independent of X. Then*

$$E([\text{vec}([L,L]^{\mathfrak{d}}), \text{vec}([L,L])^{\mathfrak{d}}]_1^{\mathfrak{d}}) = cE(\epsilon^2)\left(I_{d^2} + K_d + \text{vec}(I_d)\text{vec}(I_d)^*\right). \tag{6.3.26}$$

Moving away from a Lévy process of finite activity, a similar result holds for the following variant of type G processes, a special kind of a normal mixture.

Definition 6.3.34 (Type $\widetilde{G}$)**.** *Let L be a d-dimensional Lévy process. If there exists an $\mathbb{R}^+$-valued infinitely divisible random variable ϵ independent of a d-dimensional standard normal random variable X such that $L_1 \stackrel{\mathscr{L}}{=} \sqrt{\epsilon}X$, then L is said to be of* type $\widetilde{G}$*. (Here $\stackrel{\mathscr{L}}{=}$ denotes equality in law.)*

We have chosen the term "type $\widetilde{G}$" above, because these processes correspond to a particular case of multG laws as defined in Barndorff-Nielsen and Pérez-Abreu (2002, Def. 3.1). Actually, many interesting Lévy processes are of type $\widetilde{G}$, for instance, the multivariate symmetric GH (NIG) ones with the parameter Σ set to I_d (cf. Blæsild and Jensen (1981) or Prause (1999)). For details on distributions/Lévy processes of type G in general we refer to Barndorff-Nielsen and Pérez-Abreu (2002), Maejima and Rosiński (2002) and the references therein.

Lemma 6.3.35. *Let L be a d-dimensional Lévy process of type $\widetilde{G}$ with a finite fourth moment. Then $E([\text{vec}([L,L]^{\mathfrak{d}}), \text{vec}([L,L]^{\mathfrak{d}})]_1^{\mathfrak{d}}) = \rho_L\left(I_{d^2} + K_d + \text{vec}(I_d)\text{vec}(I_d)^*\right)$ with $\rho_L \in \mathbb{R}^+$.*

Proof. Let ϵ be as in the definition of type $\widetilde{G}$ and ν_ϵ its Lévy measure. Then by Barndorff-Nielsen and Pérez-Abreu (2002, Prop. 3.1) L has Lévy density $u(x) = \int_{\mathbb{R}^+} \phi_d(x;\tau I_d)\nu_\epsilon(d\tau)$ denoting the density of the d-dimensional normal distribution with variance Σ by $\phi_d(\cdot;\Sigma)$. Hence,

$$\begin{aligned}
E([\text{vec}([L,L]^{\mathfrak{d}}), \text{vec}([L,L]^{\mathfrak{d}})]_1^{\mathfrak{d}}) &= \int_{\mathbb{R}^d} (xx^*) \otimes (xx^*) u(x)dx \\
&= \int_{\mathbb{R}^+}\int_{\mathbb{R}^d} (xx^*) \otimes (xx^*)\phi_d(x;\tau I_d)dx\nu_\epsilon(d\tau) \\
&= \int_{\mathbb{R}^+} \tau^2\nu_\epsilon(d\tau)\left(I_{d^2} + K_d + \text{vec}(I_d)\text{vec}(I_d)^*\right)
\end{aligned}$$

using Magnus and Neudecker (1979, Th. 4.3). Now set $\rho_L := \int_{\mathbb{R}^+} \tau^2\nu_\epsilon(d\tau)$ and note that the finiteness follows from the definition of type $\widetilde{G}$ and the assumed finiteness of the fourth moment of L. □

These results motivate the following assumption.

Assumption 6.3. *The pure jump part of the driving Lévy process $(L_t)_{t\in\mathbb{R}^+}$ has a finite fourth moment, i.e. $\int_{\mathbb{R}^d}\|x\|^4\nu_L(dx)<\infty$, and there is a real constants ρ_L such that $E([\mathrm{vec}([L,L]^{\mathfrak{d}}),\mathrm{vec}([L,L]^{\mathfrak{d}})]_1^{\mathfrak{d}})=\rho_L\left(I_{d^2}+K_d+\mathrm{vec}(I_d)\mathrm{vec}(I_d)^*\right)$.*

To state our next result, we need to introduce some additional special linear operators and matrices. Define

$$\begin{aligned}&\mathbf{Q}: M_{d^2}\to M_{d^2},\\ &(\mathbf{Q}X)_{(k-1)d+l,(p-1)d+q}=X_{(k-1)d+p,(l-1)d+q}\text{ for all }\ k,l,p,q=\{1,2,\dots d\},\end{aligned}\tag{6.3.27}$$

then $\mathbf{Q}^{-1}=\mathbf{Q}$ obviously and $\mathbf{Q}(\mathrm{vec}(X)\mathrm{vec}(Z)^*)=X\otimes Z$ for all $X,Z\in\mathbb{S}_d$ (cf. Theorem 5.4.4 in Chapter 5). Furthermore, we define $\mathcal{Q}\in M_{d^4}(\mathbb{R})$ as the matrix associated to the linear map $\mathrm{vec}\circ\mathbf{Q}\circ\mathrm{vec}^{-1}$ on $\mathbb{R}^{d^4}$ and $\mathcal{K}_d\in M_d(\mathbb{R})$ as the matrix associated to the linear map $\mathrm{vec}\left(K_d\mathrm{vec}^{-1}(x)\right)$ for $x\in\mathbb{R}^{d^4}$ where $\mathrm{vec}: M_{d^2}(\mathbb{R})\to\mathbb{R}^{d^4}$. It is easy to see that both $\mathcal{Q}$ and $\mathcal{K}_d$ simply permute the entries of a vector $x\in\mathbb{R}^{d^4}$. Since thus both $\mathcal{Q}$ and $\mathcal{K}_d$ are permutation matrices, we have $\|\mathcal{Q}\|_2=\|\mathcal{K}_d\|_2=1$ where $\|\cdot\|_2$ is the operator norm associated with the usual Euclidean norm on $\mathbb{R}^{d^4}$.

Theorem 6.3.36. *Assume that Assumptions 6.2 and 6.3 hold.*

(i) If the MUCOGARCH process Y has a finite second moment for all $t\in\mathbb{R}^+$, i.e. $E(\|Y_t\|^2)<\infty\ \forall\, t\in\mathbb{R}^+$, and $t\mapsto E(\|Y_t\|^2)$ is locally bounded, then

$$\begin{aligned}\frac{d}{dt}\mathrm{vec}(E(\mathrm{vec}(Y_t)\mathrm{vec}(Y_t)^*))&=\frac{d}{dt}E(\mathrm{vec}(Y_t)\otimes\mathrm{vec}(Y_t))\\ &=\mathscr{C}\mathrm{vec}(E(\mathrm{vec}(Y_t)\mathrm{vec}(Y_t)^*))+(\sigma_L(A\otimes A)\otimes I_{d^2}+\mathscr{A}\mathcal{R})\,\mathrm{vec}(C)\otimes E(\mathrm{vec}(Y_t))\\ &\quad+(\sigma_L I_{d^2}\otimes(A\otimes A)+\mathscr{A}\mathcal{R})\,E(\mathrm{vec}(Y_t))\otimes\mathrm{vec}(C)+\mathscr{A}\mathcal{R}\mathrm{vec}(C)\otimes\mathrm{vec}(C)\end{aligned}\tag{6.3.28}$$

where

$$\mathscr{A}=(A\otimes A)\otimes(A\otimes A);\qquad \mathcal{R}=\rho_L\left(\mathcal{Q}+\mathcal{K}_d\mathcal{Q}+I_{d^4}\right)\tag{6.3.29}$$

$$\begin{aligned}\mathscr{C}:=&(B\otimes I_d+I_d\otimes B)\otimes I_{d^2}+I_{d^2}\otimes(B\otimes I_d+I_d\otimes B)\\ &+\sigma_L\left((A\otimes A)\otimes I_{d^2}+I_{d^2}\otimes(A\otimes A)\right)+\mathscr{A}\mathcal{R}.\end{aligned}\tag{6.3.30}$$

(ii) Under Assumption 6.1 the stationary second moment $E(\mathrm{vec}(Y_0)\mathrm{vec}(Y_0)^*)$ of the MUCOGARCH volatility process satisfies

$$\begin{aligned}&\mathscr{B}E(\mathrm{vec}(Y_0)\mathrm{vec}(Y_0)^*)+E(\mathrm{vec}(Y_0)\mathrm{vec}(Y_0)^*)\mathscr{B}^*\\ &\quad+(A\otimes A)\mathbf{R}E(\mathrm{vec}(Y_0)\mathrm{vec}(Y_0)^*)(A^*\otimes A^*)\\ &=-\sigma_L\left[(A\otimes A)\mathrm{vec}(C)E(\mathrm{vec}(Y_0))^*+E(\mathrm{vec}(Y_0))\mathrm{vec}(C)^*(A^*\otimes A^*)\right]\\ &\quad-(A\otimes A)\mathbf{R}\left(E(\mathrm{vec}(Y_0))\mathrm{vec}(C)^*+\mathrm{vec}(C)E(\mathrm{vec}(Y_0))^*+\mathrm{vec}(C)\mathrm{vec}(C)^*\right)(A^*\otimes A^*)\end{aligned}\tag{6.3.31}$$

with $\mathbf{R}:=\rho_L(\mathbf{Q}+K_d\mathbf{Q}+I_{d^2})$.

Provided $\mathscr{C}$ is invertible, $\mathrm{vec}(E(\mathrm{vec}(Y_0)\mathrm{vec}(Y_0)^*))$ *is given by*

$$\begin{aligned}\mathrm{vec}(E(\mathrm{vec}(Y_0)\mathrm{vec}(Y_0)^*))=-\mathscr{C}^{-1}\big[&\mathscr{A}\mathcal{R}(\mathrm{vec}(C)\otimes\mathrm{vec}(C))\\ &+(\sigma_L(A\otimes A)\otimes I_{d^2}+\mathscr{A}\mathcal{R})\,\mathrm{vec}(C)\otimes E(\mathrm{vec}(Y_0))\\ &+(\sigma_L I_{d^2}\otimes(A\otimes A)+\mathscr{A}\mathcal{R})\,E(\mathrm{vec}(Y_0))\otimes\mathrm{vec}(C)\big].\end{aligned}\tag{6.3.32}$$

Proof. From the definition of the quadratic variation (cf. Protter (2004, Section 2.6) and Lemma 4.5.11 for a special version in the context of matrix/vector multiplication) it follows that

$$\begin{aligned}
\operatorname{vec}(Y_t)\operatorname{vec}(Y_t)^* &= \operatorname{vec}(Y_0)\operatorname{vec}(Y_0)^* \\
&+ \int_0^t \Big((B \otimes I_d + I_d \otimes B) \operatorname{vec}(Y_{s-}) \operatorname{vec}(Y_{s-})^* \\
&+ \operatorname{vec}(Y_{s-}) \operatorname{vec}(Y_{s-})^* (B^* \otimes I_d + I_d \otimes B^*) \Big) ds \\
&+ \int_0^t (A \otimes A) \left((C + Y_{s-})^{1/2} \otimes (C + Y_{s-})^{1/2} \right) d\operatorname{vec}([L,L]_s^{\mathfrak{d}}) \operatorname{vec}(Y_{s-})^* \\
&+ \int_0^t \operatorname{vec}(Y_{s-}) d\operatorname{vec}([L,L]_s^{\mathfrak{d}})^* \left((C + Y_{s-})^{1/2} \otimes (C + Y_{s-})^{1/2} \right) (A^* \otimes A^*) \\
&+ [\operatorname{vec}(Y), \operatorname{vec}(Y)]_t.
\end{aligned}$$

Moreover, since $\operatorname{vec}(Y_t)$ is the sum of an absolutely continuous component and a pure jump process of finite variation, setting $\mathcal{V}_t = \left((C + Y_t)^{1/2} \otimes (C + Y_t)^{1/2}\right)$

$$\begin{aligned}
[\operatorname{vec}(Y), \operatorname{vec}(Y)^*]_t &= \int_0^t (A \otimes A) \mathcal{V}_{s-} d \left([\operatorname{vec}([L,L]^{\mathfrak{d}}), \operatorname{vec}([L,L]^{\mathfrak{d}})]_s^{\mathfrak{d}} \right) \mathcal{V}_{s-} (A^* \otimes A^*) \\
&= \int_0^t \int_{\mathbb{R}^d} (A \otimes A) \mathcal{V}_{s-} \operatorname{vec}(xx^*) \operatorname{vec}(xx^*)^* \mathcal{V}_{s-} (A^* \otimes A^*) \mu_L(ds, dx).
\end{aligned}$$

Using a Fubini argument, obvious variants of Lemmata 6.7.18 to 6.7.21 and the made Assumptions 6.2, 6.3 on the moments of ν_L we obtain

$$\begin{aligned}
E(\operatorname{vec}(Y_t)\operatorname{vec}(Y_t)^*) &= E(\operatorname{vec}(Y_0)\operatorname{vec}(Y_0)^*) \\
&+ \int_0^t \Big((B \otimes I_d + I_d \otimes B) E(\operatorname{vec}(Y_s)\operatorname{vec}(Y_s)^*) \\
&+ E(\operatorname{vec}(Y_s)\operatorname{vec}(Y_s)^*)(B^* \otimes I_d + I_d \otimes B^*) \Big) ds \\
&+ \sigma_L \int_0^t (A \otimes A) E(\operatorname{vec}(C + Y_s)\operatorname{vec}(Y_s)^*) ds \\
&+ \sigma_L \int_0^t E(\operatorname{vec}(Y_s)\operatorname{vec}(C + Y_s)^*)(A^* \otimes A^*) ds \\
&+ \int_0^t (A \otimes A) E\left(\mathcal{V}_{s-} \rho_L \left(I_{d^2} + K_d + \operatorname{vec}(I_d)\operatorname{vec}(I_d)^* \right) \mathcal{V}_{s-} \right) (A^* \otimes A^*) ds.
\end{aligned}$$

With the definition of $\mathcal{V}_t$ it follows that

$$\begin{aligned}
&E(\mathcal{V}_{s-} I_{d^2} \mathcal{V}_{s-}) = E(\mathcal{V}_s^2) = E((C + Y_s) \otimes (C + Y_{s-})) = \mathbf{Q} E(\operatorname{vec}(C + Y_s)\operatorname{vec}(C + Y_s)^*) \\
&E(\mathcal{V}_{s-} \operatorname{vec}(I_d)\operatorname{vec}(I_d)^* \mathcal{V}_{s-}) = E(\operatorname{vec}(C + Y_s)\operatorname{vec}(C + Y_s)^*) \\
&E(\mathcal{V}_{s-} K_d \mathcal{V}_{s-}) = K_d E((C + Y_s) \otimes (C + Y_s)) = K_d \mathbf{Q} E(\operatorname{vec}(C + Y_s)\operatorname{vec}(C + Y_s)^*)
\end{aligned}$$

using Magnus and Neudecker (1979, Theorem 3.1 (xii)) in the last identity. Inserting these formulae in the above result and noting that in the stationary case the integrands need to sum up to zero gives (6.3.31). Vectorizing then immediately establishes (6.3.32).

Likewise we obtain (6.3.28) in the non-stationary case by inserting the formulae above, vectorizing and differentiating. □

Remark 6.3.37. *The differential equation* (6.3.28) *is an inhomogeneous linear differential equation with constant coefficients. Hence, it is standard to obtain an explicit solution. We refrain from stating it, as the stationary case seems to be of most importance.*

Again condition (6.3.12) of Theorem 6.3.25 ensuring the existence of moments of the there obtained stationary distribution also implies invertibility of $\mathscr{C}$ under an additional technical assumption.

To state the result we set $\mathcal{S} = S \otimes S \otimes S \otimes S$ and define a new norm $\|\cdot\|_{\widetilde{B,S}}$ on $\mathbb{R}^{d^4}$ by setting $\|x\|_{\widetilde{B,S}} = \|\mathcal{S}^{-1}x\|_2$. The associated operator norm on $M_{d^4}(\mathbb{R})$ is given by $\|X\|_{\widetilde{B,S}} = \|\mathcal{S}^{-1}X\mathcal{S}\|_2$.

Lemma 6.3.38. *Assume that* (6.3.12) *is satisfied with* $k = 2$ *for the MUCOGARCH volatility process* Y *and that Assumptions 6.2 and 6.3 hold. Provided also*

$$\|\mathcal{Q} + \mathcal{K}_d\mathcal{Q} + I_{d^4}\|_{\widetilde{B,S}} \leq K_{2,B}^2 \|\mathrm{vec}(I_{d^2} + K_d + \mathrm{vec}(I_d)\mathrm{vec}(I_d)^*)\|_{\widetilde{B,S}} \tag{6.3.33}$$

holds, $\sigma(\mathscr{C}) \subset (-\infty, 0) + i\mathbb{R}$ *and* $\mathscr{C}$ *is invertible.*

Proof. We have

$$\begin{aligned}\rho_L\|\mathrm{vec}(I_{d^2} + K_d + \mathrm{vec}(I_d)\mathrm{vec}(I_d)^*)\|_{\widetilde{B,S}} &= \left\|\int_{\mathbb{R}^d} \mathrm{vec}((xx^*)\otimes(xx^*))\nu_L(dx)\right\|_{\widetilde{B,S}} \\ &\leq \int_{\mathbb{R}^d} \|\mathrm{vec}((xx^*)\otimes(xx^*))\|_{\widetilde{B,S}}\,\nu_L(dx) = \int_{\mathbb{R}^d} \|\mathrm{vec}(xx^*)\|_{B,S}^2\,\nu_L(dx), \end{aligned}\tag{6.3.34}$$

since the definition of $\|\cdot\|_{\widetilde{B,S}}$ implies $\|\mathrm{vec}((xx^*)\otimes(xx^*))\|_{\widetilde{B,S}} = \|\mathcal{S}^{-1}(x\otimes x)\otimes(x\otimes x)\|_2 = \|(S^{-1}\otimes S^{-1})(x\otimes x)\|_2^2 = \|\mathrm{vec}(xx^*)\|_{B,S}^2$ using that $\|z\otimes z\|_2 = \|z\|_2^2$ for all $z \in \mathbb{R}^{d^2}$.

For $k = 2$ Condition (6.3.12) becomes

$$\begin{aligned}&2\|S\|_2^2\|S^{-1}\|_2^2 K_{2,B}\|A\otimes A\|_{B,S}\int_{\mathbb{R}^d}\|\mathrm{vec}(xx^*)\|_{B,S}\nu_L(dx) \\ &\quad + \|S\|_2^4\|S^{-1}\|_2^4 K_{2,B}^2\|A\otimes A\|_{B,S}^2\int_{\mathbb{R}^d}\|\mathrm{vec}(xx^*)\|_{B,S}^2\nu_L(dx) < -4\lambda.\end{aligned}$$

Using (6.3.34) and results from the proof of Lemma 6.3.30 gives

$$2\sigma_L\|A\otimes A\|_{B,S} + K_{2,B}^2\|A\otimes A\|_{B,S}^2\rho_L\|\mathrm{vec}(I_{d^2} + K_d + \mathrm{vec}(I_d)\mathrm{vec}(I_d)^*)\|_{\widetilde{B,S}} < -4\lambda.$$

Combining $\|(A\otimes A)\otimes I_{d^2} + I_{d^2}\otimes(A\otimes A)\|_{\widetilde{B,S}} \leq 2\|A\otimes A\|_{B,S}$ and $\|A\otimes A\otimes A\otimes A\|_{\widetilde{B,S}} = \|A\otimes A\|_{B,S}^2$, which are elementary to prove, with (6.3.33) leads to

$$\|\mathscr{C} - \mathscr{B}\otimes I_{d^2} - I_{d^2}\otimes\mathscr{B}\|_{\widetilde{B,S}} < -4\lambda.$$

Since $\mathcal{S}^{-1}\left(\mathscr{B}\otimes I_{d^2} + I_{d^2}\otimes\mathscr{B}\right)\mathcal{S}$ is diagonal and

$$\max\left(\Re\left(\sigma\left(\mathscr{B}\otimes I_{d^2} + I_{d^2}\otimes\mathscr{B}\right)\right)\right) = 4\lambda,$$

the Bauer-Fike theorem and arguments as in the proof of Lemma 6.3.30 conclude. □

A rather unpleasant feature of this lemma is that we need the technical condition (6.3.33). The following lemma shows that it is always true if S is unitary and we believe that it is satisfied in (almost) all cases.

Lemma 6.3.39. *If S is unitary, then* (6.3.33) *is satisfied.*

Proof. That S is unitary implies that $K_{2,B} = 1$ and all the norms used are actually the Euclidean norm or the operator norm induced by it. Hence, we have to show

$$\|\mathcal{Q} + \mathcal{K}_d\mathcal{Q} + I_{d^4}\|_2 \leq \|\mathrm{vec}(I_{d^2} + K_d + \mathrm{vec}(I_d)\mathrm{vec}(I_d)^*)\|_2.$$

For $d = 1$ one calculates both sides to be equal to three.

In general we know from the fact that $\mathcal{K}_d$ and $\mathcal{Q}$ are permutation matrices that the operator norms are one. Hence, $\|\mathcal{Q} + \mathcal{K}_d\mathcal{Q} + I_{d^4}\|_2 \leq 3$. Furthermore, the entries of K_d and $\mathrm{vec}(I_d)\mathrm{vec}(I_d)^*$ are either one or zero. Therefore, $\|\mathrm{vec}(I_{d^2} + K_d + \mathrm{vec}(I_d)\mathrm{vec}(I_d)^*)\|_2 \geq \|\mathrm{vec}(I_{d^2})\|_2 = d$. This shows the inequality for $d \geq 3$.

In the remaining case $d = 2$ we have

$$\|\mathrm{vec}(I_{d^2} + K_d + \mathrm{vec}(I_d)\mathrm{vec}(I_d)^*)\|_2 \geq \|\mathrm{vec}(I_{d^2} + \mathrm{vec}(I_d)\mathrm{vec}(I_d)^*)\|_2 = \sqrt{12} > 3,$$

which again establishes the claimed inequality. □

We end our comprehensive calculations for the second order moment structure of the MUCOGARCH volatility process by turning to the stationary variance.

Corollary 6.3.40. *Assume that Assumptions 6.1, 6.2 and 6.3 hold and that $\mathscr{C}$ is invertible.*

Then the stationary variance $\mathrm{var}(\mathrm{vec}(Y_0)) = \mathrm{var}(\mathrm{vec}(V_0))$ *of the MUCOGARCH volatility process is given by*

$$\begin{aligned}\mathrm{vec}(\mathrm{var}(\mathrm{vec}(Y_0))) = -\mathscr{C}^{-1}\big[&\left(\sigma_L^2\mathscr{C}(\mathscr{B}^{-1}\otimes\mathscr{B}^{-1})\mathscr{A} + \mathscr{A}\mathcal{R}\right)(\mathrm{vec}(C)\otimes\mathrm{vec}(C))\\ &+ (\sigma_L(A\otimes A)\otimes I_{d^2} + \mathscr{A}\mathcal{R})\,\mathrm{vec}(C)\otimes E(\mathrm{vec}(Y_0))\\ &+ (\sigma_L I_{d^2}\otimes(A\otimes A) + \mathscr{A}\mathcal{R})\,E(\mathrm{vec}(Y_0))\otimes\mathrm{vec}(C)\big]. \qquad (6.3.35)\end{aligned}$$

Proof. Combine (6.3.32),

$$\mathrm{vec}(E(\mathrm{vec}(Y_0))E(\mathrm{vec}(Y_0)^*)) = \sigma_L^2(\mathscr{B}^{-1}\otimes\mathscr{B}^{-1})\mathscr{A}(\mathrm{vec}(C)\otimes\mathrm{vec}(C))$$

and the elementary formula $\mathrm{var}(\mathrm{vec}(Y_0)) = E(\mathrm{vec}(Y_0)\mathrm{vec}(Y_0)^*) - E(\mathrm{vec}(Y_0))\mathrm{vec}(Y_0)^*))^*$. □

At a first sight (6.3.35) is rather complicated. However, it can be easily implemented on a computer and for $d = 1$ it becomes formula (4.8) of Brockwell et al. (2006, Theorem 4.4) for $p = q = 1$, of course. Under specific moment assumptions on the driving Lévy process we have thus calculated the second order structure of a stationary MUCOGARCH volatility process completely.

Finally, we give conditions ensuring (asymptotic) second order stationarity starting with a precise definition.

Definition 6.3.41. *Let $X = (X_t)_{t\in\mathbb{R}^+}$ be a stochastic process in $\mathbb{R}^d$ with finite second moments. If there are $\mu \in \mathbb{R}^d$, $\Sigma \in \mathbb{S}_d^+$ and $f : \mathbb{R}^+ \to M_d(\mathbb{R})$ such that*

$$\lim_{t\to\infty} E(X_t) = \mu, \qquad \lim_{t\to\infty} \operatorname{var}(X_t) = \Sigma,$$
$$\lim_{t\to\infty} \sup_{h\in\mathbb{R}^+} \{\|\operatorname{cov}(X_{t+h}, X_t) - f(h))\|\} = 0,$$

then X is called asymptotically second order stationary *with mean μ, variance Σ and autocovariance function f.*

A stochastic process X in $M_d(\mathbb{R})$ or $\mathbb{S}_d$ is said to be asymptotically second order stationary if $\operatorname{vec}(X)$ *is asymptotically second order stationary.*

Theorem 6.3.42. *Let Assumptions 6.2 and 6.3 be satisfied. Assume further that B is diagonalizable and $\sigma(B), \sigma(\mathscr{B}), \sigma(\mathscr{C}) \subset (-\infty, 0) + i\mathbb{R}$.*

(i) If Y_0 satisfies (6.3.17) *and* (6.3.35)*, then the MUCOGARCH volatility process Y is second order stationary.*

(ii) If $E(\|Y_0\|^2) < \infty$, then the MUCOGARCH volatility process Y is asymptotically second order stationary with mean, variance and autocovariance function given by (6.3.17)*,* (6.3.35) *and* (6.3.22)*.*

Proof. (i) follows by combining Proposition 6.3.20 and Theorems 6.3.28, 6.3.32 and 6.3.36.

Regarding (ii) Proposition 6.3.20 ensures that $E(\|Y_t\|^2) < \infty$ for all $t \in \mathbb{R}^+$. The convergence of the expectation has already been noted in Remark 6.3.29, the convergence of the variance follows from (6.3.28) and the next lemma and then (6.3.23) immediately implies the convergence of the autocovariance. □

Above we need the following general lemma on differential equations.

Lemma 6.3.43. *Let $f : \mathbb{R}^+ \to \mathbb{R}^d$ be continuous and $A \in M_d(\mathbb{R})$ with $\sigma(A) \subset (-\infty, 0)+i\mathbb{R}$. If $\lim_{t\to\infty} f(t) = \xi$ with $\xi \in \mathbb{R}^d$ then for any initial value $x_0 \in \mathbb{R}^d$ the solution x to the differential equation*

$$\frac{dx(t)}{dt} = Ax(t) + f(t)$$

satisfies $\lim_{t\to\infty} x(t) = -A^{-1}\xi$.

Proof. It holds that

$$x(t) = e^{At}x_0 + \int_0^t e^{A(t-s)}f(s)ds.$$

Since

$$\lim_{t\to\infty} \int_0^t e^{A(t-s)}\xi ds = -A^{-1}\xi \text{ and } \lim_{t\to\infty} e^{At}x_0 = 0,$$

it suffices to show that $\lim_{t\to\infty} \left\| \int_0^t e^{A(t-s)}(f(s) - \xi)ds \right\| = 0$. Fix $\epsilon > 0$. There exist $t^*, t^{**} > 0$ with $t^* \le t^{**}$ such that $\|f(t) - \xi\| < \epsilon$ for all $t \ge t^*$ and

$$\left\| \int_0^{t^*} e^{A(t-s)}(f(s) - \xi)ds \right\| < \epsilon \text{ for all } t \ge t^{**}.$$

Hence,

$$\left\| \int_0^t e^{A(t-s)}(f(s)-\xi)ds \right\| \leq \left(1 + \int_{t^*}^t \|e^{A(t-s)}\|ds\right)\epsilon \leq \left(1 + \int_0^\infty \|e^{As}\|ds\right)\epsilon \; \forall\, t \geq t^{**}.$$

Since the last integral is finite and ϵ was arbitrary this concludes. □

6.4. The increments of the MUCOGARCH(1,1) process

So far we have mainly studied the MUCOGARCH volatility process Y or V, respectively. However, in practice one cannot observe the volatility but only the process G (which in a financial context, for instance, resembles log-prices) at finitely many points in time. In the following we presume that G is observed on a discrete time grid starting at zero and with fixed grid size $\Delta > 0$. It is obvious how the upcoming results of this section generalize to non-equidistant observations or the set-up considered in Klüppelberg et al. (2004, Section 5) and Brockwell et al. (2006, Section 6).

In financial time series one commonly observes that returns itself are uncorrelated, but "squared returns" (i.e. the return vector times its transpose in a multivariate setting) are considerably correlated. The following results show that the MUCOGARCH model can reproduce this feature.

We define the sequence of returns $\mathbf{G} = (\mathbf{G}_n)_{n\in\mathbb{N}}$ by setting

$$\mathbf{G}_n = \int_{(n-1)\Delta}^{n\Delta} V_{s-}^{1/2} dL_s. \tag{6.4.1}$$

Moreover, we shall throughout most of this section presume the following.

Assumption 6.4. *$C \in \mathbb{S}_d^{++}$ and Y (or equivalently V) is stationary.*

Proposition 6.4.1. *If Assumption 6.4 holds then $\mathbf{G}$ is stationary.*

Proof. Employing Theorem 6.3.22 the same arguments as for Klüppelberg et al. (2004, Corollary 3.1) show that G has stationary increments. □

In order to be able to obtain explicit expressions for the moments of $\mathbf{G}$ we need to strengthen Assumption 6.2 as follows.

Assumption 6.5. *Assumption 6.2 is satisfied and, moreover,*

$$E(L_1) = 0 \;\; \textit{and} \;\; \mathrm{var}(L_1) = (\sigma_W + \sigma_L) I_d \;\; \textit{with} \;\; \sigma_W \geq 0.$$

We start by giving conditions for the finiteness of the second moments of G and thus of $\mathbf{G}$ without requiring stationarity and explicit expressions for the moments in the stationary case.

Proposition 6.4.2. *Assume that $E(L_1) = 0$, $E(\|L_1\|^2) < \infty$, B is diagonalizable, $C \in \mathbb{S}_d^{++}$, $\lambda = \max(\Re(\sigma(B))) < 0$ and $E(\|Y_0\|) < \infty$. Then $E(\|G_t\|^2) < \infty$ for all $t \in \mathbb{R}^+$.*

If also Assumptions 6.4 and 6.5 are satisfied then the stationary sequence $\mathbf{G}$ *has the following second order structure:*

$$E(\mathbf{G}_1) = 0 \tag{6.4.2}$$

$$\mathrm{var}(\mathbf{G}_1) = (\sigma_L + \sigma_W)\Delta E(V_0) \tag{6.4.3}$$

$$\mathrm{vec}(\mathrm{var}(\mathbf{G}_1)) = (\sigma_L + \sigma_W)\Delta\mathscr{B}^{-1}(B \otimes I_d + I_d \otimes B)\mathrm{vec}(C) \tag{6.4.4}$$

$$\mathrm{acov}_{\mathbf{G}}(h) = 0 \text{ for all } h \in \mathbb{Z}\backslash\{0\}. \tag{6.4.5}$$

Proof. Proposition 6.3.20 ensures that $E(\|Y_t\|)$ and therefore $E(\|V_t\|)$ is finite and locally bounded. Since $E(\|V_t^{1/2}\|_2^2) = E(\|V_t\|_2)$, the standard L^2 stochastic integration theory (cf. Applebaum (2004, Section 4.2.1), for example) establishes $E(\|G_t\|^2) < \infty$ for all $t \in \mathbb{R}^+$.

Let now Assumptions 6.4 and 6.5 be satisfied, then (6.4.2) is clear and (6.4.5) is a straightforward consequence of the Itô isometry. The latter also implies

$$\begin{aligned}\mathrm{var}(\mathbf{G}_1) = E(\mathbf{G}_1\mathbf{G}_1^*) &= E\left(\int_0^\Delta V_{s-}^{1/2}E(L_1L_1^*)V_{s-}^{1/2}ds\right) = (\sigma_L + \sigma_W)E\left(\int_0^\Delta V_{s-}ds\right)\\ &= (\sigma_L + \sigma_W)\Delta E(V_0).\end{aligned}$$

□

Remark 6.4.3. *Straightforward extensions of the above arguments show that if* Y *is not stationary but only (asymptotically) second order stationary, then* $\mathbf{G}$ *is (asymptotically) second order stationary.*

For the squared returns $\mathbf{G}\mathbf{G}^* = (\mathbf{G}_n\mathbf{G}_n^*)_{n\in\mathbb{N}}$ we get:

Proposition 6.4.4. *Assume that* $E(L_1) = 0$, $E(\|L_1\|^4) < \infty$, B *is diagonalizable,* $C \in \mathbb{S}_d^{++}$, $\lambda = \max(\Re(\sigma(B))) < 0$ *and* $E(\|Y_0\|^2)$ *is finite. Then* $E(\|G_t\|^4) < \infty$ *and likewise* $E(\|G_tG_t^*\|^2) < \infty$ *for all* $t \in \mathbb{R}^+$.

If also Assumptions 6.4 and 6.5 are satisfied then the stationary sequence $\mathbf{G}\mathbf{G}^*$ *has the following second order structure:*

$$E(\mathbf{G}_1\mathbf{G}_1^*) = (\sigma_L + \sigma_W)\Delta E(V_0) \tag{6.4.6}$$

$$\mathrm{acov}_{\mathbf{GG}}(h) = e^{\mathscr{B}\Delta h}\mathscr{B}^{-1}\left(I_{d^2} - e^{-\mathscr{B}\Delta}\right)(\sigma_L + \sigma_W)\mathrm{cov}(\mathrm{vec}(Y_\Delta), \mathrm{vec}(\mathbf{G}_1\mathbf{G}_1^*)), \tag{6.4.7}$$

for $h \in \mathbb{N}$.

Proof. (i): We first show the finiteness of the moments. Using the Euclidean norm and its operator norm, it is clear that $E(\|G_t\|^4) < \infty$ if and only if $E(\|G_tG_t^*\|^2) < \infty$. Denoting the d components of G by G_i with $i = 1, \ldots, d$ we have that $E(\|G_t\|^4) < \infty$ is equivalent to $E(|G_{i,t}|^4) < \infty$ for all $i \in \{1, \ldots, d\}$. But the Burkholder-Davis-Gundy inequalities (see e.g. Protter (2004, p. 222)) give that $E(|G_{i,t}|^4) < \infty$ provided $E([G_i, G_i]_t^2) < \infty$. The latter is in turn ensured by $E(\|[G, G]_t\|^2) < \infty$ simultaneously for all $i \in \{1, \ldots, d\}$.

Next we observe that $[L, L]_t = \Sigma t + \int_0^t xx^*\mu_L(ds, dx)$ for some $\Sigma \in \mathbb{S}_d$ and $[G, G]_t = \int_0^t V_{s-}^{1/2}d[L, L]_sV_{s-}^{1/2}$. Hence,

$$\begin{aligned}\|[G, G]_t\|_2 &\leq \|\Sigma\|_2\int_0^t \|V_{s-}\|_2 ds + \int_0^t\int_{\mathbb{R}^d}\|V_{s-}\|_2\|xx^*\|_2\mu_L(ds, dx)\\ &\leq K\left(\int_0^t (y_{s-} + \|C\|_{B,S})ds + \int_0^t (y_{s-} + \|C\|_{B,S})d\tilde{L}_s\right)\end{aligned}$$

where $K \in \mathbb{R}^+$ is a finite constant and y and $\tilde{L}$ are the processes defined in Theorems 6.3.16 and 6.3.19. Thus it suffices to show that the expectations $E\left(\left(\int_0^t y_{s-}ds\right)^2\right)$ and $E\left(\left(\int_0^t y_{s-}d\tilde{L}_s\right)^2\right)$ are finite noting that $E(\|L_1\|^4) < \infty$ immediately implies the finiteness of the second moment of $\int_0^t \|C\|_{B,S} d\tilde{L}_s$.

$$\begin{aligned} 0 \leq & E\left(\left(\int_0^t y_{s-}ds\right)^2\right) = E\left(\int_0^t \int_0^t y_{s-}y_{u-}dsdu\right) = \int_0^t \int_0^t E\left(y_{s-}y_{u-}\right) dsdu \\ & \leq \int_0^t \int_0^t E(y_s^2)^{1/2} E(y_u^2)^{1/2} dsdu < \infty, \end{aligned}$$

because $E(\|Y_0\|^2) < \infty$ and $E(\|L_1\|^4) < \infty$ ensure that $E(y_t^2)$ is finite for all t and locally bounded in t (cf. Proposition 6.3.20 and its proof). Turning to the second term we have

$$\int_0^t y_{s-}d\tilde{L}_s = \int_0^t y_{s-}(d\tilde{L}_s - E(\tilde{L}_1)ds) + \int_0^t y_{s-}E(\tilde{L}_1)ds,$$

where the second summand has already been treated above. Hence, noting that $E(|\tilde{L}_1|^2) < \infty$ is a consequence of $E(\|L_1\|^4) < \infty$ the Itô isometry gives

$$E\left(\left(\int_0^t y_{s-}\left(d\tilde{L}_s - E(\tilde{L}_1)ds\right)\right)^2\right) = \int_0^t E(y_s^2)\mathrm{var}(\tilde{L}_1)ds < \infty,$$

which concludes.

(ii): Let now Assumptions 6.4 and 6.5 be satisfied. Then (6.4.6) has already been shown in the last proposition.

So it remains to establish (6.4.7). Let $h \in \mathbb{N}$. The definition of the quadratic (co)variation (see Lemma 4.5.11, in particular) implies

$$\begin{aligned} \mathbf{G}_{h+1}\mathbf{G}_{h+1}^* = & \int_{h\Delta}^{(h+1)\Delta} V_{s-}^{1/2} dL_s \left(\int_{h\Delta}^s dL_u^* V_{u-}^{1/2}\right) + \int_{h\Delta}^{(h+1)\Delta} \left(\int_{h\Delta}^s V_{u-}^{1/2} dL_u\right) dL_s^* V_{s-}^{1/2} \\ & + \left[\int_{h\Delta}^{(h+1)\Delta} V_{s-}^{1/2} dL_s, \int_{h\Delta}^{(h+1)\Delta} V_{s-}^{1/2} dL_s\right] \end{aligned}$$

with

$$\begin{aligned} \left[\int_{h\Delta}^{(h+1)\Delta} V_{s-}^{1/2} dL_s, \int_{h\Delta}^{(h+1)\Delta} V_{s-}^{1/2} dL_s\right] &= \int_{h\Delta}^{(h+1)\Delta} V_{s-}^{1/2} d[L,L]_s V_{s-}^{1/2} \\ &= \sigma_W \int_{h\Delta}^{(h+1)\Delta} V_{s-} ds + \int_{h\Delta}^{(h+1)\Delta} V_{s-}^{1/2} d[L,L]_s^{\mathfrak{d}} V_{s-}^{1/2}. \end{aligned}$$

Now we condition upon $\mathcal{F}_\Delta$ and denote by $Y(y, (L_r - L_{t_0})_{r \geq t_0}, t_0, t)$ with $t_0 \in \mathbb{R}^+$ the solution of

$$dY_t = (BY_{t-} + Y_{t-}B^*)dt + A(Y_{t-} + C)^{1/2} d[L,L]_t^{\mathfrak{d}} (Y_{t-} + C)^{1/2} A^*$$

for $t \geq t_0$ with $Y_{t_0} = y$. Furthermore, we denote by $E_{L,t_0}(\cdot)$ the expectation taken with respect to $(L_r - L_{t_0})_{r \geq t_0}$ only. Using Theorem 6.3.22, $E(L_1) = 0$ and that the increments of $(L_r)_{r \geq \Delta}$ are independent of $\mathcal{F}_\Delta$ one obtains

$$\begin{aligned} E\left(\int_{h\Delta}^{(h+1)\Delta} V_{s-}^{1/2} dL_s \left(\int_{h\Delta}^{s} dL_u^* V_{u-}^{1/2}\right) \middle| \mathcal{F}_\Delta\right) \\ = E\left(\int_{h\Delta}^{(h+1)\Delta} V_{s-}^{1/2} dL_s \left(\int_{h\Delta}^{s} dL_u^* V_{u-}^{1/2}\right) \middle| Y_\Delta\right) \\ = E_{L,\Delta}\left(\int_{h\Delta}^{(h+1)\Delta} \int_{h\Delta}^{s} V_{Y_\Delta,s-}^{1/2} dL_s dL_u^* V_{Y_\Delta,u-}^{1/2}\right) = 0 \end{aligned}$$

where $V_{Y_\Delta,t} := Y(Y_\Delta, (L_r - L_\Delta)_{r \geq \Delta}, \Delta, t) + C$ and likewise

$$E\left(\int_{h\Delta}^{(h+1)\Delta} \left(\int_{h\Delta}^{s} V_{u-}^{1/2} dL_u\right) dL_s^* V_{s-}^{1/2} \middle| \mathcal{F}_\Delta\right) = 0.$$

Moreover, using the moment assumptions and the compensation formula

$$\begin{aligned} E\left(\sigma_W \int_{h\Delta}^{(h+1)\Delta} V_{s-} ds + \int_{h\Delta}^{(h+1)\Delta} V_{s-}^{1/2} d[L,L]_s^{\mathfrak{d}} V_{s-}^{1/2} \middle| \mathcal{F}_\Delta\right) \\ = (\sigma_L + \sigma_W) \int_{h\Delta}^{(h+1)\Delta} E_{L,\Delta}(V_{Y_\Delta,s}) ds. \end{aligned}$$

Equation (6.3.15) implies

$$\begin{aligned} &\int_{h\Delta}^{(h+1)\Delta} E_{L,\Delta}(\mathrm{vec}(V_{Y_\Delta,s})) ds \\ &\quad = \int_{h\Delta}^{(h+1)\Delta} \left(\mathrm{vec}(C) + e^{\mathscr{B}(s-\Delta)} \left(\mathrm{vec}(Y_\Delta) + \sigma_L \mathscr{B}^{-1}(A \otimes A)\mathrm{vec}(C)\right)\right) ds \\ &\qquad - \int_{h\Delta}^{(h+1)\Delta} \sigma_L \mathscr{B}^{-1}(A \otimes A)\mathrm{vec}(C) ds \\ &\quad = \Delta \mathrm{vec}(C) + \mathscr{B}^{-1} e^{\mathscr{B}\Delta h}(I_{d^2} - e^{-\mathscr{B}\Delta}) \left(\mathrm{vec}(Y_\Delta) - E(\mathrm{vec}(Y_0))\right) + \Delta E(\mathrm{vec}(Y_0)). \end{aligned}$$

Combining the above results, we get

$$\begin{aligned} E\left(\mathrm{vec}(\mathbf{G}_{h+1}\mathbf{G}_{h+1}^*)(\mathrm{vec}(\mathbf{G}_1\mathbf{G}_1^*))^*\right) &= E\left(E\left(\mathrm{vec}(\mathbf{G}_{h+1}\mathbf{G}_{h+1}^*)\middle| \mathcal{F}_\Delta\right)(\mathrm{vec}(\mathbf{G}_1\mathbf{G}_1^*))^*\right) \\ = E(\mathrm{vec}(\mathbf{G}_1\mathbf{G}_1^*))\left(E(\mathrm{vec}(\mathbf{G}_1\mathbf{G}_1^*))\right)^* + (\sigma_L + \sigma_W) e^{\mathscr{B}\Delta h} \mathscr{B}^{-1}(I_{d^2} - e^{-\mathscr{B}\Delta}) \\ \cdot \left(E(\mathrm{vec}(Y_\Delta)(\mathrm{vec}(\mathbf{G}_1\mathbf{G}_1^*))^*) - E(\mathrm{vec}(Y_0))\left(E\left(\mathrm{vec}(\mathbf{G}_1\mathbf{G}_1^*)\right)\right)^*\right). \end{aligned}$$

Using the stationarity of Y this establishes (6.4.7). □

Thus, the squared returns $\mathbf{G}\mathbf{G}^*$ have like an ARMA(1,1) process an exponentially decreasing autocovariance function from lag one onwards.

In the univariate case Haug et al. (2007) obtained under additional assumptions on L expressions for $\mathrm{var}(\mathrm{vec}(\mathbf{G}_1\mathbf{G}_1^*))$ and $\mathrm{cov}(\mathrm{vec}(Y_\Delta), \mathrm{vec}(\mathbf{G}_1\mathbf{G}_1^*))$. As these are, however,

already rather lengthy and complicated explicit formulae, we do not calculate these values in our multivariate model. Moreover, it seems to be interesting to compare our results with those of Chapter 5 where explicit calculations for the second order structure of an OU type stochastic volatility model are presented. There independence of the volatility and the Wiener process in the log-prices eases the computations considerably compared to our MUCOGARCH setting where we have no such independence.

6.5. Illustrative examples and simulations

In this section we show some illustrative simulations of MUCOGARCH(1,1) processes in dimension $d = 2$. We restrict ourselves to driving Lévy processes having a compound Poisson jump part, as this allows us to simulate the Lévy process and the MUCOGARCH volatility process exactly and it is no crucial restriction, since for an infinite activity Lévy process one would have to use some kind of approximation with only finitely many jumps (e.g. a compound Poisson approximation or the usual Euler approximation). In the simulation of the MUCOGARCH process G itself we need to approximate the Brownian part, which is done in a standard way by $\int_{t_1}^{t_2} \sqrt{V_{s-}} dW_s \approx \sqrt{(t_2 - t_1)\sigma_W V_{t_1}} Z$ where W is the Brownian part of the Lévy process L and Z a two-dimensional standard normal random variable.

In all examples the Brownian part of L has variance I_d (i.e. $\sigma_W = 1$, see Assumption 6.5) and the jumps of L are two-dimensionally normally distributed with mean zero and variance chosen such that $\mathrm{var}(L_1) = 2I_d$ (i.e. the components of the jumps are uncorrelated and $\sigma_L = 1$, see Assumption 6.2). This means that all components of L and therefore of $[L, L]^{\mathfrak{d}}$ jump at the same time, but the jump sizes of the components of L are independent. Furthermore, Assumptions 6.2, 6.3 and 6.5 hold.

Example 1

In the first example we choose the compound Poisson jump part of the driving Lévy process to have rate four and furthermore $A = I_2$, $B = -1.6I_2$ and

$$C = \begin{pmatrix} 1 & 0 \\ 0 & 1.5 \end{pmatrix}.$$

Moreover, the initial values are chosen to be $Y_0 = 0$ (i.e. $V_0 = C$) and $G_0 = 0$. For the following pictures depicting V, G, $\mathbf{G}$ or $\mathbf{G}\mathbf{G}^*$ we simulated the processes G and V at times $0 \leq t \leq 2000$ and used $\Delta = 0.1$ for $\mathbf{G}$ and $\mathbf{G}\mathbf{G}^*$ (i.e. considered the increments over periods of length 0.1), whereas the upcoming estimation of the autocorrelation functions (acfs) is based on discrete observations by sampling every 0.1 time units from a simulation for times $0 \leq t \leq 100\,000$. Hence, in the autocorrelation functions "lag 1" corresponds to 0.1 time units.

Figure 6.1 shows the components of the MUCOGARCH log-price process G and of its stochastic volatility V. One should note that after a jump the volatility process goes rather fast towards C which can in particular be seen from Figure 6.2 depicting the components of V for $0 \leq t \leq 200$. For the variance components C provides a lower bound (1 and 1.5, respectively) and for the correlation component a kind of "mean level" (of zero). This reflects the "mean-reverting" behaviour already discussed after Definition 6.3.1. Moreover, the volatility appears to exhibit heavy-tailed behaviour and extremal clusters, as is to be

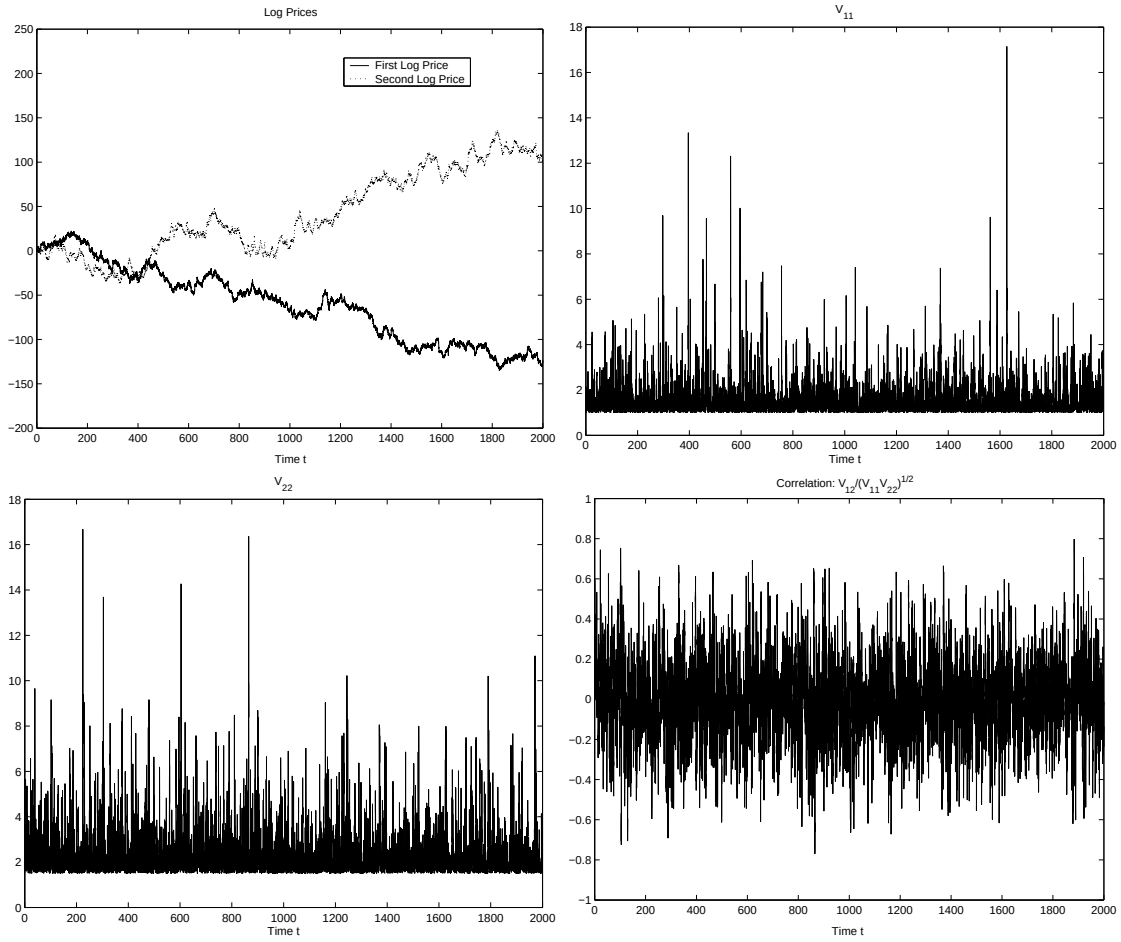


Figure 6.1.: Simulations for the first example: log price processes G_1 and G_2 (upper left), first variance process V_{11} (upper right), second variance process V_{22} (lower left) and correlation process $V_{12}/\sqrt{V_{11}V_{22}}$ (lower right).

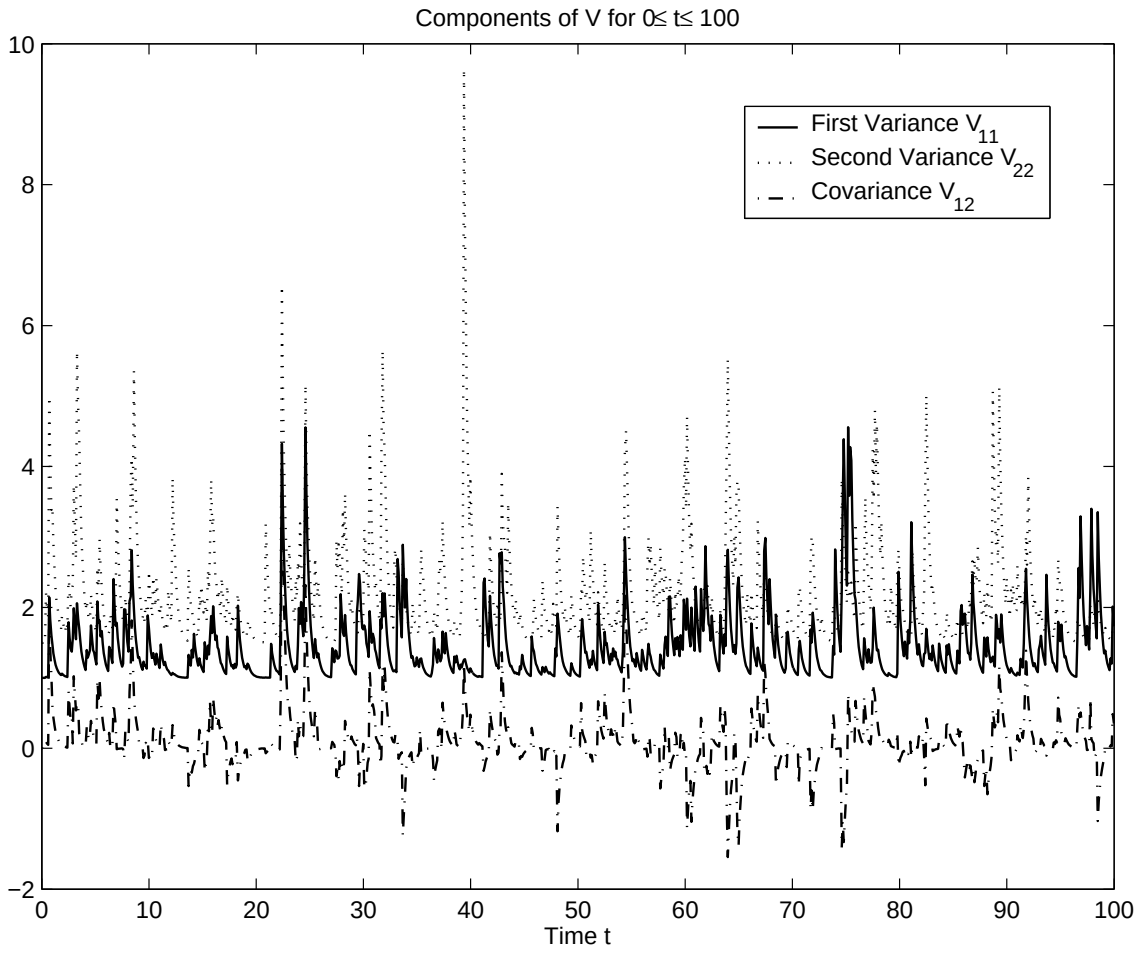


Figure 6.2.: Simulations for the first example: components of the stochastic volatility process V for $0 \leq t \leq 100$.

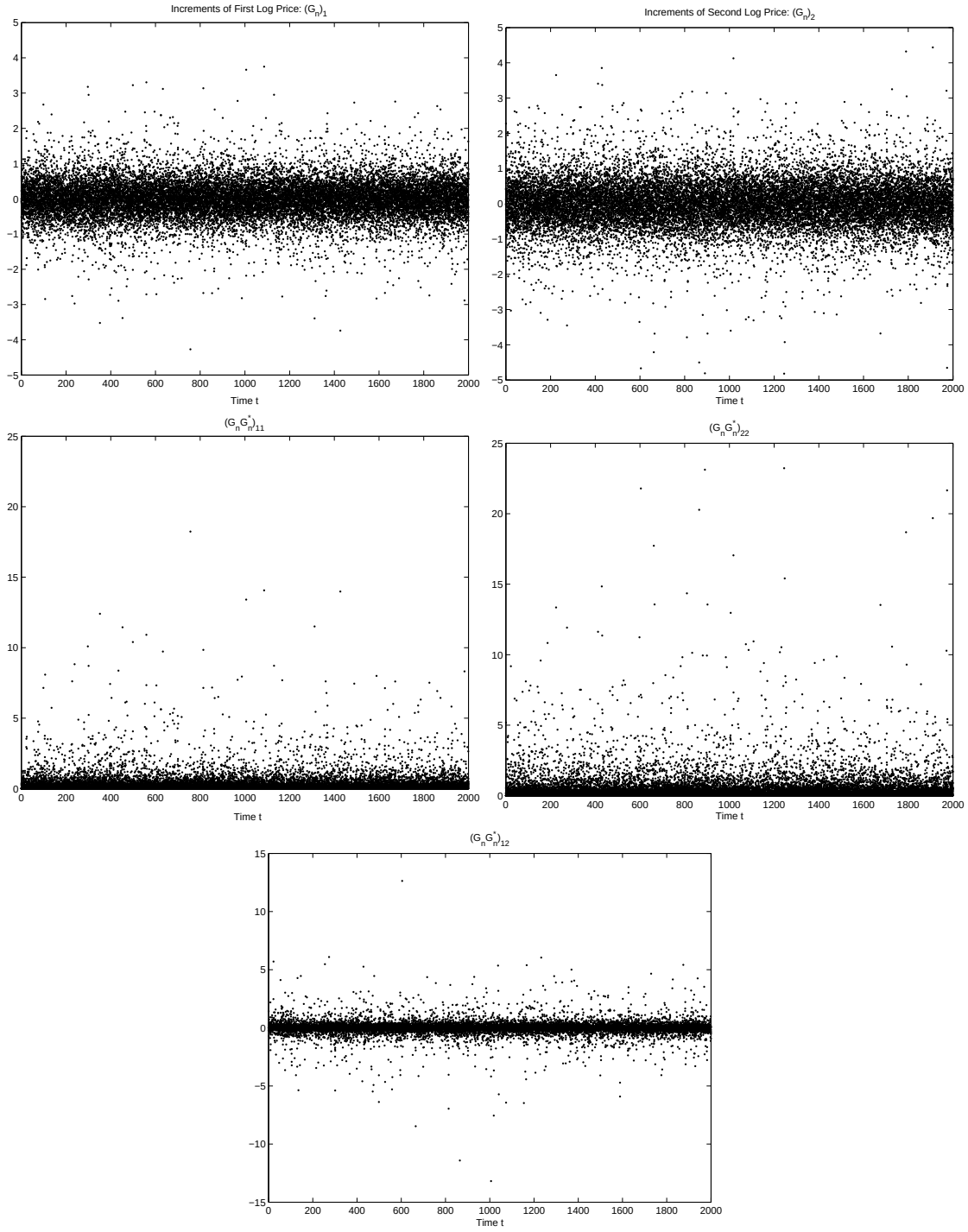


Figure 6.3.: Simulations for the first example: log price increments $\mathbf{G}_1$ (upper left) and $\mathbf{G}_2$ (upper right), components of the "squared returns" (realized variation) $(\mathbf{G}\mathbf{G}^*)_{11}$ (middle left), $(\mathbf{G}\mathbf{G}^*)_{22}$ (middle right) and $(\mathbf{G}\mathbf{G}^*)_{12}$ (lower).

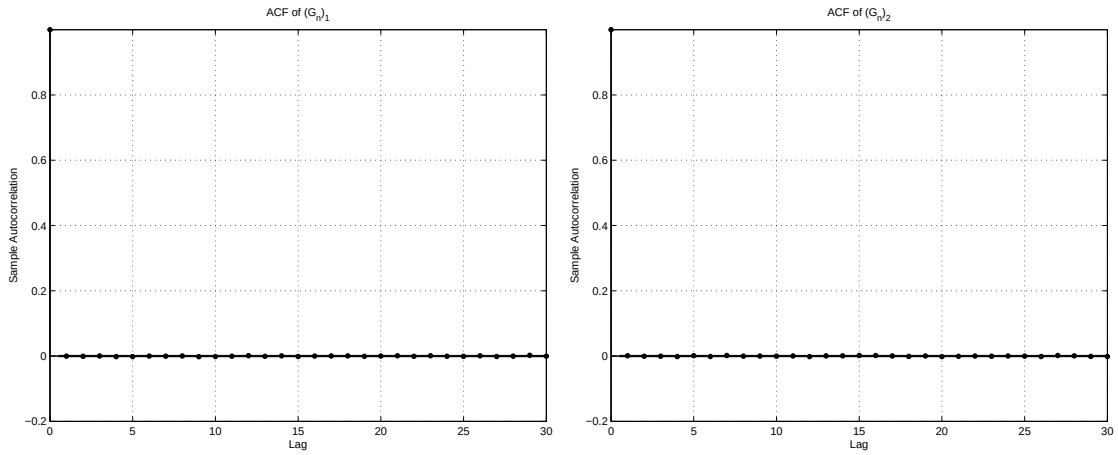


Figure 6.4.: Simulations for the first example: Autocorrelation functions (acf) of the increments of the first $\mathbf{G}_1$ (left) and second log price $\mathbf{G}_2$ (right). (Lag 1 corresponds to 0.1 time units and the horizontal lines coinciding almost with the axis are standard 95% confidence intervals for zero autocorrelation.)

expected, since in the univariate case Fasen et al. (2006) have shown this to be usually the case.

Since the matrix B is already diagonal, we can use the usual Euclidean norm for $\|\cdot\|_{B,S}$ (cf. Section 6.3.2) and obtain $\int_{\mathbb{R}^2} \left((1 + \|x\|_2^2)^2 - 1\right) \nu_L(dx) = 6$. From this it is straightforward to see that (6.3.12) is satisfied for $k = 2$ and hence there exists a stationary distribution for the volatility with finite second moments. Furthermore, Theorem 6.3.42 combined with Lemmata 6.3.30, 6.3.38 and 6.3.39 show that V is asymptotically second order stationary. This is reflected in the following: The formula of Theorem 6.3.28 implies that the (second order) stationary MUCOGARCH volatility process with the above parameters and driving Lévy process has mean

$$\begin{pmatrix} 16/11 & 0 \\ 0 & 24/11 \end{pmatrix}$$

and for our simulation (up to time 2000) the empirical mean is

$$\begin{pmatrix} 1.4368 & -0.0060 \\ -0.0060 & 2.1670 \end{pmatrix}$$

and thus already rather close to the asymptotic value. Returning to the existence of moments, elementary calculations establish that (6.3.12) is not satisfied for $k = 3$, so we cannot show that the stationary distribution of V has moments of orders higher than two. This may mean that the usual standard normal limit theory for the acf, for instance, is not valid (cf. Fasen (2007)). Yet, (CO)GARCH processes with such a behaviour are often found to be reasonable for empirical data. To obtain reliable estimates of the acfs we thus use very long simulations in the following and note that the standard confidence intervals shown later on need to be treated with care.

Also the discrete returns $\mathbf{G}$ and the squared returns $\mathbf{G}\mathbf{G}^*$, which can be understood as "realized variation", over intervals of length 0.1 display GARCH like features (stochastic volatility, heavy tails, extremal clusters), see Figure 6.3.

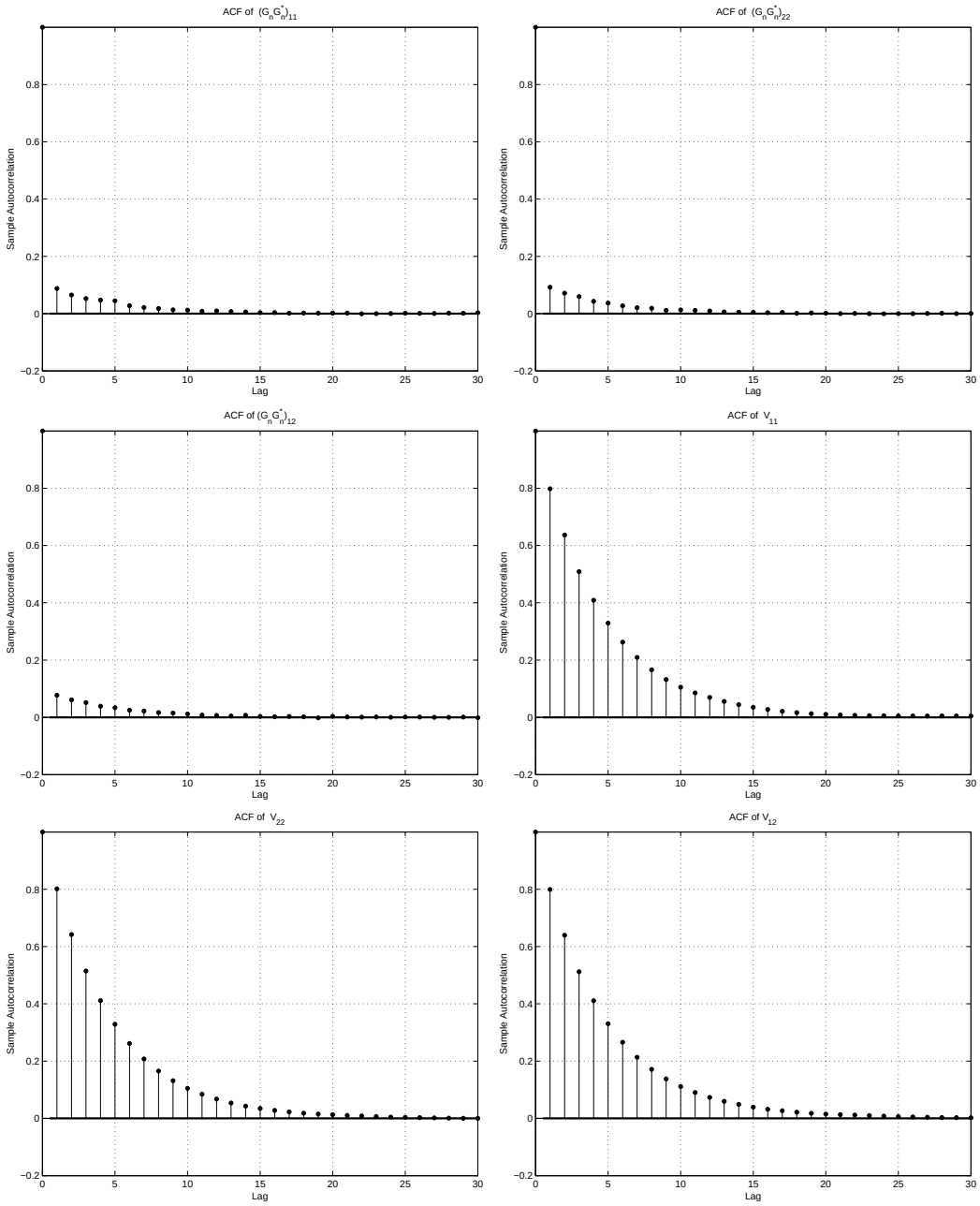


Figure 6.5.: Simulations for the first example: Acfs for the components of the squared log-price increments $(\mathbf{GG}^*)_{11}$ (upper left), $(\mathbf{GG}^*)_{22}$ (upper right), $(\mathbf{GG}^*)_{12}$ (middle left) and of the volatility process V_{11} (middle right), V_{22} (lower left), V_{12} (lower right). (Lag 1 corresponds to 0.1 time units and the horizontal lines coinciding almost with the axis are standard 95% confidence intervals for zero autocorrelation.)

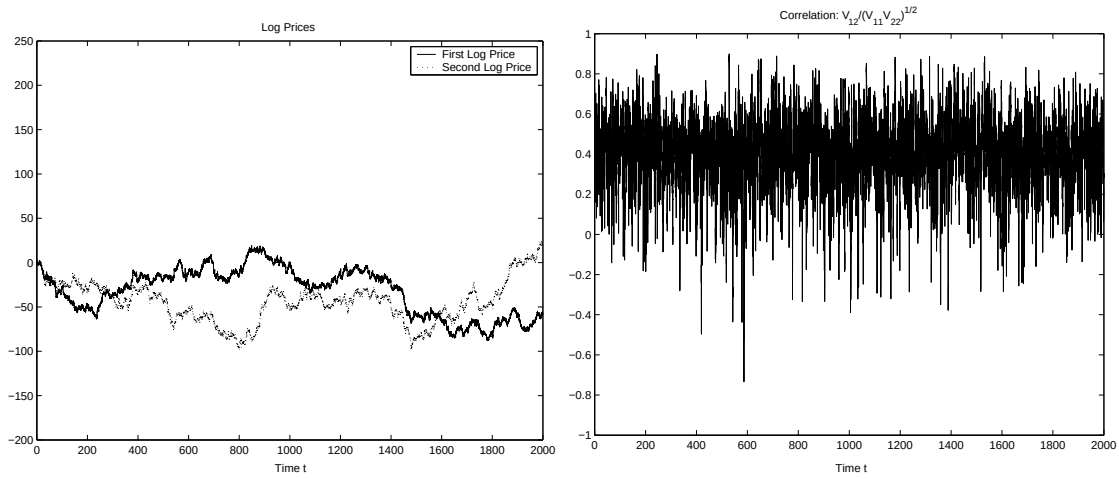


Figure 6.6.: Simulations for the second example: log price processes G_1 and G_2 (left) and correlation process $V_{12}/\sqrt{V_{11}V_{22}}$ (right).

Finally, the increments $\mathbf{G}$ of the MUCOGARCH process exhibit zero autocorrelation, whereas the squared increments $\mathbf{GG}^*$ (from lag one onwards) and the volatility process V exhibit exponentially decreasing autocorrelations, as is illustrated in Figures 6.4 and 6.5 (note that we refrain from depicting the off-diagonal elements of the autocovariance matrix functions). As usual this is much clearer to see for the stochastic volatility process V, where the exponential decay starts immediately at lag 0, but for $\mathbf{GG}^*$ the autocorrelations are also clearly non-zero for the first lags and decay exponentially from lag 1 onwards. This resembles formulae (6.3.22) and (6.4.7) which imply that $\operatorname{acov}_V(h) = e^{-2.2I_4 h}\operatorname{var}(\operatorname{vec}(V_0))$ for $h \in \mathbb{R}^+$ and $\operatorname{acov}_{\mathbf{GG}^*}(h) = e^{-2.2I_4\Delta h}K$ for $h \in \mathbb{N}$ and some $K \in M_d(\mathbb{R})$. These formulae also imply that all acfs decay with the same rate and thus the acfs are actually the same for all components of $\mathbf{GG}^*$ or V, respectively. The autocovariance functions, however, are different, since the components have different variances, as can be seen by using Corollary 6.3.40 or estimating the variances from the simulated data.

Example 2

Let us now change C to

$$C = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1.5 \end{pmatrix}.$$

and keep everything else as in the first example. Thus we can see the effect of changing the lower bound on the covariance matrix recalling from Proposition 6.3.5 that $V_t \geq C$ for all $t \in \mathbb{R}$, since we start with a positive semi-definite Y_0. The above chosen matrix C corresponds to a correlation of c. 0.41.

Figure 6.6 shows the components of the MUCOGARCH log-price process G and the (normalized) covariance component V_{12} of its stochastic volatility. The qualitative behaviour is very similar to the first example except that the correlation tends to be higher and the price processes tend to move more often in the same direction. However, this tendency is not very strong, as is to be expected for moderately correlated processes. Changing C does not affect condition (6.3.12) and hence there again exists a stationary distribution for

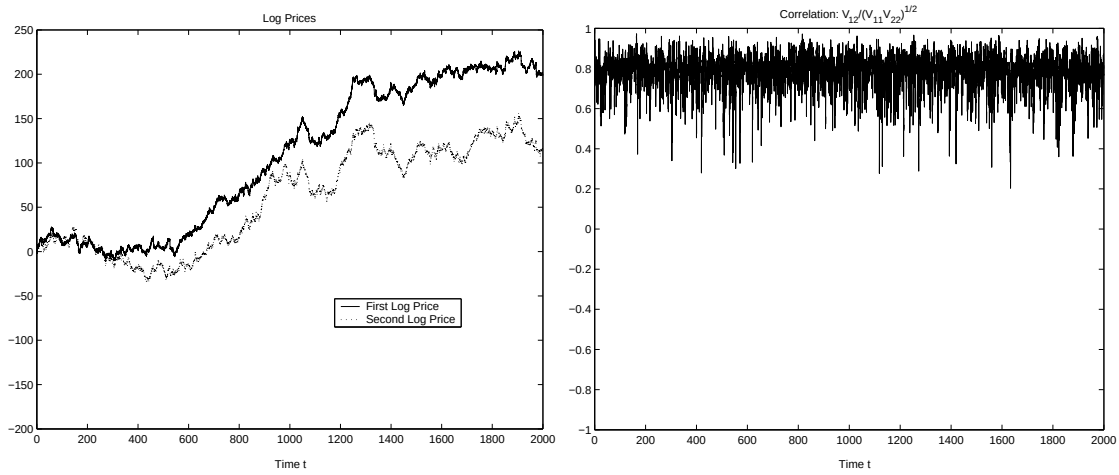


Figure 6.7.: Simulations for the third example: log price processes G_1 and G_2 (left) and correlation process $V_{12}/\sqrt{V_{11}V_{22}}$ (right).

the volatility with finite second moments. Likewise one again has asymptotic second order stationarity.

In this and the following examples, we only show pictures were the qualitative behaviour is different compared to Example 1. The returns $\mathbf{G}$ and the squared returns $\mathbf{G}\mathbf{G}^*$ over intervals of length 0.1 again display GARCH like features (stochastic volatility, heavy tails, extremal clusters) and the behaviour of the autocorrelation functions is completely the same as in the first example.

Example 3

To see the effect of further increasing the correlation we change now C to

$$C = \begin{pmatrix} 1 & 0.97 \\ 0.97 & 1.5 \end{pmatrix}.$$

and keep everything else as before. This choice of C corresponds to a correlation of c. 0.79

The features of the MUCOGARCH process and its volatility process are basically the same as in the previous two examples. However, as is to be seen from Figure 6.7, the stochastic correlation now tends to be rather high and thus the two log-price processes evolve very much alike. Again the acfs are the same as in Example 1.

Example 4

With this example we illustrate the effect of a change in the parameter B. Hence, we choose all parameters as in the first example except for

$$B = \frac{128}{85}\begin{pmatrix} -21/16 & 1 \\ 1 & -81/16. \end{pmatrix}$$

The matrix B is symmetric which implies that we again do not need to use some $\|\cdot\|_{B,S}$-norm but can simply work with $\|\cdot\|_2$ again. Its eigenvalues are -1.6 and -8 and hence

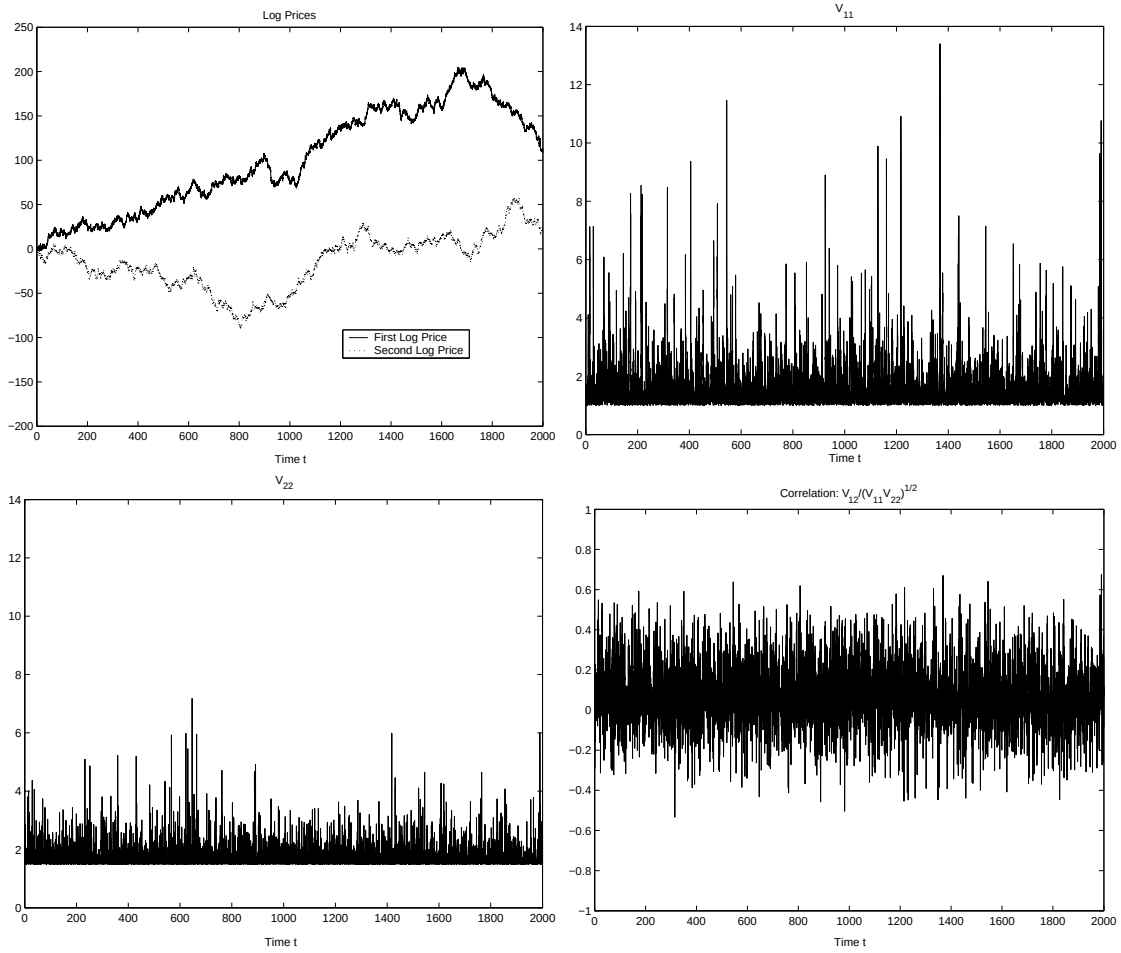


Figure 6.8.: Simulations for the fourth example: log price processes G_1 and G_2 (upper left), first variance process V_{11} (upper right), second variance process V_{22} (lower left) and correlation process $V_{12}/\sqrt{V_{11}V_{22}}$ (lower right).

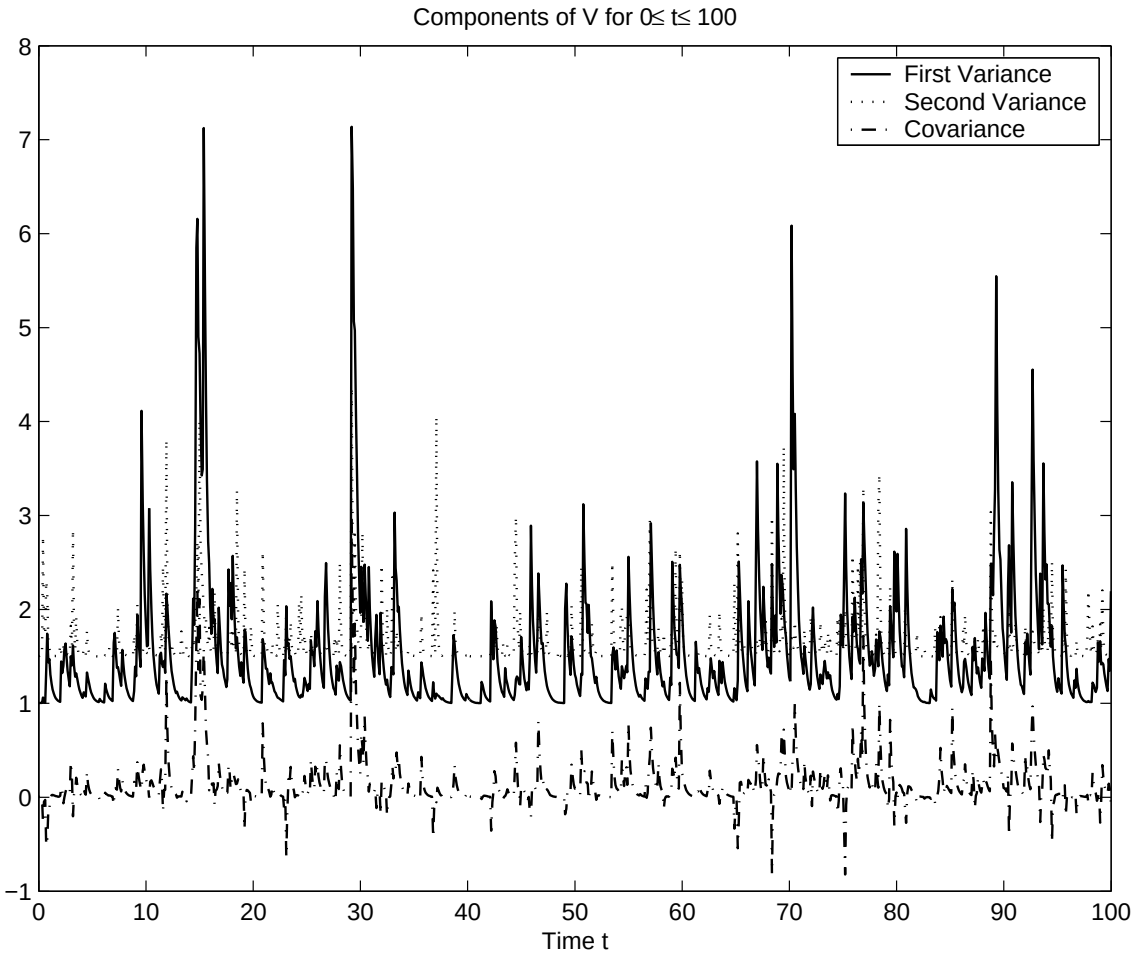


Figure 6.9.: Simulations for the fourth example: components of the stochastic volatility process V for $0 \leq t \leq 100$.

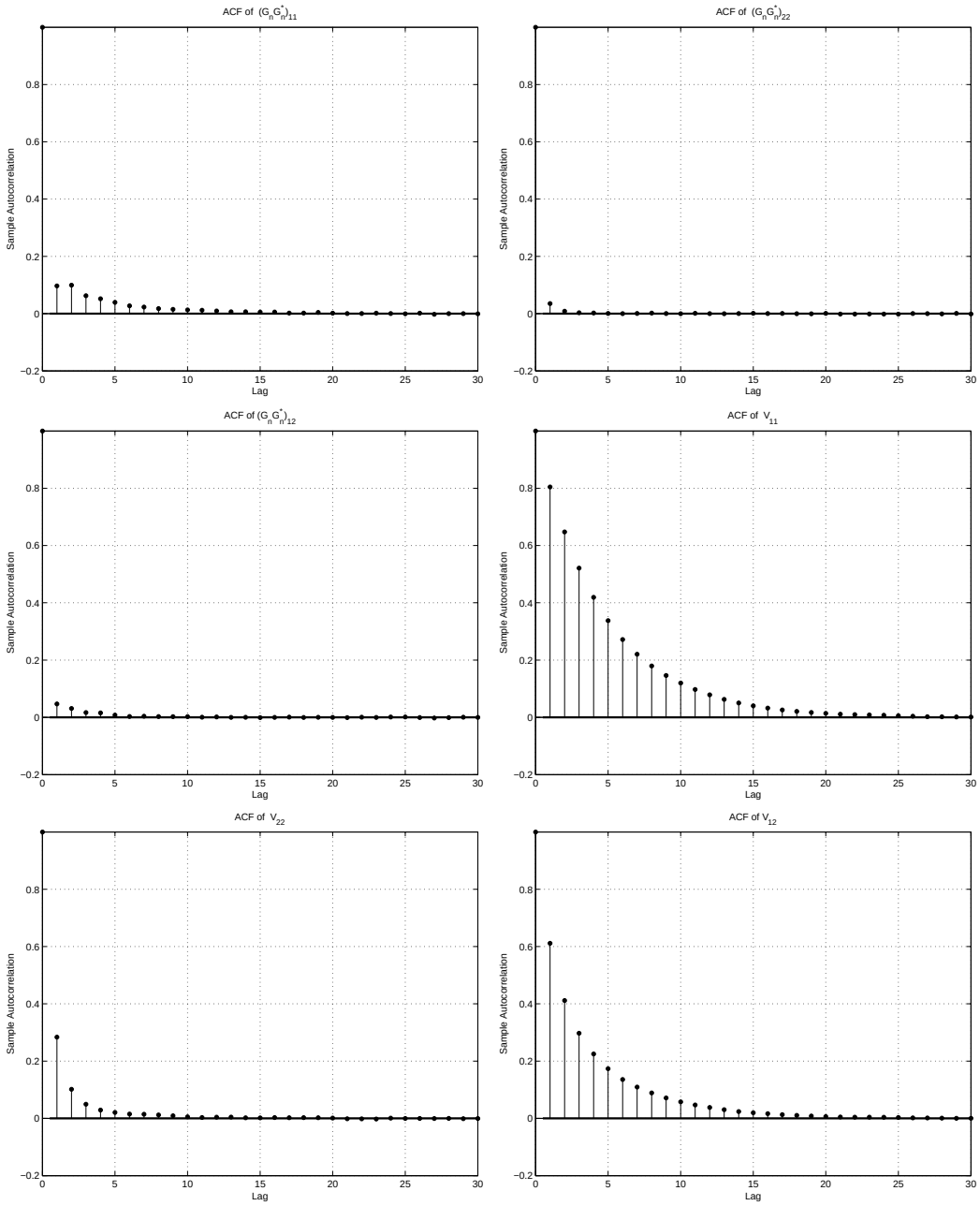


Figure 6.10.: Simulations for the fourth example: Acfs for the components of the squared log-price increments $(\mathbf{GG}^*)_{11}$ (upper left), $(\mathbf{GG}^*)_{22}$ (upper right), $(\mathbf{GG}^*)_{12}$ (middle left) and of the volatility process V_{11} (middle right), V_{22} (lower left), V_{12} (lower right). (Lag 1 corresponds to 0.1 time units and the horizontal lines coinciding almost with the axis are standard 95% confidence intervals for zero autocorrelation.)

condition (6.3.12) is clearly again satisfied for $k = 2$. Thus we have the existence of a stationary distribution for the volatility with finite second moments and likewise we again have asymptotic second order stationarity.

As is to be seen from Figures 6.8 and 6.10, the behaviour of the MUCOGARCH process and its volatility process is principally the same as in the previous examples. The main difference is that the exponential decay of the individual components takes part at different rates now which is best seen from Figure 6.9 depicting the components of the volatility process for $0 \leq t \leq 100$. This is also reflected in the acfs (see Figure 6.10) and it is the reason behind the second variance process tending to be smaller than the first one despite having the larger lower bound as given in C. Calculating $e^{Bt} \begin{pmatrix} v_{11} & v_{12} \\ v_{12} & v_{22} \end{pmatrix} e^{B^*t}$ explicitly shows that the decay of the first variance component is dominated by $e^{-(16/5)t}$ whereas e^{-16t} dominates the decay of the second variance component. So B mainly affects the decay to C (as long as there is no jump) and the autocorrelation structures.

Example 5

Finally, we analyse the influence of the parameter A. To do this we choose all parameters as in the first example except for

$$A = 0.6 I_2.$$

Once more condition (6.3.12) is clearly satisfied for $k = 2$. Thus we have the existence of a stationary distribution for the volatility with finite second moments. Moreover, we also have asymptotic second order stationarity.

As is to be seen from Figures 6.11 and 6.12, the qualitative behaviour of the MUCOGARCH process and its volatility process is principally the same as in the previous examples, but jumps are considerably smaller, as are the values of the volatility process and the (squared) increments of the log-price processes. To be precise compared to the previous examples the jumps are scaled down to 0.36 of their size. So the processes appear to be much "tamer". This is also reflected in a lower expected volatility

$$E(V_0) \approx 1.13C = \begin{pmatrix} 1.13 & 0 \\ 0 & 1.695 \end{pmatrix}.$$

Moreover, the autocorrelation functions decay faster now than in Examples 1 to 3, since $B \otimes I_4 + I_4 \otimes B + \sigma_L A \otimes A = -(71/25) I_4$.

So A affects the jump sizes and the autocorrelation structures in particular.

6.6. Possible extension to higher orders

So far we have only considered multivariate COGARCH(p, q) processes of orders $p = q = 1$. Now we discuss how one may use multivariate CARMA processes in order to define multivariate COGARCH processes of general orders (p, q) with $p, q \in \mathbb{N}$, $q \geq p$. However, we do not know how to restrict the parameters in order to ensure the positive-definiteness of the volatility process for $q > 1$ so far.

6.6.1. Multivariate CARMA processes

As the name "continuous time ARMA" (CARMA) already suggests, these processes are the continuous time analogue of the well-known ARMA processes. For a review of the existing

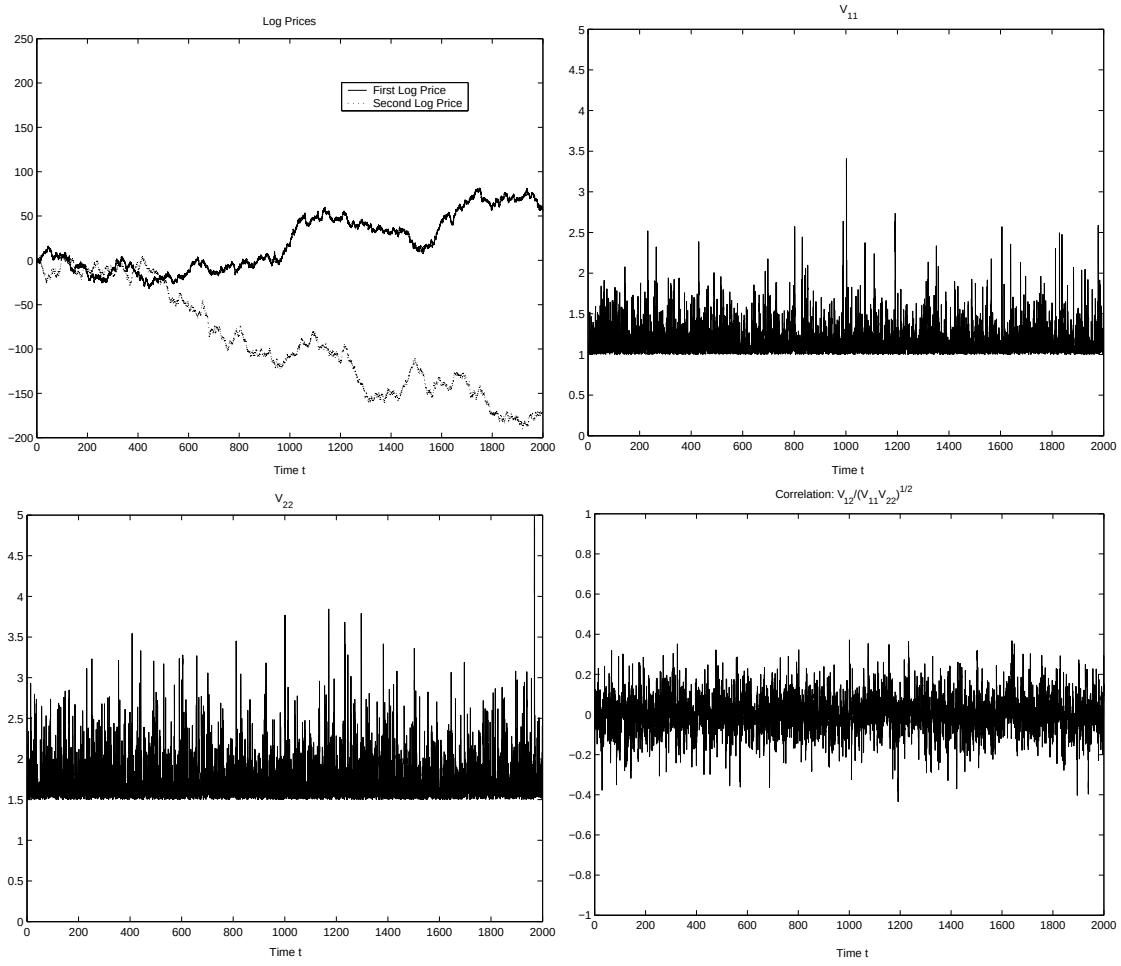


Figure 6.11.: Simulations for the fifth example: log price processes G_1 and G_2 (upper left), first variance process V_{11} (upper right), second variance process V_{22} (lower left) and correlation process $V_{12}/\sqrt{V_{11}V_{22}}$ (lower right).

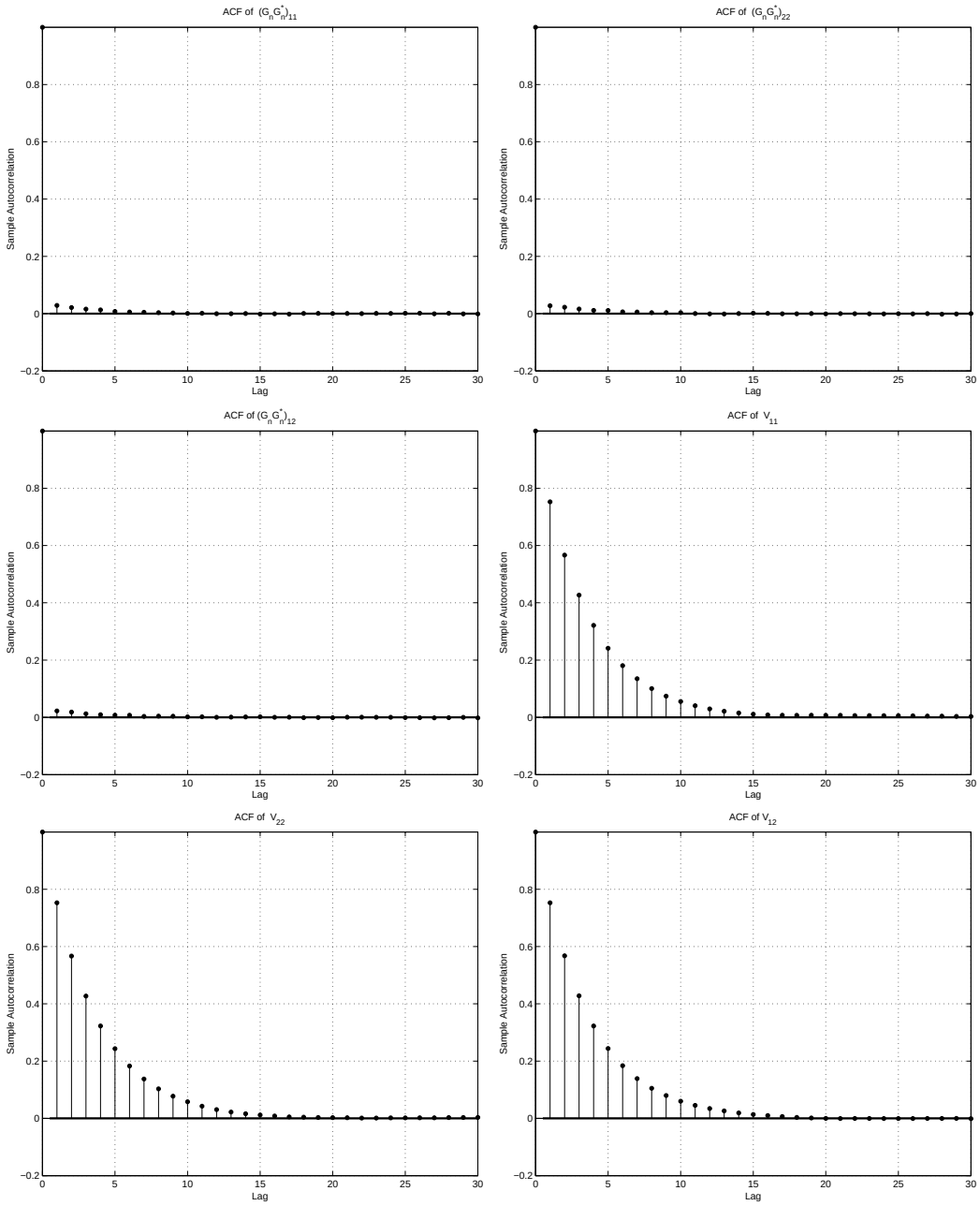


Figure 6.12.: Simulations for the fifth example: Acfs for the components of the squared log-price increments $(\mathbf{G}\mathbf{G}^*)_{11}$ (upper left), $(\mathbf{G}\mathbf{G}^*)_{22}$ (upper right), $(\mathbf{G}\mathbf{G}^*)_{12}$ (middle left) and of the volatility process V_{11} (middle right), V_{22} (lower left), V_{12} (lower right). (Lag 1 corresponds to 0.1 time units and the horizontal lines coinciding almost with the axis are standard 95% confidence intervals for zero autocorrelation.)

literature see the references given in Chapter 2. The d-dimensional CARMA(q,p) processes are defined in that chapter and can be viewed as the solution to the formal differential equation:

$$Q(\mathcal{D})X_t = P(\mathcal{D})\mathcal{D}L_t \tag{6.6.1}$$

where $L = (L_t)_{t\in\mathbb{R}}$ is a Lévy process in $\mathbb{R}^d$ and $\mathcal{D}$ the differential operator with respect to t.

$$\begin{aligned} Q(z) &= z^q + Q_1 z^{q-1} + Q_2 z^{q-2} + \ldots + Q_q && (6.6.2)\\ P(z) &= P_1 z^p + P_2 z^{p-1} + \ldots + P_{p+1} && (6.6.3)\end{aligned}$$

with $P_1,\ldots,P_{p+1},Q_1,\ldots,Q_q \in M_d(\mathbb{R})$ and $Q_q, P_1 \neq 0$ are referred to as the autoregressive and moving average polynomials, respectively. In order to be able to define CARMA processes properly one needs $q > p$ and that the zeros of $\det(Q(z))$ have all strictly negative real part. Then the CARMA(q,p) process X is defined as the unique stationary solution of

$$\begin{aligned} X_t &= (I_d, 0, \ldots, 0)Y_t && (6.6.4)\\ dY_t &= \begin{pmatrix} 0 & I_d & 0 & \cdots & 0 \\ 0 & 0 & I_d & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & I_d \\ -Q_q & -Q_{q-1} & -Q_{q-2} & \cdots & -Q_1 \end{pmatrix} Y_t dt + D dL_t, && (6.6.5)\end{aligned}$$

where $D = (D_1^*, D_2^*, \ldots, D_q^*)^*$ is a $d \times d$ matrix with elements

$$D_{q-j} = -\sum_{i=1}^{q-j-1} Q_i D_{q-j-i} + P_{p+1-j}$$

(setting $P_i = 0$ for $i \leq 0$). The process Y is usually called state space representation. In the univariate case $d = 1$ the representation by (6.6.4), (6.6.5) can be replaced by

$$\begin{aligned} X_t &= (P_{p+1}, P_p, \ldots, P_{p-q+2})\tilde{Y}_t && (6.6.6)\\ d\tilde{Y}_t &= \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ -Q_q & -Q_{q-1} & -Q_{q-2} & \cdots & -Q_1 \end{pmatrix} \tilde{Y}_t dt + \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix} dL_t. && (6.6.7)\end{aligned}$$

(Setting $P_i = 0$ for $i \leq 0$)

The equivalence (in L^2, i.e. the driving Lévy process L has to have a finite second moment) of the two representations of univariate CARMA processes has been shown in Chapter 2 using Fourier methods. The following result provides an alternative approach for $p = 0$ and $(q,p) = (2,1)$ and does not need a finite variance of L. Therefore, it will later on generalize immediately to COGARCH(p,q) processes

Lemma 6.6.1. *a) Representations (6.6.4), (6.6.5) and (6.6.6), (6.6.7) are equivalent for a one-dimensional CARMA(q,0) process. In particular,* $\tilde{Y}_t = Y_t/P_1$.

b) Assume that we are given a one-dimensional CARMA(2,1) process with the autoregressive polynomial $Q(z)$ and the moving average polynomial $P(z)$ having no common zeros. Then representations (6.6.4), (6.6.5) and (6.6.6), (6.6.7) are equivalent. In particular, $Y_t = S\tilde{Y}_t$ with

$$S = \begin{pmatrix} P_2 & P_1 \\ -Q_2P_1 & -Q_1P_1 + P_2 \end{pmatrix}$$

being invertible.

Proof. The proof of a) is immediate by substituting the variables as prescribed by the relationship $\tilde{Y}_t = Y_t/P_1$.

Turning to b) we first observe using $P_1 \neq 0$:

$$det(S) = P_2^2 - Q_1P_1P_2 + Q_2P_1^2 = P_1^2\left((P_2/P_1)^2 - Q_1(P_2/P_1) + Q_2\right) = P_1^2Q(-P_2/P_1)$$

But this is different from zero, as $-P_2/P_1$ is the unique zero of $P(z)$ and $Q(z), P(z)$ have no zeros in common. Hence, S is invertible with

$$S^{-1} = \frac{1}{P_2^2 - Q_1P_1P_2 + Q_2P_1^2}\begin{pmatrix} -Q_1P_1 + P_2 & -P_1 \\ Q_2P_1 & P_2 \end{pmatrix}.$$

Rewrite now (6.6.6) and (6.6.7) to

$$X_t = (P_2, P_1)S^{-1}S\tilde{Y}_t \tag{6.6.8}$$

$$dS\tilde{Y}_t = S\begin{pmatrix} 0 & 1 \\ -Q_2 & -Q_1 \end{pmatrix}S^{-1}S\tilde{Y}_t dt + S\begin{pmatrix} 0 \\ 1 \end{pmatrix}dL_t. \tag{6.6.9}$$

Substituting $Y_t = S\tilde{Y}_t$ and multiplying everything out the last two equations become (6.6.4) and (6.6.5). This shows the equivalence of the two representations. □

Note, however, that it is not possible to extend the result to general dimension d, as the proof above needs multiplicative commutativity of the ARMA coefficients P_i, Q_j and matrices are in general not commutative.

6.6.2. Multivariate COGARCH(p, q) processes

In order to define a continuous time multivariate GARCH process of general order (p, q) the main idea is again to replace the noise ϵ of a general order multivariate GARCH process by the increments of a multivariate Lévy process L and the ARMA structure of the volatility process by a multivariate CARMA structure as just recalled.

Regarding the multivariate GARCH model in its vectorized form, observe that (6.2.6) can be rewritten as

$$\text{vec}(\Sigma_n) = \text{vec}(C) + \sum_{i=1}^{p}\tilde{A}_i(\Sigma_{n-i}^{1/2} \otimes \Sigma_{n-i}^{1/2})\text{vec}(\epsilon_{n-i}\epsilon_{n-i}^*) + \sum_{j=1}^{q}\tilde{B}_j\text{vec}(\Sigma_{n-j}). \tag{6.6.10}$$

This shows that the dynamics of (the vectorized) Σ are those of a multivariate ARMA(q, $p-1$) process, which is "self-exciting" in the sense that the noise is given by

$$\left((\Sigma_{n-1}^{1/2} \otimes \Sigma_{n-1}^{1/2})\text{vec}(\epsilon_{n-1}\epsilon_{n-1}^*)\right)_{n\in\mathbb{Z}}.$$

Replacing (like in the univariate case) the ARMA$(q, p-1)$ structure with a CARMA$(q, p-1)$ one and using $(V_{t-}^{1/2} \otimes V_{t-}^{1/2})\text{vec}(d[L, L^{\mathfrak{d}}]_t)$ as "noise" where L is a Lévy process, leads from the discrete vec representation to a continuous time GARCH process G with parameters $A_1, \ldots, A_p, B_1, \ldots, B_q \in M_{d^2}(\mathbb{R}), C \in \mathbb{S}_d^+$ given by:

$$
\begin{aligned}
dG_t &= V_{t-}^{1/2} dL_t && (6.6.11)\\
V_t &= C + (I_d, 0, \ldots, 0)Y_t && (6.6.12)\\
d\text{vec}(Y_t) &= \mathbf{B}\text{vec}(Y_{t-})dt + D(V_{t-}^{1/2} \otimes V_{t-}^{1/2})\text{vec}(d[L, L]_t^{\mathfrak{d}}), && (6.6.13)\\
\mathbf{B} &:= \begin{pmatrix} 0 & I_{d^2} & 0 & \cdots & 0 \\ 0 & 0 & I_{d^2} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & I_{d^2} \\ -B_q & -B_{q-1} & -B_{q-2} & \cdots & -B_1 \end{pmatrix}, && (6.6.14)
\end{aligned}
$$

where $D = (D_1^*, D_2^*, \ldots, D_q^*)^*$ is a $d^2q \times d^2$ matrix with elements

$$
D_{q-j} = -\sum_{i=1}^{q-j-1} B_i D_{q-j-i} + A_{p-j} \text{ (setting } A_i = 0 \text{ for } i \leq 0).
$$

Finally, observe that the state space representation Y assumes values in $M_{dq,d}(\mathbb{R})$ and thereby vec(Y) lives on $\mathbb{R}^{d^2q}$.

An important question is when $V_{t-}^{1/2}$ is well defined, i.e. when V_{t-} is positive semi-definite. A path-wise analysis considering the jumps and the time in between jumps separately shows that

$$
D_1(\text{vec}(\mathbb{S}_d^+)) \subseteq (\mathbb{S}_d^+) \text{ and } (I_{d^2}, 0, \ldots, 0)e^{\mathbf{B}t}\left(\text{vec}(\mathbb{S}_d^+) \times \mathbb{R}^{(q-1)d^2}\right) \subseteq \text{vec}(\mathbb{S}_d^+)\ \forall\, t \in \mathbb{R}^+
$$

are sufficient to ensure $V_t \geq C$ for all $t \in \mathbb{R}^+$ provided $V_0 \geq C$. This result is comparable to Brockwell et al. (2006, Theorem 5.1).

Yet, practical conditions ensuring this have yet to be developed. For the time being we can only assume that the parameters are chosen such that V_{t-} is always positive semi-definite. Characterizing the feasible parameter sets ensuring positivity is already in the univariate case rather involved (cf. Brockwell et al. (2006), Tsai and Chan (2005) and Tsai and Chan (2006)), in particular for general order (p, q).

At a first sight we have a different model than Brockwell et al. (2006) in the univariate case. Yet, this is due to the fact that we used the multivariate CARMA representation of Chapter 2. However, under L^2 conditions this CARMA representation gives as shown in Chapter 2 the same CARMA process in the univariate case as the representation used by Brockwell et al. (2006). This leads us to conjecture:

Conjecture: *In the univariate case the above defined COGARCH(p, q) process is the same as the one defined in Brockwell et al. (2006), i.e. our processes* $(G_t)_{t\in\mathbb{R}^+}$ *and* $(V_t)_{t\in\mathbb{R}^+}$ *agree with their processes* $(G_t)_{t\in\mathbb{R}^+}$ *and* $(V_{t+})_{t\in\mathbb{R}^+}$.

Unfortunately, the proof of the equivalence of the CARMA representations given in Chapter 2 uses Fourier methods and thus cannot be extended to the COGARCH case. However, we have the following.

Proposition 6.6.2. *The above conjecture is at least true for* $(p,q) = (2,2)$ *and* $(p,q) = (1,q)$ *with* $q \in \mathbb{N}$.

Proof. Obvious from Lemma 6.6.1 and its proof. □

We conjecture Lemma 6.6.1 (b) holds for general orders with an appropriate matrix S, but so far we have not been able to prove this.

6.7. Appendix to Chapter 6

6.7.1. Stochastic differential equations on open sets

In this appendix we derive general results for solutions to stochastic differential equations which are only defined on an open subset U of $\mathbb{R}^d$.

6.7.1.1. Existence and uniqueness of solutions

It is common knowledge to experts in stochastic differential equations that one needs a local Lipschitz property to obtain existence and uniqueness up to the first exit time of the set U and that a linear growth condition is needed to ensure that no explosions happen (see e.g. Jacod (2007)). However, since this issue seems not to have been dealt with in any sufficiently detailed manner in the literature, we present the relevant results in detail in the following. Note that one can, of course, immediately replace $\mathbb{R}^d$ by $M_d(\mathbb{R})$, $\mathbb{S}_d$ or any other finite-dimensional vector space over $\mathbb{R}$ and that "solution" implicitly means a strong, adapted and càdlàg solution.

Thus let $U \subseteq \mathbb{R}^d$ be an open set, $f : U \to M_{dm}(\mathbb{R})$ a function and $Z = (Z_t)_{t\in\mathbb{R}^+}$ a semi-martingale with values in $\mathbb{R}^m$. Then we are interested in the stochastic differential equation

$$dX_t = f(X_{t-})dZ_t,\ t \in \mathbb{R}^+ \tag{6.7.1}$$

with initial value X_0 being an $\mathcal{F}_0$-measurable random variable with values in $\mathbb{R}^d$.

For the relevant standard results on stochastic differential equations defined on $\mathbb{R}^d$ we refer to any of the standard texts, Ikeda and Watanabe (1989), Jacod and Shiryaev (2003), Métivier (1982), Métivier and Pellaumail (1980b) or Protter (2004), for instance.

Definition 6.7.1. *Let* $(U, \|\cdot\|_U), (V, \|\cdot\|_V)$ *be two normed spaces and* $W \subseteq U$ *be open. Then a function* $f : W \to V$ *is called* locally Lipschitz*, if for every* $x \in W$ *there exists an open neighbourhood* $\mathcal{U}(x) \subset W$ *and a constant* $C(x) \in \mathbb{R}^+$ *such that*

$$\|f(z) - f(y)\|_V \leq C(x)\|z - y\|_U \ \forall\, z, y \in \mathcal{U}(x).$$

$C(x)$ *is said to be a local Lipschitz coefficient.*

If there is a $K \in \mathbb{R}^+$ *such that* $C(x) = K$ *can be chosen for all* $x \in W$*,* f *is called* (globally or uniformly) Lipschitz*.*

The following implication is standard.

Proposition 6.7.2. *Let* $(U, \|\cdot\|_U), (V, \|\cdot\|_V)$ *be two normed spaces,* $W \subseteq U$ *be open and* $K \subset W$ *be compact. If a function* $f : W \to V$ *is locally Lipschitz, then* f *is Lipschitz on* K*, i.e. there exists a constant* $C(K) \in \mathbb{R}^+$ *such that* $\|f(x)-f(y)\|_V \leq C(K)\|x-y\|_U \forall x, y \in K$,

The main result on the existence of unique solutions of stochastic differential equations on open sets is given by the following theorem. The condition imposed on the open set U below implies that U is connected, clearly a natural condition, and is obviously satisfied by most open sets one might be interested in.

Theorem 6.7.3. *Let $U \subseteq \mathbb{R}^d$ be an open set and $(U_n)_{n\in\mathbb{N}}$ a sequence of convex closed sets such that $U_n \subset U$, $U_n \subseteq U_{n+1}$ for all $n \in \mathbb{N}$ and $\bigcup_{n\in\mathbb{N}} U_n = U$. Assume that $f : U \to M_{dm}(\mathbb{R})$ is a locally Lipschitz function and $Z = (Z_t)_{t\in\mathbb{R}^+}$ is an m-dimensional semimartingale. Then for each U-valued $\mathcal{F}_0$-measurable initial value X_0 there exist a stopping time T and a unique U-valued solution X to the stochastic differential equation*

$$dX_t = f(X_{t-})dZ_t \tag{6.7.2}$$

up to the time T, i.e. on the stochastic interval $[0, T)$.

On $T < \infty$ we have that either X hits the boundary ∂U of U at T, i.e. $X_{T-} \in \partial U$, jumps out of the set U, i.e. $X_{T-} \in U$ and $X_{T-} + f(X_{T-})\Delta Z_T \in \mathbb{R}^d\backslash U$, or explodes, i.e. $\limsup_{t\to T, t<T} \|X_t\| = \infty$. If f satisfies the linear growth condition

$$\|f(x)\|^2 \leq K(1 + \|x\|^2) \tag{6.7.3}$$

with some constant $K \in \mathbb{R}^+$, then no explosion can occur.

Proof. The idea of the following proof is to first obtain unique solutions on the sets $U_n \cap \{x \in \mathbb{R}^d : \|x\| \leq n\}$ up to stopping times T_n. This sequence of solutions is then utilized as in standard analysis to establish the existence of a unique solution on the whole of U up to a stopping time.

In the following we presume that $\|\cdot\|$ denotes the Euclidean norm and the induced operator norm, as we shall need a Hilbert space structure. Note that due to the equivalence of all norms on finite dimensional spaces this can be done without loss of generality.

First we set $W_n = U_n \cap \{x \in \mathbb{R}^d : \|x\| \leq n\}$ for all $n \in \mathbb{N}$. It is immediate that all W_n are compact convex sets, $W_{n+1} \supseteq W_n$ and $\bigcup_{n\in\mathbb{N}} W_n = U$.

Let $\Pi_{W_n} : \mathbb{R}^d \to W_n$ for all $n \in \mathbb{N}$ denote the unique orthogonal projection onto the set W_n (see e.g. Werner (2002, Satz V.3.2)). For later use recall that Π_{W_n} is a contraction, i.e. $\|\Pi_{W_n}(x) - \Pi_{W_n}(y)\| \leq \|x - y\| \; \forall\, x, y \in \mathbb{R}^d$ (Werner (2002, p. 215)).

Consider now the functions $f_n := f \circ \Pi_{W_n}$ defined on $\mathbb{R}^d$ and the stochastic differential equations

$$dX_{n,t} = f_n(X_{n,t-})dZ_t,\; t \in \mathbb{R}^+ \;\; X_{n,0} = X_0, \tag{6.7.4}$$

on $\mathbb{R}^d$. It is clear that any solution $X_{n,t}$ to (6.7.4) must equal a solution X_t to (6.7.2), as long as it stays in W_n, and vice versa, since f and f_n agree on W_n.

Since the function f is Lipschitz on the compact sets W_n, we have from Proposition 6.7.2

$$\|f_n(x) - f_n(y)\| \leq C(W_n)\|\Pi_{W_n}(x) - \Pi_{W_n}(y)\| \leq C(W_n)\|x - y\| \;\; \forall\, x, y \in \mathbb{R}^d$$

and hence the functions f_n are globally Lipschitz on $\mathbb{R}^d$. Moreover, the global Lipschitz property implies linear growth. Hence, by Métivier and Pellaumail (1980b, Chapter 3, Sections 6.10, 7.2) or Protter (2004, Chapter V, Theorem 7) the stochastic differential equations (6.7.4) have unique solutions $X_n = (X_{n,t})_{t\in\mathbb{R}^+}$ defined for all times $t \in \mathbb{R}^+$.

Setting $T_n := \inf\{t \in \mathbb{R}^+ : X_{n,t} \notin W_n\}$ gives a sequence of stopping times using Protter (2004, Chapter I, Theorem 3), since $\mathbb{R}^d\backslash W_n$ is open. Of course, we have that X_n stays in

W_n for all $t \in [0, T_n)$, where in the case $T_n = 0$ we understand $[0, T_n) = \emptyset$, and that X_n provides a solution to (6.7.2) on $[0, T_n)$. Moreover, from the increasingness of the sequence of sets $(W_n)_{n\in\mathbb{N}}$ it is clear that for any $n, k \in \mathbb{N}$ the processes X_n and X_{n+k} agree, as long as they stay in W_n, i.e. for all $t \in [0, T_n)$. (Actually an elementary argument gives even $X_{n,T_n} := X_{n,T_n-} + f(X_{n,T_n-})\Delta Z_{T_n} = X_{n+k,T_n}$ for all $k \in \mathbb{N}$, but this value may already be outside of W_n.) This implies that T_n is an increasing sequence of stopping times and that the following is well-defined:

$$X_t := \begin{cases} X_0 & \text{for } t = 0, \\ X_{1,t} & \text{for } t \in (0, T_1), \\ X_{n,t} & \text{for } t \in [T_{n-1}, T_n), n \in \mathbb{N}\backslash\{1\}. \end{cases} \tag{6.7.5}$$

Since $(T_n)_{n\in\mathbb{N}}$ is an increasing sequence, $T := \lim_{n\to\infty} T_n$ is a well-defined, possibly infinite, stopping time. Due to the above construction $X = (X_t)_{t\in[0,T)}$ is a solution of (6.7.2) on $[0, T)$. Therefore the existence has been established.

The uniqueness of the solution X to (6.7.2) in U defined until T is clear, since until $\tilde{T}_n = \inf\{t \in \mathbb{R}^+ : X_t \notin W_n\}$ any solution must agree with the unique solution X_n to (6.7.4) for all $n \in \mathbb{N}$.

It remains to verify the limiting behaviour of X at $t \to T$ for $T < \infty$. But from our construction of the process X it is clear that $T < \infty$ can only happen, if the process X hits the boundary of U, jumps out of U or explodes at time T (all for the first time, of course), as otherwise we would still have to be in the interior of some set W_n with the process X_n at time T and would thus stay in W_n in the time interval $[T, T + \delta)$ for some stopping time $T + \delta > T$ due to the right continuity of X_n, which is a contradiction to the definitions of T_n and T.

That under the linear growth condition (6.7.3) the process X does not explode in finite time can be shown using the arguments of Métivier and Pellaumail (1980a) (see also Métivier and Pellaumail (1980b)). □

Remark 6.7.4. *(i) The stopping time T, until which a solution exists, cannot be zero. The latter would only be possible, if the process could jump out of U immediately after time 0, but this is not possible, since both the driving semi-martingale Z and the solution X are càdlàg and thus in particular continuous from the right.*

(ii) It is obvious that the above proof can easily be extended to non-autonomous random f, i.e. functions f depending also on time t and being random, as considered in Métivier and Pellaumail (1980b), for instance. We have opted not to state our theorem in this generality, as the generalization should be obvious and would only make the above proof notationally very cumbersome. The same comment applies to SDEs in general Hilbert spaces.

A combination of Theorems II.6 and V.7 in Protter (2004) implies immediately the following result.

Proposition 6.7.5. *Assume that for the sequence of stopping times $(T_n)_{n\in\mathbb{N}}$ constructed in the proof of Theorem 6.7.3 one has that $\lim_{n\to\infty} T_n = \infty$ almost surely. Then the unique solution X on $\mathbb{R}^+$ to the SDE (6.7.2) is a semi-martingale.*

The following convergence result adds some more insight into the nature of the approximation done.

Proposition 6.7.6. *Assume that for the sequence of stopping times $(T_n)_{n\in\mathbb{N}}$ constructed in the proof of Theorem 6.7.3 one has that $\lim_{n\to\infty} T_n = \infty$ almost surely. Then the sequence of processes $(X_{n,t})_{t\in\mathbb{R}^+}$ converges uniformly on compacts in probability (ucp) to $(X_t)_{t\in\mathbb{R}^+}$ for $n \to \infty$.*

Proof. For any $t \in \mathbb{R}^+$ and $\epsilon > 0$ the construction implies $P(\sup_{0\le s\le t} \|X_s - X_{n,s}\| > \epsilon) \le P(T_n \le t)$. Observing that $\lim_{n\to\infty} P(T_n \le t) = 0$ concludes. □

6.7.1.2. Markovian properties

The aim of this subsection is to show that the solution of the SDE (6.7.2) exhibits important Markovian properties, if the driving semi-martingale Z is actually a Lévy process L.

Throughout the remainder of this section we again assume that $U \subseteq \mathbb{R}^d$ is an open set such that there is a sequence of convex closed sets $(U_n)_{n\in\mathbb{N}}$ with $U_n \subset U$, $U_n \subseteq U_{n+1}$ for all $n \in \mathbb{N}$ and $\bigcup_{n\in\mathbb{N}} U_n = U$. Furthermore, $f : U \to M_{dm}(\mathbb{R})$ is a locally Lipschitz function and $L = (L_t)_{t\in\mathbb{R}^+}$ is an m-dimensional Lévy process. Then we consider the stochastic differential equation

$$dX_t = f(X_{t-})dL_t \tag{6.7.6}$$

and assume that for any initial value $X_0 = x \in U$ the stopping time T (which depends on x, of course) constructed in the proof of Theorem 6.7.3 is infinite (almost surely). Furthermore, we assume that the sequence $W_n = U_n \cap \{x \in \mathbb{R}^d : \|x\| \le n\}$ satisfies $W_n \subset W_{n+1}^\circ$ for all $n \in \mathbb{N}$ where W_n° denotes the open kernel of W_n.

Let us briefly recall some standard notions regarding Markov properties. For further details on Markov processes we refer to any of the standard books (e.g. Dynkin (1965), Ethier and Kurtz (1986), Gihman and Skorohod (1975)).

Definition 6.7.7. *Let $Z = (Z_t)_{t\in\mathbb{R}^+}$ be a process with values in U which is adapted to a filtration $(\mathcal{F}_t)_{t\in\mathbb{R}^+}$.*

(i) Z is called a Markov process *with respect to $(\mathcal{F}_t)_{t\in\mathbb{R}^+}$, if*

$$E(g(Z_u)|\mathcal{F}_t) = E(g(Z_u)|Z_t)$$

for all $t \in \mathbb{R}^+, u \ge t$ and $g : U \to \mathbb{R}$ bounded and Borel measurable.

(ii) Let Z be a Markov process and define for all $s, t \in \mathbb{R}^+, s \le t$ the transition functions $P_{s,t}(Z_s, g) = E(g(Z_t)|\mathcal{F}_s)$ with $g : U \to \mathbb{R}$ bounded and Borel measurable. If $P_{s,t} = P_{0,t-s} =: P_{t-s}$ for all $s, t \in \mathbb{R}^+, s \le t$, Z is said to be a time homogeneous Markov process.

(iii) A time homogeneous Markov process is called a strong Markov *process, if $E(g(Z_{T+s})|\mathcal{F}_T) = P_s(Z_T, g)$ for all $g : U \to \mathbb{R}$ bounded and Borel measurable and a.s. finite stopping times T.*

As stated in the beginning, we are assuming given an appropriate filtered probability space $(\Omega, \mathcal{F}, P, (\mathcal{F}_t)_{t\in\mathbb{R}^+})$. In order to be able to have arbitrary initial conditions we need to enlarge our probability space as in Protter (2004, p. 293). Define thus

$$\overline{\Omega} = U \times \Omega, \qquad \overline{\mathcal{F}}_t^0 = \boldsymbol{\sigma}(\mathcal{U} \times \mathcal{F}_t), \qquad \overline{P}^y = \delta_y \times P$$

where $\mathcal{U}$ is the Borel σ-algebra of U, $\boldsymbol{\sigma}(\mathcal{U} \times \mathcal{F}_t)$ is the σ-algebra generated by $\mathcal{U} \times \mathcal{F}_t$ and δ_y the Dirac measure with respect to $y \in U$. In order to have the usual conditions satisfied

again, let $\overline{\mathcal{F}}_t = \bigcap_{u>t} \overline{\mathcal{F}}^0_u$. A random variable Z defined on Ω is extended to $\overline{\Omega}$ by setting $Z((y,\omega)) = Z(\omega)$ for all $(y,\omega) \in \overline{\Omega}$. Furthermore, we define the random variable X_0 by $X_0((y,\omega)) = y$.

Theorem 6.7.8. *Under the above assumptions the unique solution $(X_t)_{t\in\mathbb{R}^+}$ to*

$$X_t = X_0 + \int_0^t f(X_{t-})dL_t$$

is a temporally homogeneous strong Markov process on U under all $\overline{P}^y$ with $y \in U$.

Proof. Using the following lemmata ensuring the continuity of the flow and the joint measurability, the proof of Protter (2004, Theorem V.32) generalizes in a straightforward manner to our set-up noting that the independence of the components of the Lévy process assumed there is actually not needed in the proof. □

In the following we denote the space of all càdlàg functions from $\mathbb{R}^+$ to U by D_U and equip this space with the topology of uniform convergence on compacts. For an SDE $X_t = x + \int_0^t f(X_{s-})dZ_s$ having a unique solution $(X_t)_{t\in\mathbb{R}^+}$ in U for all $x \in U$ the function $\phi : U \to D_U$ mapping the initial condition to the solution is called the flow of the SDE.

Lemma 6.7.9. *Under the above assumptions the flow of*

$$X_t = x + \int_0^t f(X_{t-})dL_t$$

is almost surely a continuous mapping from U to D_U (with respect to the topology of uniform convergence on compacts).

Proof. The proof follows by the same arguments as in Protter (2004, Theorem V.38) using the functions f_n constructed in the proof of Theorem 6.7.3 instead of g_l, the sets W_n° instead of Λ_l and observing that $T_n^\circ(x) := \inf\{t \in \mathbb{R}^+ : X_{n,t}(x) \notin W_n^\circ \text{ or } X_{n,t-}(x) \notin W_n^\circ\}$ increases almost surely to infinity for all $x \in U$. To see the last claim, it suffices to note that by the above made assumptions the sequence $T_n(x) := \inf\{t \in \mathbb{R}^+ : X_{n,t}(x) \notin W_n\}$ increases almost surely to infinity and $T_n(x) \le T_{n+1}^\circ(x) \le T_{n+1}(x)$. □

In the next lemma we denote by $X = X(x,t,\omega)$ the mapping from $U \times \mathbb{R}^+ \times \Omega$ to U which maps (x,t,ω) to the solution of $X_t = x + \int_0^t f(X_{t-})dL_t$ at time $t \in \mathbb{R}^+$ for $\omega \in \Omega$. Moreover, $\mathcal{B}(\mathbb{R}^+)$ stands for the Borel σ-algebra over $\mathbb{R}^+$.

Lemma 6.7.10. *Under the above assumptions there exists a version of $X : U\times\mathbb{R}^+\times\Omega \to U$, $(x,t,\omega) \mapsto X(x,t,\omega)$ that is $\boldsymbol{\sigma}(\mathcal{U} \times \mathcal{B}(\mathbb{R}^+) \times \mathcal{F})$-measurable and such that $X(x,\cdot,\cdot)$ is an adapted càdlàg solution to $X_t = x + \int_0^t f(X_{t-})dL_t$ for each $x \in U$.*

Proof. Protter (2004, Theorem V.31) implies the existence of mappings $X_n(x,t,\omega)$ which are $\boldsymbol{\sigma}(\mathcal{U} \times \mathcal{B}(\mathbb{R}^+) \times \mathcal{F})$-measurable and such that $X_n(x,\cdot,\cdot)$ is an adapted càdlàg solution to $X_{n,t} = x + \int_0^t f_n(X_{n,t-})dL_t$ for each $x \in U$. Combining Proposition 6.7.6 with Protter (2004, Theorem IV.62) concludes. □

Another important property of Markov processes is the weak Feller property. Since there are various variants of the definition in the literature, we start with our definition. The definition we shall give agrees with that of da Prato and Zabczyk (1996) and it has also been given in Dynkin (1965) and Gihman and Skorohod (1975), for instance.

For a Markov process Z in U, $x \in U$ and a set $A \in \mathcal{U}$ we set $P_s(x, A) = E(I_A(Z_s)|Z_0 = x)$ and note that this is the transition probability for the Markov process X to move from the initial state x to the set A at time $s \in \mathbb{R}^+$. $P_s(\cdot, \cdot)$ is referred to as the Markov transition function. Moreover we denote by $B_b(U)$ the space of all measurable and bounded functions $f : U \to \mathbb{R}$ and by $C_b(U)$ the set of all functions $f : U \to \mathbb{R}$ that are bounded and continuous.

The Markov transition function $P_s(\cdot, \cdot)$ gives rise to a Markov (transition) semi-group of bounded linear operators $P_s : B_b(U) \to B_b(U)$ with $s \in \mathbb{R}^+$ by setting $P_s f(x) = \int_U f(y) P_s(x, dy)$ for all $f \in B_b(U)$ and $x \in U$. Moreover, we define $\mathcal{M}_1(U)$ to be the set of all probability measures on $(U,\mathcal{U})$ and operators P_s^* for $s \in \mathbb{R}^+$ on $\mathcal{M}_1(U)$ by setting $P_s^* \mu(A) = \int_U P_s(x, A)\mu(dx)$.

Definition 6.7.11. *Let $(P_s)_{s\in\mathbb{R}^+}$ be the transition semi-group of a time homogeneous Markov process Z on U.*

(i) $(P_s)_{s\in\mathbb{R}^+}$ (respectively the associated Markov process Z) is called stochastically continuous*, if*

$$\lim_{t\to 0, t\geq 0} P_t(x, U(x)) = 1$$

for all $x \in U$ and open neighbourhoods $U(x)$ of x.

(ii) A stochastically continuous semi-group $(P_s)_{s\in\mathbb{R}^+}$ (respectively its associated Markov process Z) is called weakly Feller*, if*

$$P_s(C_b(U)) \subseteq C_b(U) \text{ for all } s \in \mathbb{R}^+.$$

(iii) A probability measure $\mu \in \mathcal{M}_1(U)$ is said to be an invariant (stationary) *measure for the Markovian semi-group $(P_s)_{s\in\mathbb{R}^+}$ (respectively for its associated Markov process Z), if $P_s^* \mu = \mu$ for all $s \in \mathbb{R}^+$.*

Remark 6.7.12. *Regarding our general definitions for Markov processes the open set $U \subseteq \mathbb{R}^d$ can be replaced by any Polish space.*

Having set up the necessary technical details we can return to our stochastic differential equation.

Proposition 6.7.13. *Under the above assumptions the solution to*

$$X_t = x + \int_0^t f(X_{t-}) dL_t$$

with $x \in U$ is a weak Feller process on U.

Proof. Above we have already shown that the process is a time homogeneous Markov process. In the following we denote by $(P_s)_{s\in\mathbb{R}^+}$ its transition semi-group. Since the paths of X are càdlàg, the process is stochastically continuous. Let now f be a function in $C_b(U)$ and $t \in \mathbb{R}^+$. Then $X_t(x)$ is a.s. a continuous function of the initial value x by Lemma 6.7.9. Hence, $f(X_t(x))$ is a.s. a continuous function of x. Thus the boundedness of f ensures that $P_t f(x) = E(f(X_t(x)))$ is a continuous function of $x \in U$ and obviously $\sup_{x\in U} |P_t f(x)| \leq \sup_{x\in U} |f(x)|$ for all $t \in \mathbb{R}^+$. □

6.7.2. Lipschitz and growth properties of the SDE defining the MUCOGARCH(1,1) volatility process

This appendix deals with some matrix analytic results analysing the Lipschitz properties of the map $V \mapsto V^{1/2} \otimes V^{1/2}$ used in the definition of the MUCOGARCH(1,1) volatility process. We denote by $\|\cdot\|_2$ the operator norm associated with the usual Euclidean norm on $\mathbb{R}^d$.

Lemma 6.7.14 (Bhatia (1997, Problem I.6.11))**.** *For all $A, B \in M_d(\mathbb{R})$ we have*

$$\|A \otimes A - B \otimes B\|_2 \leq 2 \max\{\|A\|_2, \|B\|_2\} \|A - B\|_2.$$

In particular, the mapping $\otimes : M_d(\mathbb{R}) \to M_{d^2}(\mathbb{R})$, $X \mapsto X \otimes X$ is uniformly Lipschitz on any set of the form $\{x \in M_d(\mathbb{R}) : \|x\| \leq c\}$ with $c > 0$.

The proof is obvious from the ideas given there.

Lemma 6.7.15 (Bhatia (1997, p. 305))**.** *Let $A, B \in \mathbb{S}_d^+$ and $a > 0$ such that $A, B \geq aI_d$. Then*

$$\|A^{1/2} - B^{1/2}\|_2 \leq \frac{1}{2\sqrt{a}} \|A - B\|_2.$$

Hence, the mapping $\mathbb{S}_d^+ \to \mathbb{S}_d^+$, $X \mapsto X^{1/2}$ is uniformly Lipschitz on any set of the form $\{x \in \mathbb{S}_d^+ : x \geq cI_d\} \subset \mathbb{S}_d^{++}$ with $c > 0$.

For a variant of the above statement see Horn and Johnson (1991, p. 557).

Lemma 6.7.16. *Consider the map $F : \mathbb{S}_d^+ \to \mathbb{S}_d^+$, $X \mapsto X^{1/2} \otimes X^{1/2} = (X \otimes X)^{1/2}$. F is continuous and uniformly Lipschitz on any set of the form $\{x \in \mathbb{S}_d^+ : x \geq cI, \|x\| \leq \tilde{c}\}$ with $c, \tilde{c} > 0$. Moreover, we have that $\|A^{1/2} \otimes A^{1/2}\|_2 = \|A\|_2$ for all $A \in \mathbb{S}_d$.*

Proof. The identity $X^{1/2} \otimes X^{1/2} = (X \otimes X)^{1/2}$ is an immediate consequence of basic properties of the tensor product (see Horn and Johnson (1991, Ch. 4)) and the continuity of F follows from the one of the tensor product and the positive definite square root (see Horn and Johnson (1991, Theorem 6.2.37)). The Lipschitz property follows from a combination of the previous two lemmas. Finally, $\|A^{1/2} \otimes A^{1/2}\|_2 = \|A\|_2$ is established by noting that $\|A^{1/2} \otimes A^{1/2}\|_2 = \|A^{1/2}\|_2^2$ (cf. Bhatia (1997, p. 15)) and $\|A^{1/2}\|_2 = \|A\|_2^{1/2}$. The latter follows immediately from the fact that $\|A\|_2 = \sqrt{\rho(A^*A)} = \rho(A)$ for all $A \in \mathbb{S}_d^+$. □

Finally we show that the global Lipschitz property is not satisfied for this map, not even if we restrict it to sets being bounded away from zero.

Lemma 6.7.17. *For the map F defined in the previous lemma there exists no finite $K \in \mathbb{R}^+$ such that*

$$\|F(x) - F(y)\| \leq K\|x - y\| \tag{6.7.7}$$

for all $x, y \in \mathbb{S}_d^{++}$. The same holds for all $x, y \in \{z \in \mathbb{S}_d : z \geq C\}$ with arbitrary $C \in \mathbb{S}_d^{++}$.

Proof. From the following proof it is clear that we can take $d = 2$ w.l.o.g. Let $x = \operatorname{diag}(x_1, x_2)$ and $y = \operatorname{diag}(y_1, y_2)$ with $x_1, x_2, y_1, y_2 \in \mathbb{R}^{++}$ and $\operatorname{diag}(x_1, x_2)$ being as usual the diagonal matrix with diagonal entries x_1 and x_2. We have $F(x) = \operatorname{diag}(x_1, \sqrt{x_1 x_2}, \sqrt{x_1 x_2}, x_2)$. Assume (6.7.7) is true with a finite $K \in \mathbb{R}^+$. Then there is a finite $k \in \mathbb{R}^+$

such that $|\sqrt{x_1x_2} - \sqrt{y_1y_2}| \leq k\,(|x_1 - y_1| + |x_2 - y_2|)$ for all $x_1, x_2, y_1, y_2 \in \mathbb{R}^{++}$. Choosing $x_2 = y_2 = 1$ this gives $|\sqrt{x_1} - \sqrt{y_1}| \leq k|x_1 - y_1|$ for all $x_1, y_1 \in \mathbb{R}^{++}$ which is a contradiction to the well-known fact that the square root is not globally Lipschitz on $\mathbb{R}^{++}$. Regarding the case $x, y \in \{z \in \mathbb{S}_d : z \geq C\}$ we can w.l.o.g. restrict ourselves to $C = cI_d$ with $c \in \mathbb{R}^{++}$. Choosing $x_2 = y_2$, $x_1 = 9c$ and $y_1 = 4c$ gives $|\sqrt{cx_2}| \leq 5kc$. As x_2 can be taken arbitrarily large this is a contradiction. □

6.7.3. Auxiliary results regarding the calculation of moments

In the following we present some elementary but important lemmata used in the calculation of the second order structure of MUCOGARCH(1,1) processes. Observe that we state some of them for processes in $\mathbb{R}^d$ for notational ease, but the generalization to the matrix case is immediate.

The first three lemmata are stated without a proof, because they follow immediately from the stochastic continuity properties of Lévy processes.

Lemma 6.7.18. *Let $s, t \in \mathbb{R}^+$ with $t > s$ and let $(X_r)_{r\in\mathbb{R}^+}$ be an adapted càdlàg process in $\mathbb{R}^d$ which satisfies $\Delta X_r = A_r \Delta L_r$ for all $r \in \mathbb{R}^+$ with $(A_r)_{r\in\mathbb{R}^+}$ being a process taking values in $M_d(\mathbb{R})$ such that A_r is $\mathcal{F}_{r-}$ measurable and $(L_r)_{r\in\mathbb{R}^+}$ being an $\mathbb{R}^d$-valued Lévy process. (Here $\mathcal{F}_{r-}$ denotes the σ-algebra generated by $\cup_{0\leq u<r}\mathcal{F}_r$.) Let furthermore $f : \mathbb{R}^d \to \mathbb{R}^d$ be a Borel-measurable function such that $E(\|f(X_r)\|^2)) < \infty$ for all $r \in \mathbb{R}^+$. Assume finally that $E(\|X_r\|^2) < \infty$ for all $r \in \mathbb{R}^+$. Then $E(f(X_s)X_{t-}^*) = E(f(X_s)X_t^*)$ and these expectations are finite.*

Applied to the MUCOGARCH(1,1) volatility process this lemma and an obvious variant of it imply immediately the following, since $\Delta Y_t = A(C + Y_{t-})\Delta L_t(\Delta L_t)^*(C + Y_{t-})A^*$ and $\sum_{0\leq s\leq t} \Delta L_t(\Delta L_t)^*$ is a Lévy process.

Lemma 6.7.19. *Let $s, t \in \mathbb{R}^+$ with $t > s$ and let Y be the volatility process of the MUCOGARCH(1,1) process as given in (6.3.4). Be $f : W \to W$ a Borel-measurable function defined on $W = \mathbb{S}_d$ or $W = \mathbb{R}^{d^2}$, respectively, such that $E(\|f(Y_r)\|^2)) < \infty$ or $E(\|f(\mathrm{vec}(Y_r))\|^2)) < \infty$, respectively, for all $r \in \mathbb{R}^+$. Assume furthermore that the process Y has a finite second moment.*

Then $E(f(Y_s)Y_{t-}) = E(f(Y_s)Y_t)$ or $E(f(\mathrm{vec}(Y_s))\mathrm{vec}(Y_{t-})^*) = E(f(\mathrm{vec}(Y_s))\mathrm{vec}(Y_t)^*)$, respectively, and these expectations are finite.

Lemma 6.7.20. *Let $t \in \mathbb{R}^+$ and let Y be the volatility process of the MUCOGARCH(1,1) process as given in (6.3.4). Be f a measurable (with respect to the Borel-σ-algebra) function defined on $\mathbb{S}_d$ such that $E(\|f(Y_t)\|)) < \infty$.*

Then $E(f(Y_{t-})) = E(f(Y_t))$ and this expectation is finite.

Finally let us state a version of the so called compensation formula (cf. Kyprianou (2006, Section 4.3.2), for instance) which is suitable for our purposes.

Lemma 6.7.21. *Assume that $(X_t)_{t\in\mathbb{R}^+}$ is an adapted càdlàg process in $M_d(\mathbb{R})$ satisfying $E(\|X_t\|) < \infty$ for all $t \in \mathbb{R}^+$, $t \mapsto E(\|X_t\|)$ is locally bounded and that $(L_t)_{t\in\mathbb{R}^+}$ is a driftless pure jump Lévy process in $\mathbb{R}^d$ of finite variation with finite expectation $E(L_1)$. Then $E\left(\int_0^t X_{s-}dL_s\right) = \int_0^t E(X_{s-})E(L_1)ds$ for $t \in \mathbb{R}^+$.*

Proof. Since $L_t = \int_0^t \int_{\mathbb{R}^d} z\mu_L(ds, dz)$ the compensation formula (see Kyprianou (2006, Section 4.3.2)) implies

$$\begin{aligned} E\left(\int_0^t X_{s-} dL_s\right) &= E\left(\int_0^t \int_{\mathbb{R}^d} X_{s-} z\mu_L(ds, dz)\right) = E\left(\int_0^t X_{s-} \int_{\mathbb{R}^d} z\nu_L(dz)ds\right) \\ &= E\left(\int_0^t X_{s-} E(L_1)ds\right). \end{aligned}$$

Observing that $\int_0^t E(\|X_{s-}\|)\|E(L_1)\|ds$ is finite for every $t \in \mathbb{R}^+$ an application of Fubini's theorem concludes the proof. □

7. Absolute Moments of Generalized Hyperbolic Distributions and Approximate Scaling of Normal Inverse Gaussian Lévy Processes[1]

7.1. Introduction

The generalized hyperbolic (GH) distribution was introduced in Barndorff-Nielsen (1977) in connection to a study of the grain-size distribution of wind-blown sand. Since then it has been used in many different areas. Before outlining the contents of the present chapter, we now give a brief overview of various applications of the GH laws and the associated Lévy processes.

The original paper by Barndorff-Nielsen (1977) was focused on the special case of the hyperbolic law. That law and its applicability have been further discussed, inter alia, in Barndorff-Nielsen, Blæsild, Jensen and Sørensen (1983), Barndorff-Nielsen, Blæsild, Jensen and Sørensen (1985) (particle size distributions of sand), Barndorff-Nielsen and Christiansen (1988), Hartmann and Bowman (1993), Sutherland and Lee (1994) and references therein (coastal sediments), Xu, Durst and Tropea (1993) and their list of references (fluid sprays). In Barndorff-Nielsen (1982) the appearance of the three dimensional hyperbolic law in relativistic statistical physics was pointed out. Other areas, where hyperbolic distributions have been employed, include biology (e.g. Blæsild (1981)) and primary magnetization of lava flows (cf. Kristjansson and McDougall (1982)). Furthermore, in Barndorff-Nielsen, Jensen and Sørensen (1989) the hyperbolic distribution is employed to model wind shear data of landing aircrafts parsimoniously. See also Barndorff-Nielsen (1979) for applications in turbulence.

Moreover, Barndorff-Nielsen, Blæsild and Schmiegel (2004) recently demonstrated (following an indication in Barndorff-Nielsen (1998a)) that the normal inverse Gaussian (NIG) law, another important special case of the GH law, is capable of describing velocity data from turbulence experiments with high accuracy. Eriksson, Forsberg and Ghysels (2004) employ the NIG distribution to approximate other (unknown) probability distributions.

In recent years many authors have successfully fitted generalized hyperbolic distributions and in particular normal inverse Gaussian (NIG) laws to returns in financial time series; see Eberlein and Keller (1995), Prause (1997, 1999), Barndorff-Nielsen (1997), Barndorff-Nielsen and Shephard (2001a, 2001b, 2007) and references therein, Schoutens (2003). Benth and Šaltytė-Benth (2005) have recently put forth a model for Norwegian temperature data driven by a GH Lévy process.

[1] The contents of this chapter appeared in Barndorff-Nielsen, O.E. and Stelzer, R. (2005), Absolute Moments of Generalized Hyperbolic Distributions and Approximate Scaling of Normal Inverse Gaussian Lévy Processes, *Scand. J. Statist.*, **32**(4), 617–637

This has, in particular, led to modelling the time dynamics of financial markets by stochastic processes using generalized hyperbolic or normal inverse Gaussian laws and associated Lévy processes as building blocks (e.g. Rydberg (1997), Bibby and Sørensen (1997), Barndorff-Nielsen (1998b), Rydberg (1999), Prause (1999), Raible (2000), Barndorff-Nielsen (2001a), Barndorff-Nielsen and Shephard (2001a, 2007) and references therein, Eberlein (2001), Schoutens (2003), Cont and Tankov (2004), Rasmus, Asmussen and Wiktorsson (2004), Emmer and Klüppelberg (2004) and Mencia and Sentana (2004)).

One of the reasons, why the GH distribution is used in such a variety of situations, is that it is not only flexible enough to fit many different data sets well, but also is rather tractable analytically and many important properties (density, characteristic function, cumulant transform, Lévy measure, ...) are known. Some of these properties are recalled in Section 7.2. Yet, until now, no details on absolute moments of arbitrary order $r > 0$ are known, except for $r = 1$. Thus we derive, in Section 7.3, formulae for the (absolute) moments of arbitrary order $r > 0$ of the generalized hyperbolic distribution in terms of moments of the corresponding symmetric GH law. For μ–centred (absolute) moments, i.e. moments centred at the location parameter μ, we are able to give explicit formulae using Bessel functions. From these general formulae we will then, as special cases, obtain formulae for the absolute moments of the NIG law and NIG Lévy process. We especially focus on the NIG case because of its wide applicability and tractability. In particular, the NIG Lévy process has a marginal NIG distribution at all times, an appealing feature not shared by the general GH Lévy process.

Due to Kolmogorov's famous laws for homogeneous and isotropic turbulence (see, for instance, Frisch (1995)) scaling is an important issue when considering turbulence data and models. Ongoing research indicates that the time transformation carried out in Barndorff-Nielsen et al. (2004) leads to a process with NIG marginals and very strong apparent scaling.

The question of the possible relevance of scaling in finance was raised by Mandelbrot (1963) and has since been discussed by a number of authors, see, in particular, Müller, Dacorogna, Olsen, Pictet, Schwarz and Morgenegg (1990), Guillaume et al. (1997) and Mandelbrot (1997). More recently the question was taken up by Barndorff-Nielsen and Prause (2001), who showed that an NIG Lévy process may exhibit moment behaviour which is very close to scaling. However, they solely studied the first absolute moment and obtained analytic results only in the case of a symmetric NIG Lévy process. As part of this chapter we generalize their findings to the skewed case and higher order moments.

Based on the explicit general formulae for μ-centred moments of the NIG law, resp. NIG Lévy process, we are able to deduce analytic results for the approximate scaling of an NIG Lévy process, namely explicit expressions for the derivative of the logarithm of the μ-centred absolute moments of the NIG Lévy process with respect to the logarithm of time.

In the final sections we discuss the numerical implementation of the formulae obtained and give numerical examples for the apparent scaling present in NIG Lévy processes.

Our results show that, in particular, the occurrence of empirical scaling laws is not bound to necessitate the use of self-similar or even multifractal processes for modelling, since this type of behaviour is already exhibited by such simple a model as an NIG Lévy process, when looking at the relevant time horizons. For a survey of the theory and (approximate) occurrence of self-similarity and scaling see Embrechts and Maejima (2002).

Throughout this chapter $\mathbb{R}_{>0}$ denotes the strictly positive real numbers and $\mathbb{R}_{\geq 0}$ the

positive real numbers including zero.

7.2. Generalized hyperbolic and inverse Gaussian distributions

In this chapter we consider the class of one-dimensional generalized hyperbolic (GH) distributions and the subclass of normal inverse Gaussian (NIG) distributions in particular. The GH law was, as already noted, introduced in Barndorff-Nielsen (1977) and its properties were further studied in Barndorff-Nielsen (1978a), Barndorff-Nielsen and Blæsild (1981) and Blæsild and Jensen (1981). Some recent results, in particular regarding the multivariate GH laws, can be found in Prause (1999), Eberlein (2001), Eberlein and Hammerstein (2004) and Mencia and Sentana (2004).

We denote the (one-dimensional) GH distribution by $GH(\nu, \alpha, \beta, \mu, \delta)$ and characterize it via its probability density given by:

$$\begin{aligned} p(x;\nu,\alpha,\beta,\mu,\delta) &= \frac{\bar{\gamma}^{\nu}\bar{\alpha}^{1/2-\nu}}{\sqrt{2\pi}\delta K_{\nu}(\bar{\gamma})}\left(1+\frac{(x-\mu)^2}{\delta^2}\right)^{\nu/2-1/4} \\ &\quad \cdot K_{\nu-1/2}\left(\bar{\alpha}\sqrt{1+\frac{(x-\mu)^2}{\delta^2}}\right)e^{\beta(x-\mu)} \end{aligned} \tag{7.2.1}$$

for $x \in \mathbb{R}$ and where the parameters satisfy $\nu \in \mathbb{R}$, $0 \leq |\beta| < \alpha$, $\mu \in \mathbb{R}$, $\delta \in \mathbb{R}_{>0}$, and $\gamma := \sqrt{\alpha^2-\beta^2}$, $\bar{\alpha} := \delta\alpha$, $\bar{\beta} := \delta\beta$, $\bar{\gamma} := \delta\gamma$. Here α can be interpreted as a shape, β as a skewness, μ as a location and δ as a scaling parameter; finally, ν characterizes subclasses and primarily influences the tail behaviour.

Furthermore, $K_{\nu}(\cdot)$ denotes the modified Bessel function of the third kind and order $\nu \in \mathbb{R}$. For a comprehensive discussion of Bessel functions of complex arguments see Watson (1952). Jørgensen (1982) contains an appendix listing important properties of Bessel functions of the third kind and related functions. Most of these properties can also be found in standard reference books like Gradshteyn and Ryzhik (1965) or Bronstein, Semendjaev, Mühlig and Musiol (2000). For the following we need to know that K_{ν} is defined on the positive half plane $D = \{z \in \mathbb{C} : \Re(z) > 0\}$ of the complex numbers and is holomorphic on D. From Watson (1952, p. 182) or Jørgensen (1982, p. 170) we have the representation

$$K_{\nu}(z) = \frac{1}{2}\int_0^{\infty} y^{\nu-1}e^{-\frac{1}{2}z(y+y^{-1})}dy, \tag{7.2.2}$$

which shows the strict positivity of K_{ν} on $\mathbb{R}_{>0}$. The substitution $x := y^{-1}$ immediately gives $K_{-\nu} = K_{\nu}$. Furthermore, $K_{\nu}(z)$ is obviously monotonically decreasing in z on $\mathbb{R}_{>0}$. From the alternative representation

$$K_{\nu}(z) = \int_0^{\infty} e^{-z\cosh(t)}\cosh(\nu t)dt \tag{7.2.3}$$

(cf. Watson (1952, p. 181)) one reads off that, for fixed $z \in \mathbb{R}_{>0}$, $K_{\nu}(z)$ is strictly increasing in ν for $\nu \in \mathbb{R}_{\geq 0}$.

There are several popular subclasses contained within the GH laws. For $\nu = 1$ the hyperbolic and for $\nu = -1/2$ the normal inverse Gaussian distributions are obtained. The normal, exponential, Laplace, Variance-Gamma and Student-t distributions are among many others

limiting cases of the GH distribution (cf. Eberlein and Hammerstein (2004) for a comprehensive analysis).

Alternatively one often uses

$$\rho := \frac{\beta}{\alpha} = \frac{\bar{\beta}}{\bar{\alpha}},\ \xi := (1+\bar{\gamma})^{-1/2} \text{ and } \chi := \rho\xi$$

to parametrize the GH law, since these quantities are invariant under location-scale changes and χ, resp. ξ, can be interpreted as a skewness, resp. kurtosis, measure. Moreover, the parameter restrictions imply $0 < |\chi| < \xi < 1$. For fixed ν this gives rise to the use of *shape triangles* as a graphical tool to study generalized hyperbolic distributions (see e.g. Barndorff-Nielsen et al. (1983, 1985), Barndorff-Nielsen and Christiansen (1988), Rydberg (1997) and Prause (1999)).

A very useful representation in law of the generalized hyperbolic distribution can be given using the generalized inverse Gaussian distribution. The generalized inverse Gaussian distribution $GIG(\nu, \delta, \gamma)$ with parameters $\nu \in \mathbb{R}$, $\gamma, \delta \in \mathbb{R}_{\geq 0}$ and $\gamma+\delta > 0$ is the distribution on $\mathbb{R}_{>0}$ which has probability density function

$$\begin{aligned} p(x;\nu,\delta,\gamma) &= \frac{(\gamma/\delta)^\nu}{2K_\nu(\delta\gamma)} x^{\nu-1} \exp\left(-\frac{1}{2}(\delta^2 x^{-1} + \gamma^2 x)\right) \\ &= \frac{\bar{\gamma}^\nu}{2K_\nu(\bar{\gamma})} \delta^{-2\nu} x^{\nu-1} \exp\left(-\frac{1}{2}(\delta^2 x^{-1} + \bar{\gamma}^2 \delta^{-2} x)\right). \end{aligned} \tag{7.2.4}$$

For more information on the GIG law we refer to Jørgensen (1982) and for an interpretation in terms of hitting times to Barndorff-Nielsen, Blæsild and Halgreen (1978). The following normal variance-mean mixture representation of the generalized hyperbolic law holds.

Lemma 7.2.1. *Let $X \sim GH(\nu, \alpha, \beta, \mu, \delta)$, $V \sim GIG(\nu, \delta, \gamma)$ with $\gamma = \sqrt{\alpha^2 - \beta^2}$ and $\varepsilon \sim N(0,1)$, where V and ε are independent, then:*

$$X \stackrel{D}{=} \mu + \beta V + \sqrt{V}\varepsilon.$$

(For a general overview over normal variance-mean mixtures see Barndorff-Nielsen, Kent and Sørensen (1982).)

Furthermore, the cumulant function of the generalized hyperbolic law $X \sim GH(\nu, \alpha, \beta, \mu, \delta)$ is given by

$$\mathrm{K}(\theta \ddagger X) = \frac{\nu}{2}\log\left(\frac{\gamma}{\alpha^2 - (\beta+\theta)^2}\right) + \log\left(\frac{K_\nu\left(\delta\sqrt{\alpha^2-(\beta+\theta)^2}\right)}{K_\nu\left(\delta\sqrt{\alpha^2-\beta^2}\right)}\right) + \theta\mu. \tag{7.2.5}$$

Obviously $\mathrm{K}(\theta \ddagger X)$ is defined for all $\theta \in \mathbb{R}$ with $|\beta+\theta| < \alpha$. From this fact and Barndorff-Nielsen (1978b, Corollary 7.1) we immediately obtain:

Lemma 7.2.2. *Assume $X \sim GH(\nu, \alpha, \beta, \mu, \delta)$. Then $X \in L^p$ for all $p > 0$, i.e. $E(|X|^p)$ exists for all $p > 0$.*

One immediately calculates the expected value and variance of a GH distributed random variate X to be

$$\begin{aligned} E(X) &= \mu + \beta \frac{\delta K_{\nu+1}(\bar{\gamma})}{\gamma K_\nu(\bar{\gamma})}, \\ Var(X) &= \delta^2 \left(\frac{K_{\nu+1}(\bar{\gamma})}{\bar{\gamma} K_\nu(\bar{\gamma})} + \frac{\beta^2}{\gamma^2} \left(\frac{K_{\nu+2}(\bar{\gamma})}{K_\nu(\bar{\gamma})} - \left(\frac{K_{\nu+1}(\bar{\gamma})}{K_\nu(\bar{\gamma})} \right)^2 \right) \right). \end{aligned}$$

Higher order cumulants can also be calculated, but the expressions become more and more complicated.

In the GH law the existence of moments of all orders is combined with semi-heavy tails

$$p(x; \nu, \alpha, \beta, 0, \delta) \sim C|x|^{\nu-1} \exp\left((\mp\alpha + \beta)x\right) \text{ as } x \to \pm\infty \tag{7.2.6}$$

for some constant C.

In Figure 7.1[2] the densities of NIG distributions fitted to turbulent velocity increments (from data set I of Barndorff-Nielsen et al. (2004); for a detailed description of the data see that paper) at different lags are plotted on a logarithmic scale together with histograms of the original data. The graphs were obtained using the programme "hyp" (Blæsild and Sørensen (1992)). They exemplify the rich variety of distributional shapes one can already get from the NIG law. Letting ν vary offers the possibility to get even more different shapes (see e.g. Eberlein and Özkan (2003) for some hyperbolic fits). Note especially the marked differences in the centre of the distributions and the difference in how the asymptote given by (7.2.6) is approached. For a lag of 9000 the shape is already very close to the quadratic one of the Gaussian law.

Moreover, the GH law is infinitely divisible (in fact, self-decomposable) and leads thus to an associated Lévy process. Yet, the GH distribution is not closed under convolution, but the NIG distribution has this property, so that all marginal distributions of a Lévy process associated to an NIG distribution belong to the NIG class.

The above facts and the analytical tractability due the existence of explicit expressions for the density, the cumulant function and related functions makes using the GH law appealing in many different areas, as already pointed out in the introduction.

7.3. Moments and absolute moments of GH laws

In this section we give expressions for different (absolute) moments of arbitrary GH distributions in terms of moments of corresponding symmetric GH distributions. Based upon this we obtain explicit expressions for μ-centred (absolute) moments of GH distributions, employing the variance-mean mixture representation.

Theorem 7.3.1. *Let $X \sim GH(\nu, \alpha, \beta, \mu, \delta)$, $Y \sim GH(\nu, \alpha, 0, \mu, \delta)$, then for every $r > 0$*

[2]The graphs in this figure have been kindly provided by Jürgen Schmiegel, Århus.

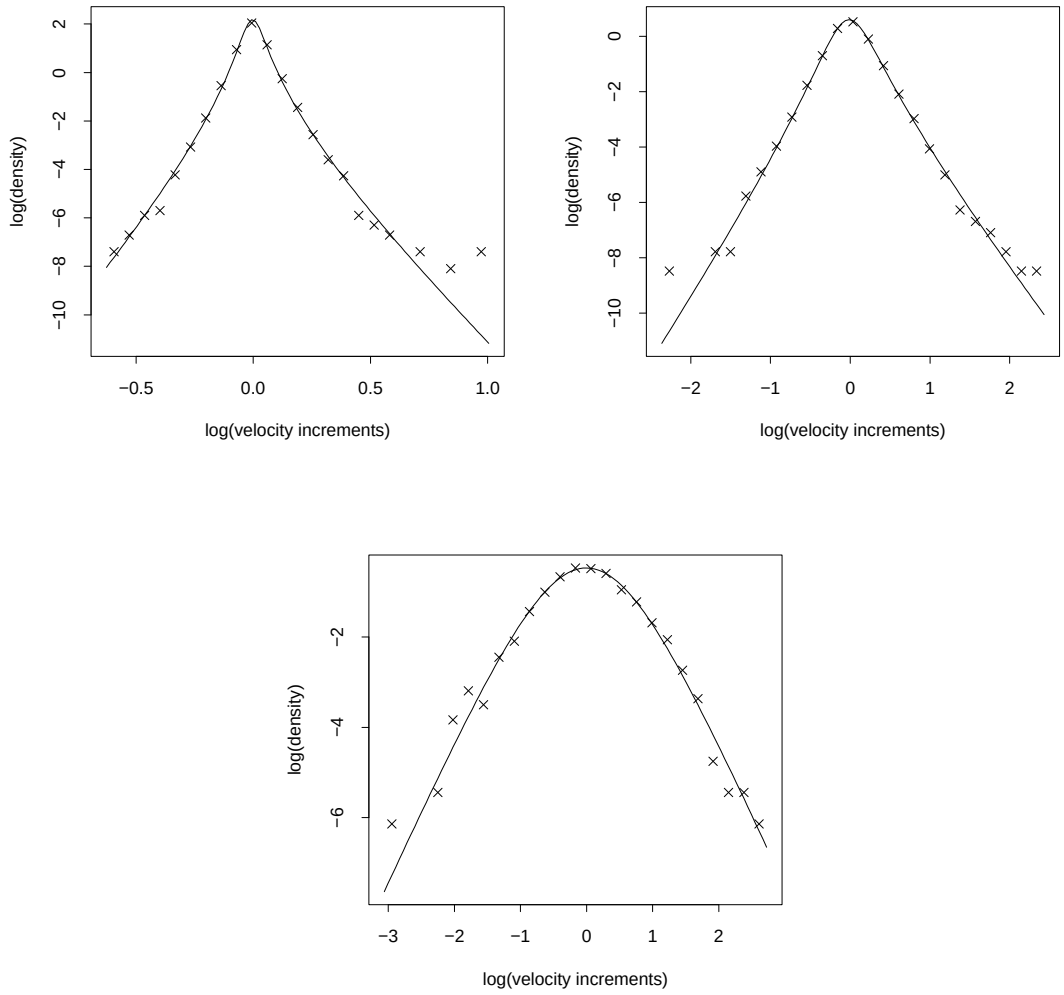


Figure 7.1.: Log-density (solid line) of NIG distribution fitted to velocity increments at lags 12 (upper left), 500 (upper right) and 9000 (lower centre) and log-histogram (only top end points are given (×))

and $n \in \mathbb{N}$*:*

$$(i)\quad E(X^n) = \left(\frac{\bar\gamma}{\bar\alpha}\right)^\nu \frac{K_\nu(\bar\alpha)}{K_\nu(\bar\gamma)} \sum_{k=0}^\infty \frac{\beta^k}{k!} E(Y^n (Y-\mu)^k)$$

$$(ii)\quad E(|X|^r) = \left(\frac{\bar\gamma}{\bar\alpha}\right)^\nu \frac{K_\nu(\bar\alpha)}{K_\nu(\bar\gamma)} \sum_{k=0}^\infty \frac{\beta^k}{k!} E(|Y|^r (Y-\mu)^k)$$

$$(iii)\quad E((X-\mu)^n) = \left(\frac{\bar\gamma}{\bar\alpha}\right)^\nu \frac{K_\nu(\bar\alpha)}{K_\nu(\bar\gamma)} \sum_{k=0}^\infty \frac{\beta^{2k+m}}{(2k+m)!} E((Y-\mu)^{2k+m+n})$$

$$(iv)\quad E(|X-\mu|^r) = \left(\frac{\bar\gamma}{\bar\alpha}\right)^\nu \frac{K_\nu(\bar\alpha)}{K_\nu(\bar\gamma)} \sum_{k=0}^\infty \frac{\beta^{2k}}{(2k)!} E(|Y-\mu|^{2k+r}),$$

where $m := n \bmod 2$*. All moments above are finite.*

Note that from the cumulant function we have

$$E(Y) = \mu \text{ and } E(X) = \mu + \beta\delta K_{\nu+1}(\bar\gamma)/(\gamma K_\nu(\bar\gamma)),$$

as stated before. Hence, we have that $E((Y-\mu)^r)$ are central moments, whereas $E((X-\mu)^r)$ are in general just μ–centred moments. Note also that $\operatorname{sgn} E((X-\mu)^n)) = \operatorname{sgn}\beta$ for all odd n.

Proof. We will only prove (ii), since the proofs of the other formulae proceed along the same lines, except that to obtain (iii) and (iv) one notes in the final step that odd central moments of Y vanish, since the distribution of Y is symmetric around μ.

The series representation of the exponential function gives

$$\begin{aligned} E(|X|^r) &= \int_{\mathbb{R}} \frac{\bar\gamma^\nu \bar\alpha^{1/2-\nu}}{\sqrt{2\pi}\delta K_\nu(\bar\gamma)} \left(1+\frac{(x-\mu)^2}{\delta^2}\right)^{\nu/2-1/4} \\ &\quad \cdot K_{\nu-1/2}\left(\bar\alpha\sqrt{1+\frac{(x-\mu)^2}{\delta^2}}\right) e^{\beta(x-\mu)}|x|^r dx \\ &= \int_{\mathbb{R}} \sum_{k=0}^\infty \frac{\bar\gamma^\nu \bar\alpha^{1/2-\nu}}{\sqrt{2\pi}\delta K_\nu(\bar\gamma)} \left(1+\frac{(x-\mu)^2}{\delta^2}\right)^{\nu/2-1/4} \\ &\quad \cdot K_{\nu-1/2}\left(\bar\alpha\sqrt{1+\frac{(x-\mu)^2}{\delta^2}}\right) \frac{\beta^k}{k!}(x-\mu)^k |x|^r dx. \end{aligned}$$

The integrals exist (cf. Lemma 7.2.2) and the same is true with β changed to $-\beta$. This implies that the integrals

$$\int_\mu^\infty \frac{\bar\gamma^\nu \bar\alpha^{1/2-\nu}}{\sqrt{2\pi}\delta K_\nu(\bar\gamma)} \left(1+\frac{(x-\mu)^2}{\delta^2}\right)^{\nu/2-1/4} K_{\nu-1/2}\left(\bar\alpha\sqrt{1+\frac{(x-\mu)^2}{\delta^2}}\right) e^{|\beta(x-\mu)|}|x|^r dx$$

and

$$\int_{-\infty}^\mu \frac{\bar\gamma^\nu \bar\alpha^{1/2-\nu}}{\sqrt{2\pi}\delta K_\nu(\bar\gamma)} \left(1+\frac{(x-\mu)^2}{\delta^2}\right)^{\nu/2-1/4} K_{\nu-1/2}\left(\bar\alpha\sqrt{1+\frac{(x-\mu)^2}{\delta^2}}\right) e^{|\beta(x-\mu)|}|x|^r dx$$

and hence the integral

$$\int_{\mathbb{R}} \frac{\bar{\gamma}^\nu \bar{\alpha}^{1/2-\nu}}{\sqrt{2\pi}\delta K_\nu(\bar{\gamma})} \left(1+\frac{(x-\mu)^2}{\delta^2}\right)^{\nu/2-1/4} K_{\nu-1/2}\left(\bar{\alpha}\sqrt{1+\frac{(x-\mu)^2}{\delta^2}}\right) e^{|\beta(x-\mu)|}|x|^r dx$$

exist. Using the last one as majorant, Lebesgue's convergence theorem gives

$$\begin{aligned} E(|X|^r) =& \sum_{k=0}^{\infty} \int_{\mathbb{R}} \frac{\bar{\gamma}^\nu \bar{\alpha}^{1/2-\nu}}{\sqrt{2\pi}\delta K_\nu(\bar{\gamma})} \left(1+\frac{(x-\mu)^2}{\delta^2}\right)^{\nu/2-1/4} \\ & \cdot K_{\nu-1/2}\left(\bar{\alpha}\sqrt{1+\frac{(x-\mu)^2}{\delta^2}}\right) \frac{\beta^k}{k!}(x-\mu)^k |x|^r dx \\ =& \sum_{k=0}^{\infty} \frac{\beta^k}{k!} \left(\frac{\bar{\gamma}}{\bar{\alpha}}\right)^\nu \frac{K_\nu(\bar{\alpha})}{K_\nu(\bar{\gamma})} \int_{\mathbb{R}} \frac{\bar{\alpha}^\nu \bar{\alpha}^{1/2-\nu}}{\sqrt{2\pi}\delta K_\nu(\bar{\alpha})} \left(1+\frac{(x-\mu)^2}{\delta^2}\right)^{\nu/2-1/4} \\ & \cdot K_{\nu-1/2}\left(\bar{\alpha}\sqrt{1+\frac{(x-\mu)^2}{\delta^2}}\right) (x-\mu)^k |x|^r dx. \end{aligned}$$

From this we immediately conclude

$$E(|X|^r) = \left(\frac{\bar{\gamma}}{\bar{\alpha}}\right)^\nu \frac{K_\nu(\bar{\alpha})}{K_\nu(\bar{\gamma})} \sum_{k=0}^{\infty} \frac{\beta^k}{k!} E\left(|Y|^r (Y-\mu)^k\right).$$

□

Corollary 7.3.2. *Let $X \sim GH(\nu,\alpha,\beta,\mu,\delta)$, $V \sim GIG(\nu,\delta,\alpha)$ and $\varepsilon \sim N(0,1)$ with V and ε independent, then for every $r>0$ and $n \in \mathbb{N}$:*

$$(i)\quad E((X-\mu)^n) = \left(\frac{\bar{\gamma}}{\bar{\alpha}}\right)^\nu \frac{K_\nu(\bar{\alpha})}{K_\nu(\bar{\gamma})} \sum_{k=0}^{\infty} \frac{\beta^{2k+m}}{(2k+m)!} E\left(V^{k+(m+n)/2}\right) E\left(\epsilon^{2k+m+n}\right)$$

$$(ii)\quad E(|X-\mu|^r) = \left(\frac{\bar{\gamma}}{\bar{\alpha}}\right)^\nu \frac{K_\nu(\bar{\alpha})}{K_\nu(\bar{\gamma})} \sum_{k=0}^{\infty} \frac{\beta^{2k}}{(2k)!} E\left(V^{k+r/2}\right) E\left(|\epsilon|^{2k+r}\right),$$

where $m := n \bmod 2$.

Proof. Combine Theorem 7.3.1 with Lemma 7.2.1. □

Note that we obtain the (absolute) moments of X provided $\mu = 0$ and the (absolute) central moments if $\beta = 0$. For $\beta = 0$ the above series are in fact just a single term or vanish completely.

Using the explicit expressions for the moments of GIG and normal laws, given in Appendix 7.8.1, it is now straightforward to obtain more explicit expressions for the μ–centred (absolute) moments of GH laws.

Theorem 7.3.3. *Let $X \sim GH(\nu,\alpha,\beta,\mu,\delta)$, then for every $r>0$ and $n \in \mathbb{N}$:*

$$(i)\quad E((X-\mu)^n) = \frac{2^{\lceil \frac{n}{2} \rceil} \bar{\gamma}^\nu \delta^{2\lceil \frac{n}{2} \rceil} \beta^m}{\sqrt{\pi} K_\nu(\bar{\gamma}) \bar{\alpha}^{\nu+\lceil \frac{n}{2} \rceil}} \sum_{k=0}^{\infty} \frac{2^k \bar{\beta}^{2k} \Gamma\left(k+\lceil \frac{n}{2} \rceil + \frac{1}{2}\right)}{\bar{\alpha}^k (2k+m)!} K_{\nu+k+\lceil \frac{n}{2} \rceil}(\bar{\alpha})$$

$$(ii)\quad E(|X-\mu|^r) = \frac{2^{\frac{r}{2}} \bar{\gamma}^\nu \delta^r}{\sqrt{\pi} K_\nu(\bar{\gamma}) \bar{\alpha}^{\nu+\frac{r}{2}}} \sum_{k=0}^{\infty} \frac{2^k \bar{\beta}^{2k} \Gamma\left(k+\frac{r}{2}+\frac{1}{2}\right)}{\bar{\alpha}^k (2k)!} K_{\nu+k+\frac{r}{2}}(\bar{\alpha}),$$

where $m := n \bmod 2$.

Proof. Combine Corollary 7.3.2 with Lemmata 7.8.1 and 7.8.2 observing that we have $(n+m) \bmod 2 = 0$ and $(m+n)/2 = (n \bmod 2 + n)/2 = \lceil n/2 \rceil$. □

The absolute convergence of the series on the right hand sides is obviously implied by the finiteness of $E((X-\mu)^n)$, resp. $E(|X-\mu|^r)$, and the positivity of all terms involved. Yet, one can also immediately give an analytic argument, which adds further insight into the convergence behaviour and is useful when one implements the above formulae on a computer (see Section 7.6). Let

$$a_k := \frac{2^k \bar{\beta}^{2k} \Gamma\left(k + \frac{r}{2} + \frac{1}{2}\right)}{\bar{\alpha}^k (2k)!} K_{\nu+k+\frac{r}{2}}(\bar{\alpha}).$$

From $K_\nu(z) \sim \sqrt{(\pi/2)} 2^\nu \nu^{\nu-1/2} e^{-\nu} z^{-\nu}$ for $\nu \to \infty$ (Ismail (1977), Jørgensen (1982, p. 171)) we obtain

$$\frac{a_{k+1}}{a_k} \sim \frac{4\bar{\beta}^2 \left(k + \frac{r}{2} + \frac{1}{2}\right) \left(k + \nu + \frac{r}{2}\right) \left(1 + \frac{1}{k+\nu+\frac{r}{2}}\right)^{k+\nu+\frac{r+1}{2}}}{\bar{\alpha}^2 e (2k+2)(2k+1)} \overset{k\to\infty}{\to} \left(\frac{\bar{\beta}}{\bar{\alpha}}\right)^2 < 1 \qquad (7.3.1)$$

and thus the quotient criterion from standard analysis implies absolute convergence. Lemma 7.8.4, which we give in Appendix 7.8.2.2, and its proof add some further insight into the behaviour of the series.

As a side result of Theorem 7.3.3 we also obtain two identities for modified Bessel functions of the third kind.

Corollary 7.3.4. *Let* $x, y, z \in \mathbb{R}_{>0}$ *s.t.* $z = \sqrt{x^2 - y^2}$ *and* $\nu \in \mathbb{R}$. *Then*

$$(i) \quad K_\nu(z) = \frac{z^\nu}{x^\nu} \sum_{k=0}^{\infty} \frac{1}{2^k \cdot k!} \frac{y^{2k}}{x^k} K_{\nu+k}(x)$$

$$(ii) \quad zK_\nu(z) + y^2 K_{\nu+1}(z) = \frac{z^{\nu+1}}{x^\nu} \sum_{k=0}^{\infty} \frac{2k+1}{2^k \cdot k!} \frac{y^{2k}}{x^k} K_{\nu+k}(x).$$

Proof. Combine Theorem 7.3.3 with

$$E(X) = \frac{\delta\bar{\beta}}{\bar{\gamma}} \frac{K_{\nu+1}(\bar{\gamma})}{K_\nu(\bar{\gamma})} \text{ and } E(X^2) = \delta^2 \left(\frac{K_{\nu+1}(\bar{\gamma})}{\bar{\gamma} K_\nu(\bar{\gamma})} + \frac{\bar{\beta}^2}{\bar{\gamma}^2} \frac{K_{\nu+2}(\bar{\gamma})}{K_\nu(\bar{\gamma})} \right)$$

for $X \sim GH(\nu, \alpha, \beta, 0, \delta)$ and use $\Gamma(n+1/2) = (2n)!\sqrt{\pi}/(2^{2n} \cdot n!)$. Finally identify x, y, z, ν with $\bar{\alpha}, \bar{\beta}, \bar{\gamma}, \nu + 1$. □

7.4. Moments of NIG laws

We now turn to the normal inverse Gaussian subclass of the generalized hyperbolic law. For an overview see especially Barndorff-Nielsen (1998b). Recall that the $NIG(\alpha, \beta, \mu, \delta)$ law with $0 \le |\beta| < \alpha$, $\mu \in \mathbb{R}$ and $\delta \in \mathbb{R}_{>0}$ is the special case of the $GH(\nu, \alpha, \beta, \mu, \delta)$ law given by $\nu = -1/2$, as already mentioned when summarizing the properties of the GH law previously. Hence, our above calculations for (μ–centred) moments immediately lead to the following.

Corollary 7.4.1. *Let $X \sim NIG(\alpha, \beta, \mu, \delta)$, $Y \sim NIG(\alpha, 0, \mu, \delta)$, then for every $r > 0$ and $n \in \mathbb{N}$:*

$$(i) \quad E(X^n) = e^{\bar{\gamma}-\bar{\alpha}} \sum_{k=0}^{\infty} \frac{\beta^k}{k!} E(Y^n (Y-\mu)^k)$$

$$(ii) \quad E(|X|^r) = e^{\bar{\gamma}-\bar{\alpha}} \sum_{k=0}^{\infty} \frac{\beta^k}{k!} E(|Y|^r (Y-\mu)^k)$$

$$(iii) \quad E((X-\mu)^n) = \frac{2^{\lceil \frac{n}{2} \rceil + \frac{1}{2}} \delta^{2\lceil \frac{n}{2} \rceil} \beta^m}{\pi \bar{\alpha}^{\lceil \frac{n}{2} \rceil - \frac{1}{2}}} e^{\bar{\gamma}} \sum_{k=0}^{\infty} \frac{2^k \bar{\beta}^{2k} \Gamma\left(k + \lceil \frac{n}{2} \rceil + \frac{1}{2}\right)}{\bar{\alpha}^k (2k+m)!} K_{k+\lceil \frac{n}{2} \rceil - \frac{1}{2}}(\bar{\alpha})$$

$$(iv) \quad E(|X-\mu|^r) = \frac{2^{\frac{r+1}{2}} \delta^r}{\pi \bar{\alpha}^{\frac{r-1}{2}}} e^{\bar{\gamma}} \sum_{k=0}^{\infty} \frac{2^k \bar{\beta}^{2k} \Gamma\left(k + \frac{r+1}{2}\right)}{\bar{\alpha}^k (2k)!} K_{k+\frac{r-1}{2}}(\bar{\alpha}),$$

where $m := n \bmod 2$. All moments above are finite.

Proof. Follows immediately from Theorems 7.3.1 and 7.3.3 by using the identity $K_{1/2}(z) = K_{-1/2}(z) = \sqrt{(\pi/2)} z^{-1/2} e^{-z}$ (see e.g. Jørgensen (1982, p. 170)). □

Formulae (iii) and (iv) for r equal to an even natural number can be given more explicitly using

$$K_{n+\frac{1}{2}}(z) = K_{\frac{1}{2}}(z) \left(1 + \sum_{i=1}^{n} \frac{(n+i)!}{i!(n-i)!} 2^{-i} z^{-i}\right) \tag{7.4.1}$$

for all $n \in \mathbb{N}$ (see e.g. Jørgensen (1982, p. 170)). But in order to avoid making the above formulae even more complex, we omit this.

7.5. Moments of NIG Lévy processes and their time-wise behaviour

Based on the above results our aim now is to generalize the findings of Barndorff-Nielsen and Prause (2001) regarding the time-wise approximate scaling behaviour of NIG Lévy processes.

7.5.1. Moments of NIG Lévy processes

Let $Z(t)$, $t \in \mathbb{R}_{>0}$, be the $NIG(\alpha, \beta, \mu, \delta)$ Lévy process, i.e. the Lévy process for which $Z(1) \sim NIG(\alpha, \beta, \mu, \delta)$. Owing to the closedness under convolution of the NIG law, the marginal distribution of the NIG Lévy process at an arbitrary time $t \in \mathbb{R}_{>0}$ is given by $NIG(\alpha, \beta, t\mu, t\delta)$. For more background on NIG Lévy processes see in particular Barndorff-Nielsen (1998b). From our previous results we can immediately infer:

Corollary 7.5.1. *Let $Z(t)$, $t \in \mathbb{R}_{>0}$, be an $NIG(\alpha, \beta, \mu, \delta)$ Lévy process, then for every $r > 0$ and $n \in \mathbb{N}$:*

$$(i)\quad E((Z(t)-\mu t)^n) = \frac{2^{\lceil \frac{n}{2} \rceil + \frac{1}{2}} \delta^{2\lceil \frac{n}{2} \rceil} \beta^m}{\pi \bar{\alpha}^{\lceil \frac{n}{2} \rceil - \frac{1}{2}}} e^{t\bar{\gamma}} \sum_{k=0}^{\infty} \frac{2^k \bar{\beta}^{2k} \Gamma\left(k + \lceil \frac{n}{2} \rceil + \frac{1}{2}\right)}{\bar{\alpha}^k (2k+m)!} t^{k+\lceil \frac{n}{2} \rceil + \frac{1}{2}} \cdot K_{k+\lceil \frac{n}{2} \rceil - \frac{1}{2}}(t\bar{\alpha})$$

$$(ii)\quad E(|Z(t)-\mu t|^r) = \frac{2^{\frac{r+1}{2}} \delta^r}{\pi \bar{\alpha}^{\frac{r-1}{2}}} e^{t\bar{\gamma}} \sum_{k=0}^{\infty} \frac{2^k \bar{\beta}^{2k} \Gamma\left(k + \frac{r+1}{2}\right)}{\bar{\alpha}^k (2k)!} t^{k+(r+1)/2} K_{k+\frac{r-1}{2}}(t\bar{\alpha})$$

where $m := n \bmod 2$.

In Appendix 7.8.2.2 it is shown that the moments above are analytic functions of time (Lemma 7.8.4). This fact is later needed to calculate derivatives of log moments.

7.5.2. Scaling and apparent scaling

Before we now turn to discussing the scaling properties of an NIG Lévy process, let us briefly state what scaling precisely means. Let $X(t)$ be some stochastic process. We say some moment of X obeys a *scaling law*, if the logarithm of this moment is an affine function of log time, i.e., for the r-th absolute moment, $\ln E(|X(t)|^r) = s_r \ln t + c_r$ for some constants $s_r, c_r \in \mathbb{R}$. Here s_r is called the *scaling coefficient*. If all (absolute) moments of X, or at least those one is interested in, follow a scaling law, we say that the process itself obeys one. For example, in the case of Brownian motion $X(t)$ with drift μ we know from $X(t) - \mu t \stackrel{D}{=} \sqrt{t}(X(1) - \mu)$ that $\ln E(|X(t) - \mu t|^r) = (r/2) \ln t +$ constant for all $r > 0$, i.e. all absolute moments exhibit scaling. More generally all self-similar processes, e.g. the strictly α-stable Lévy processes (cf. Samorodnitsky and Taqqu (1994, Chapter 7) and Sato (1999, Chapter 3)), obey a scaling law. When looking only at small changes in time the *local scaling* behaviour is determined by $d \ln E(|X(t)|^r)/d \ln t$ (in the case of the r-th absolute moment). In the presence of scaling the latter derivative is constant and equals the value of the scaling coefficient. Provided some log moment of a process $X(t)$ exhibits a very close to affine dependence on log time over some time horizon of interest, we speak of *approximate* or *apparent scaling*. This is equivalent to the local scaling varying only little over the time spans considered. When working with real empirical data, it is often not possible to distinguish between apparent and strict scaling due to the randomness of the available observations. Hence, it is of interest, from a statistical point of view, whether some given theoretical process shows approximate scaling.

7.5.3. The time-wise behaviour of μ-centred moments

Let us now examine the scaling behaviour exhibited by the $NIG(\alpha, \beta, \mu, \delta)$ Lévy process $Z(t)$. For the following discussion of the time dependence of $E(|Z(t) - \mu t|^r)$ we will abbreviate the time independent terms:

$$c(r) := \frac{2^{\frac{r+1}{2}} \delta^r}{\pi \bar{\alpha}^{\frac{r-1}{2}}} \tag{7.5.1}$$

$$a_k(r) := \frac{2^k \bar{\beta}^{2k} \Gamma\left(k + \frac{r+1}{2}\right)}{\bar{\alpha}^k (2k)!} \tag{7.5.2}$$

If we define

$$\psi(t) := \exp(t\bar{\gamma}) \sum_{k=0}^{\infty} a_k(r) t^{k+(r+1)/2} K_{k+(r-1)/2}(t\bar{\alpha}) \tag{7.5.3}$$

and

$$\phi(t) := \ln \psi(e^t), \tag{7.5.4}$$

we have from Corollary 7.5.1 that

$$E(|Z(t) - \mu t|^r) = c(r) \cdot \psi(t) \tag{7.5.5}$$

and

$$\ln E(|Z(t) - \mu t|^r) = \ln c(r) + \phi(\ln t). \tag{7.5.6}$$

Thus:

$$\frac{d \ln E(|Z(t) - \mu t|^r)}{d \ln t} = \phi'(\ln t) \tag{7.5.7}$$

Lemma 7.5.2. *Let $\phi : \mathbb{R}_{>0} \to \mathbb{R}$ be defined by (7.5.4), then*

$$\phi'(t) = 1 + \bar{\gamma} e^t - \bar{\alpha} e^t \frac{\sum_{k=0}^{\infty} a_k(r) e^{tk} K_{k+(r-3)/2}(e^t \bar{\alpha})}{\sum_{k=0}^{\infty} a_k(r) e^{tk} K_{k+(r-1)/2}(e^t \bar{\alpha})}. \tag{7.5.8}$$

Proof. Using

$$K_\nu'(z) = -K_{\nu-1}(z) - \nu z^{-1} K_\nu(z) \tag{7.5.9}$$

(see e.g. Jørgensen (1982, p. 170) or Bronstein et al. (2000, p. 528)), we obtain for $\psi(t)$ as defined in equation (7.5.3):

$$\begin{aligned}
\psi'(t) &= \exp(t\bar{\gamma}) \left(\bar{\gamma} \sum_{k=0}^{\infty} a_k(r) t^{k+(r+1)/2} K_{k+(r-1)/2}(t\bar{\alpha}) + \sum_{k=0}^{\infty} a_k(r) \left(k + \frac{r+1}{2} \right) \right. \\
&\quad \cdot t^{k+(r-1)/2} K_{k+(r-1)/2}(t\bar{\alpha}) - \sum_{k=0}^{\infty} a_k(r) t^{k+(r+1)/2} \bar{\alpha} \\
&\quad \left. \cdot \left(K_{k+(r-3)/2}(t\bar{\alpha}) + \left(k + \frac{r-1}{2} \right) (t\bar{\alpha})^{-1} K_{k+\frac{r-1}{2}}(t\bar{\alpha}) \right) \right) \\
&= \bar{\gamma}\psi(t) + t^{-1}\psi(t) - \bar{\alpha} e^{\bar{\gamma} t} \sum_{k=0}^{\infty} a_k(r) t^{k+(r+1)/2} K_{k+(r-3)/2}(t\bar{\alpha}).
\end{aligned}$$

That we may interchange differentiation and summation above is an immediate consequence of Lemma 7.8.4 and Weierstraß's theorem for sequences of holomorphic functions (see Appendix 7.8.2.1). Hence, we get from (7.5.4)

$$\phi'(t) = \frac{e^t \psi'(e^t)}{\psi(e^t)} = 1 + \bar{\gamma} e^t - \bar{\alpha} e^t \frac{\sum_{k=0}^{\infty} a_k(r) e^{t(k+(r+1)/2)} K_{k+(r-3)/2}(e^t \bar{\alpha})}{\sum_{k=0}^{\infty} a_k(r) e^{t(k+(r+1)/2)} K_{k+(r-1)/2}(e^t \bar{\alpha})}.$$

□

Now we can formulate our main result on the scaling behaviour of NIG-Lévy-Processes.

Theorem 7.5.3. *Let $Z(t)$, $t \in \mathbb{R}_{>0}$, be an $NIG(\alpha, \beta, \mu, \delta)$ Lévy process, then*

$$\frac{d \ln E(|Z(t) - \mu t|^r)}{d \ln t} = 1 + \bar{\gamma} t - \bar{\alpha} t \frac{\sum_{k=0}^{\infty} a_k(r) t^k K_{k+(r-3)/2}(\bar{\alpha} t)}{\sum_{k=0}^{\infty} a_k(r) t^k K_{k+(r-1)/2}(\bar{\alpha} t)}$$

for every $r > 0$.

Proof. The result follows by combining Lemma 7.5.2 and (7.5.7). □

When comparing the above results with Barndorff-Nielsen and Prause (2001) note that they looked at the derivatives with respect to $\ln(\bar{\alpha} t)$, whereas we look at the derivative with respect to $\ln t$. The difference is related to the fact that Barndorff-Nielsen and Prause (2001) only consider the case $\beta = 0$. In the general case the parameters $\bar{\alpha}$ and $\bar{\beta}$ of the marginals at time t are both scaled with t. Hence, it is most natural and convenient to consider the change of the log moments versus the change of log time directly.

The expression for the local scaling behaviour derived in Theorem 7.5.3 is in general not constant in time, hence, the absolute μ-centred moments of an NIG Lévy process do not obey a strict scaling law. Later we shall see from numerical examples that apparent scaling is common. If we look at the symmetric NIG Lévy process, i.e. $\beta = 0$, the above formula becomes

$$\frac{d \ln E(|Z(t) - \mu t|^r)}{d \ln t} = 1 + \bar{\alpha} t - \bar{\alpha} t \frac{K_{(r-3)/2}(\bar{\alpha} t)}{K_{(r-1)/2}(\bar{\alpha} t)}. \tag{7.5.10}$$

From this one deducts using $K_\nu = K_{-\nu}$ that the second μ-centred moment obeys a scaling law with slope one, which is the same as for Brownian motion.

The aggregational Gaussianity of NIG Lévy processes (due to the central limit theorem the marginal distribution at time t of any Lévy process with finite second moment gets more and more Gaussian as t increases) becomes visible in the asymptotic scaling of the symmetric case for large times. For $r = 1$ it was already noted in Barndorff-Nielsen and Prause (2001) that the local scaling approaches $1/2$ for $t \to \infty$ and hence for large t the first absolute μ-centred moment of the process seems to scale like Brownian motion. Using formula (ii) in Corollary 7.5.1, which for $\beta = 0$ becomes

$$E(|Z(t) - \mu t|^r) = \frac{2^{\frac{r+1}{2}} \delta^r}{\pi \bar{\alpha}^{\frac{r-1}{2}}} \exp(t\bar{\alpha}) \, \Gamma\left(\frac{r+1}{2}\right) t^{(r+1)/2} K_{\frac{r-1}{2}}(t\bar{\alpha}), \tag{7.5.11}$$

and $K_\nu(x) \sim \sqrt{(\pi/2)} x^{-1/2} e^{-x}$ for $x \to \infty$ (cf. Jørgensen (1982, p. 171) or Bronstein et al. (2000)) we get that $\ln E(|Z(t) - \mu t|^r) \sim (r/2) \ln t + c$ for $t \to \infty$, where $c \in \mathbb{R}$ is a constant. Note that, as usual, "$\sim$" denotes asymptotic equivalence. Hence, $E(|Z(t) - \mu t|^r)$ obeys a scaling law with slope $r/2$ for $t \to \infty$, i.e. the symmetric NIG Lévy process approaches the exact scaling behaviour of Brownian motion. This result can also be easily deduced from (7.5.10) using an asymptotic expansion of $K_{(r-3)/2}(z)/K_{(r-1)/2}(z)$ for $z \to +\infty$ (see e.g. Jørgensen (1982, p. 173)).

Studying the limiting behaviour of the absolute μ-centred moments analytically for $\beta \neq 0$ seems hardly possible. Yet, numerical studies indicate that a skewed NIG Lévy process does not scale like Brownian motion for large times in general. For example, when computing $d \ln E((Z(t))^2)/d \ln t$ of the $NIG(100, 30, 0, 0.001)$ Lévy process for times from $1/2$ to 1024

the values increase monotonically from 1.004 to 6.268. If we are, however, close to the symmetric case, i.e. if $|\beta|/\alpha$ is small, then basically the same approximate scaling behaviour is obtained as in the symmetric case. This can, in particular, be seen in the numerical data presented in Section 7, where for large times the value of $d \ln E(|Z(t) - \mu t|^r)/d \ln t$ is very close to the Brownian motion scaling slope of $r/2$.

To see from (7.5.11) what happens in the symmetric case for $t \searrow 0$ we employ the fact that

$$K_\nu(x) \sim \begin{cases} \Gamma(\nu) 2^{\nu-1} x^{-\nu} & \text{for } \nu > 0, x \searrow 0 \\ -\ln x & \text{for } \nu = 0, x \searrow 0 \end{cases} \tag{7.5.12}$$

(see e.g. Jørgensen (1982, p. 171)). For $r = 1$ and $t \searrow 0$ we obtain that $\ln E(|Z(t) - \mu t|)$ becomes $\ln t + \bar{\alpha} e^{\ln t} + \ln(-\ln(t\bar{\alpha})) + c$ with $c \in \mathbb{R}$ being a constant. From this we conclude that for small values of t the first absolute μ-centred moment approximately scales with slope one, as already noted in Barndorff-Nielsen and Prause (2001). The same asymptotic scaling slope of one holds for $r > 1$, since $\ln E(|Z(t) - \mu t|^r) \sim \ln t + \bar{\alpha} e^{\ln t} + c(r)$ for $t \searrow 0$. Yet, a different result is obtained for $0 < r < 1$. In this case one obtains again using (7.5.12) and the identity $K_{-\nu} = K_\nu$ that $\ln E(|Z(t) - \mu t|^r) \sim r \ln t + \bar{\alpha} e^{\ln t} + c(r)$ and so there is asymptotic scaling with slope r.

7.6. Notes on the numerical implementation

We will now briefly discuss some issues related to the implementation of formula (ii) in Corollary 7.5.1 and Theorem 7.5.3 on a computer. Similar results hold for formula (i) of Corollary 7.5.1. First note that (ii) in Corollary 7.5.1 can be reexpressed using (7.5.2) as:

$$E(|Z(t) - \mu t|^r) = \left(\frac{2\delta^2 t}{\bar{\alpha}}\right)^{r/2} \frac{\sqrt{2t\bar{\alpha}}}{\pi} \exp(t\bar{\gamma}) \sum_{k=0}^{\infty} a_k(r) t^k K_{k+\frac{r-1}{2}}(t\bar{\alpha}). \tag{7.6.1}$$

The value of the infinite series can only be approximated. Yet, note that the analytic convergence discussion of the series in Section 7.3, especially formula (7.3.1), implies asymptotically geometric convergence of this series, which is the faster, the smaller $|\beta|$ is relatively to α. We suggest to compute the individual summands recursively as discussed below, add them up and stop, when summands become negligible compared to the current value of the approximation. To calculate the individual summands recursively note that

$$a_0(r) = \Gamma\left(\frac{r+1}{2}\right) \tag{7.6.2}$$

and

$$a_k(r) t^k = \frac{2\bar{\beta}^2 (k + (r-1)/2)}{\bar{\alpha}(2k-1)(2k)} t \cdot a_{k-1}(r) t^{k-1}, \tag{7.6.3}$$

which is obtained using the functional equation $\Gamma(z+1) = z\Gamma(z)$ of the Gamma function, and that the recursion formula for Bessel functions (Jørgensen (1982, p. 170)) gives

$$K_{k+\frac{r-1}{2}}(t\bar{\alpha}) = 2 \cdot \left(k - 1 + \frac{r-1}{2}\right) (t\bar{\alpha})^{-1} K_{k-1+\frac{r-1}{2}}(t\bar{\alpha}) + K_{k-2+\frac{r-1}{2}}(t\bar{\alpha}). \tag{7.6.4}$$

The latter formula implies that we can calculate the values of the Bessel functions needed from a two term recursion, for which we only need to calculate $K_{-1+(r-1)/2}(t\bar{\alpha})$ and

$K_{(r-1)/2}(t\bar{\alpha})$ as starting values. Hence, the calculation of the value of the series involves, apart from basic manipulations, only one evaluation of the Gamma function and two of the Bessel functions.

The series in the denominator in Theorem 7.5.3 is the series just discussed above and the numerator is of the same type, only the index of the Bessel functions is changed, and can hence be calculated analogously. Actually, both series can be calculated simultaneously using only the recursion for $a_k(r)t^k$ and the two term recursion for the Bessel functions described above.

There is, however, one possible problem when using the two term recursion. If the starting values are zeros up to numerical precision, then only zeros will be calculated as summands. For example when using Matlab and the built in function for K_ν one gets $K_0(z) = 0$ for $z > 697$. Hence, one needs to take care of this possible case. Provided the recursion works, the numerical results obtained are usually almost identical to the numerical results one gets when using a built in Bessel function routine of e.g. Matlab for each summand, but the recursion may save computing power. Furthermore, it should now be obvious, how numerical evaluations of the formulae for μ-centred (absolute) moments of GH laws given in Theorem 7.3.3 can be organized efficiently.

The Matlab code we used to produce the numerical results in this chapter is available from `www.ma.tum.de/stat/Papers`. It is based upon the above considerations and can be used to compute μ-centred moments of the NIG distribution/Lévy process and the derivatives of the log moments with respect to log time.

7.7. Apparent scaling behaviour of NIG Lévy processes

The aim of this section is to show that NIG Lévy processes may well exhibit a behaviour very close to strict scaling over a wide range of orders of moments. We exemplify the possible apparent scaling of absolute μ-centred moments of NIG Lévy processes using the parameters from Barndorff-Nielsen and Prause (2001). They considered the USD/DEM exchange rate from the whole of 1996, contained in the HFDF96 data set from Olsen & Associates, and fitted an NIG Lévy process to the log returns by maximum likelihood estimation. The estimates obtained based on the three hour log returns are $\alpha = 415.9049$, $\beta = 1.512$, $\delta = 0.0011$ and $\mu = 0.000026$. For further details on the data, the estimation procedure and the relevance for finance we refer the interested reader to the paper by Barndorff-Nielsen and Prause. Note especially that, as is typical for returns of exchange rate series, μ is very close to zero and therefore there is practically no difference between moments and μ–centred moments. Figure 7.2 (left), which depicts the logarithm of the first absolute μ–centred moment versus the logarithm of time in seconds, is therefore optically indistinguishable from the figure in Barndorff-Nielsen and Prause (2001) showing the first absolute moment calculated via numerical integration. The estimated regression line of the log moments against log time, fitted by least squares, has slope 0.5863, which is slightly higher than the slope 0.5705 reported in Barndorff-Nielsen and Prause (2001), and $d \ln E(|Z(t) - \mu t|)/d \ln t$ decreases from 0.7853 to 0.5011 over the time interval depicted, which is 5.625 minutes to 32 days. This is significantly different from the Brownian motion case, where it is exactly 1/2 (cf. above). The behaviour of $d \ln E(|Z(t) - \mu t|)/d \ln t$ over the time interval considered indicates that for $t \to \infty$ the slope asymptotically becomes about 1/2, the exact Gaussian scaling coefficient. This is related to the fact that $|\beta|$ is relatively small, as already pointed

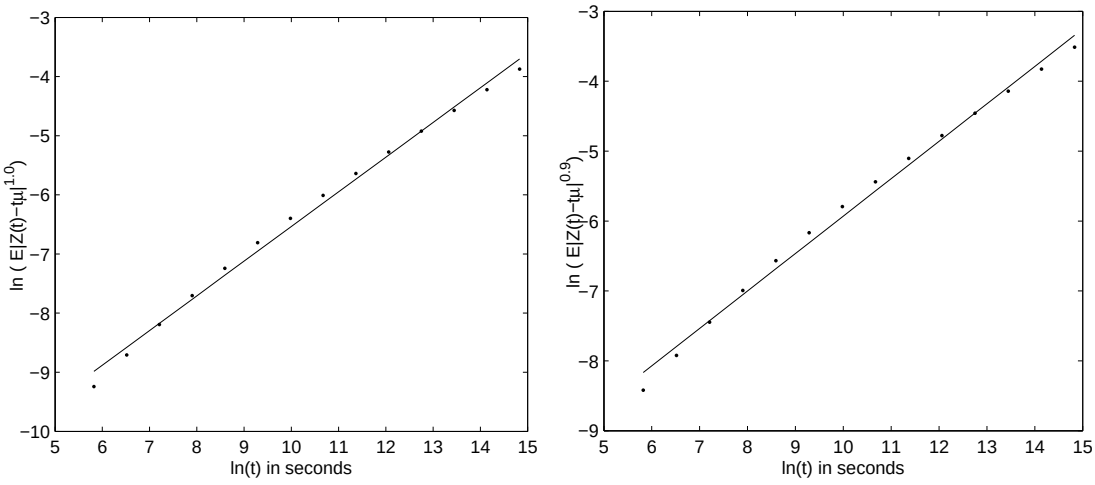


Figure 7.2.: Approximate scaling law for the first (left) and 0.9th μ-centred absolute moment of the NIG Lévy process fitted to the USD/DEM exchange rate: moments ($\bullet$) and regression line log moments against log time

out earlier in the discussion of the scaling asymptotics.

With our results obtained above it is possible to study the behaviour of moments other than the first. Figures 7.2 (right), 7.3 (left), 7.3 (right) and 7.4 (left) show the time behaviour of the 0.9th, 1.1th, 0.5th and 1.5th μ–centred absolute moments over the same time horizon. All figures exhibit apparent scaling, which improves with the order of the moment. The fitted regression lines have slope 0.53535, 0.63536, 0.31322 and 0.81327 respectively, which are all higher than the corresponding values for Brownian motion, which are 0.45, 0.55, 0.25 and 0.75. The values of $d\ln E(|Z(t)-\mu t|^r)/d\ln t$ decrease from 0.7316 to 0.4509, 0.8316 to 0.5512, 0.4499 to 0.2503 and 0.9499 to 0.7515 respectively. So again they seem to converge to some value close to the Brownian motion scaling slope.

Figure 7.4 (right) shows that the second μ–centred moment seems to exhibit perfect linear scaling. Yet, there is in fact no strict scaling law holding. The values of the regression coefficient 1.0001 and $d\ln E(|Z(t)-\mu t|^2)/d\ln t$ are very close to one with $d\ln E(|Z(t)-\mu t|^2)/d\ln t$ increasing very slowly from 1 to 1.0015. Such a result is to be expected, since $|\beta|$ is small (compared to α) and for $\beta=0$ we have that the variance, which is in this case identical to the second μ-centred moment, obeys a strict scaling law with slope one, as for Brownian motion.

The third μ–centred absolute moment still exhibits apparent scaling behaviour with a regression slope of 1.2966, but the values of $d\ln E(|Z(t)-\mu t|^3)/d\ln t$ are now increasing from 1.0134 to 1.5007 rather than decreasing and the slope is lower than the scaling coefficient 1.5 for Brownian motion. However, $d\ln E(|Z(t)-\mu t|^3)/d\ln t$ still seems to converge to some value close to 3/2 at large times. It generally seems to be the case that $d\ln E(|Z(t)-\mu t|^r)/d\ln t$ increases with time for $r>2$, whereas it decreases for $r<2$. Actually, further calculations indicate that this change takes place marginally below 2 at about 1.9995. Some more numerical calculations hint that in the symmetric case $\ln E(|Z(t)-\mu t|^r)$ is concave as a function of $\ln t$ for $0<r\leq 2$ and convex for $r\geq 2$.

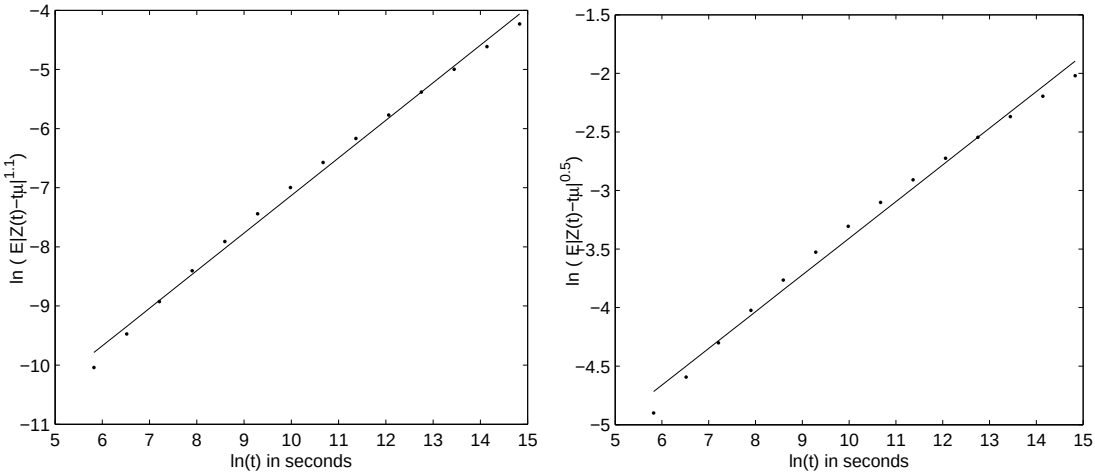


Figure 7.3.: Approximate scaling law for the 1.1th (left) and 0.5th μ-centred absolute moment: moments ($\bullet$) and regression line log moments against log time

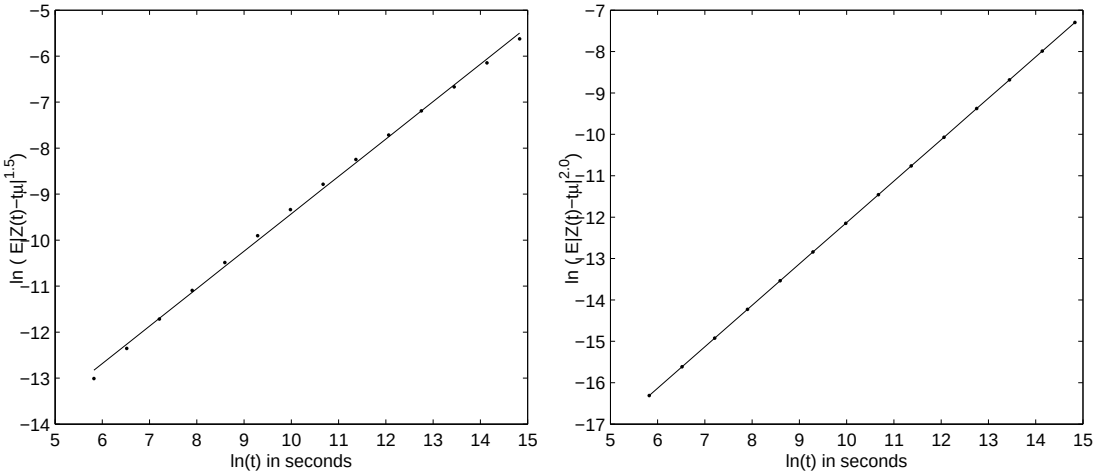


Figure 7.4.: Approximate scaling law for the 1.5th (left) and second μ-centred absolute moment: moments ($\bullet$) and regression line log moments against log time

For very high values of r, i.e. ten and greater, no apparent linear scaling is observed over the time horizon considered. Looking at the slopes of the apparent scaling of the μ-centred absolute moments (of orders 0.2 to 4, for instance, as we have done in further calculations), the relationship between scaling coefficient and order is apparently not simply linear as, for example, in the case of an α-stable Lévy process (see Samorodnitsky and Taqqu (1994, Chapter 7) and Sato (1999, Chapter 3)), but a concave one.

Our results obtained above, show that NIG Lévy processes may exhibit something close to scaling. These findings are particularly interesting, since both in finance, especially when dealing with foreign exchange returns, and turbulence there is on the one hand empirical and for turbulence also theoretical evidence of scaling laws, and on the other hand models based on NIG Lévy processes have been put forth in the literature. Compared to Brownian motion the NIG Lévy process does in general not exhibit exact linear scaling, but approximate scaling over wide time horizons and for a practically interesting range of (absolute) moments is demonstrated here.

7.8. Appendix to Chapter 7

7.8.1. Moments of GIG and normal laws

For completeness we provide below the well-known formulae for the moments of the GIG and normal laws.

For the GIG law the following result is given in Jørgensen (1982, p. 13), who uses a slightly different parametrization.

Lemma 7.8.1. *Let $X \sim GIG(\nu, \delta, \gamma)$ with $\delta, \gamma > 0$. Then*

$$E(X^r) = \left(\frac{\delta}{\gamma}\right)^r \frac{K_{\nu+r}(\bar{\gamma})}{K_\nu(\bar{\gamma})}$$

for every $r > 0$.

Proof.

$$\begin{aligned} E(X^r) &= \int_0^\infty \frac{\bar{\gamma}^\nu}{2K_\nu(\bar{\gamma})}\delta^{-2\nu}x^{\nu+r-1}\exp\left(-\frac{1}{2}\bar{\gamma}\left((\bar{\gamma}\delta^{-2}x)^{-1}+\bar{\gamma}\delta^{-2}x\right)\right)dx \overset{y:=\bar{\gamma}\delta^{-2}x}{=} \\ &= \frac{\bar{\gamma}^{-r}\delta^{2r}}{K_\nu(\bar{\gamma})}\frac{1}{2}\int_0^\infty y^{\nu+r-1}\exp\left(-\frac{1}{2}\bar{\gamma}(y^{-1}+y)\right)dy = \left(\frac{\delta}{\gamma}\right)^r \frac{K_{\nu+r}(\bar{\gamma})}{K_\nu(\bar{\gamma})}, \end{aligned}$$

where in the last step we employed the integral representation of $K_{\nu+r}$ stated earlier when introducing the modified Bessel function of the third kind (Equation (7.2.2)). □

The absolute moments of the normal distribution $N(0,1)$ are well known and given in many standard texts on probability theory, viz.:

Lemma 7.8.2. *Let $X \sim N(0,1)$ and $r > 0$ then*

$$E(|X|^r) = \frac{2^{r/2}\Gamma\left(\frac{r+1}{2}\right)}{\sqrt{\pi}}.$$

Proof.

$$\begin{aligned} E(|X|^r) &= (2\pi)^{-1/2} \int_{\mathbb{R}} |x|^r e^{-\frac{x^2}{2}} dx = \left(\frac{2}{\pi}\right)^{1/2} \int_0^\infty x^r e^{-\frac{x^2}{2}} dx \overset{t:=\frac{x^2}{2}}{=} \\ &= \frac{2^{r/2}}{\sqrt{\pi}} \int_0^\infty t^{\frac{r+1}{2}-1} e^{-t} dt = \frac{2^{r/2}}{\sqrt{\pi}} \Gamma\left(\frac{r+1}{2}\right) \end{aligned}$$

□

7.8.2. Analyticity of the moments of an NIG Lévy process as a function of time

In this appendix we show that the μ-centred (absolute) moments of an NIG Lévy process Z, given in Corollary 7.5.1, are analytic functions of time. To this end we employ some complex function theory, so we start with a brief review of the needed result, viz. Weierstraß's convergence theorem.

7.8.2.1. Convergence of sequences of holomorphic functions

Recall that a function $f : D \to \mathbb{C}$, $D \subseteq \mathbb{C}$, is called holomorphic, if it is complex differentiable on D, i.e.

$$\lim_{h \in \mathbb{C}, h \to 0} \frac{f(z+h) - f(z)}{h}$$

exists for all $z \in D$ (confer textbooks on complex function theory, e.g. Remmert (1991) or Freitag and Busam (2000), for a thorough discussion of holomorphicity and related concepts). Holomorphic functions have many useful properties that real differentiable functions lack in general and thus it is often preferable to use holomorphic functions when possible. One of the nice implications of holomorphicity is that any once complex differentiable function is automatically infinitely often complex differentiable and another is that locally uniform convergence commutes with differentiation:

Theorem 7.8.3 (Weierstraß's convergence Theorem)**.** *Let $f_n : D \to \mathbb{C}$, $n \in \mathbb{N}$, be a sequence of holomorphic functions, defined on an open subset $D \subseteq \mathbb{C}$, which converges locally uniform to a function $f : D \to \mathbb{C}$. Then f is holomorphic on D and for every $k \in \mathbb{N}$ the sequence of k-th derivatives $f_n^{(k)}$ converges locally uniform to $f^{(k)}$ on D.*

For a proof and related results see one of the books mentioned above. The crucial difference to the real differentiable case is that complex differentiation has an integral representation.

7.8.2.2. Holomorphicity of some series

The following lemma gives in particular that the μ-centred (absolute) moments of an NIG Lévy process Z (cf. Corollary 7.5.1) are analytic functions of time.

Lemma 7.8.4. *Let $\bar{\alpha} > 0$, $|\bar{\beta}| < \bar{\alpha}$, $1 < \epsilon < \bar{\alpha}^2/|\bar{\beta}|^2$, $\nu \in \mathbb{R}$, $r > 0$, $n \in \mathbb{N}$, $m = n \bmod 2$, $D = \{z \in \mathbb{C} : \Re(z) > 0, |z| < \epsilon \Re(z)\}$,*

$$f : D \to \mathbb{C}, z \mapsto \sum_{k=0}^{\infty} \frac{2^k \bar{\beta}^{2k} \Gamma\left(k + \frac{r}{2} + \frac{1}{2}\right)}{\bar{\alpha}^k (2k)!} z^{k+(r+1)/2} K_{\nu+k+\frac{r}{2}}(z\bar{\alpha})$$

and

$$g : D \to \mathbb{C}, z \mapsto \sum_{k=0}^{\infty} \frac{2^k \bar{\beta}^{2k} \Gamma\left(k + \lceil \frac{n}{2} \rceil + \frac{1}{2}\right)}{\bar{\alpha}^k (2k+m)!} z^{k+\lceil \frac{n}{2} \rceil + \frac{1}{2}} K_{\nu+k+\lceil \frac{n}{2} \rceil}(z\bar{\alpha}).$$

Then both series are locally uniformly convergent and f, g are holomorphic on D.

Note that ν is $-1/2$ in the series of Corollary 7.5.1.

Proof. It is sufficient to show the locally uniform convergence, since this implies the holomorphicity via Weierstraß's convergence theorem (see Appendix 7.8.2.1). Furthermore it is obvious that the result for g follows from the one for f.

Let us now prove the uniform convergence of the series $f(z) = \sum_{k=0}^{\infty} f_k(z)$ on $D \cap \{z \in \mathbb{C} : a < \Re(z) < b\}$ for arbitrary $0 < a < b < \infty$, where

$$f_k(z) = \frac{2^k \bar{\beta}^{2k} \Gamma\left(k + \frac{r}{2} + \frac{1}{2}\right)}{\bar{\alpha}^k (2k)!} z^{k+(r+1)/2} K_{\nu+k+\frac{r}{2}}(z\bar{\alpha}).$$

An immediate consequence of the integral representation for K_ν given in (7.2.2) is $|K_\nu(z)| \leq K_\nu(\Re(z))$ for $z \in D$ and thus

$$\begin{aligned} |f_k(z)| &= \left| \frac{2^k \bar{\beta}^{2k} \Gamma\left(k + \frac{r}{2} + \frac{1}{2}\right)}{\bar{\alpha}^k (2k)!} z^{k+(r+1)/2} K_{\nu+k+\frac{r}{2}}(z\bar{\alpha}) \right| \\ &\leq \frac{2^k |\bar{\beta}|^{2k} \Gamma\left(k + \frac{r}{2} + \frac{1}{2}\right)}{\bar{\alpha}^k (2k)!} (\epsilon \Re(z))^{k+(r+1)/2} K_{\nu+k+\frac{r}{2}}(\Re(z)\bar{\alpha}) \\ &\overset{x:=\Re(z)}{\leq} \frac{2^k |\bar{\beta}|^{2k} \Gamma\left(k + \frac{r}{2} + \frac{1}{2}\right)}{\bar{\alpha}^k (2k)!} (\epsilon x)^{k+(r+1)/2} K_{\nu+\frac{r}{2}+k}(x\bar{\alpha}) \end{aligned}$$

for all $k \in \mathbb{N}_0$. Note that we defined $x \in (a, b)$ to be the real part of z. Using equation (7.5.9) we obtain

$$\frac{d}{dx} x^{k+\nu+\frac{r}{2}} K_{\nu+\frac{r}{2}+k}(x\bar{\alpha}) = -\bar{\alpha} x^{k+\nu+\frac{r}{2}} K_{\nu+\frac{r}{2}+k-1}(x\bar{\alpha}) < 0.$$

This implies for $x \in (a, b)$:

$$x^{k+(r+1)/2} K_{\nu+\frac{r}{2}+k}(x\bar{\alpha}) \leq d a^{k+(r+1)/2} K_{\nu+\frac{r}{2}+k}(a\bar{\alpha})$$

where $d := a^{\nu-1/2} \cdot \max\{a^{-\nu+1/2}, b^{-\nu+1/2}\}$. Applying this inequality to the above expression, we get for all $k \in \mathbb{N}_0$

$$|f_k(z)| \leq \frac{2^k d |\bar{\beta}|^{2k} \Gamma\left(k + \frac{r}{2} + \frac{1}{2}\right)}{\bar{\alpha}^k (2k)!} (\epsilon a)^{k+(r+1)/2} K_{\nu+\frac{r}{2}+k}(a\bar{\alpha}).$$

From the finiteness of the $r/2$th absolute moment of the $GH(\nu, a\bar{\alpha}, a|\bar{\beta}|\sqrt{\epsilon}, 0, 1)$ law and Theorem 7.3.3 follows that

$$\begin{aligned} &\sum_{k=0}^{\infty} \frac{2^k d |\bar{\beta}|^{2k} \Gamma\left(k + \frac{r}{2} + \frac{1}{2}\right)}{\bar{\alpha}^k (2k)!} (\epsilon a)^{k+(r+1)/2} K_{\nu+\frac{r}{2}+k}(a\bar{\alpha}) \\ = \; & d(\epsilon a)^{\frac{r+1}{2}} \sum_{k=0}^{\infty} \frac{2^k (|\bar{\beta}| a\sqrt{\epsilon})^{2k} \Gamma\left(k + \frac{r}{2} + \frac{1}{2}\right)}{(a\bar{\alpha})^k (2k)!} K_{\nu+\frac{r}{2}+k}(a\bar{\alpha}) \end{aligned}$$

converges absolutely. Hence, the uniform convergence of $\sum_{k=0}^{\infty} |f_k(z)|$ on $D \cap \{z \in \mathbb{C} : a < \Re(z) < b\}$ is established. □

A. Linear Operators on the Real Symmetric Matrices whose Exponentials Preserve the Inertia

A.1. Introduction

Let $\mathbb{S}_d$ be the space of real symmetric $d \times d$ matrices and denote for a matrix $A \in \mathbb{S}_d$ its inertia by $\mathrm{In}(A) = (r, s, t)$ where the triplet of natural numbers r, s, t means that A has r strictly positive, s strictly negative and t zero eigenvalues (counted with their respective algebraic multiplicities).

In this appendix we show that the exponential group generated by a linear operator $\mathbf{B} : \mathbb{S}_d \to \mathbb{S}_d$ preserves the inertia, i.e. $\mathrm{In}(e^{\mathbf{B}t}A) = \mathrm{In}(A)$ for all $t \in \mathbb{R}$ and $A \in \mathbb{S}_d$, and that $e^{\mathbf{B}t}$ maps the positive (semi)-definite matrices onto themselves for all $t \in \mathbb{R}$, if and only if there is a $d \times d$ matrix B such that $\mathbf{B}$ is representable as $X \mapsto BX + XB^T$. That the exponential of linear operators of this form preserves the inertia is not hard to be seen (cf. the upcoming Theorem A.2.2) and well-known. Thus the main contribution of this appendix is to establish the converse. The proof given is based on results on the linear preservers of the inertia and the positive (semi-)definite matrices, respectively. For an overview over linear preserver problems and references to the original literature see Pierce et al. (1992), in particular its third chapter Loewy (1992), or Li and Pierce (2001).

The complete characterization of linear operators having the above mentioned properties is of particular interest in applications, since one often wants to construct models for the time evolution of a positive definite matrix based on (stochastic) differential equations whose solution involves the exponential of a given linear operator.

As regards stochastic processes, Ornstein-Uhlenbeck type processes taking values in the positive (semi-)definite matrices were introduced in Chapter 4. These processes can be used as a flexible and very tractable model for the stochastic evolution of some covariance matrix in continuous time, as can be seen from Chapter 5 where we presented a detailed analysis of the second-order moment structure and an application to the modelling of financial data. In Chapter 4 we used linear operators $\mathbf{B}$ on the symmetric $d \times d$ matrices of the form $X \mapsto BX + XB^T$ for some $d \times d$ matrix B in order to construct Ornstein-Uhlenbeck type processes in the positive (semi-)definite matrices. The results of this appendix show that these are indeed the only linear operators possible when one demands that $\exp(\mathbf{B}t)$ maps the positive semi-definite cone onto itself at all times.

Furthermore linear operators of the type $X \mapsto BX + XB^T$ have also been used in ordinary differential equations in order to ensure positive (semi-)definiteness of the solution (see Dragan, Freiling, Hochhaus and Morozan (2004), for instance). For example, the elementary differential equation

$$x'(t) = \mathbf{B}x(t)$$

in the symmetric $d \times d$ matrices has the solution $x(t) = e^{\mathbf{B}t}x_0$. In this set-up demanding that $e^{\mathbf{B}t}$ maps the positive semi-definite matrices onto themselves thus means that $x(t)$ varies over all of the positive semi-definite cone when x_0 does so.

Notation

Throughout this appendix we write $\mathbb{R}^+$ for the positive real numbers including zero and we denote the set of real $d \times d$ matrices by $M_d(\mathbb{R})$ and the group of invertible $d \times d$ matrices by $GL_d(\mathbb{R})$, the linear subspace of symmetric matrices by $\mathbb{S}_d$, the positive semidefinite cone by $\mathbb{S}_d^+$ and the open positive definite cone by $\mathbb{S}_d^{++}$. I_d stands for the $d \times d$ identity matrix. Finally, A^T is the transposed of a matrix $A \in M_d(\mathbb{R})$ and the exponential of a matrix or linear operator A is denoted by $\exp(A)$ or e^A.

Moreover, for a matrix A we denote by A_{ij} the element in the i-th row and j-th column and this notation is extended to matrix-valued functions in a natural way. The standard basis matrices (i.e. the matrices which have only zero entries except for a one in the i-th row and j-th column) of $M_d(\mathbb{R})$ are denoted by $E^{(ij)}$ for $i, j = 1, 2, \ldots, d$.

A.2. Exponential inertia preservers

Definition A.2.1 (Exponential inertia preserver)**.** *A linear operator $\mathbf{B} : \mathbb{S}_d \to \mathbb{S}_d$ is said to be an* exponential inertia preserver*, if $\mathrm{In}(e^{\mathbf{B}t}A) = \mathrm{In}(A)$ for all $t \in \mathbb{R}$ and $A \in \mathbb{S}_d$.*

We start by recalling a well-known result showing that a special class of linear operators are exponential inertia preservers.

Theorem A.2.2. *Assume that a linear operator $\mathbf{B} : \mathbb{S}_d \to \mathbb{S}_d$ can be represented as $X \mapsto BX + XB^T$ for some $B \in M_d(\mathbb{R})$. Then $\mathbf{B}$ is an exponential inertia preserver.*

Proof. It is easy to see that $e^{\mathbf{B}t}A = e^{Bt}Ae^{B^Tt}$ using e.g. Horn and Johnson (1991, pp. 255 and 440). Thus the result follows immediately from the general results on inertia preservers (see Loewy (1992) or Li and Pierce (2001) and the original articles cited therein), as the exponential of a matrix is necessarily invertible. □

It is important to note that the linear operator $\mathbf{B}$ above is uniquely characterized by the matrix B.

Proposition A.2.3. *Let $B, C \in M_d(\mathbb{R})$. Then the linear operators $\mathbf{B} : \mathbb{S}_d \to \mathbb{S}_d$, $X \mapsto BX + XB^T$ and $\mathbf{C} : \mathbb{S}_d \to \mathbb{S}_d$, $X \mapsto CX + XC^T$ are the same, if and only if $B = C$.*

Moreover, for any operator $\mathbf{B} : \mathbb{S}_d \to \mathbb{S}_d$, $X \mapsto BX + XB^T$ the matrix B is already uniquely identified by the values $\{\mathbf{B}E^{(ii)}\}_{i=1,\ldots,d}$.

Proof. It suffices to show the second claim that $\{\mathbf{B}E^{(ii)}\}_{i=1,\ldots,d}$ already uniquely characterizes B. It is easy to see that

$$(BE^{(ii)} + E^{(ii)}B^T)_{kl} = \begin{cases} 2B_{ii} & \text{for } k = l = i \\ B_{il} & \text{for } k = i \text{ and } k \neq l \\ B_{ik} & \text{for } l = i \text{ and } k \neq l \\ 0 & \text{otherwise} \end{cases}$$

Thus $\mathbf{B}E^{(ii)}$ uniquely characterizes the $i - th$ column of B and, hence, $\{\mathbf{B}E^{(ii)}\}_{i=1,\ldots,d}$ uniquely characterizes B. □

In the following we establish the converse result that all exponential inertia preservers are of the form given in Theorem A.2.2.

Lemma A.2.4. *Let $\mathbf{B} : \mathbb{S}_d \to \mathbb{S}_d$ be a linear operator and assume that there exists an $\epsilon > 0$ and a function $D :]-\epsilon, \epsilon[\to M_d(\mathbb{R})$ such that $e^{\mathbf{B}t}A = D(t)AD(t)^T$ for all $t \in]-\epsilon, \epsilon[$ and $A \in \mathbb{S}_d$. Then there exists an $\widetilde{\epsilon} > 0$, a continuously differentiable function $\widetilde{D} :]-\epsilon, \epsilon[\to M_d(\mathbb{R})$ and a unique matrix $B \in M_d(\mathbb{R})$ such that $e^{\mathbf{B}t}A = \widetilde{D}(t)A\widetilde{D}(t)^T$ for all $t \in]-\widetilde{\epsilon}, \widetilde{\epsilon}[$ and $A \in \mathbb{S}_d$ and such that $\mathbf{B}A = BA + AB^T$ for all $A \in \mathbb{S}_d$.*

Proof. We first show the existence of a matrix $\widetilde{D}(t)$ with the stated properties. Observe that we obtain as a side result that provided $\left(\exp(\mathbf{B}t)E^{(11)}\right)_{11} \neq 0$ the operator $\exp(\mathbf{B}t)$ is already identified by the values $\exp(\mathbf{B}t)E^{(11)}$ and $\exp(\mathbf{B}t)(E^{(1,j)}+E^{(j,1)})$ with $j = 2, 3, \ldots, d$ and the fact that it can be represented as $X \mapsto D(t)XD(t)^T$.

Note that $D(t)XD(t)^T = (-D(t))X(-D(t))^T$ for all $X \in \mathbb{S}_d$, so the matrix $D(t)$ can only be unique up to a multiplication by minus one. Elementary calculations give

$$\exp(\mathbf{B}t)E^{(11)} = \begin{pmatrix} D_{11}^2 & D_{11}D_{21} & D_{11}D_{31} & \cdots & D_{11}D_{d1} \\ * & * & * & \cdots & * \end{pmatrix} \tag{A.2.1}$$

and

$$\exp(\mathbf{B}t)\left(E^{(1j)} + E^{(j1)}\right) = \begin{pmatrix} D_{11}D_{1j} + D_{11}D_{1j} & D_{1j}D_{21} + D_{11}D_{2j} & \cdots & D_{1j}D_{d1} + D_{11}D_{dj} \\ * & * & \cdots & * \end{pmatrix} \tag{A.2.2}$$

for $j = 2, 3, \ldots, d$. Here D_{kl} denotes $D_{kl}(t)$ for notational convenience and $*$ represents entries which are of no interest in the following. It is easy to see that the above equations uniquely characterize the matrix $D(t)$ up to the sign of $D_{11}(t)$, as long as $\left(\exp(\mathbf{B}t)E^{(11)}\right)_{11} = D_{11}(t)^2 \neq 0$.

Since $\exp(\mathbf{B} \cdot 0)$ is the identity on $\mathbb{S}_d$ and thus $\left(\exp(\mathbf{B} \cdot 0)E^{(11)}\right)_{11} = 1$, the continuity of $t \mapsto \exp(\mathbf{B}t)$ ensures that there exists an $\widetilde{\epsilon} > 0$ with $\widetilde{\epsilon} \leq \epsilon$ such that $\left(\exp(\mathbf{B}t)E^{(11)}\right)_{11} > 0$ for all $t \in]-\widetilde{\epsilon}, \widetilde{\epsilon}[$. Thus it follows from (A.2.1) and (A.2.2) that $\widetilde{D}(t) \in M_d(\mathbb{R})$ defined by

$$\widetilde{D}_{11}(t) = \sqrt{\left(\exp(\mathbf{B}t)E^{(11)}\right)_{11}} \tag{A.2.3}$$

$$\widetilde{D}_{i1}(t) = \frac{\left(\exp(\mathbf{B}t)E^{(11)}\right)_{i1}}{\widetilde{D}_{11}(t)} \quad \text{for } i = 2, 3, \ldots, d \tag{A.2.4}$$

$$\widetilde{D}_{1j}(t) = \frac{\left(\exp(\mathbf{B}t)(E^{(1j)} + E^{(j1)})\right)_{11}}{2\widetilde{D}_{11}(t)} \quad \text{for } j = 2, 3, \ldots, d \tag{A.2.5}$$

$$\widetilde{D}_{ij}(t) = \frac{\left(\exp(\mathbf{B}t)(E^{(1j)} + E^{(j1)})\right)_{1i}}{\widetilde{D}_{11}(t)} - \frac{\left(\exp(\mathbf{B}t)(E^{(1j)} + E^{(j1)})\right)_{11}\left(\exp(\mathbf{B}t)E^{(11)}\right)_{i1}}{2\widetilde{D}_{11}(t)^3} \quad \text{for } i, j = 2, 3, \ldots, d \tag{A.2.6}$$

is well-defined for all $t \in]-\widetilde{\epsilon}, \widetilde{\epsilon}[$ and satisfies

$$\exp(\mathbf{B}t)A = \widetilde{D}(t)A\widetilde{D}(t)^T \; \forall\, t \in]-\epsilon, \epsilon[\;\text{ and }\; A \in \mathbb{S}_d.$$

The continuous differentiability of $t \mapsto \exp(\mathbf{B}t)$ and the one of the square root function on $\mathbb{R}^+\backslash\{0\}$ imply together with (A.2.3) to (A.2.6) and the strict positivity of $\left(\exp(\mathbf{B}\cdot t)E^{(11)}\right)_{11}$ that the map $]-\epsilon,\epsilon[\to M_d(\mathbb{R})$, $t \mapsto \widetilde{D}(t)$ is continuously differentiable.

Using the notion of Fréchet derivatives (see Rudin (1976) or Bhatia (1997, Section X.4) for a review in connection with matrix analysis) and denoting the linear operators on $\mathbb{S}_d$ by $L(\mathbb{S}_d)$, it follows immediately from

$$(X+H)A(X+H)^T = XAX^T + XAH^T + HAX^T + HAH^T \; \forall\, A \in \mathbb{S}_d$$

for $X, H \in M_d(\mathbb{R})$ that the map $f : M_d(\mathbb{R}) \to L(\mathbb{S}_d), X \mapsto f(X)$ with $f(X)A = XAX^T$ for all $A \in \mathbb{S}_d$ is continuously differentiable and the derivative $\mathcal{D}f(X)$ is given by the linear map $M_d(\mathbb{R}) \mapsto L(\mathbb{S}_d)$, $H \mapsto \mathcal{D}f(X)(H)$ with $\mathcal{D}f(X)(H)A = XAH^T + HAX^T$ for $A \in \mathbb{S}_d$. Thus

$$\begin{aligned}\left(\frac{d}{dt}\exp(\mathbf{B}t)\right)A &= \left(\frac{d}{dt}f\left(\widetilde{D}(t)\right)\right)A = \mathcal{D}f(\widetilde{D}(t))\left(\frac{d}{dt}\widetilde{D}(t)\right)A \\ &= \widetilde{D}(t)A\left(\frac{d}{dt}\widetilde{D}(t)\right)^T + \left(\frac{d}{dt}\widetilde{D}(t)\right)A\widetilde{D}(t)^T\end{aligned}$$

for all $t \in]-\epsilon,\epsilon[$ and $A \in \mathbb{S}_d$.

Since $\frac{d}{dt}\exp(\mathbf{B}t) = \exp(\mathbf{B}t)\mathbf{B}$, it follows that $\mathbf{B} = \exp(-\mathbf{B}t)\frac{d}{dt}\exp(\mathbf{B}t)$. Moreover, it is easy to see $\widetilde{D}(t) \in GL_d(\mathbb{R})$ and $\exp(\mathbf{B}t)^{-1}A = \exp(-\mathbf{B}t)A = \widetilde{D}(t)^{-1}A\widetilde{D}(t)^{-T}$ for $A \in \mathbb{S}_d$. Hence

$$\mathbf{B}A = \widetilde{D}(t)^{-1}\left(\widetilde{D}(t)A\left(\frac{d}{dt}\widetilde{D}(t)\right)^T + \left(\frac{d}{dt}\widetilde{D}(t)\right)A\widetilde{D}(t)^T\right)\widetilde{D}(t)^{-T} \tag{A.2.7}$$

$$= A\left(\widetilde{D}(t)^{-1}\frac{d}{dt}\widetilde{D}(t)\right)^T + \left(\widetilde{D}(t)^{-1}\frac{d}{dt}\widetilde{D}(t)\right)A \tag{A.2.8}$$

for $A \in \mathbb{S}_d$ and all $t \in]-\epsilon,\epsilon[$. As by construction $\widetilde{D}(0) = I_d$, setting $B = \frac{d}{dt}\widetilde{D}(t)\Big|_{t=0}$ concludes the proof now noting that Proposition A.2.3 ensures the uniqueness of $B \in M_d(\mathbb{R})$. □

For $d \geq 3$ we can use the above Lemma to fully characterize exponential inertia preservers.

Theorem A.2.5. *Let $d \in \mathbb{N}$ with $d \geq 3$. Then the following holds:*

(i) A linear mapping $\mathbf{B} : \mathbb{S}_d \to \mathbb{S}_d$ is an exponential inertia preserver, if and only if there exists a matrix $B \in M_d(\mathbb{R})$ such that $\mathbf{B}X = BX + XB^T$ for all $X \in \mathbb{S}_d$.

(ii) A linear mapping $\mathbf{B} : \mathbb{S}_d \to \mathbb{S}_d$ is an exponential inertia preserver, if there exists an $\epsilon > 0$ such that $\mathrm{In}(e^{\mathbf{B}t}A) = \mathrm{In}(A)$ for all $A \in \mathbb{S}_d$ and $t \in]-\epsilon,\epsilon[$.

That we need to restrict ourselves to $d \geq 3$ is clear from the proof below and Johnson and Pierce (1985, Remark 1.3).

Proof. (i): The 'if' part is given in Theorem A.2.2. Regarding the 'only if' part we have that for $d \geq 3$ all linear preservers $\mathbf{B} : \mathbb{S}_d \to \mathbb{S}_d$ on $\mathbb{S}_d$ of the inertia class $(d-1,1,0)$ are of the form $X \mapsto CXC^T$ for some $C \in M_d(\mathbb{R})$ (cf. Loewy (1992) and references therein). Hence, there is a function $D(t) : \mathbb{R} \to M_d(\mathbb{R})$ such that $e^{\mathbf{B}t}A = D(t)AD(t)^T$ for all $t \in \mathbb{R}$ and $A \in \mathbb{S}_d$. Lemma A.2.4 immediately concludes now.

(ii): This follows from Theorem A.2.2 and Lemma A.2.4 in a straightforward manner. □

Moreover, we can now also characterize the linear operators whose exponential group maps the positive semi-definite matrices onto themselves.

Corollary A.2.6. *Let $d \in \mathbb{N}$. Then the following holds:*

(i) A linear mapping $\mathbf{B} : \mathbb{S}_d \to \mathbb{S}_d$ satisfies $e^{\mathbf{B}t}(\mathbb{S}_d^+) = \mathbb{S}_d^+$ for all $t \in \mathbb{R}$, if and only if there exists a matrix $B \in M_d(\mathbb{R})$ such that $\mathbf{B}X = BX + XB^T$ for all $X \in \mathbb{S}_d$.

(ii) A linear mapping $\mathbf{B} : \mathbb{S}_d \to \mathbb{S}_d$ satisfies $e^{\mathbf{B}t}(\mathbb{S}_d^+) = \mathbb{S}_d^+$ for all $t \in \mathbb{R}$, if there exists an $\epsilon > 0$ such that $e^{\mathbf{B}t}(\mathbb{S}_d^+) = \mathbb{S}_d^+$ for all $t \in]-\epsilon, \epsilon[$.

Proof. Since all linear maps that map $\mathbb{S}_d^+$ onto itself are of the form $X \mapsto CXC^T$ for some $C \in M_d(\mathbb{R})$ (cf. Loewy (1992) or the original article Schneider (1965)), the proof is analogous to the one of the last theorem. □

$\mathbb{S}_d^+$ can be replaced by $\mathbb{S}_d^{++}$ in the above corollary. Yet we cannot extend the result to the case $e^{\mathbf{B}t}(\mathbb{S}_d^+) \subset \mathbb{S}_d^+$, as there are linear operators $\mathbf{C}$ such that $\mathbf{C}(\mathbb{S}_d^+) \subset \mathbb{S}_d^+$ which are not representable by $X \mapsto CXC^T$ for some $C \in M_d(\mathbb{R})$ (cf. Choi (1975)).

Furthermore note that from the characterization of linear preservers of various other fixed inertia classes (see Loewy (1992) for an overview) many results analogous to the above ones follow immediately, since we obviously have the following general result:

Theorem A.2.7. *Let $\mathfrak{P}$ be some property of linear maps on $\mathbb{S}_d$ and assume that a linear map $\mathbf{C} : \mathbb{S}_d \to \mathbb{S}_d$ has the property $\mathfrak{P}$, if and only if there is a matrix $C \in M_d(\mathbb{R})$ such that $\mathbf{C}A = CAC^T$ for all $A \in \mathbb{S}_d$. Then the following holds:*

(i) The exponential $e^{\mathbf{B}t}$ of a linear map $\mathbf{B} : \mathbb{S}_d \to \mathbb{S}_d$ has the property $\mathfrak{P}$ for all $t \in \mathbb{R}$, if and only if there exists a matrix $B \in M_d(\mathbb{R})$ such that $\mathbf{B}X = BX + XB^T$ for all $X \in \mathbb{S}_d$.

(ii) The exponential $e^{\mathbf{B}t}$ of a linear map $\mathbf{B} : \mathbb{S}_d \to \mathbb{S}_d$ has the property $\mathfrak{P}$ for all $t \in \mathbb{R}$, if there exists an $\epsilon > 0$ such that $e^{\mathbf{B}t}$ has the property $\mathfrak{P}$ for all $t \in]-\epsilon, \epsilon[$.

B. On the Relation Between the vec and BEKK Multivariate GARCH Models

B.1. Introduction

Multivariate GARCH models have been studied intensively in recent years and many different specifications have been used in the literature (cf. Bauwens et al. (2006) for a comprehensive overview and Boussama (1998, 2006) for a detailed discussion on strict stationarity and geometric ergodicity). In this appendix we present some results on the relationship between the vec and BEKK models. These models have been presented and analysed in detail in Engle and Kroner (1995). In that paper it has been noted that all BEKK models are representable as vec ones, but regarding the converse it has only been shown that all diagonal vec models are representable as diagonal BEKK ones and stated that the BEKK parametrization "eliminates very few if any interesting models allowed by the vec representation". However, apart from the recent paper by Scherrer and Ribarits (2007), which came to our attention only after finishing the work on the present appendix, no further results on the relationship between the two models seem to have been obtained since then, nor are there simple and tractable examples of vec models which are not representable in the BEKK form to be found in the literature.

Applying long known results from linear algebra in a straightforward manner, we show in this appendix that in dimension two the models are actually equivalent and that all vec-models not representable in the simplest BEKK form with invertible parameter matrices exhibit necessarily some degeneracy, viz. that one of the matrices appearing in the vec model is degenerated in the sense that it maps the vectorized positive semi-definite matrices to a strict subset of themselves. Finally, we present an example of a vec model with no BEKK representation in dimension three. Comparing our results to those of Scherrer and Ribarits (2007) they have shown the equivalence in dimension two using semi-definite programming, whereas we note that it is an immediate consequence of a long known result in linear algebra. The linear algebra literature we are referring to seems not to have been used in connection to GARCH models before. But it is obviously intimately connected to multivariate GARCH models and should be useful to obtain other results as well. For an example of a vec model having no BEKK representation Scherrer and Ribarits (2007) refer to Ribarits (2006). The example presented in that thesis on page 61 (stated in a transformed way only) is of a rather complicated structure and it is argued by numerical optimization and not an analytical proof that it gives an admissible vec term which cannot be represented in the BEKK form. In contrast to this we present a very simple example with interesting properties which is analysed completely analytically.

The remainder of this appendix is organized as follows. We briefly state the necessary definitions of multivariate GARCH models in the next section and then present our results in Section B.3.

Regarding notation we denote the set of real $d \times d$ matrices by $M_d(\mathbb{R})$, the group of

invertible $d \times d$ matrices by $GL_d(\mathbb{R})$, the linear subspace of symmetric matrices by $\mathbb{S}_d$ and the positive semi-definite cone by $\mathbb{S}_d^+$. Finally, A^T is the transposed of a matrix $A \in M_d(\mathbb{R})$.

B.2. Multivariate GARCH processes

The well-known single dimensional GARCH(p,q) model introduced in Bollerslev (1986) is defined via an i.i.d. sequence $(\epsilon_n)_{n\in\mathbb{N}}$ and the equations

$$X_n = \sqrt{\sigma_n^2}\epsilon_n \tag{B.2.1}$$

$$\sigma_n^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i X_{n-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{n-j}^2 \tag{B.2.2}$$

for $n \in \mathbb{N}$. Moreover, the initial values $\sigma_0^2, \sigma_{-1}^2, \ldots, \sigma_{1-q}^2$ and the parameters $\alpha_1, \ldots, \alpha_p$, $\beta_1, \ldots, \beta_q$ are non-negative and $\alpha_0 > 0$. $X = (X_n)_{n\in\mathbb{N}}$ is referred to as a GARCH(p,q) process and σ^2 is its latent conditional variance process.

When one moves from a scalar X to a d-dimensional X, the variance process σ^2 becomes a $d\times d$ covariance matrix process Σ and one uses the vec (or alternatively vech) transformation in order to specify the model. The vec transformation maps the $d \times d$ matrices bijectively to $\mathbb{R}^{d^2}$ by stacking the columns of a matrix below one another. This leads to the vec-model (Engle and Kroner (1995)) which is given by:

$$X_n = \Sigma_n^{1/2}\epsilon_n \tag{B.2.3}$$

$$\mathrm{vec}(\Sigma_n) = \mathrm{vec}(C) + \sum_{i=1}^{p} \tilde{A}_i \mathrm{vec}(X_{n-i}X_{n-i}^T) + \sum_{j=1}^{q} \tilde{B}_j \mathrm{vec}(\Sigma_{n-j}). \tag{B.2.4}$$

for $n \in \mathbb{N}$ where $(\epsilon_n)_{n\in\mathbb{N}}$ is now an $\mathbb{R}^d$-valued i.i.d. sequence and $\Sigma_n^{1/2}$ denotes the unique positive semi-definite matrix whose square is Σ_n, i.e. $\Sigma_n^{1/2} \in \mathbb{S}_d^+$ and $\Sigma_n^{1/2}\Sigma_n^{1/2} = \Sigma_n$. To ensure the positive semi-definiteness of the process Σ the initial values and C have to be positive semi-definite and $\tilde{A}_1, \ldots, \tilde{A}_p, \tilde{B}_1, \ldots, \tilde{B}_q$ need to be $d^2 \times d^2$ matrices mapping the vectorized positive semi-definite matrices into themselves.

For notational convenience we shall not only use the vec-model in the following, but also an obviously equivalent specification defined directly on the symmetric matrices. This model, referred to as the "general d-dimensional GARCH(p,q) model" in the following, is given by

$$X_n = \Sigma_n^{1/2}\epsilon_n \tag{B.2.5}$$

$$\Sigma_n = C + \sum_{i=1}^{p} A_i X_{n-i}X_{n-i}^T + \sum_{j=1}^{q} B_j \Sigma_{n-j}. \tag{B.2.6}$$

The only difference to the vec-model is that $A_1, \ldots, A_p$ and $B_1, \ldots, B_q$ are now linear operators from $\mathbb{S}_d$ to $\mathbb{S}_d$ that map the positive semi-definite $d \times d$ matrices into themselves, i.e. $A_i(\mathbb{S}_d^+) \subseteq \mathbb{S}_d^+$ and $B_j(\mathbb{S}_d^+) \subseteq \mathbb{S}_d^+$ for $i = 1, \ldots, p$ and $j = 1, \ldots, q$.

The restrictions on the linear operators A_i and B_j (or $\tilde{A}_i$ and $\tilde{B}_j$ in the vec-model) necessary to ensure positive semi-definiteness led to the introduction of the so-called BEKK

model (see again Engle and Kroner (1995)), which automatically ensures positive semi-definiteness:

$$X_n = \Sigma_n^{1/2}\epsilon_n \tag{B.2.7}$$

$$\Sigma_n = C + \sum_{i=1}^{p}\sum_{k=1}^{l_i} \bar{A}_{i,k} X_{n-i} X_{n-i}^T \bar{A}_{i,k}^T + \sum_{j=1}^{q}\sum_{r=1}^{s_j} \bar{B}_{j,r}\Sigma_{n-j}\bar{B}_{j,r}^T, \tag{B.2.8}$$

where $\bar{A}_{i,k}, \bar{B}_{j,r}$ are now arbitrary elements of $M_d(\mathbb{R})$. It is standard that the BEKK model is equivalent to the vec model with $\tilde{A}_i = \sum_{k=1}^{l_i} \bar{A}_{i,k} \otimes \bar{A}_{i,k}$ and $\tilde{B}_j = \sum_{r=1}^{s_j} \bar{B}_{j,r} \otimes \bar{B}_{j,r}$ with $\otimes$ denoting the tensor (Kronecker) product.

B.3. The relationship between the vec and BEKK model

From the definitions of the models it is clear that studying the relationships between the vec (or general) multivariate GARCH and the BEKK model further is intrinsically related to characterizing the linear operators on $\mathbb{S}_d$ that map the positive semi-definite matrices into themselves. The latter has been studied for a long time in linear algebra under the general topic "Linear Preserver Problems" (see, for instance, the overview articles Pierce et al. (1992) and Li and Pierce (2001)). From the results obtained there we need the following:

Proposition B.3.1. *Let $\mathbf{A} : \mathbb{S}_d \to \mathbb{S}_d$ be a linear operator. Then:*

(i) $\mathbf{A}(\mathbb{S}_d^+) = \mathbb{S}_d^+$, if and only if there exists a matrix $B \in GL_d(\mathbb{R})$ such that $\mathbf{A}$ can be represented as $X \mapsto BXB^T$.

(ii) For $d = 2$, $\mathbf{A}(\mathbb{S}_d^+) \subseteq \mathbb{S}_d^+$, if and only if there is an $r \in \mathbb{N}$ and $B_1, B_2, \ldots, B_r \in M_d(\mathbb{R})$ such that $\mathbf{A}$ can be represented as

$$X \mapsto \sum_{i=1}^{r} B_i X B_i^T.$$

Proof. (a) was initially proved in Schneider (1965), alternatively a more general proof in a Hilbert space context may be found in Li et al. (2003). (b) was established in Størmer (1963) (cf. also Loewy (1992)). □

From this we can immediately infer the relations between the general (or equivalently vec) multivariate GARCH model and the BEKK model:

Theorem B.3.2. *(i) For $d \leq 2$ the general (or* vec*) multivariate GARCH and the BEKK model are equivalent.*

(ii) Every general multivariate GARCH(p,q) model satisfying $A_i(\mathbb{S}_d^+) = \mathbb{S}_d^+$ and $B_j(\mathbb{S}_d^+) = \mathbb{S}_d^+$ for $i = 1, 2, \ldots, p$ and $j = 1, 2, \ldots, q$ can be represented as a BEKK(p,q) model with $l_i = s_j = 1 \,\forall\, i, j$ and $\bar{A}_{i,1}, \bar{B}_{j,1} \in GL_d(\mathbb{R})$.

For the vec model $A_i(\mathbb{S}_d^+) = \mathbb{S}_d^+$ and $B_j(\mathbb{S}_d^+) = \mathbb{S}_d^+$ for $i = 1, 2, \ldots, p$ and $j = 1, 2, \ldots, q$ translates into demanding that $\tilde{A}_i$ and $\tilde{B}_j$ map the vectorized positive semi-definite matrices onto themselves.

The above result means that when a general (or vec) multivariate GARCH model does not have a BEKK representation with $l_i = s_j = 1 \; \forall \, i, j$ and invertible $\bar{A}_{i,1}, \bar{B}_{j,1}$ it has to be the case that A_i or B_j map the positive semi-definite matrices into a strict subset of themselves for some $i = 1, 2, \ldots, p$ or $j = 1, 2, \ldots, q$. This is a somehow degenerated situation, since it may imply that the distribution of $\Sigma_n - C$ for all $n \in \mathbb{N}$ (and thus any limiting or stationary distribution) is concentrated on a subset of the positive semi-definite matrices.

Let us now turn to providing an example for a vec model that cannot be represented in the BEKK form. Consider $d = 3$ and the linear operator $D : \mathbb{S}_d \to \mathbb{S}_d$ given by

$$\begin{pmatrix} x_{11} & x_{12} & x_{13} \\ x_{12} & x_{22} & x_{23} \\ x_{13} & x_{23} & x_{33} \end{pmatrix} \mapsto \begin{pmatrix} x_{11} + 2x_{22} & -x_{12} & -x_{13} \\ -x_{12} & x_{22} + 2x_{33} & -x_{23} \\ -x_{13} & -x_{23} & x_{33} + 2x_{11}. \end{pmatrix} \tag{B.3.1}$$

It has been shown by Choi (1975) that $D(\mathbb{S}_d^+) \subseteq \mathbb{S}_d^+$ and that there exist no $r \in \mathbb{N}$ and $E_1, E_2, \ldots, E_r \in M_3(\mathbb{R})$ such that

$$DX = \sum_{i=1}^{r} E_i X E_i^T \;\text{ for all }\; X \in \mathbb{S}_d.$$

Hence, using D as some A_i or B_j gives a general three-dimensional multivariate GARCH model having no BEKK representation. Clearly this means that the corresponding vec models have no BEKK representation. Since the operator D was defined on $\mathbb{S}_d$ only and not on $M_d(\mathbb{R})$, the corresponding 9×9 matrix in the vec model is not unique, but the corresponding vec models are unique. One of the possible 9×9 matrices the operator D corresponds to is

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 2 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 \\ 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

It should be noted that D is an invertible linear operator, as can easily be seen, and that Proposition B.3.1 implies $D(\mathbb{S}_d^+) \subset \mathbb{S}_d^+$. An example for a positive semi-definite matrix not being the image of another positive semi-definite matrix under D is

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} = D \begin{pmatrix} 1/9 & 0 & 0 \\ 0 & 4/9 & 0 \\ 0 & 0 & -2/9 \end{pmatrix}.$$

So we have given an example showing the following:

Proposition B.3.3. *For $d \geq 3$ there exist general (or* vec*) multivariate GARCH models that cannot be represented in the BEKK form.*

Bibliography

Ando, T. (1988). Comparison of norms $|||f(a) - f(b)|||$ and $|||f(|a - b|)|||$, *Math. Z.* **197**: 403–409.

Applebaum, D. (2004). *Lévy Processes and Stochastic Calculus*, Vol. 93 of *Cambridge Studies in Advanced Mathematics*, Cambridge University Press, Cambridge.

Arató, M. (1982). *Linear Stochastic Systems with Constant Coefficients*, Vol. 45 of *Lecture Notes in Control and Information Sciences*, Springer, Berlin.

Ash, R. B. and Gardner, M. F. (1975). *Topics in Stochastic Processes*, Academic Press, New York.

Bardina, X. and Jolis, M. (1997). An extension of Ito's formula for elliptic diffusion processes, *Stochastic Process. Appl.* **69**: 83–109.

Barndorff-Nielsen, O. E. (1977). Exponentially decreasing distributions for the logarithm of particle size, *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **353**: 401–419.

Barndorff-Nielsen, O. E. (1978a). Hyperbolic distributions and distributions on hyperbolae, *Scand. J. Statist.* **5**: 151–157.

Barndorff-Nielsen, O. E. (1978b). *Information and Exponential Families in Statistical Theory*, John Wiley and Sons, Chicester.

Barndorff-Nielsen, O. E. (1979). Models for non-Gaussian variation with applications to turbulence, *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **368**: 501–520.

Barndorff-Nielsen, O. E. (1982). The hyperbolic distribution in statistical physics, *Scand. J. Statist.* **9**: 43–46.

Barndorff-Nielsen, O. E. (1997). Normal inverse Gaussian distributions and stochastic volatility modelling, *Scand. J. Statist.* **24**: 1–13.

Barndorff-Nielsen, O. E. (1998a). Probability and statistics; selfdecomposability, finance and turbulence, *in* L. Accardi and C. C. Heyde (eds), *Proceedings of the conference "Probability towards 2000" held at Columbia University, New York, 2-6 October 1995*, Springer, Berlin, pp. 47–57.

Barndorff-Nielsen, O. E. (1998b). Processes of normal inverse Gaussian type, *Finance Stoch.* **2**: 41–68.

Barndorff-Nielsen, O. E. (2001a). Modelling by Lévy processess, *in* I. V. Basawa, C. C. Heyde and R. L. Taylor (eds), *Selected Proceedings of the Symposium on Inference for Stochastic Processes*, Vol. 37 of *Lecture Notes - Monograph Series*, Institute of Mathematical Statistics, Hayworth, California.

Barndorff-Nielsen, O. E. (2001b). Superposition of Ornstein–Uhlenbeck type processes, *Theory Probab. Appl.* **45**: 175–194.

Barndorff-Nielsen, O. E. and Blæsild, P. (1981). Hyperbolic distributions and ramifications, *in* C. Taillie, G. P. Patil and B. A. Baldessari (eds), *Statistical distributions in scientific work - Proceedings of the NATO Advanced Study Institute held at the Université degli Studi di Trieste, Triest, Italy, July 10 - August 1, 1980*, Vol. 4, D. Reidel, Dordrecht, Holland, pp. 19–44.

Barndorff-Nielsen, O. E., Blæsild, P. and Halgreen, C. (1978). First hitting time models for the generalized inverse Gaussian distribution, *Stochastic Process. Appl.* **7**: 49–54.

Barndorff-Nielsen, O. E., Blæsild, P., Jensen, J. L. and Sørensen, M. (1983). The fascination of sand - with three appendices by R.A. Bagnold, *Research Report 93*, Department of Theoretical Statistics, Institute of Mathematics, University of Århus, Denmark.

Barndorff-Nielsen, O. E., Blæsild, P., Jensen, J. L. and Sørensen, M. (1985). The fascination of sand, *in* A. C. Atkinson and S. E. Fienberg (eds), *A celebration of statistics - The ISI centenary volume*, Springer, New York, pp. 57–87.

Barndorff-Nielsen, O. E., Blæsild, P. and Schmiegel, J. (2004). A parsimonious and universal description of turbulent velocity increments, *Eur. Phys. J. B Condens. Matter Phys.* **41**: 345–363.

Barndorff-Nielsen, O. E. and Christiansen, C. (1988). Erosion, deposition and size distribution of sand, *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **417**: 335–352.

Barndorff-Nielsen, O. E., Graversen, S. E., Jacod, J., Podolskij, M. and Shephard, N. (2006). A central limit theorem for realised power and bipower variations of continuous semimartingales, *in* Y. Kabanov, R. Lipster and J. Stoyanov (eds), *From Stochastic Calculus to Mathematical Finance - The Shiryaev Festschrift*, Springer, Heidelberg, pp. 33–68.

Barndorff-Nielsen, O. E., Graversen, S. E., Jacod, J. and Shephard, N. (2006). Limit theorems for bipower variation in financial econometrics, *Econometric Theory* **22**: 677–719.

Barndorff-Nielsen, O. E., Jensen, J. L. and Sørensen, M. (1989). Wind shear and hyperbolic distributions, *Boundary-Layer Meteorology* **49**: 417–431.

Barndorff-Nielsen, O. E., Jensen, J. L. and Sørensen, M. (1998). Some stationary processes in discrete and continuous time, *Adv. in Appl. Probab.* **30**: 989–1007.

Barndorff-Nielsen, O. E., Kent, J. T. and Sørensen, M. (1982). Normal variance-mean mixtures and z distributions, *Int. Stat. Rev.* **50**: 145–159.

Barndorff-Nielsen, O. E., Nicolato, E. and Shephard, N. (2002). Some recent developments in stochastic volatility modelling, *Quantitative Finance* **2**: 11–23.

Barndorff-Nielsen, O. E., Pedersen, J. and Sato, K. (2001). Multivariate subordination, self-decomposability and stability, *Adv. in Appl. Probab.* **33**: 160–187.

Barndorff-Nielsen, O. E. and Pérez-Abreu, V. (2002). Extensions of type G and marginal infinite divisibility, *Theory Probab. Appl.* **47**: 202–218.

Barndorff-Nielsen, O. E. and Pérez-Abreu, V. (2007). Matrix subordinators and related Upsilon transformations, *Theory Probab. Appl.* . to appear.

Barndorff-Nielsen, O. E. and Prause, K. (2001). Apparent scaling, *Finance Stoch.* **5**: 103–113.

Barndorff-Nielsen, O. E. and Schmiegel, J. (2004). Lévy-based tempo-spatial modelling; with applications to turbulence, *Uspekhi Mat. Nauk* **59**: 65–91.

Barndorff-Nielsen, O. E. and Shephard, N. (2001a). Modelling by Lévy processess for financial econometrics, *in* O. E. Barndorff-Nielsen, T. Mikosch and S. I. Resnick (eds), *Lévy Processes – Theory and Applications*, Birkhäuser, Basel, pp. 283–318.

Barndorff-Nielsen, O. E. and Shephard, N. (2001b). Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial econometrics (with discussion), *J. R. Stat. Soc. B Statist. Methodol.* **63**: 167–241.

Barndorff-Nielsen, O. E. and Shephard, N. (2002). Econometric analysis of realized volatility and its use in estimating stochastic volatility models, *J. R. Stat. Soc. B Statist. Methodol.* **64**: 252–280.

Barndorff-Nielsen, O. E. and Shephard, N. (2003). Integrated OU processes and non-Gaussian OU-based stochastic volatility models, *Scand. J. Statist.* **30**: 277–295.

Barndorff-Nielsen, O. E. and Shephard, N. (2004). Power and bipower variation with stochastic volatility and jumps (with discussion), *Journal of Financial Econometrics* **2**: 1–48.

Barndorff-Nielsen, O. E. and Shephard, N. (2007). *Financial Volatility in Continuous Time*, Cambridge University Press, Cambridge. to appear.

Barndorff-Nielsen, O. E. and Stelzer, R. (2005). Absolute moments of generalized hyperbolic distributions and approximate scaling of normal inverse Gaussian Lévy processes, *Scand. J. Statist.* **32**: 617–637.

Barndorff-Nielsen, O. E. and Stelzer, R. (2007). Positive-definite matrix processes of finite variation, *Probab. Math. Statist.* **27**: 3–43.

Bauwens, L., Laurent, S. and Rombouts, J. V. K. (2006). Multivariate GARCH models: A survey, *J. Appl. Econometrics* **21**: 79–109.

Benth, F. E. and Šaltytė-Benth, J. (2005). Stochastic modelling of temperature variations with a view towards weather derivatives, *Appl. Math. Fin.* **12**: 53–85.

Berman, A. (1988). Complete positivity, *Linear Algebra Appl.* **107**: 57–63.

Berman, A. and Shaked-Monderer, N. (2003). *Completely Positive Matrices*, World Scientific, Singapore.

Bhatia, R. (1997). *Matrix Analysis*, Vol. 169 of *Graduate Texts in Mathematics*, Springer, New York.

Bibby, B. M. and Sørensen, M. (1997). A hyperbolic diffusion model for stock prices, *Finance Stoch.* **1**: 25–41.

Bichteler, K. (2002). *Stochastic Integration with Jumps*, Vol. 89 of *Encyclopedia of Mathematics and its Applications*, Cambridge University Press, Cambridge.

Billingsley, P. (1999). *Convergence of Probability Measures*, Wiley, New York.

Birman, M. S., Koplienko, L. S. and Solomjak, M. Z. (1975). Estimates of the spectrum of a difference of fractional powers of selfadjoint operators, *Izv. Vysš. Učebn. Zaved. Matematika* **154**: 3–10. English translation: Soviet Math. (Iz. VUZ) 19 (1975), no. 3, 1–6.

Bladt, M. and Sørensen, M. (2005). Statistical inference for discretely observed Markov jump processes, *J. R. Stat. Soc. Ser. B Stat. Methodol.* **67**: 395–410.

Blæsild, P. (1981). On the two-dimensional hyperbolic distribution and some related distributions; with an application to Johannsen's bean data, *Biometrika* **68**: 251–263.

Blæsild, P. and Jensen, J. L. (1981). Multivariate distributions of hyperbolic type, *in* C. Taillie, G. P. Patil and B. A. Baldessari (eds), *Statistical distributions in scientific work - Proceedings of the NATO Advanced Study Institute held at the Université degli Studi di Trieste, Triest, Italy, July 10 - August 1, 1980*, Vol. 4, D. Reidel, Dordrecht, Holland, pp. 45–66.

Blæsild, P. and Sørensen, M. K. (1992). 'hyp' - a computer programme for analyzing data by means of the hyperbolic distribution, *Research Report 248*, Department of Theoretical Statistics, University of Århus, Århus, Denmark.

Bollerslev, T. (1986). Generalized autoregeressive conditional heteroskedasticity, *J. Econometrics* **31**: 307–327.

Bollerslev, T., Engle, R. F. and Wooldridge, J. M. (1988). A capital asset pricing model with time varying covariances, *Journal of Political Economy* **96**: 116–131.

Bougerol, P. and Picard, N. (1992). Stationarity of GARCH processes and of some nonnegative time series, *J. Econometrics* **52**: 115–127.

Boussama, F. (1998). *Ergodicité, mélange et estimation dans les modèles GARCH*, PhD Thesis, Université Paris 7, Paris, France.

Boussama, F. (2006). Ergodicité des chaînes de Markov à valeurs dans une variété algébrique: application aux modèles GARCH multivariés, *C. R. Math. Acad. Sci. Paris* **343**: 275–278.

Brockwell, P., Chadraa, E. and Lindner, A. (2006). Continuous time GARCH processes, *Ann. Appl. Probab.* **16**: 790–826.

Brockwell, P. J. (2001a). Continuous-time ARMA processes, *in* D. N. Shanbhag and C. R. Rao (eds), *Stochastic Processes: Theory and Methods*, Vol. 19 of *Handbook of Statistics*, Elsevier, Amsterdam, pp. 249–276.

Brockwell, P. J. (2001b). Lévy-driven CARMA processes, *Ann. Inst. Statist. Math.* **52**: 1–18.

Brockwell, P. J. (2004). Representations of continuous-time ARMA processes, *J. Appl. Probab.* **41A**: 375–382.

Brockwell, P. J. and Davis, R. A. (1991). *Time Series: Theory and Methods*, 2nd edn, Springer, New York.

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

Bronstein, I., Semendjaev, K., Mühlig, H. and Musiol, G. (2000). *Taschenbuch der Mathematik*, 5th edn, Verlag Harri Deutsch, Thun.

Bru, M.-F. (1991). Wishart processes, *J. Theoret. Probab.* **4**: 725– 751.

Chandrasekharan, K. (1989). *Classical Fourier Transforms*, Springer, Berlin.

Chib, S., Nardari, F. and Shephard, N. (2006). Analysis of high-dimensional multivariate stochastic volatility models, *J. Econometrics* **134**: 341–371.

Chiu, T. Y. M., Leonard, T. and Tsui, K.-W. (1996). The matrix-logarithmic covariance model, *J. Amer. Statist. Assoc.* **91**: 198–210.

Choi, M.-D. (1975). Positive semidefinite biquadratic forms, *Linear Algebra Appl.* **12**: 95–100.

Chojnowska-Michalik, A. (1987). On processes of Ornstein-Uhlenbeck type in Hilbert space, *Stochastics* **21**: 251–286.

Cont, R. and Tankov, P. (2004). *Financial Modelling with Jump Processes*, CRC Financial Mathematical Series, Chapman & Hall, London.

Corradi, V. (2000). Reconsidering the continuous time limit of the GARCH(1,1) process, *J. Econometrics* **96**: 145–153.

Culver, W. J. (1966). On the existence and uniqueness of the real logarithm of a matrix, *Proc. Amer. Math. Soc.* **17**: 1146–1151.

da Prato, G. and Zabczyk, J. (1996). *Ergodicity for Infinite Dimensional Systems*, Vol. 229 of *London Mathematical Society Lecture Notes Series*, Cambridge University Press, Cambridge.

Davydov, Y. A. (1973). Mixing conditions for Markov chains, *Theory Probab. Appl.* **18**: 312–328.

Doob, J. L. (1944). The elementary Gaussian processes, *Ann. Math. Stat.* **25**: 229–282.

Doob, J. L. (1953). *Stochastic Processes*, John Wiley & Sons, New York.

Doukhan, P. (1994). *Mixing*, Vol. 85 of *Lecture Notes in Statistics*, Springer, New York.

Dragan, V., Freiling, G., Hochhaus, A. and Morozan, T. (2004). A class of nonlinear differential equations on the space of symmetric matrices, *Electron. J. Differential Equations* **2004**(96): 1–48.

Drost, F. C. and Nijman, T. E. (1993). Temporal aggregation of GARCH processes, *Econometrica* **61**: 909–927.

Duda, J. (2007). Cone monotone mappings: Continuity and differentiability, *Nonlinear Anal.* . to appear.

Dynkin, E. B. (1965). *Markov Processes Volume I*, Vol. 121 of *Grundlehren der Mathematischen Wissenschaften*, Springer, Berlin.

Eberlein, E. (2001). Application of generalized hyperbolic Lévy motions to finance, *in* O. E. Barndorff-Nielsen, T. Mikosch and S. I. Resnick (eds), *Lévy Processes – Theory and Applications*, Birkhäuser, Basel, pp. 283–318.

Eberlein, E. and Hammerstein, E. A. v. (2004). Generalized hyperbolic and inverse Gaussian distribution: Limiting cases and approximation of processes, *in* R. Dalang, M. Dozzi and F. Russo (eds), *Seminar on stochastic analysis, random fields and applications IV; Centro Stefano Franscini, Ascona, May 2002*, Vol. 58 of *Progress in probability*, Birkhäuser, Basel, pp. 221–264.

Eberlein, E. and Keller, U. (1995). Hyperbolic distributions in finance, *Bernoulli* **1**: 281–299.

Eberlein, E. and Özkan, F. (2003). Time consistency of Lévy models, *Quant. Finance* **3**: 40–50.

Embrechts, P. and Maejima, M. (2002). *Selfsimilar processes*, Princeton Series in Applied Mathematics, Princeton University Press, Princeton, New Jersey.

Emmer, S. and Klüppelberg, C. (2004). Optimal portfolios when stock prices follow an exponential Lévy process, *Finance Stoch.* **8**: 17–44.

Engle, R. F. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation, *Econometrica* **50**: 987–1008.

Engle, R. F. and Kroner, K. F. (1995). Multivariate simultaneous generalized ARCH, *Econometric Theory* **11**: 122–150.

Eriksson, A., Forsberg, L. and Ghysels, E. (2004). Approximating the probability distribution of functions of random variables: A new approach, *Serie Scientifique 21*, CIRANO, Quebec, Canada.

Ethier, S. N. and Kurtz, T. G. (1986). *Markov Processes: Characterization and Convergence*, Wiley Series in Probability and Mathematical Statistics, Wiley, New York.

Fasen, V. (2007). Asymptotic results for sample autocovariance functions and extremes of integrated generalized Ornstein-Uhlenbeck processes, *Preprint*, ORIE, Cornell University, Ithaca, NY. available from: `www.ma.tum.de/stat`.

Fasen, V., Klüppelberg, C. and Lindner, A. (2006). Extremal behavior of stochastic volatility models, *in* M. R. Grossinho, A. N. Shiryaev, M. Eaquivel and P. E. Oliviera (eds), *Stochastic Finance*, Springer, New York, pp. 107–155.

Freitag, E. and Busam, R. (2000). *Funktionentheorie I*, 3rd edn, Springer, Berlin.

Frisch, U. (1995). *Turbulence - The legacy of A.N. Kolmogorow*, Cambridge University Press, Cambridge.

Genon-Catalot, V., Jeantheau, T. and Larédo, C. (2000). Stochastic volatility as hidden markov models and statistical applications, *Bernoulli* **6**: 1051–1079.

Ghomrasni, R. and Peskir, G. (2003). Local time-space calculus and extensions of Ito's formula, *in* J. Hoffmann-Jørgensen, M. B. Marcus and J. A. Wellner (eds), *High Dimensional Probability III*, Vol. 55 of *Progress in Probability*, Birkhäuser, Basel, pp. 177–192.

Gihman, I. I. and Skorohod, A. V. (1974). *The Theory of Stochastic Processes I*, Vol. 210 of *Grundlehren der Mathematischen Wissenschaften*, Springer, Berlin.

Gihman, I. I. and Skorohod, A. V. (1975). *The Theory of Stochastic Processes II*, Vol. 218 of *Grundlehren der Mathematischen Wissenschaften*, Springer, Berlin.

Gourieroux, C. (2006). Continuous time Wishart process for stochastic risk, *Econometric Rev.* **25**: 177–217.

Gourieroux, C., Jasiak, J. and Sufana, R. (2004). The Wishart autoregressive process of multivariate stochastic volatility, *Document de travail du CREST 2004-32*, CREST, Malakoff, France.

Gradshteyn, I. and Ryzhik, I. W. (1965). *Tables of Integrals, Series and Products*, Academic Press, New York.

Guillaume, D. M., Dacorogna, M. M., Davé, R. D., Müller, U. A., Olsen, R. B. and Pictet, O. V. (1997). From the bird's eye to the microscope: a survey of new stylized facts of the intra-daily foreign exchange markets, *Finance Stoch.* **1**: 95–129.

Gupta, A. K. and Nagar, D. K. (2000). *Matrix Variate Distributions*, Chapman and Hall, Boca Raton.

Hafner, C. M. (2003). Fourth moment structure of multivariate GARCH models, *Journal of Financial Econometrics* **1**: 26–54.

Hafner, C. M. (2007). Temporal aggregation of multivariate GARCH processes, *J. Econometrics* . to appear.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators, *Econometrica* **50**: 1029–1054.

Hartmann, D. and Bowman, D. (1993). Efficiency of the log-hyperbolic distribution – a case study: Pattern of sediment sorting in a small tidal-inlet – Het Zwin, The Netherlands, *Journal of Coastal Research* **9**: 1044–1053.

Harvey, A. C., Ruiz, E. and Shephard, N. (1994). Multivariate stochastic variance models, *Rev. Econom. Stud.* **61**: 247–264.

Haug, S. (2007). *Exponential COGARCH and other continuous time models – with applications to high-frequency data*, PhD thesis, Centre for Mathematical Sciences, Munich University of Technology, Garching bei München.

Haug, S. and Czado, C. (2006). A fractionally integrated ECOGARCH process, *Discussion Paper 484*, SFB 386, Munich. available from: `www.stat.uni-muenchen.de/sfb386/`.

Haug, S. and Czado, C. (2007). An exponential continuous time GARCH process, *J. Appl. Probab.* **44**. to appear.

Haug, S., Klüppelberg, C., Lindner, A. and Zapp, M. (2007). Method of moment estimation in the COGARCH(1,1) model, *Econom. J.* **10**: 320–341.

Haug, S. and Stelzer, R. (2007). First jump approximation of a Lévy driven SDE process and an application to multivariate ECOGARCH processes. In Preparation.

Horn, R. A. and Johnson, C. R. (1985). *Matrix Analysis*, Cambridge University Press, Cambridge.

Horn, R. A. and Johnson, C. R. (1991). *Topics in Matrix Analysis*, Cambridge University Press, Cambridge.

Hubalek, F. and Nicolato, E. (2005). On multivariate extensions of Lévy-driven Ornstein-Uhlenbeck type stochastic volatility models and multi-asset options. In preparation.

Ikeda, N. and Watanabe, S. (1989). *Stochastic differential equations and diffusion processes*, Vol. 24 of *North Holland Mathematical Library*, 2nd edn, North-Holland, Amsterdam.

Ismail, M. E. H. (1977). Integral representations and complete monotonicity of various quotients of Bessel functions, *Canad. J. Math.* **29**: 1198–1207.

Jacod, J. (1979). *Calcul Stochastique et Problèmes de Martingales*, Vol. 714 of *Lecture Notes in Mathematics*, Springer, Berlin.

Jacod, J. (2007). Inference for stochastic processes, *in* Y. Ait-Sahalia and L. P. Hansen (eds), *Handbook of Financial Econometrics*, Elsevier, Oxford. to appear.

Jacod, J. and Shiryaev, A. N. (2003). *Limit Theorems for Stochastic Processes*, Vol. 288 of *Grundlehren der Mathematischen Wissenschaften*, 2nd edn, Springer, Berlin.

Johnson, C. R. and Pierce, S. (1985). Linear maps on Hermitian matrices: The stabilizer of an inertia class, *Canad. Math. Bull.* **28**: 401–404.

Jones, R. H. and Ackerson, L. M. (1990). Serial correlation in unequally spaced longitudinal data, *Biometrika* **77**: 721–731.

Jørgensen, B. (1982). *Statistical Properties of the Generalized Inverse Gaussian Distribution*, Lecture Notes in Statistics, Springer, New York.

Jurek, Z. J. and Mason, D. J. (1993). *Operator-limit Distributions in Probability Theory*, John Wiley & Sons, New York.

Kallsen, J. and Vesenmayer, B. (2006). COGARCH as a continuous-time limit of GARCH(1,1), *Preprint*, Centre for Mathematical Sciences, Munich University of Technology, Garching.

Karandikar, R. L. (1982a). A.s. approximation results for multiplicative stochastic integrals, *Séminaire de Probabilities (Strasbourg), tome 16*, Vol. 920 of *Lecture Notes in Mathematics*, Springer, Berlin, pp. 384–391.

Karandikar, R. L. (1982b). Multiplicative decomposition of non-singular matrix valued continuous semimartingales, *Ann. Probab.* **10**: 1088–1091.

Karandikar, R. L. (1991). Multiplicative decomposition of nonsingular matrix valued semimartingales, *Séminaire de Probabilites (Strasbourg), tome 25*, Vol. 1485 of *Lecture Notes in Mathematics*, Springer, Berlin, pp. 262–269.

Kawakatsu, H. (2006). Matrix exponential GARCH, *J. Econometrics* **134**: 95–128.

Kessler, M. and Rahbek, A. (2004). Identification and inference for multivariate cointegrated and ergodic Gaussian diffusions, *Stat. Inference Stoch. Process.* **7**: 137–151.

Klüppelberg, C., Lindner, A. and Maller, R. (2004). A continuous-time GARCH process driven by a Lévy process: Stationarity and second-order behaviour, *J. Appl. Probab.* **41**: 601–622.

Klüppelberg, C., Lindner, A. and Maller, R. (2006). Continuous time volatility modelling: COG-ARCH versus Ornstein-Uhlenbeck models, *in* Y. Kabanov, R. Lipster and J. Stoyanov (eds), *The Shiryaev Festschrift: From Stochastic Calculus to Mathematical Finance*, Springer, Berlin, pp. 393–419.

Krengel, U. (1985). *Ergodic Theorems*, Walter de Gruyter, Berlin.

Kristjansson, L. and McDougall, I. (1982). Some aspects of the late tertiary geomagnetic field in Iceland, *The Geophysical Journal of the Royal Astronomical Society* **68**: 273–294.

Kurtz, T. and Protter, P. (1991). Weak limit theorems for stochastic integrals and stochastic differential equations, *Ann. Probab.* **19**: 1035–1070.

Kurtz, T. and Protter, P. (1996). Weak convergence of stochastic integrals and differential equations, *in* D. Talay and L. Tularo (eds), *Probabilistic Models for Nonlinear Partial Differential Equations*, Vol. 1627 of *Lecture Notes in Mathematics*, Springer, pp. 1–41.

Kyprianou, A. E. (2006). *Introductory Lectures on Fluctuations of Lévy Processes with Applications*, Universitext, Springer, Berlin.

Lang, S. (1993). *Complex Analysis*, Vol. 103 of *Graduate Texts in Mathematics*, 3rd edn, Springer, New York.

Lévy, P. (1948). The arithmetical character of the Wishart distribution, *Proc. Cambdrige Philos. Soc.* **44**: 123–124.

Li, C.-K. and Pierce, S. (2001). Linear preserver problems, *Amer. Math. Monthly* **108**: 591–605.

Li, C.-K., Rodman, L. and Semrl, P. (2003). Linear maps on selfadjoint operators preserving invertibility, positive definiteness, numerical range, *Canad. Math. Bull.* **46**: 216–228.

Lindberg, C. (2005). *Portfolio Optimization and Statistics in Stochastic Volatility Markets*, PhD thesis, Department of Mathematical Sciences, Division of Mathematical Statistics, Chalmers University of Technology and Göteborg University, Göteborg, Sweden.

Lindner, A. and Maller, R. (2005). Lévy integrals and the stationarity of generalised Ornstein-Uhlenbeck processes, *Stochastic Process. Appl.* **115**: 1701–1722.

Loéve, M. (1977). *Probability Theory I*, Vol. 45 of *Graduate Texts in Mathematics*, 4th edn, Springer, New York.

Loewy, R. (1992). A survey of linear preserver problems - chapter 3: Inertia preservers, *Linear Multilinear Algebra* **33**: 22–30.

Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis*, Springer, Berlin.

Maejima, M. and Rosiński, J. (2002). Type G distributions on $\mathbb{R}^d$, *J. Theoret. Probab.* **15**: 323–341.

Magnus, J. R. and Neudecker, H. (1979). The commutation matrix: Some properties and applications, *Ann. Statist.* **7**: 381–394.

Maller, R. A., Müller, G. and Szimayer, A. (2006). GARCH modelling in continuous time for irregularly spaced time series data, *Preprint*, Centre for Mathematical Sciences, Munich University of Technology, Garching.

Mandelbrot, B. B. (1963). The variation of certain speculative prices, *Journal of Business* **36**: 394–419.

Mandelbrot, B. B. (1997). *Fractals and scaling in finance*, Springer, New York.

Marcus, M. B. and Rosinski, J. (2005). Continuity and boundedness of infinitely divisible processes: a Poisson point process approach, *J. Theoret. Probab.* **18**: 109 – 160.

Marquardt, T. and Stelzer, R. (2007). Multivariate CARMA processes, *Stochastic Process. Appl.* **117**: 96–120.

Masuda, H. (2004). On multidimensional Ornstein-Uhlenbeck processes driven by a general Lévy process, *Bernoulli* **10**: 97–120.

Maxfield, J. E. and Minc, H. (1962). On the equation $X'X = A$, *Proc. Edinburgh Math. Soc.* **13**: 125–129.

Mencia, F. J. and Sentana, E. (2004). Estimation and testing of dynamic models with generalized hyperbolic innovations, *CEMFI Working Paper 0411*, CEMFI, Madrid, Spain.

Métivier, M. (1982). *Semimartingales: a Course on Stochastic Processes*, Vol. 2 of *De Gruyter Studies in Mathematics*, Walter de Gruyter, Berlin.

Métivier, M. and Pellaumail, J. (1980a). On a stopped Doob's inequality and general stochastic equations, *Ann. Probab.* **8**: 96–114.

Métivier, M. and Pellaumail, J. (1980b). *Stochastic Integration*, Academic Press, New York.

Mossberg, M. and Larsson, E. K. (2004). Fast and approximative estimation of continuous-time stochastic signals from discrete-time data, *IEEE International Conference on Acoustics, Speech, and Signal Processing, 2004. Proceedings. (ICASSP '04)*, Vol. 2, Montreal, QC, May 17-21, 2004, pp. 529–532.

Müller, G. (2006). MCMC estimation of the COGARCH(1,1) model, *Preprint*, Centre for Mathematical Sciences, University of Technology, Garching. available from: `www.ma.tum.de/stat`.

Müller, U. A., Dacorogna, M. M., Olsen, R. B., Pictet, O. V., Schwarz, M. and Morgenegg, C. (1990). Statistical study of foreign exchange rates, empirical evidence of a price scaling law and intraday analysis, *Journal of Banking and Finance* **14**: 1189–1208.

Nelson, D. B. (1990). ARCH models as diffusion approximations, *J. Econometrics* **45**: 7–38.

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach, *Econometrica* **59**: 347–370.

Øksendal, B. (1998). *Stochastic Differential Equations – An Introduction with Applications*, 5th edn, Springer.

Östermark, R. (2001). Genetic modelling of multivariate EGARCHX-processes: evidence on the international asset return signal response mechanism, *Comput. Statist. Data Anal.* **38**: 71–93.

Pérez-Abreu, V. and Rocha-Arteaga, A. (2005). Covariance-parameter Lévy proceses in the space of trace-class operators, *Inf. Dimens. Anal. Quantum Probab. Relat. Topics* **8**: 33–54.

Peskir, G. (2005). A change-of-variables formula with local time on curves, *J. Theoret. Probab.* **18**: 499–535.

Picard, J. (1996). On the existence of smooth densities for jump processes, *Probab. Theory Related Fields* **105**: 481–511.

Pierce, S., Lim, M. H., Loewy, R., Li, C. K., Tsing, N. K., McDonald, B. and Beasley, L. (1992). A survey of linear preserver problems, *Linear Multilinear Algebra* **33**: 1–129.

Pigorsch, C. and Stelzer, R. (2007). A multivariate generalization of the Ornstein-Uhlenbeck stochastic volatility model, *Preprint*, Centre for Mathematical Science, Munich University of Technology, Garching, Germany. available from: `www.ma.tum.de/stat`.

Pollard, D. (1984). *Convergence of Stochastic Processes*, Springer, New York.

Prause, K. (1997). Modelling financial data using generalized hyperbolic distributions, *FDM Preprint* **48**: 1–11.

Prause, K. (1999). *The Generalized Hyperbolic Model: Estimation, Financial Derivatives and Risk Measures*, Dissertation, Mathematische Fakultät, Albert-Ludwigs-Universität Freiburg i. Br., Freiburg, Germany.

Protter, P. (1978). $\mathscr{H}^p$ stability of solutions of stochastic differential equations, *Z. Wahrscheinlichkeitstheorie verw. Gebiete* **44**: 337–352.

Protter, P. (2004). *Stochastic Integration and Differential Equations*, Vol. 21 of *Stochastic Modelling and Applied Probability*, 2nd edn, Springer-Verlag, New York.

Raible, S. (2000). *Lévy Processes in Finance: Theory, Numerics and Empirical Facts*, Dissertation, Mathematische Fakultät, Albert-Ludwigs-Universität Freiburg i. Br., Freiburg, Germany.

Rajput, B. S. and Rosinski, J. (1989). Spectral representations of infinitely divisible processes, *Probab. Theory Related Fields* **82**: 451–487.

Ran, A. C. M. and Reurings, M. C. B. (2002). The symmetric linear matrix equation, *Electron. J. Linear Algebra* **9**: 93–107.

Rasmus, S., Asmussen, S. and Wiktorsson, M. (2004). Pricing of some exotic options with NIG Lévy input, *in* M. Bubak, G. D. v. Albada, P. M. A. Sloot and J. Dongarra (eds), *Computational science – ICCS 2004: 4th international conference, Krakow, Poland. Proceedings, part IV*, Vol. 3039 of *Lecture notes in computer science*, Springer, pp. 795–802.

Reiß, M., Riedle, M. and van Gaans, O. (2006). Delay differential equations driven by Lévy processes: Stationarity and Feller properties, *Stochastic Process. Appl.* **116**: 1409–1432.

Reiß, M., Riedle, M. and van Gaans, O. (2007). On Émery's inequality and a variation-of-constants formula, *Stoch. Anal. Appl.* **25**: 353–379.

Remmert, R. (1991). *Theory of complex functions*, Vol. 122 of *Readings in Mathematics*, Springer, New York.

Ribarits, E. M. (2006). *Multivariate Modelling of Financial Time Series*, PhD thesis, Institute for Mathematical Methods in Economics, Vienna University of Technology, Vienna, Austria.

Rocha-Arteaga, A. (2006). Subordinators in a class of Banach spaces, *Random Oper. Stochastic Equations* **14**: 1–14.

Rogers, L. C. G. and Williams, D. (2000). *Diffusions, Markov Processes and Martingales - Volume 2: Itô Calculus*, 2nd edn, Cambridge University Press, Cambridge.

Rozanov, Y. A. (1967). *Stationary Random Processes*, Holden-Day, San Francisco.

Rudin, W. (1976). *Principles of Mathematical Analysis*, 3rd edn, McGraw-Hill, Singapore.

Rydberg, T. H. (1997). The normal inverse Gaussian Lévy process: Simulation and approximation, *Comm. Statist. Stochastic Models* **13**: 887–910.

Rydberg, T. H. (1999). Generalized hyperbolic diffusions with applications towards finance, *Math. Finance* **9**: 183–201.

Samorodnitsky, G. and Taqqu, M. S. (1994). *Stable non-Gaussian random processes*, Stochastic modeling, Chapman & Hall/CRC, Boca Raton.

Sato, K. (1999). *Lévy Processes and Infinitely Divisible Distributions*, Vol. 68 of *Cambridge Studies in Advanced Mathematics*, Cambridge University Press.

Sato, K. (2006). Addtitive processes and stochastic integrals, *Illinois J. Math.* **50**: 825–851.

Sato, K. and Yamazato, M. (1984). Operator-selfdecomposable distributions as limit distributions of processes of Ornstein-Uhlenbeck type, *Stochastic Process. Appl.* **17**: 73–100.

Scherrer, W. and Ribarits, E. (2007). On the parametrization of multivariate GARCH models, *Econometric Theory* **23**: 464–484.

Schneider, H. (1965). Positive operators and an inertia theorem, *Numer. Math.* **7**: 11–17.

Schoutens, W. (2003). *Lévy Processes in Finance - Pricing Financial Derivatives*, Wiley, Chicester.

Skorohod, A. V. (1991). *Random Processes with Independent Increments*, Vol. 47 of *Mathematics and Its Applications (Soviet Series)*, Kluwer, Dordrecht.

Sørensen, M. (2000). Prediction-based estimating functions, *Econom. J.* **3**: 123–147.

Stelzer, R. (2006). Linear operators on the real symmetric matrices whose exponential preserves the inertia, *Preprint*, Centre for Mathematical Sciences, Munich University of Technology, Garching. available from: `www.ma.tum.de/stat`.

Stelzer, R. (2007a). Multivariate continuous time Lévy-driven GARCH processes. In Preparation.

Stelzer, R. (2007b). On the relation between the vec and BEKK multivariate GARCH models, *Preprint*, Centre for Mathematical Sciences, Munich University of Technology, Garching. available from `www.ma.tum.de/stat`.

Størmer, E. (1963). Positive linear maps of operator algebras, *Acta Math.* **110**: 233–278.

Sutherland, R. A. and Lee, C.-T. (1994). Application of the log-hyperbolic distribution to Hawai'ian beach sands, *Journal of Coastal Research* **10**: 251–262.

Szimayer, A. and Maller, R. A. (2007). Finite approximation schemes for Lévy processes, and their application to optimal stopping problems, *Stochastic Process. Appl.* **117**: 1422–1447.

Todorov, V. and Tauchen, G. (2006). Simulation methods for Lévy-driven CARMA stochastic volatility models, *J. Bus. Econom. Statist.* **24**: 450–469.

Tsai, H. and Chan, K. S. (2005). A note on non-negative continuous time processes, *J. R. Stat. Soc. B Statist. Methodol.* **67**: 589–597.

Tsai, H. and Chan, K. S. (2006). A note on the non-negativity of continuous-time ARMA and GARCH processes, *Technical Report 359*, Department of Statistics & Actuarial Science, University of Iowa, Iowa City.

Tse, Y. and Hackard, J. C. (2004). Can island provide liquidity and price discovery in the dark?, *Review of Quantitative Finance and Accounting* **23**: 149–166.

Watson, G. N. (1952). *A Treatise on the Theory of Bessel Functions*, reprinted 2nd edn, Cambridge University Press.

Wei, W. W. S. (1990). *Time Series Analysis*, Addison-Wesley, Redwood City, California.

Werner, D. (2002). *Funktionalanalysis*, 4th edn, Springer, Berlin.

Wolfe, S. J. (1982). On a continuous analogue of the stochastic difference equation $X_n = \rho X_{n-1} + B_n$, *Stochastic Process. Appl.* **12**: 301–312.

Xu, C. (2004). Completely positive matrices, *Linear Algebra Appl.* **379**: 319–327.

Xu, T. H., Durst, F. and Tropea, C. (1993). The three-parameter log-hyperbolic distribution and its application to particle sizing, *Atomization and Sprays* **3**: 109–124.

Yamazato, M. (1983). Absolute continuity of operator-selfdecomposable distributions on $\mathbb{R}^d$, *J. Multivariate Anal.* **13**: 550–560.

Yang, S.-Y. and Doong, S.-C. (2004). Price and volatility spillovers between stock prices and exchange rates: Empirical evidence from the G-7 countries, *International Journal of Business and Economics* **3**: 139–153.

Yosida, K. (1965). *Functional Analysis*, Vol. 123 of *Die Grundlehren der mathematischen Wissenschaften*, Springer, Berlin.

General Notation

Symbols involving letters in alphabetical order:

$\mathcal{B}(\cdot)$	Borel σ-algebra
$\mathbb{C}$	complex numbers
$GL_n(\cdot)$	invertible $n \times n$ matrices
$\Im$	imaginary part
i	imaginary unit
I_n	$n \times n$ identity matrix
$I_A(\cdot)$	indicator function of the set A
$\log^+(\cdot)$	maximum of logarithm and zero, i.e. $\log^+(a) = \max\{\log(a), 0\}$
$L^p(\cdot)$	space of p-times integrable functions/random variables
$M_{m,n}(\cdot)$	$m \times n$ matrices
$M_n(\cdot)$	$n \times n$ matrices
$\mathbb{N}$	natural numbers
$\mathbb{N}_0$	natural numbers and zero
$O(\cdot)$	of the same order, e.g. $a(t) = O(b(t))$, iff $\lim_{t\to\infty} \frac{a(t)}{b(t)} = C \in \mathbb{R}^+ \backslash \{0\}$
$o(\cdot)$	of smaller order, e.g. $a(t) = o(b(t))$, iff $\lim_{t\to\infty} \frac{a(t)}{b(t)} = 0$
$P(\cdot)$	probability
$\Re$	real part
$\mathbb{R}$	real numbers
$\mathbb{R}^+$	non-negative real numbers
$\mathbb{R}^{++}$	positive real numbers (zero excluded)
$\mathbb{S}_n$	symmetric $n \times n$ matrices
$\mathbb{S}_n^+$	positive semi-definite $n \times n$ matrices
$\mathbb{S}_n^{++}$	positive definite $n \times n$ matrices
$\cdot^T$	transposed of a matrix/vector
$\mathbb{Z}$	whole numbers

Other symbols in alphabetical order of their meaning:

$\cdot^*$	adjoint of a linear operator, Hermitian of a matrix/vector
$[\cdot, \cdot]$	closed interval
$\sim$	distributed as, asymptotically equal
$\emptyset$	empty set
$1_A(\cdot)$	indicator function of the set A
$\vee$	maximum, e.g. $a \vee b = \max\{a, b\}$
$\wedge$	minimum, e.g. $a \wedge b = \min\{a, b\}$
$\|\cdot\|$	norm
$(\cdot, \cdot)$	open interval

Abbreviations

a.s.	almost sure
a.e.	almost everywhere
cf.	confer
ch.	chapter
cor.	corollary
def.	definition
e.g.	for example
eq.	equation
et al.	et alii
etc.	et cetera
f	and the following one
ff	and the following ones
i.e.	that is (id est)
iff	if and only if
i.i.d.	independent and identically distributed
p.	page
pp.	pages
prop.	proposition
resp.	respectively
rv	random variable
SDE	stochastic differential equation
s.t.	such that
th.	theorem
viz.	namely, that is to say (vide licet)
w.l.o.g.	without loss of generality
w.r.t.	with respect to