

Stephan Haug

**Exponential COGARCH and
other continuous time models**
with applications to high frequency data.

December 2006

Center for Mathematical Sciences
Munich University of Technology
85747 Garching bei München

Technische Universität München
Zentrum Mathematik
Lehrstuhl für Mathematische Statistik

**Exponential COGARCH and other
continuous time models**
with applications to high frequency data

Stephan Haug

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. Jan Kallsen
Prüfer der Dissertation: 1. Univ.-Prof. Claudia Czado, Ph.D.
2. Prof. Peter Brockwell, Ph.D.,
Colorado State University, Fort Collins, USA

Die Dissertation wurde am 07.12.2006 bei der Technischen Universität eingereicht
und durch die Fakultät für Mathematik am 23.03.2007 angenommen.

Abstract: We address several approaches to modelling high frequency financial data in continuous time. Besides considering estimation for the existing continuous time GARCH(1, 1) (COGARCH) process we will propose three new models. At first we suggest a method of moment estimator for the parameters of the COGARCH(1, 1) process. We show that the resulting estimators are consistent and asymptotically normal and investigate the empirical quality in a simulation study based on the compound Poisson and Variance Gamma driven COGARCH(1, 1) model. The model is also fitted to high-frequency financial data from the New York Stock Exchange. In the following chapter we develop the first new model, an exponential COGARCH(p, q) process. We investigate stationarity and moment properties of the model. An instantaneous leverage effect can be shown if $p = q$. The first steps in estimating this new model are taken by proposing a quasi-maximum likelihood type estimator for the parameters of a compound Poisson ECOGARCH(1, 1) process. To account for the strong persistence in volatility, which is sometimes observed in empirical data, we develop a fractionally integrated ECOGARCH(p, d, q) process. Similarly to the short memory case we investigate stationarity and moment properties of the model. It is also shown that the long memory effect introduced in the log-volatility propagates to the volatility process. Finally considering absolute log returns as a proxy for stochastic volatility, the influence of explanatory variables on absolute log returns of ultra high frequency data is analysed. In particular we propose a new mixed effect model class for the absolute log returns. Explanatory variable information is used to model the fixed effects, whereas the error is decomposed in a non-negative Lévy driven continuous time ARMA process and a market microstructure noise component. The parameters are estimated in a state space approach with application of the Kalman filter. In a small simulation study the performance of the estimators is investigated. The model is applied to IBM trade data and the influence of bid-ask spread and duration is quantified on a daily basis.

*To Nici
for sharing
so many
office hours
with me.*

Contents

Introduction	1
1 Preliminaries	6
1.1 Lévy processes	6
1.2 Stochastic integration	17
1.2.1 General theory	17
1.2.2 Itô's formula	21
1.3 State space models	23
1.3.1 Discrete time state space model	24
1.3.2 Kalman filter and Gaussian likelihood	25
1.4 Continuous time autoregressive moving average processes	27
1.4.1 Lévy driven CARMA(p, q) processes	27
1.4.2 Lévy driven FICARMA(p, d, q) processes	32
1.5 Continuous time GARCH(1, 1) processes	34
2 Estimating the COGARCH(1, 1)	42
2.1 Identifiability of the model parameters	42
2.2 The estimation algorithm	44
2.3 Asymptotic properties of the moment estimators	46
2.4 Examples of COGARCH(1, 1) processes	51
2.4.1 Compound Poisson COGARCH(1, 1)	51
2.4.2 Variance Gamma COGARCH(1, 1)	54
2.5 Simulation study	55
2.5.1 Estimation results for the compound Poisson COGARCH(1, 1)	56
2.5.2 Estimation results for the Variance Gamma COGARCH(1, 1)	56
2.5.3 Estimation of the volatility σ_t^2	58
2.6 Real data analysis	59
3 Exponential continuous time GARCH process	64
3.1 The discrete time EGARCH process	65
3.2 Exponential COGARCH	66

3.3	Second order properties of the volatility process	75
3.4	Second order properties of the return process	78
3.4.1	Moments and autocovariance function of the return process	78
3.4.2	Leverage effect	81
4	QMLE of compound Poisson ECOGARCH(1,1)	84
4.1	Quasi MLE of compound Poisson ECOGARCH(1,1)	84
4.2	Simulation study	87
4.2.1	Leverage case with Gaussian jump distribution	88
4.2.2	Leverage case with student- t jump distribution	90
4.2.3	Non-leverage case	92
4.3	Prediction	93
4.3.1	One-step ahead prediction of the volatility	93
4.3.2	Prediction interval for the log-price	98
4.4	Analysis of General Motors stock prices	99
5	Fractionally integrated ECOGARCH process	104
5.1	Fractionally integrated exponential COGARCH	105
5.2	Second order properties of the volatility process	110
5.3	Second order properties of the return process	116
6	Mixed effect models for absolute log returns of ultra high frequency data	118
6.1	A mixed effect regression model for irregularly spaced data	120
6.1.1	Regression mean specification	120
6.1.2	Correlated residuals	121
6.1.3	A generalised regression model with CARMA(p,q) random effects	122
6.2	Parameter Estimation	122
6.2.1	Direct approach	125
6.2.2	State space approach	126
6.3	Simulation results	128
6.4	Application	128
6.4.1	Explorative data analysis	130
6.4.2	Estimation results	131
6.4.3	Analysis of the correlation structure	132
6.4.4	Prediction	133
6.4.5	Prediction results	135
	Conclusion and outlook	139

Appendix	140
A Basics on stochastic processes and related topics	140
A.1 Definitions and review of basic results	140
A.2 Auxiliary results	144
B Variogram for irregularly spaced time series	150
<i>References</i>	152
<i>Index</i>	159

Notation

Throughout the text stochastic processes like $(X_t)_{t \geq 0}$ will frequently just be referred to as X . The abbreviation a.s. denotes almost surely. Convergence in distribution, probability and almost surely is written as \xrightarrow{d} , $\xrightarrow{\mathbb{P}}$ and $\xrightarrow{\text{a.s.}}$, respectively. Equality in distribution is denoted by $\stackrel{d}{=}$.

We denote $d \times m$ matrices either by $\mathcal{A} = (A_{i,j})$ or bold capital Greek letters like $\Sigma = (\Sigma_{i,j})$ for $1 \leq i \leq d, 1 \leq j \leq m$. The space of real $d \times m$ matrices is denoted by $\mathbb{M}_{d,m}(\mathbb{R})$. The transpose of a matrix \mathcal{A} is written as \mathcal{A}^T and the identity in $\mathbb{M}_{d,d}$ as \mathcal{I}_d . Vectors are denoted by bold capital Roman letters like \mathbf{A} or bold small Greek letters like $\boldsymbol{\theta}$.

Let (S, \mathcal{F}) be a measurable space. Then the indicator function of a set $A \in \mathcal{F}$ is denoted by χ_A and defined for $x \in S$ by

$$\chi_A(x) := \begin{cases} 1, & \text{if } x \in A \\ 0, & \text{otherwise} \end{cases} .$$

Let (X, \mathcal{T}) be a topological space. The complement of a set $A \subseteq X$ is denoted by A^c and the closure in the topology \mathcal{T} by \overline{A} .

The Borel σ -algebra of \mathbb{R}^d is denoted by $\mathcal{B}(\mathbb{R}^d)$. $Leb(A)$ is the Lebesgue measure of a set A and $Leb(dx)$ is written dx . For $\mathbf{x} \in \mathbb{R}^d$, $\delta_{\mathbf{x}}(B)$ will be the Dirac measure of a set B , that is

$$\delta_{\mathbf{x}}(B) = \begin{cases} 1, & \text{if } \mathbf{x} \in B \\ 0, & \text{otherwise} \end{cases} .$$

Let f be a continuously differentiable function from \mathbb{R}^d to \mathbb{R}^m . Then the total derivative of $f = (f^1, \dots, f^m)^T$ at $\mathbf{x} \in \mathbb{R}^d$ is written as

$$\partial_{\mathbf{x}} f(\mathbf{x}) = \begin{pmatrix} \partial_1 f^1(\mathbf{x}) & \cdots & \partial_d f^1(\mathbf{x}) \\ \vdots & & \vdots \\ \partial_1 f^m(\mathbf{x}) & \cdots & \partial_d f^m(\mathbf{x}) \end{pmatrix} \in \mathbb{M}_{m,d}(\mathbb{R}),$$

where $\partial_i f^j(\mathbf{x})$ is the i -th first order partial derivative of $f^j(\mathbf{x})$ for $1 \leq i \leq d$, $1 \leq j \leq m$. Let f, g be function from \mathbb{R} to \mathbb{R} . Then by $f \sim g$ as $x \rightarrow \infty$ it is meant that $\lim_{x \rightarrow \infty} f(x)/g(x) = 1$.

Introduction

Over the years finance has proceeded from analysing monthly, to weekly, to daily data and is now confronted with the problem of designing models for intradaily data. Hence the frequency of available data has enormously increased over the last years. This process has come to an end, as observed by Engle (2000), since in almost all cases all transactions are recorded nowadays and nothing more can be expected. He called this limiting frequency "*ultra high frequency*". The most eminent characteristic of this ultra high frequency, or also called tick-by-tick, data is the irregular spacing of the observations in time. If one aggregates the data up to fixed time intervals, there is truly a loss of information, which should be avoided. Therefore it seems natural to model the price processes in continuous time. Of course newly proposed models should also reflect as many as possible of the other so-called *stylised facts* of financial data (cf. Cont (2001)), such as absence of autocorrelation of returns despite being dependent, volatility clustering on high levels and the leverage effect. A leverage effect in the econometric literature means the phenomenon of negative correlation between current returns and future volatility.

These findings suggest that a suitable model is one with stochastic volatility. Barndorff-Nielsen and Shephard (2001) proposed such a model, where the volatility process $(\sigma_t^2)_{t \geq 0}$ is described by an Ornstein-Uhlenbeck (OU) type process, which is driven by a Lévy process L with increasing sample path. It is defined by the stochastic differential equations

$$\begin{aligned}dX_t &= (\mu + \beta\sigma_t^2)dt + \sigma_t dW_t + \rho d\bar{L}_{\lambda t}, \\d\sigma_t^2 &= -\lambda\sigma_t^2 dt + dL_{\lambda t}, \quad \lambda > 0,\end{aligned}$$

for each $t \geq 0$, where W is a Brownian motion independent of L and σ_0^2 , $\bar{L}_t := L_t - \mathbb{E}(L_t)$ for each $t \geq 0$ and μ, β and ρ are constants. The process X is understood as the log-price process of an asset. For $\rho < 0$ the model is able to describe a leverage effect, but in the way that an increment in the volatility will lead to a negative shock in the log-price process and not the other way round, which seems more natural.

Another approach in modeling asset prices in continuous time are GARCH type models. In the discrete time context the ARCH (autoregressive conditionally heteroscedastic) and GARCH (generalised ARCH) models proposed by Engle (1982)

and Bollerslev (1986), respectively, have become most popular, since they capture many of the stylised facts. Different attempts have been made to define a continuous time process sharing the same features as the discrete time GARCH process. One approach could be summarised as diffusion approximation to the discrete time GARCH process as presented in Nelson (1990). This idea however leads to a stochastic volatility model driven by two independent Brownian motions, which are asymptotically non equivalent to GARCH models as shown in Wang (2002). The GARCH model in contrast is driven by only one source of randomness. Therefore Klüppelberg et al. (2004) recently introduced a new continuous time GARCH(1, 1) (COGARCH) process driven by a single Lévy process L . The model is defined by the stochastic differential equations

$$\begin{aligned} dG_t &= \sigma_t dL_t, \\ d\sigma_{t+}^2 &= (\beta - \eta\sigma_t^2) dt + \varphi \sigma_t^2 d[L, L]_t^d, \end{aligned}$$

for each $t \geq 0$, where β, η and φ are constants and $[L, L]_t^d = \sum_{0 \leq s \leq t} (\Delta L_s)^2$. The asset log-price is described by the stochastic process G and its increments $G^{(r)} = G_{t+r} - G_t$, $r > 0$, correspond to the discrete time GARCH(1, 1) process. One clearly recognises that the driving process of the volatility is similar to the discrete time case. This model was further extended by Brockwell et al. (2006) to a COGARCH(p, q) process with $p \leq q$. In analogy to the discrete time case the model is not able to describe a leverage effect without further extensions. Different solutions to this problem have been proposed in discrete time. On the one hand it was tried to keep the GARCH model and just extend it to incorporate a leverage effect, as done e.g. in Glosten et al. (1993). Another attempt was taken by Nelson (1991), who introduced the exponential GARCH (EGARCH) process. It is not directly a GARCH model, since it models the log-volatility process with the innovations of the returns as driving noise sequence in contrast to the GARCH process, but it shares important properties with it and additionally is able to describe a leverage effect. To our knowledge there exists no continuous time extension of the EGARCH process up to now. Such an extension will be given in this thesis.

Another stylised fact, the sometimes observed long-run volatility persistence, is not taken into account by any of the models addressed so far. Numerous discrete time long memory models have been introduced like the integrated GARCH (IGRACH) process of Engle and Bollerslev (1986), the fractionally integrated GARCH (FIGARCH) process of Baillie et al. (1996) or the fractionally integrated EGARCH (FIEGARCH) process of Bollerslev and Mikkelsen (1996). The FIGARCH process has to be treated carefully since the existence of a stationary version is not clear; see Section 4 in Mikosch and Stărică (2000). Although in the continuous time setting there have been suggested a number of long memory models. Comte and Renault (1998) defined a continuous time stochastic volatility model by specifying the log-volatility process as an OU process driven by a fractional Brownian motion.

Brockwell and Marquardt (2005) proposed to model the stochastic volatility as a non-negative fractionally integrated continuous time ARMA (CARMA) process. Another non-Gaussian continuous time SV model with long memory was introduced by Anh et al. (2002), where they define their model via the Green function solution of a fractional differential equation driven by a Lévy process. All of these are stochastic volatility type models. Continuous time GARCH like models, which are able to describe a long memory effect in the volatility process, do not exist as far as we know. Again we will propose such an extension.

When dealing with ultra high frequency data one is provided with additional information about each trade, which should be taken into account as pointed out in Engle (2000). Such information consists of the bid and ask prices quoted at the transaction time, the volume of the trade or the duration since the last trade. Using this information in a sensible way can provide more insight into the market microstructure as shown in Engle (2000). We will also address this topic at the end of the thesis, which is organised as follows.

This thesis is divided into six chapters, whereas Chapter 2, 3, 5 and 6 are based on the papers Haug, Klüppelberg, Lindner, and Zapp (2007), Haug and Czado ((2006a), (2006b), (2006c)), respectively.

The framework of the thesis will be set in Chapter 1. After a short introduction to the theory of Lévy processes, stochastic integration with respect to martingale valued random measures will be defined. Further we give a short review about the discrete time state space representation of time series and how the Kalman filter can be applied to compute Gaussian likelihoods. Afterwards the continuous time ARMA process, as well as the fractionally integrated CARMA process are introduced and discussed shortly. Finally the continuous time GARCH(1, 1) process of Klüppelberg et al. (2004) is presented, some useful properties stated and an expression for the fourth moment of the log-return process derived.

In Chapter 2 we develop a method of moment estimator for the parameters of the COGARCH(1, 1) process. The estimator requires the second and fourth moment of the log-return process, as well as the autocorrelation function of the squared log-returns. It is shown that under these conditions the parameters are identifiable. Before we consider the asymptotic properties of the estimator, a strong mixing condition for an equidistant sequence of log-returns is derived. It is then shown that the resulting estimators are consistent and asymptotically normal. The small sample properties are investigated in a simulation study based on the compound Poisson and Variance Gamma driven COGARCH(1, 1) model. The model is also fitted to high-frequency financial data from the New York Stock Exchange (NYSE) and Nasdaq. In this real data example the need for modeling the asymmetry in financial log-returns is highlighted. At the end of the chapter a possible solution to that problem is presented for the COGARCH(1, 1) model.

Our aim in Chapter 3 is to propose a continuous time parameter model, which shares the main features of the discrete time EGARCH model. The new model will

be called an *exponential COGARCH*(p, q) (ECOGARCH) process, and models the stock log-prices. The increments (log-returns) of the process will correspond to the discrete time EGARCH process. The ECOGARCH process is defined in such a way that it is able to describe the asymmetry in financial data, i.e. the leverage effect, directly. After investigating stationarity conditions, a strong mixing property of the log-return process is shown. The second order structure of the volatility and the log-return process is analysed and moment expressions derived, which depend on the characteristic triplet of the driving Lévy process. An instantaneous leverage effect can be shown if $p = q$.

The first steps concerning inference in this new model are taken in Chapter 4 by proposing a quasi-maximum likelihood type estimator for the parameters of a compound Poisson ECOGARCH(1, 1) process. The empirical quality of the estimator is tested in a simulation study. We also consider the problem of predicting the volatility process and the construction of a prediction interval for the log-price process. In the end the model is fitted to ultra high-frequency financial data of the NYSE.

To account for the strong persistence in volatility, which is sometimes observed in empirical data, we develop in Chapter 5 a *fractionally integrated ECOGARCH*(p, d, q) (FIECOGARCH) process. Similarly to the short memory case we investigate stationarity and moment properties of the model. It is also shown that the long memory effect introduced in the log-volatility process propagates to the volatility process.

Finally considering absolute log returns as a proxy for the instantaneous volatility, the influence of explanatory variables on absolute log returns of ultra high frequency data is analysed in Chapter 6. In particular we propose a new mixed effect model class for the absolute log returns. Explanatory variable information, such as the bid-ask spread and the duration, is used to model the fixed effects, whereas the error is decomposed in a non-negative Lévy driven CARMA process and a market microstructure noise component. The parameters are estimated in a state space approach with application of an augmented Kalman filter. In a small simulation study the performance of the estimators is investigated. For checking the correlation in the random effect we will use the variogram, which is suitable for irregularly spaced time series. The model is applied to IBM trade data and the influence of bid-ask spread and duration on the instantaneous volatility is quantified on a daily basis.

Acknowledgements: First of all I would like to thank my advisor Prof. Claudia Czado for having confidence in me and offering me a position, for all her support and advice over the years and for her willingness to let me find my way.

Also, I would like to thank Prof. Claudia Klüppelberg and Prof. Alexander Lindner for helpful discussions, from which I benefited a lot, and enjoyable collaboration. For serving as a referee of this thesis and introducing the topic of continuous time ARMA processes to me in a lecture here at TUM in the winter term 2001/02 my thanks go to Prof. Peter Brockwell.

I also would like to thank my colleagues at Prof. Claudia Klüppelberg's research group for a very pleasant working atmosphere. Special thanks also go to Dr. B for providing essential assistance in the first years of my undergraduate studies.

Finally, I would like to thank my family for providing the opportunity for me to actually do all this and Andrea for all her loving support over the years.

Financial support by the Deutsche Forschungsgemeinschaft through the Sonderforschungsbereich 386 is gratefully acknowledged.

Chapter 1

Preliminaries

The aim of the first part of this chapter is to provide the basic facts about univariate Lévy processes and the stochastic integration with respect to them. These results are mainly taken from Applebaum (2004) and Protter (2004). We also like to refer to Sato (1999), Kyprianou (2006) and Cont and Tankov (2004) among others for general treatment of Lévy processes and/or stochastic calculus.

The second part introduces the time series class of continuous time autoregressive moving average processes and the continuous time GARCH(1,1) process after a short overview about state space models for time series.

1.1 Lévy processes

In this thesis we will always assume to work with a complete, filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$ with right-continuous filtration $(\mathcal{F}_t)_{t \geq 0}$ such that \mathcal{F}_0 contains all \mathbb{P} -null sets of \mathcal{F} . In that case the filtered probability space is said to satisfy the usual conditions. We start with the definition of a Lévy process on \mathbb{R} .

Definition 1.1.1 *Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$ be a filtered probability space satisfying the usual conditions. An adapted process $L := (L_t)_{t \geq 0}$ with $L_0 = 0$ a.s. is a \mathbb{R} valued Lévy process if*

(L1) *L has independent increments; i.e. $L_t - L_s$ is independent of \mathcal{F}_s , $0 \leq s < t < \infty$,*

(L2) *L has stationary increments; i.e. $L_t - L_s$ has the same distribution as L_{t-s} , $0 \leq s < t < \infty$,*

(L3) *L is continuous in probability; i.e. $\mathbb{P}(|L_{t+s} - L_s| > \epsilon) \rightarrow 0$ as $t \rightarrow 0$.*

For each $t > 0$ the distribution of L_t will be in the class of infinitely divisible distributions, which are defined in the following.

Definition 1.1.2 Let X be a random variable taking values in \mathbb{R} . We say that X has an infinitely divisible distribution if, for each $n \in \mathbb{N}$, there exist i.i.d. random variables $Y_1^{(n)}, \dots, Y_n^{(n)}$ such that

$$X \stackrel{d}{=} Y_1^{(n)} + \dots + Y_n^{(n)}.$$

If L is a Lévy process, then the law of L_t is infinitely divisible for each $t \geq 0$, since for each $n \in \mathbb{N}$ we have

$$L_t = L_{t/n} + (L_{2t/n} - L_{t/n}) + \dots + (L_t - L_{(n-1)t/n}),$$

which implies together with the stationarity and independence of the increments the assumption.

From Theorem 8.1 in Sato (1999) we know that the characteristic function of every infinitely divisible distribution can be written in the *Lévy-Khinchine representation*, which is then called the *Lévy-Khinchine formula*. The Lévy-Khinchine formula for the Lévy process L at time $t \geq 0$ is then given by

$$\mathbb{E}(e^{iuL_t}) = \exp\{t\psi_L(u)\}, \quad t \geq 0,$$

with

$$\psi_L(u) := iu\gamma_L - \frac{u^2}{2}\sigma_L^2 + \int_{\mathbb{R}} [e^{iux} - 1 - iux\chi_{(-1,1)}(x)] \nu_L(dx), \quad (1.1)$$

for $u \in \mathbb{R}$ and $\gamma_L \in \mathbb{R}$, $\sigma_L^2 > 0$ and ν_L is a measure on \mathbb{R} which satisfies

$$\nu_L(\{0\}) = 0 \quad \text{and} \quad \int_{\mathbb{R}} \min(|x|^2, 1)\nu_L(dx) < \infty. \quad (1.2)$$

The measure ν_L is called the *Lévy measure* of L . In the sequel we will see that for $A \in \mathbb{R}$, $\nu_L(A)$ is the expected number, per unit time, of jumps with size belonging to A . The map ψ_L is referred to as the *Lévy symbol* or *Lévy exponent*. The triplet $(\gamma_L, \sigma_L^2, \nu_L)$ satisfying (1.1) uniquely determines the distribution of L and is called the *characteristic triplet* of L .

Remark 1.1.3 Observe that the integral in (1.1) is well-defined, since

$$|e^{iux} - 1 - iux\chi_{(-1,1)}(x)| \leq \frac{1}{2}u^2x^2\chi_{(-1,1)}(x) + 2\chi_{(-1,1)^c}(x),$$

which follows from Lemma 8.6 in Sato (1999), and hence the integrand is integrable with respect to ν_L by (1.2).

Before we proceed with our short overview four examples of Lévy processes occurring in the subsequent chapters will now be presented. Remember that the processes in all the examples satisfy the conditions (L1) – (L3). In the examples we will speak of the jump of a Lévy process at time $t \geq 0$, which is defined as $\Delta L_t := L_t - L_{t-}$ and L_{t-} is the left limit of the sample path of L at time $t \geq 0$, with convention $L_{0-} := 0$. Such a limit exists as we will see right after the examples.

Example 1.1.4 (*Brownian motion*) The first example concerns the only Lévy process with continuous sample path. Such a process is called Brownian motion. Let $\sigma_B^2 > 0$ then we call a Lévy process $B = (B_t)_{t \geq 0} \in \mathbb{R}$ a Brownian motion if

- $B_t \sim N(0, t\sigma_B^2)$ for each $t \geq 0$,
- B has continuous sample path.

Hence the Lévy symbol of a Brownian motion is given by

$$\psi_B(u) = -\frac{u^2}{2}\sigma_B^2$$

for each $u \in \mathbb{R}$. The characteristic triplet is therefore equal to $(0, \sigma_B^2, 0)$. In case $\sigma_B^2 = 1$, B is called a standard Brownian motion. Now let $\gamma_B \in \mathbb{R}$ and let B be a Brownian motion in \mathbb{R} , then the process $\tilde{B}_t := \gamma_B t + B_t$, $t \geq 0$, is called a Brownian motion with drift and the triplet of \tilde{B} is given by $(\gamma_B, \sigma_B^2, 0)$. The simulated sample path of a standard Brownian motion can be seen in Figure 1.1.

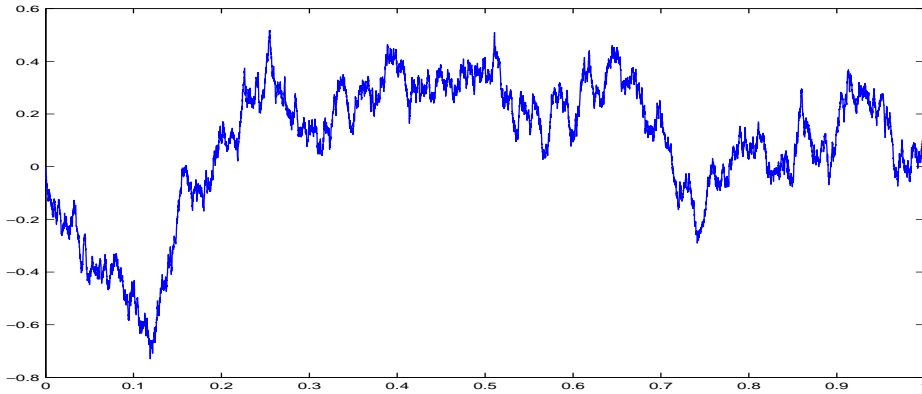


Figure 1.1: Simulated sample path of standard Brownian motion over the time interval $[0, 1]$.

For a detailed discussion of Brownian motion we like to refer to Karatzas and Shreve (1991) among many others.

Example 1.1.5 (Poisson process) Let $(T_i)_{i \in \mathbb{N}}$ be a strictly increasing sequence of stopping times, then the counting process N defined by

$$N_t := \sum_{i=1}^{\infty} \chi_{[T_i, \infty)}(t)$$

for each $t \geq 0$ is an adapted processes and if further N has independent and stationary increments it is a Poisson process. For each $t \geq 0$ and some $\lambda > 0$ N_t is Poisson distributed with parameter λt (c.f. Theorem I.23 in Protter (2004)). The parameter λ will be called the jump rate of N . The sample path of a Poisson process is piecewise constant with discontinuities of size one at the random time points $(T_i)_{i \in \mathbb{N}}$. A simulated sample path of a Poisson process with parameter λ equal to 10 is shown in Figure 1.2. The T_i will be referred to as the jump times of the process. Let $\delta_1(\cdot)$ be the Dirac measure, then the characteristic function of N_t is of the form

$$\begin{aligned} \mathbb{E}(e^{iuN_t}) &= \exp \left\{ t \int_{\mathbb{R}} (e^{iux} - 1 - iux\chi_{(-1,1)}(x)) \lambda \delta_1(dx) \right\} \\ &= \exp \{ \lambda t (e^{iu} - 1) \} \end{aligned}$$

and the characteristic triplet is hence given by $(0, 0, \lambda \delta_1)$. From the characteristic function it follows immediately that $\mathbb{E}(N_t) = \lambda t$ and $\text{Var}(N_t) = \lambda t$.

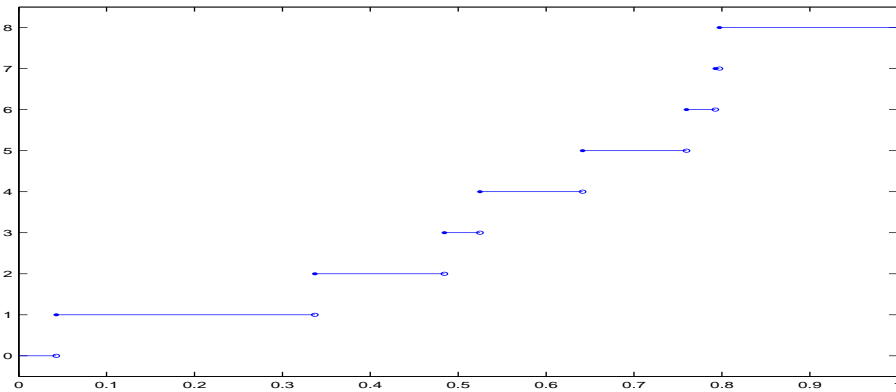


Figure 1.2: Simulated sample path of a Poisson process with parameter $\lambda = 10$ over the time interval $[0, 1]$.

Example 1.1.6 (*Compound Poisson process*) If we want to generalise the Poisson process with respect to the jump sizes we will get a compound Poisson process. Let $(Z_i)_{i \in \mathbb{N}} \in \mathbb{R}$ be a sequence of i.i.d. random variables independent of a Poisson process N . The compound Poisson process C is then defined as

$$C_t := \sum_{i=1}^{N_t} Z_i$$

for each $t \geq 0$. The rate λ of N will also be called the jump rate of C . The sample path of the compound Poisson process is therefore still piecewise constant with discontinuities at random time points $(T_i)_{i \in \mathbb{N}}$, but now the size of the jumps is itself random. The compound Poisson process C has characteristic function for all $u \in \mathbb{R}$

$$\begin{aligned} \mathbb{E}(e^{iuC_t}) &= \sum_{k=0}^{\infty} \mathbb{E}(\exp[iu(Z_1 + \dots + Z_{N_t})] | N_t = k) e^{-\lambda t} \frac{(\lambda t)^k}{k!} \\ &= e^{-\lambda t} \sum_{k=0}^{\infty} \frac{(\lambda t \phi_Z(u))^k}{k!} = \exp\{t\lambda(\phi_Z(u) - 1)\} \\ &= \exp\left\{t \int_{\mathbb{R}} (e^{iux} - 1) \lambda F_Z(dx)\right\} \\ &=: \exp\left\{t \left[iu\gamma_C + \int_{\mathbb{R}} (e^{iux} - 1 - iux\chi_{(-1,1)}(x)) \lambda F_Z(dx) \right]\right\}, \end{aligned} \tag{1.3}$$

where $\gamma_C = \int_{|x| < 1} x \lambda F_Z(dx)$, ϕ_Z is the characteristic function and F_Z the law of Z_1 . Thus the characteristic triplet is given by $(\gamma_C, 0, \lambda F_Z)$. As an illustration we plotted in Figure 1.3 the sample path of a compound Poisson process with standard normally distributed jumps and jump rate $\lambda = 20$ over the time interval $[0, 1]$.

From (1.2) it follows that for a Lévy measure ν_L we have $\nu_L((-\epsilon, \epsilon)^c) < \infty$ for all $\epsilon > 0$. But of course we can have $\nu_L(\mathbb{R}) = \infty$. Such Lévy processes are called *infinite activity* Lévy processes in contrast to *finite activity* ones. In the last two examples we presented Lévy processes with finite activity. The last example introduces a Lévy process with infinite activity.

Example 1.1.7 (*Variance Gamma process*) In the last example we want to introduce the Variance Gamma (VG) process. It can be defined as a time-changed Brownian motion with drift. The time change is done with respect to a Gamma process which is defined as a process $(T_t^G)_{t \geq 0}$ with parameters $a, b > 0$, such that each $T_t^G \sim \Gamma(at, b)$ is Gamma distributed with density

$$f_{T_t^G}(x) = \frac{b^{at}}{\Gamma(at)} x^{at-1} e^{-bx},$$

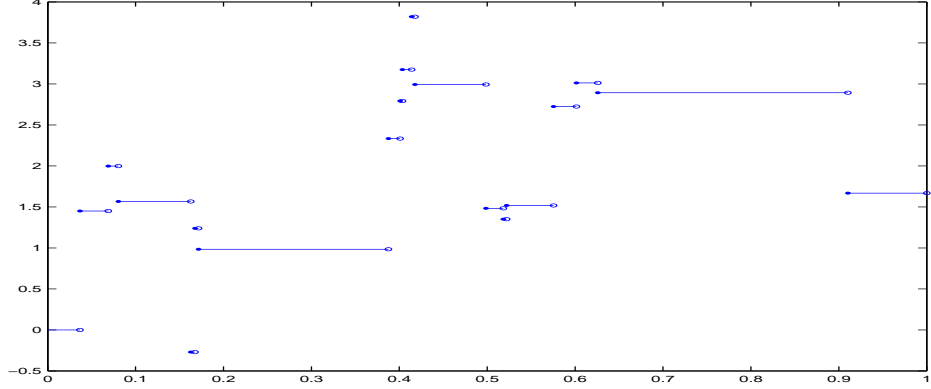


Figure 1.3: Simulated sample path of a compound Poisson process with parameter $\lambda = 20$ and standard normally distributed jumps Z_i over the time interval $[0, 1]$.

for $x \geq 0$, where $\Gamma(\cdot)$ denotes the Gamma function. Now let B be a standard Brownian motion, $\sigma > 0$, $\theta \in \mathbb{R}$ and let $a = 1/\tau > 0$ and $b = 1/\tau > 0$ be the parameters of the Gamma process T^G . Then the VG process V can be defined as

$$V_t := \theta T_t^G + \sigma B_{T_t^G}, \quad (1.4)$$

for each $t \geq 0$. The characteristic function of V_T for $t \geq 0$ is then given by

$$\mathbb{E}(e^{iuV_t}) = \left(1 - iu\theta\tau + \frac{1}{2}\sigma^2\tau u^2\right)^{-t/\tau},$$

for $u \in \mathbb{R}$. If we parameterise the Variance Gamma process by

$$C = 1/\tau, \quad G = \left(\sqrt{\frac{1}{4}\theta^2\tau^2 + \frac{1}{2}\sigma^2\tau} - \frac{1}{2}\theta\tau\right)^{-1} \quad \text{and} \quad M = \left(\sqrt{\frac{1}{4}\theta^2\tau^2 + \frac{1}{2}\sigma^2\tau} + \frac{1}{2}\theta\tau\right)^{-1}$$

then the Lévy measure ν_V can be expressed as

$$\nu_V(dx) = \begin{cases} C \exp(Gx)|x|^{-1}dx, & x < 0 \\ C \exp(-Mx)x^{-1}dx, & x > 0 \end{cases}.$$

The mean and variance of the VG process V at time $t \geq 0$ are given by

$$\mathbb{E}(V_t) = \theta t \quad \text{and} \quad \text{Var}(V_t) = (\theta^2\tau + \sigma^2)t. \quad (1.5)$$

For a detailed discussion of the VG process we refer to Madan et al. (1998).

The Lévy measure has infinite mass $\nu_V(\mathbb{R}) = \infty$ and hence the VG process has infinitely many jumps in any finite time interval. If we want to simulate the sample path of a VG process one should be aware of the fact that with any computer simulation we can only describe a finite number of jumps in any given path. Taking this into account, the simulation of a VG process is particularly simple. By (1.4) we get a sample path of a VG process over a grid $\{i\Delta t, i = 0, 1, \dots\}$ by sampling a standard Brownian motion and a Gamma process over this grid. If we can simulate random variables $g_i \sim \Gamma(\Delta t/\tau, 1/\tau)$ the Gamma process can be approximated by $T_{i\Delta t}^G = T_{(i-1)\Delta t}^G + g_i, i \geq 1$, and $T_0^G = 0$. By simulating normal random variables $N_i \sim N(0, T_{i\Delta t}^G - T_{(i-1)\Delta t}^G)$ one can construct a time changed Brownian motion by adding up the N_i . The VG process can then be simulated by (1.4). For a detailed description we refer to Chapter 8.4 in Schoutens (2003). The sample path of a Variance Gamma processes over the time interval $[0, 1]$ with parameters $\sigma = 0.3, \theta = -0.03$ and $\tau = 0.01$ is shown in Figure 1.4. For this simulation the grid size was $\Delta t = 10^{-5}$.

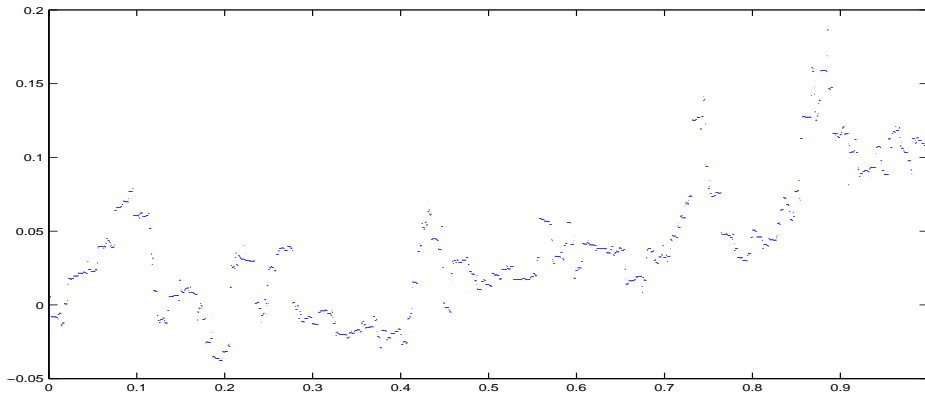


Figure 1.4: Simulated sample path of VG process with parameters $\sigma = 0.3, \theta = -0.03$ and $\tau = 0.01$ over the time interval $[0, 1]$.

From Theorem 2.1.7 in Applebaum (2004) we know that every Lévy process has a càdlàg modification that is itself a Lévy process. In the following we will always assume to work with this càdlàg modification. Hence the jumps of L are well-defined.

Below we concentrate on the jump process $\Delta L := (\Delta L_t)_{t \geq 0}$ of the Lévy process L . Therefore one can define a random measure counting the jumps of L of a specific size. More precisely, for $0 \leq t < \infty$ and $A \in \mathcal{B}(\mathbb{R})$ define

$$N_L(t, A) = \#\{0 \leq s < t; \Delta L_s \in A\} = \sum_{0 \leq s < t} \chi_A(\Delta L_s).$$

The measure N_L is a Poisson random measure (see Chapter 2.3.1 in Applebaum (2004) for details), in particular this implies that:

- (1) For each $t > 0, \omega \in \Omega$, $N_L(t, \cdot)(\omega)$ is a counting measure on $\mathcal{B}(\mathbb{R} - \{0\})$.
- (2) For each A bounded away from 0 ($N_L(t, A), t \geq 0$) is a Poisson process with intensity $\mathbb{E}(N_L(1, A))$, where $A \in \mathcal{B}(\mathbb{R})$ is said to be bounded away from 0 if $0 \notin \bar{A}$.

Observe that for A bounded away from 0, $N_L(t, A) < \infty$ a.s. for all $t \geq 0$. Since if we do not count jumps of size 0 the set $\{s \geq 0 : \Delta L_s \neq 0\}$ of jump times is countable due to the càdlàg sample path of L (see Theorem 2.8.1 in Applebaum (2004)). Hence the sequence of hitting times of the set A defined by $T_1^A := \inf\{s > 0 : \Delta L_s \in A\} > 0$ and $T_n^A := \inf\{s > T_{n-1}^A : \Delta L_s \in A\}$, for $n > 1$, has no accumulation point, which implies $\lim_{n \rightarrow \infty} T_n^A = \infty$ a.s.. Thus we get for each $t \geq 0$

$$N_L(t, A) = \sum_{0 \leq s \leq t} \chi_A(\Delta L_s) = \sum_{n=1}^{\infty} \chi_{(0, t]}(T_n^A) < \infty \quad a.s..$$

The intensity of the Poisson process $(N_L(t, A), t \geq 0)$, $A \in \mathcal{B}(\mathbb{R} - \{0\})$, describes the expected number, per unit time, of jumps with size belonging to A , and defines a measure on $\mathcal{B}(\mathbb{R} - \{0\})$ which is equal to ν_L (cf. Theorem I.35 in Protter (2004)), i.e.

$$\nu_L(\cdot) = \mathbb{E}(N_L(1, \cdot)).$$

We further define for each $t \geq 0$ and A bounded away from 0 the *compensated Poisson random measure* by

$$\tilde{N}_L(t, A) = N_L(t, A) - t\nu_L(A).$$

Notice that $\tilde{N}_L(t, A)$ is a martingale-valued measure on $\mathcal{B}(\mathbb{R} - \{0\})$, as we will show now. Let $(\mathcal{F}_t)_{t \geq 0}$ be the filtration generated by $(\tilde{N}_L(t, A))_{t \geq 0}$. Then for $0 \leq s \leq t$ we get

$$\begin{aligned} \mathbb{E}(\tilde{N}_L(t, A) | \mathcal{F}_s) &= \mathbb{E}(N_L(s, A) + N_L(t, A) - N_L(s, A) | \mathcal{F}_s) - t\nu_L(A) \\ &= N_L(s, A) + (t - s)\nu_L(A) - t\nu_L(A) = \tilde{N}_L(s, A) \end{aligned} \quad (1.6)$$

since $(N_L(t, A))_{t \geq 0}$ is a Poisson process with intensity $\mathbb{E}(N_L(1, A)) = \nu_L(A)$.

In the following we state the *Lévy-Itô decomposition* (see e.g. Theorem 2.4.16 of Applebaum (2004)) of the sample path of L into continuous and jump part. For this we will need one further concept; namely integration with respect to a Poisson random measure.

Let $f : \mathbb{R} \rightarrow \mathbb{R}$ be a measurable function and $A \in \mathcal{B}(\mathbb{R})$ bounded away from 0. Then the *Poisson integral* of f over the set A can be defined by

$$\int_A f(x)N_L(t, dx) := \sum_{0 \leq s \leq t} f(\Delta L_s)\chi_A(\Delta L_s)$$

(see e.g. Chapter 2.3.2 in Applebaum (2004) or Theorem I.36 in Protter (2004)).

The jumps of L up to time $t \geq 0$ of magnitude greater or equal to one can be expressed as the Poisson integral

$$\int_{|x| \geq 1} xN_L(t, dx) = \sum_{0 \leq s \leq t} \Delta L_s \chi_{(-1,1)^c}(\Delta L_s). \quad (1.7)$$

The process $\left(\int_{|x| \geq 1} xN_L(t, dx)\right)_{t \geq 0}$ is a compound Poisson process with rate $\nu_L((-1,1)^c)$ and jump distribution $\nu_L((-1,1)^c)^{-1}\nu_L(dx)$ defined on $(-1,1)^c$ (see e.g. Lemma 2.8 in Kyprianou (2006)). Now a new stochastic process L^b can be defined by

$$L_t^b := L_t - \int_{|x| \geq 1} xN_L(t, dx)$$

for $t \geq 0$. By Theorem 2.4.7 and 2.4.8 in Applebaum (2004) we know that L^b is again a Lévy process with moments of all orders. The existence of moments results from the fact that L^b has only jumps, which are bounded, more precisely bounded by one. The centered version of L^b can be further decomposed into a martingale L^c with continuous sample path and a martingale

$$L_t^d = \int_{|x| < 1} x\tilde{N}_L(t, dx), \quad (1.8)$$

for each $t \geq 0$, which are independent Lévy processes. Observe that in (1.8) the integration is with respect to the compensated Poisson random measure. This has to be the case since L can have infinitely many small jumps and their sum does not necessarily converge. Actually the integral in (1.8) is defined as the L^2 limit of the martingales

$$M_t^n := \sum_{k=1}^n \int_{A_k} x(N_L(t, dx) - t\nu_L(dx)), \quad t \geq 0, \quad (1.9)$$

which have (see e.g. equation (2.9) in Applebaum (2004)) the characteristic function

$$\mathbb{E}(e^{iuM_t^n}) = \exp \left(t \sum_{k=1}^n \int_{A_k} [e^{iux} - 1 - iux] \nu_L(dx) \right), \quad (1.10)$$

where $A_k := \{x \in \mathbb{R} : \frac{1}{k+1} < |x| < \frac{1}{k}\}$. In particular one can find a subsequence converging a.s., uniformly in t on any bounded time interval. For a proof of this result see e.g. Theorem I.41 in Protter (2004) or Theorem 2.10 in Kyprianou (2006). The continuous part L^c can be identified as a Brownian motion (see e.g. Theorem 2.4.15 in Applebaum (2004)). Hence for each $t \geq 0$ we can write $L_t^b - \mathbb{E}(L_t^b) = L_t^c + L_t^d =: B_t + L_t^d$. Observe that (1.9) and (1.7) are independent, since N_L has independent counts on disjoint sets (cf. Theorem 2.4.6 in Applebaum (2004)). This implies that (1.8) and (1.7) are independent processes.

The full *Lévy-Itô decomposition* then says the following. If L is a Lévy process, then there exists $\gamma_L \in \mathbb{R}$, a Brownian motion B with variance σ_L^2 and an independent Poisson random measure N_L on $\mathbb{R}_+ \times (\mathbb{R} - \{0\})$ such that, for each $t \geq 0$,

$$\begin{aligned} L_t &= L_t^b + \int_{|x| \geq 1} x N_L(t, dx) = B_t + L_t^d + \mathbb{E}(L_t^b) + \int_{|x| \geq 1} x N_L(t, dx) \\ &= B_t + \int_{|x| < 1} x (N_L(t, dx) - t \nu_L(dx)) + t \mathbb{E} \left(L_1 - \int_{|x| \geq 1} x N_L(1, dx) \right) \\ &\quad + \int_{|x| \geq 1} x N_L(t, dx) \\ &= B_t + \int_{|x| < 1} x \tilde{N}_L(t, dx) + \gamma_L t + \int_{|x| \geq 1} x N_L(t, dx), \end{aligned} \quad (1.11)$$

where $\gamma_L = \mathbb{E} \left(L_1 - \int_{|x| \geq 1} x N_L(1, dx) \right)$. Given (1.11) one can derive the Lévy-Khintchine formula for the characteristic function of L at time $t \geq 0$

$$\begin{aligned} \mathbb{E}(e^{iuL_t}) &= \mathbb{E} \left(\exp \left\{ iu(\gamma_L t + B_t + L_t^d + \int_{|x| \geq 1} x N_L(t, dx)) \right\} \right) \\ &= e^{iu\gamma_L t} \mathbb{E}(e^{iuB_t}) \mathbb{E}(e^{iuL_t^d}) \mathbb{E} \left(\exp \left\{ iu \int_{|x| \geq 1} x N_L(t, dx) \right\} \right) \end{aligned}$$

$$\begin{aligned}
&= e^{iu\gamma_L t} e^{-t\frac{u^2}{2}\sigma_L^2} \exp \left\{ t \int_{|x|<1} [e^{iux} - 1 - iux] \nu_L(dx) \right\} \\
&\quad \times \exp \left\{ t \int_{|x|\geq 1} [e^{iux} - 1] \nu_L(dx) \right\} \\
&= \exp \left\{ t \left(iu\gamma_L - \frac{u^2}{2}\sigma_L^2 + \int_{\mathbb{R}} [e^{iux} - 1 - iux\chi_{(-1,1)}(x)] \nu_L(dx) \right) \right\},
\end{aligned}$$

where the second equality is due to the independence of the summands in (1.11) and the third follows from $B_t \sim N(0, t\sigma_L^2)$, taking limits in (1.10) for getting $\mathbb{E}(e^{iuL_t^d})$ and the fact that $\left(\int_{|x|\geq 1} xN_L(t, dx)\right)_{t\geq 0}$ is a compound Poisson process together with (1.3).

The existence of moments of L depends on the behaviour of the *large jumps*. Necessary and sufficient conditions for L to have finite mean and variance are e.g. given in Example 25.12 in Sato (1999), which are

$$\int_{|x|\geq 1} |x|\nu_L(dx) < \infty \quad \text{and} \quad \int_{|x|\geq 1} x^2\nu_L(dx) < \infty,$$

respectively. If these conditions are met, then one gets by differentiating $\mathbb{E}(e^{iuL_t})$ (see e.g. Proposition 2.5(ix) in Sato (1999))

$$\mathbb{E}(L_t) = t \left(\int_{|x|\geq 1} x\nu_L(dx) + \gamma_L \right) \quad (1.12)$$

and

$$\text{Var}(L_t) = t \left(\sigma_L^2 + \int_{\mathbb{R}} x^2\nu_L(dx) \right). \quad (1.13)$$

Throughout this text we will only deal with square integrable Lévy processes. In case we further assume $\mathbb{E}(L_1) = 0$, then it follows from (1.12) that $\gamma_L = -\int_{|x|\geq 1} x\nu_L(dx)$ and the Lévy-Itô decomposition (1.11) simplifies to

$$L_t = B_t + \int_{\mathbb{R}-\{0\}} x\tilde{N}_L(t, dx). \quad (1.14)$$

If the Lévy process L is of finite variation (see also Definition A.13), which will be the case if and only if $\sigma_L^2 = 0$ and $\int_{|x|<1} |x|\nu_L(dx) < \infty$ (cf. Proposition 3.9 in Cont and Tankov (2004)), the Lévy-Itô decomposition will have the form

$$L_t = \gamma_L t + \int_{|x|<1} x(N_L(t, dx) - t\nu_L(dx)) + \int_{|x|\geq 1} xN_L(t, dx) \quad (1.15)$$

$$= \gamma_L^* t + \int_{\mathbb{R}-\{0\}} xN_L(t, dx) = \gamma^* t + \sum_{0\leq s\leq t} \Delta L_s, \quad (1.16)$$

where $\gamma_L^* := \gamma_L - \int_{|x|<1} x\nu_L(dx)$. The characteristic function can be expressed as

$$\begin{aligned} \mathbb{E}(e^{iuLt}) &= \exp\left\{t\left[iu\gamma + \int_{|x|<1} (e^{iux} - 1 - iux)\nu_L(dx) + \int_{|x|\geq 1} (e^{iux} - 1)\nu_L(dx)\right]\right\} \\ &= \exp\left\{t\left[iu\gamma^* + \int_{\mathbb{R}} (e^{iux} - 1)\nu_L(dx)\right]\right\}, \end{aligned}$$

which follows from (1.15) and the expression for the characteristic function of L_t^d .

Up to now we have only considered one-sided Lévy processes defined on the positive half real line. But it is also possible to define an extension of L to the whole real line by taking a second Lévy process $\tilde{L}_{t\geq 0}$ independent of L but with the same distribution as L , and setting

$$L_t^* := L_t\chi_{[0,\infty)}(t) - \tilde{L}_{-t}\chi_{(-\infty,0)}(t), \quad t \in \mathbb{R}. \quad (1.17)$$

L^* is then a Lévy process defined on the whole real line with càdlàg sample path. For ease of notation we will denote one-sided as well as two-sided Lévy processes by L .

1.2 Stochastic integration

This section deals with the stochastic integration of predictable processes with respect to Lévy processes. We restrict the discussion to Lévy processes as integrator, since this will be sufficient for this thesis. The class of integrators can of course be extended and we refer for a comprehensive treatment of stochastic integration to Chapter 4 in Applebaum (2004) or Chapter II and IV in Protter (2004).

In case the Lévy process L has almost surely trajectories of finite variation, the stochastic integral can be defined as a path-by-path Stieltjes integral, cf. Chapter I.7 in Protter (2004). This concept will not work for Lévy processes of unbounded variation, e.g. the Brownian motion, as shown in Chapter I.8 in Protter (2004). The main idea to overcome this problem is due to Kiyosi Itô, which was to restrict the integrands to those that are predictable. This will be explained in the following.

1.2.1 General theory

From the Lévy-Itô decomposition (1.11) we see that the aim of this section has to be to define integrals of the form

$$\int_0^t \mathbf{F}(s, \omega) dB_s, \quad \int_0^t \int_A \mathbf{H}(s, x, \omega) \tilde{N}_L(ds, dx) \quad \text{and} \quad \int_0^t \int_A \mathbf{K}(s, x, \omega) N_L(ds, dx),$$

for each $t \geq 0$, where L is a \mathbb{R} -valued Lévy process with Lévy-Itô decomposition (1.11), \mathbf{F} , \mathbf{H} and \mathbf{K} are \mathbb{R}^m valued random mappings and $A \in \mathcal{B}(\mathbb{R})$ is bounded away

from 0. Since the integration will be understood componentwise, we just consider the univariate case from now on.

The Brownian motion B as well as the compensated Poisson random measure \tilde{N}_L are both martingale-valued random measures (see also Definition A.12). Let $\delta_0(\cdot)$ be the Dirac measure, then define

$$M_B(t, A) := B_t \delta_0(A),$$

for each $t \geq 0$ and $A \in \mathcal{B}(\{0\})$, and

$$M_N(t, A) := \tilde{N}_L(t, A),$$

for each $t \geq 0$ and $A \in \mathcal{B}(\mathbb{R} - \{0\})$. Then it is clear that M_B is a martingale-valued random measure on $\mathbb{R}_+ \times \{0\}$ since B is a martingale and in (1.6) it was shown that M_N is one on $\mathbb{R}_+ \times \mathbb{R} - \{0\}$. In the following we will just work with one martingale-valued random measure M on $\mathbb{R}_+ \times E$, where $E \in \mathcal{B}(\mathbb{R})$, such that

$$(M1) \quad M(0, A) = 0, \quad a.s.,$$

$$(M2) \quad M(t, A) - M(s, A) \text{ is independent of } \mathcal{F}_s \text{ for } t > s,$$

$$(M3) \quad \text{there exists a } \sigma\text{-finite measure } \mu \text{ on } E \text{ for which}$$

$$\mathbb{E}(M(t, A)^2) = t\mu(A)$$

for all $0 \leq s < t < \infty$, $A \in \mathcal{B}(E)$. Observe that M_B and M_N satisfy (M1)-(M3) with $E = \{0\}$ and $\mu(\{0\}) = \sigma_L^2$ for M_B and $E = \mathbb{R} - \{0\}$ and $\mu(\cdot) = \nu_L(\cdot)$ for M_N . Before we can define the stochastic integral with respect to martingale valued random measures, we have to specify the class of integrands. In the first step these will be all random mappings $F : [0, T] \times E \times \Omega \rightarrow \mathbb{R}$ such that

- F is predictable
- $\int_0^T \int_E \mathbb{E}(|F(t, x, \omega)|^2) \mu(dx) dt < \infty$ for fixed $T > 0$.

We denote the linear space of all equivalence classes of such mappings, which coincide almost everywhere with respect to $Leb \times \mu \times \mathbb{P}$ by $\mathcal{H}_2(T, E)$, where Leb denotes the Lebesgue measure. One can define an inner product on $\mathcal{H}_2(T, E)$ by

$$\langle F, G \rangle_{T, \mu} := \int_0^T \int_E \mathbb{E}(F(s, x, \omega)G(s, x, \omega)) \mu(dx) ds,$$

for each $F, G \in \mathcal{H}_2(T, E)$. The inner product defines a norm $\|\cdot\|_{T, \mu}$, making $\mathcal{H}_2(T, E)$ a normed space and actually a Hilbert space, since it is a closed subspace of the complete space $L^2([0, T] \times E \times \Omega, Leb \times \mu \times \mathbb{P})$ (cf. Lemma 4.1.3 in Applebaum

(2004)). In the Brownian motion case where $E = \{0\}$ and $\mu(\{0\}) = 1$ we denote $\mathcal{H}_2(T, E)$ just by $\mathcal{H}_2(T)$, with norm $\|F\|_T = \sqrt{\int_0^T \mathbb{E}(|F(s, \omega)|^2) ds}$, for $F \in \mathcal{H}_2(T)$.

A random mapping $F \in \mathcal{H}_2(T, E)$ will be called *simple* if, for some $m, k \in \mathbb{N}$, there exists $0 \leq t_1 \leq \dots \leq t_{m+1} = T$ and disjoint Borel sets $A_j \subset E$, with $\mu(A_j) < \infty$, $j = 1, \dots, k$, such that

$$F(t, x, \omega) = \sum_{i=1}^m \sum_{j=1}^k F_{i,j}(\omega) \chi_{(t_i, t_{i+1}]}(t) \chi_{A_j}(x), \quad (1.18)$$

for each $t \in [0, T]$, $x \in E$ and where each $F_{i,j}(\omega) \in b\mathcal{F}_{t_i}$ is \mathcal{F}_{t_i} measurable and bounded. Observe that $F(t, x, \omega)$ is left continuous and measurable with respect to the product σ -algebra $\mathcal{B}(E) \otimes \mathcal{F}_t$, hence F is predictable. The linear subspace of all simple mappings in $\mathcal{H}_2(T, E)$ will be referred to as $S(T, E)$. The stochastic integral $I_T(F)(\omega) = \int_0^T \int_E F(s, x, \omega) M(ds, dx)$ of a simple mapping $F \in S(T, E)$ with respect to the martingale-valued random measure M is then defined as the random variable

$$I_T(F)(\omega) := \sum_{i=1}^m \sum_{j=1}^k F_{i,j}(\omega) [M(t_{i+1}, A_j) - M(t_i, A_j)], \quad (1.19)$$

for each $T > 0$. Notice that due to the predictability of F and the martingale property of M , $F_{i,j}$ will be independent of $M(t_{i+1}, A_j) - M(t_i, A_j)$ for each $j \in \{1, \dots, k\}$, which implies that the martingale property is preserved. Here one recognises the importance of predictability of the integrand. For further properties of I_T see Theorem 4.2.3 in Applebaum (2004). The quantity in (1.19) defines a linear mapping $F \mapsto I_T(F)$ from $S(T, E)$ into $L^2(\Omega, \mathcal{F}, \mathbb{P})$. Since $L^2(\Omega, \mathcal{F}, \mathbb{P})$ is a Hilbert space with norm $\|I_T(F)(\omega)\|_2 := \sqrt{\mathbb{E}(I_T(F)^2(\omega))}$ one obtains the following relation

$$\|I_T(F)(\omega)\|_2^2 = \mathbb{E}(I_T(F)^2(\omega)) = \int_0^T \int_E \mathbb{E}(|F(t, x, \omega)|^2) \mu(dx) dt = \|F\|_{T, \mu}^2, \quad (1.20)$$

showing that I_T is an isometry. For a proof of (1.20) see Lemma 4.2.2 in Applebaum (2004) or Lemma A.18. But since $S(T, E)$ is dense in $\mathcal{H}_2(T, E)$ (see e.g. Lemma 4.1.4 in Applebaum (2004)) I_T can be extended to an isometric embedding of $\mathcal{H}_2(T, E)$ into $L^2(\Omega, \mathcal{F}, \mathbb{P})$. We will denote this extension also by $I_T(F)(\omega) := \int_0^T \int_E F(t, x, \omega) M(dt, dx)$ and call it the *stochastic integral* of $F \in \mathcal{H}_2(T, E)$. In particular this means that for any $F \in \mathcal{H}_2(T, E)$ one can find a sequence $(F_n)_{n \in \mathbb{N}} \in S(T, E)$ of simple mappings such that $\lim_{n \rightarrow \infty} \|F - F_n\|_{T, \mu} = 0$ and

$$\lim_{n \rightarrow \infty} \left\| \int_0^T \int_E F(t, x, \omega) M(dt, dx) - \int_0^T \int_E F_n(t, x, \omega) M(dt, dx) \right\|_2 = 0.$$

Since $\chi_{(a,b)}\chi_A F \in \mathcal{H}_2(T, E)$ for $0 \leq a \leq b \leq T$, $A \in \mathcal{B}(E)$ and $F \in \mathcal{H}_2(T, E)$ we can define $I_{(a,b),A}(F)(\omega) := I_T(\chi_{(a,b)}\chi_A F)(\omega)$. If $\|F\|_{t,\mu} < \infty$ for all $t \geq 0$ one can define a stochastic process $(I_t(F)(\omega))_{t \geq 0}$ through $t \mapsto \int_0^t \int_E F(t, x, \omega) M(dt, dx)$.

If we consider again the random measure M_B and some $F \in \mathcal{H}(T)$ we will denote the stochastic integral $I_T(F)$ by $\int_0^T F(s) dB_s$.

The stochastic integral has the following useful properties: Let $F, G \in \mathcal{H}_2(T, E)$ and $\alpha, \beta \in \mathbb{R}$, then

$$(I1) \quad I_T(\alpha F + \beta G) = \alpha I_T(F) + \beta I_T(G)$$

(I2)

$$\mathbb{E}(I_T(F)) = 0 \quad (1.21)$$

$$\mathbb{E}(I_T(F)^2) = \int_0^T \int_E \mathbb{E}(|F(t, x)|^2) \mu(dx) dt. \quad (1.22)$$

(I3) $(I_t(F)(\omega))_{t \geq 0}$ is \mathcal{F}_t -adapted and a square integrable martingale.

For a proof of these properties see e.g. Theorem 4.2.3 in Applebaum (2004).

Further we would like to mention that the class of integrands can be extended if one defines the stochastic integral through convergence in probability and not in L^2 . In that case all mappings $F : [0, T] \times E \times \Omega \rightarrow \mathbb{R}$ will be integrable with respect to the martingale valued measure M , which satisfy:

- F is predictable
- $\mathbb{P}(\int_0^T \int_E |F(t, x, \omega)|^2 \mu(dx) dt < \infty) = 1$, for fixed $T > 0$.

The linear space of all equivalence classes of such mappings will be denoted by $\mathcal{P}(T, E)$. For a detailed treatment of this topic we refer to Chapter 4.2.2 in Applebaum (2004). This completes our discussion of integration with respect to martingale-valued measures.

Remember that we also wanted to define integration with respect to a Poisson random measure N_L over a set $A \in \mathcal{B}(\mathbb{R})$ which is bounded away from 0. Since N_L is not a martingale-valued random measure the above considerations can not be applied here, but actually the definition of the integral is much easier. Let $A \in \mathcal{B}(\mathbb{R})$, which is bounded away from 0 then all mappings $K : [0, T] \times A \times \Omega \rightarrow \mathbb{R}$ are integrable with respect to N_L if they are predictable and we define

$$\int_0^T \int_A K(t, x, \omega) N_L(dt, dx) := \sum_{0 \leq u \leq T} K(u, \Delta L_u, \omega) \chi_A(\Delta L_u) \quad (1.23)$$

as a random finite sum. The finiteness follows from the predictability of K and the fact that the Lévy process L has only finitely many jumps of size belonging to A in the time interval $[0, T]$, if A is bounded away from 0.

Notice that in the following we will omit the dependence on $\omega \in \Omega$ of the stochastic integral.

1.2.2 Itô's formula

In this section we want to introduce the *Itô formula*, which is the stochastic analogue of the Fundamental Theorem of Calculus. Let $D_1 := \{y \in \mathbb{R} : |y| < 1\}$ be the open unit ball and $\mathbf{Y} := (Y^1, \dots, Y^m)^T \in \mathbb{R}^m$ a Lévy-type stochastic integral with stochastic differential

$$\begin{aligned} d\mathbf{Y}_t &= \mathbf{G}(t)dt + \mathbf{F}(t)dB_t + \int_{|x|<1} \mathbf{H}(t, x)\tilde{N}_L(dt, dx) \\ &\quad + \int_{|x|\geq 1} \mathbf{K}(t, x)N_L(dt, dx), \end{aligned} \quad (1.24)$$

where $\mathbf{G} = (G^1, \dots, G^m)^T$, $\mathbf{F} = (F^1, \dots, F^m)^T$, $\mathbf{H} = (H^1, \dots, H^m)^T$ and $\mathbf{K} = (K^1, \dots, K^m)$ are \mathbb{R}^m valued random mappings with $|G^i|^{1/2}, F^i \in \mathcal{H}_2(T)$, $H^i \in \mathcal{H}_2(T, D_1 - \{0\})$ and K^i predictable for each $i = 1, \dots, m$. Notice that one can generalise (1.24) by considering $|G^i|^{1/2}, F^i \in \mathcal{P}_2(T)$ and $H^i \in \mathcal{H}(T, D_1 - \{0\})$ (cf. equation (4.13) in Applebaum (2004)).

In Definition A.15 we defined the quadratic variation process $[Y^i, Y^j]$ of Y^i and Y^j , $1 \leq i, j \leq m$. From (A.2) we know that the quadratic variation can be decomposed into a continuous and pure jump part, i.e.

$$[Y^i, Y^j]_t = [Y^i, Y^j]_t^c + \sum_{0 \leq u \leq t} \Delta Y_u^i \Delta Y_u^j.$$

$[Y^i, Y^j]^c$ is part of the Itô formula which will now be presented in the following theorem. From equation (4.15) in Applebaum (2004) we know that for a Lévy-type stochastic integral (1.24) the quadratic variation of the components Y^i and Y^j , $1 \leq i, j \leq m$, can be expressed as

$$\begin{aligned} [Y^i, Y^j]_t &= \sigma_L^2 \int_0^t F^i(u)F^j(u)du + \int_0^t \int_{|x|<1} H^i(u, x)H^j(u, x)N_L(du, dx) \\ &\quad + \int_0^t \int_{|x|\geq 1} K^i(u, x)K^j(u, x)N_L(du, dx), \end{aligned} \quad (1.25)$$

which follows from the formal product relations between differentials:

$$dB_t dB_t = \sigma_L^2 dt \quad \text{and} \quad N_L(dt, dx)N_L(dt, dy) = N_L(dt, dx)\delta_0(x - y), \quad (1.26)$$

whereas all other products of differentials vanish.

Theorem 1.2.1 (*Itô formula*) *Let $\mathbf{Y} \in \mathbb{R}^m$ be a Lévy-type stochastic integral of the form (1.24) and $f \in C^2(\mathbb{R}^m)$. Then we have for $t \geq 0$*

$$\begin{aligned} f(\mathbf{Y}_t) - f(\mathbf{Y}_0) &= \sum_{i=1}^m \int_0^t \frac{\partial f}{\partial y_i}(\mathbf{Y}_{u-})dY_u^i + \frac{1}{2} \sum_{1 \leq i, j \leq m} \int_0^t \frac{\partial^2 f}{\partial y_i \partial y_j}(\mathbf{Y}_{u-})d[Y^i, Y^j]_u^c \\ &\quad + \sum_{0 \leq u \leq t} \left\{ f(\mathbf{Y}_u) - f(\mathbf{Y}_{u-}) - \sum_{i=1}^m \frac{\partial f}{\partial y_i}(\mathbf{Y}_{u-})\Delta Y_u^i \right\}. \end{aligned} \quad (1.27)$$

For a proof of Itô's formula see for example Theorem 4.4.10 in Applebaum (2004). In the following example we shall show two applications of Itô's formula. The first one is integration by parts and in the second example it is shown how to find the solution of a system of linear stochastic differential equations (see Chapter 6 of Applebaum (2004) for details on stochastic differential equations).

Example 1.2.2 (i) *Let X and Y be two univariate Lévy-type stochastic integrals of the form (1.24). Define a mapping $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ by $f(x, y) = xy$. Then applying (1.27) we get for all $t \geq 0$*

$$\begin{aligned}
X_t Y_t - X_0 Y_0 &= \int_0^t X_{u-} dY_u + \int_0^t Y_{u-} dX_u + \int_0^t d[X, Y]_u^c \\
&\quad + \sum_{0 \leq u \leq t} \{X_u Y_u - X_{u-} Y_{u-} - Y_{u-} (X_u - X_{u-}) - X_{u-} (Y_u - Y_{u-})\} \\
&= \int_0^t X_{u-} dY_u + \int_0^t Y_{u-} dX_u + [X, Y]_t^c \\
&\quad + \sum_{0 \leq u \leq t} (X_u - X_{u-})(Y_u - Y_{u-}) \\
&= \int_0^t X_{u-} dY_u + \int_0^t Y_{u-} dX_u + [X, Y]_t. \tag{1.28}
\end{aligned}$$

In particular we get

$$[X, X]_t = X_t^2 - X_0^2 - 2 \int_0^t X_{u-} dX_u. \tag{1.29}$$

(ii) *Let's consider the following system of linear stochastic differential equations, which is driven by a one dimensional Lévy process $(L_t)_{t \geq 0}$*

$$d\mathbf{X}_t = \mathbf{B}\mathbf{X}_t dt + \boldsymbol{\sigma} dL_t, \quad 0 \leq t < \infty, \tag{1.30}$$

where $\mathbf{B} \in \mathbb{M}_{m,m}(\mathbb{R})$ and $\boldsymbol{\sigma} \in \mathbb{R}^m$ are non-random. Notice that (1.30) is of the appropriate form (1.24), since

$$d\mathbf{X}_t = (\mathbf{B}\mathbf{X}_t + \gamma_L)dt + \boldsymbol{\sigma} dB_t + \int_{|x| < 1} \boldsymbol{\sigma} x \tilde{N}(dt, dx) + \int_{|x| \geq 1} \boldsymbol{\sigma} x N(dt, dx),$$

if L has Lévy-Itô decomposition (1.11). A solution $\mathbf{X} \in \mathbb{R}^m$ exists if all the eigenvalues of the matrix \mathbf{B} have negative real part (see e.g. Theorem 6.7 in Karatzas and Shreve (1991)) and can be found with the help of the Itô formula. Let's define the mapping $f : \mathbb{R}^{m+1} \rightarrow \mathbb{R}^m$, by $f(t, \mathbf{x}) := e^{-\mathbf{B}t} \mathbf{x}$ for $t \in \mathbb{R}$. The exponential matrix $e^{\mathbf{B}u}$ is defined by $e^{\mathbf{B}u} := \sum_{k=0}^{\infty} \frac{(\mathbf{B}u)^k}{k!}$. Let $\mathbf{e}_i \in \mathbb{R}^m$, $i = 1, \dots, m$, be the i -th unit vector,

then (1.27) tells us

$$\begin{aligned}
e^{-\mathcal{B}t}\mathbf{X}_t - \mathbf{X}_0 &= \int_0^t -\mathcal{B}e^{-\mathcal{B}u}\mathbf{X}_u du + \sum_{i=1}^m \int_0^t e^{-\mathcal{B}u} \frac{\partial}{\partial x_i} \mathbf{X}_u dX_u^i \\
&\quad + \frac{1}{2} \sum_{1 \leq i, j \leq m} \int_0^t e^{-\mathcal{B}u} \frac{\partial^2}{\partial x_i \partial x_j} \mathbf{X}_u d[X^i, X^j]_u^c \\
&\quad + \sum_{0 \leq u \leq t} \left\{ e^{-\mathcal{B}u} \mathbf{X}_u - e^{-\mathcal{B}u} \mathbf{X}_{u-} - \sum_{i=1}^m e^{-\mathcal{B}u} \frac{\partial}{\partial x_i} \mathbf{X}_u \Delta X_u^i \right\} \\
&= \int_0^t -\mathcal{B}e^{-\mathcal{B}u}\mathbf{X}_u du + \sum_{i=1}^m \int_0^t e^{-\mathcal{B}u} \mathbf{e}_i ((\mathcal{B}\mathbf{X}_u)^i du + \sigma^i dL_u) \\
&\quad + \sum_{0 \leq u \leq t} \left\{ e^{-\mathcal{B}u} \mathbf{X}_u - e^{-\mathcal{B}u} \mathbf{X}_{u-} - e^{-\mathcal{B}u} \sum_{i=1}^m \mathbf{e}_i (X_u^i - X_{u-}^i) \right\} \\
&= \int_0^t -\mathcal{B}e^{-\mathcal{B}u}\mathbf{X}_u du + \int_0^t e^{-\mathcal{B}u} (\mathcal{B}\mathbf{X}_u du + \sigma dL_u) \\
&\quad + \sum_{0 \leq u \leq t} \left\{ e^{-\mathcal{B}u} \mathbf{X}_u - e^{-\mathcal{B}u} \mathbf{X}_{u-} - e^{-\mathcal{B}u} (\mathbf{X}_u - \mathbf{X}_{u-}) \right\} \\
&= \int_0^t e^{-\mathcal{B}u} \sigma dL_u.
\end{aligned}$$

Notice that all terms including partial second derivatives with respect to t vanish due to (1.26). Therefore the solution of (1.30) is given by

$$\mathbf{X}_t = e^{\mathcal{B}t} \mathbf{X}_0 + \int_0^t e^{\mathcal{B}(t-u)} \sigma dL_u.$$

1.3 State space models

The purpose of this section is to give a short introduction to the state space approach to time series analysis. We will concentrate on the discrete time state space model. Time series analysis based on state space models has been extensively studied over the recent years. This is not surprising since the state space model is applicable to a large number of time series models (see e.g. Chapter 12.1 in Brockwell and Davis (1991) for examples). Another reason is the Kalman filter which is a powerful tool for prediction and smoothing of time series as we will see later on. For a detailed account of this subject we refer to the books of Durbin and Koopman (2001), Harvey (1990), Jones (1993), Brockwell and Davis (1991) and Hannan and Deistler (1988).

1.3.1 Discrete time state space model

The general state space form can be written in a variety of ways. We shall use the following notation:

$$\mathbf{Y}_t = \mathcal{G}_t \boldsymbol{\alpha}_t + \boldsymbol{\varepsilon}_t, \quad (1.31)$$

$$\boldsymbol{\alpha}_{t+1} = \mathcal{F}_t \boldsymbol{\alpha}_t + \boldsymbol{\eta}_t, \quad (1.32)$$

for $t \in \mathbb{N}$. The vectors $\mathbf{Y}_t \in \mathbb{R}^p$ and $\boldsymbol{\alpha}_t \in \mathbb{R}^m$ are called the *observation* and *state* vector of the state space system. Before we continue the necessary assumptions underlying the state space model defined by the *state equation* (1.32) and *observation equation* (1.31) are given:

(S1) (1.31) and (1.32) can be parametrised by a vector $\boldsymbol{\theta} \in \mathbb{R}^q$

(S2) $(\mathcal{F}_t(\boldsymbol{\theta}))_{t \in \mathbb{N}} \in \mathbb{R}^{m \times m}$ and $(\mathcal{G}_t(\boldsymbol{\theta}))_{t \in \mathbb{N}} \in \mathbb{R}^{p \times m}$ are sequences of parameter matrices.

(S3) $\begin{pmatrix} \boldsymbol{\varepsilon}_t \\ \boldsymbol{\eta}_t \end{pmatrix}_{t \in \mathbb{N}} \in \mathbb{R}^{p+m}$ is an orthogonal sequence of random vectors.

(S4) The initial state $\boldsymbol{\alpha}_1$ is orthogonal to $(\boldsymbol{\varepsilon}_t^T, \boldsymbol{\eta}_t^T)_{t \in \mathbb{N}}$.

(S5) The errors $\boldsymbol{\varepsilon}_t \in \mathbb{R}^p$ and $\boldsymbol{\eta}_t \in \mathbb{R}^m$ satisfy

$$\begin{aligned} \mathbb{E}(\boldsymbol{\varepsilon}_t^T, \boldsymbol{\eta}_t^T) &= \mathbf{0}, \\ \mathbb{E} \left[\begin{pmatrix} \boldsymbol{\varepsilon}_t \\ \boldsymbol{\eta}_t \end{pmatrix} (\boldsymbol{\varepsilon}_t^T, \boldsymbol{\eta}_t^T) \right] &= \begin{pmatrix} \mathcal{R}_t(\boldsymbol{\theta}) & 0 \\ 0 & \mathcal{Q}_t(\boldsymbol{\theta}) \end{pmatrix}, \quad t \in \mathbb{N}, \end{aligned}$$

where $(\mathcal{R}_t(\boldsymbol{\theta}))_{t \in \mathbb{N}} \in \mathbb{R}^{p \times p}$ and $(\mathcal{Q}_t(\boldsymbol{\theta}))_{t \in \mathbb{N}} \in \mathbb{R}^{m \times m}$ are sequences of parameter matrices.

The assumption of uncorrelated error processes $\boldsymbol{\varepsilon}$ and $\boldsymbol{\eta}$ can be relaxed (cf. Chapter 3 in Hannan and Deistler (1988)), but in our context zero correlation will always be the case. In a lot of applications the matrices $\mathcal{F}_t(\boldsymbol{\theta})$, $\mathcal{G}_t(\boldsymbol{\theta})$, $\mathcal{Q}_t(\boldsymbol{\theta})$ and $\mathcal{R}_t(\boldsymbol{\theta})$ are independent of time. Compare for example Example 12.1.5 in Brockwell and Davis (1991) for the state space representation of a causal discrete time ARMA(p, q) process. In Remark 1.4.3 we will give an example of a time-dependent state space model.

The term *observation equation* indicates that \mathbf{Y} models the observed data which depends on a latent state process $\boldsymbol{\alpha}$. In many applications the state process is of interest. The best linear estimate in mean square sense of $\boldsymbol{\alpha}$ at time t based on observations \mathbf{Y} can be found with the Kalman filter recursion.

In general the Kalman filter is concerned with three different types of problems. The *prediction*, *filtering* and *smoothing* problem, which are defined as the

estimation of the state α_t in terms of $\mathbf{Y}_1, \dots, \mathbf{Y}_{t-s}$, $t \geq s > 0$, $\mathbf{Y}_1, \dots, \mathbf{Y}_t$ and $\mathbf{Y}_1, \dots, \mathbf{Y}_{t+s}$, $s > 0$, respectively.

In our applications we will mainly deal with the prediction problem. If we speak of the Kalman filter in the following we will mean the one-step prediction problem.

As we have already said, the Kalman filter computes an estimate $\alpha_{t+1|t}(\boldsymbol{\theta})$ of α_{t+1} , which is linear in $\mathbf{Y}_1, \dots, \mathbf{Y}_t$ and best in the mean square sense. In particular the i -th component of the one-step ahead prediction $\alpha_{t+1|t}(\boldsymbol{\theta})$ of α_{t+1} is the projection of the i -th component, $i = 1, \dots, m$, onto the linear span of all of the components of $\mathbf{Y}_1, \dots, \mathbf{Y}_t$ (cf. also Chapter 2.7 and Definition 12.2.2 in Brockwell and Davis (1991)). If all the components of $\alpha_{t+1}, \mathbf{Y}_1, \dots, \mathbf{Y}_t$ are jointly normally distributed then

$$\alpha_{t+1|t}(\boldsymbol{\theta}) = \mathbb{E}(\alpha_{t+1} | Y_1, \dots, Y_t),$$

for each $t \geq 1$ (see Chapter 2.7 and Remark 3 in Chapter 12.2 in Brockwell and Davis (1991)).

1.3.2 Kalman filter and Gaussian likelihood

Suppose we have a state space model (1.31) and (1.32) which satisfies (S1)-(S5). The Kalman filter is a recursive algorithm which has to be initialised. There are different ways to get initial values of the recursion. Since we will always deal throughout the thesis with a stationary state space model, reasonable initial values for the best linear prediction $\alpha_{t+1|t}(\boldsymbol{\theta})$ of α_{t+1} and the error covariance matrix

$$\mathcal{P}_{t+1}(\boldsymbol{\theta}) = \mathbb{E}[(\alpha_{t+1} - \alpha_{t+1|t})(\alpha_{t+1} - \alpha_{t+1|t})^T]$$

are the mean $\mathbb{E}(\alpha_1)$ and covariance $\text{Cov}(\alpha_1)$ of the stationary distribution of α . For other initialisation techniques we refer to Chapter 5 of Durbin and Koopman (2001).

Then the best linear prediction $\alpha_{t+1|t}(\boldsymbol{\theta})$ and the error covariance matrix $\mathcal{P}_{t+1}(\boldsymbol{\theta})$ are uniquely determined by the initial conditions

$$\alpha_{1|0}(\boldsymbol{\theta}) = \mathbb{E}(\alpha_1) \quad \text{and} \quad \mathcal{P}_1(\boldsymbol{\theta}) = \mathbb{E}[(\alpha_1 - \mathbb{E}(\alpha_1))(\alpha_1 - \mathbb{E}(\alpha_1))^T]$$

and the filtering equations for $t \geq 1$:

$$\begin{aligned} \Omega_t(\boldsymbol{\theta}) &= \mathcal{G}_t(\boldsymbol{\theta})\mathcal{P}_t(\boldsymbol{\theta})\mathcal{G}_t^T(\boldsymbol{\theta}) + \mathcal{R}_t(\boldsymbol{\theta}), \\ \mathcal{K}_t(\boldsymbol{\theta}) &= \mathcal{F}_t(\boldsymbol{\theta})\mathcal{P}_t(\boldsymbol{\theta})\mathcal{G}_t^T(\boldsymbol{\theta})\Omega_t^{-1}(\boldsymbol{\theta}), \\ \mathcal{P}_{t+1}(\boldsymbol{\theta}) &= \mathcal{F}_t(\boldsymbol{\theta})\mathcal{P}_t(\boldsymbol{\theta})\mathcal{F}_t^T(\boldsymbol{\theta}) + \mathcal{Q}_t(\boldsymbol{\theta}) - \mathcal{K}_t(\boldsymbol{\theta})\Omega_t(\boldsymbol{\theta})\mathcal{K}_t(\boldsymbol{\theta})^T, \\ \alpha_{t+1|t}(\boldsymbol{\theta}) &= \mathcal{F}_t(\boldsymbol{\theta})\alpha_{t|t-1}(\boldsymbol{\theta}) + \mathcal{K}_t(\boldsymbol{\theta})(\mathbf{Y}_t - \mathcal{G}_t(\boldsymbol{\theta})\alpha_{t|t-1}(\boldsymbol{\theta})). \end{aligned}$$

For a proof of the Kalman recursion see e.g. Theorem 3.2.1 in Hannan and Deistler (1988). Observe that $\Omega_t^{-1}(\boldsymbol{\theta})$ is any generalised inverse of $\Omega_t(\boldsymbol{\theta})$ (see also Remark 5 on p. 475 in Brockwell and Davis (1991)).

Next we want to show an application of the Kalman filter in context of maximum likelihood estimation. Assume the state space model (1.31), (1.32) is parametrised by the vector $\boldsymbol{\theta} \in \mathbb{R}^q$. The parameters can then be estimated by maximising the Gaussian log-likelihood

$$\begin{aligned} \log L(\boldsymbol{\theta}, \mathbf{Y}_1, \dots, \mathbf{Y}_n) &= -\frac{np}{2} \log(2\pi) \\ &\quad - \frac{1}{2} \log(\det(\mathbf{C}(\boldsymbol{\theta})) - \frac{1}{2}(\mathbf{Y} - \boldsymbol{\mu}(\boldsymbol{\theta}))^T \mathbf{C}^{-1}(\boldsymbol{\theta})(\mathbf{Y} - \boldsymbol{\mu}(\boldsymbol{\theta})), \end{aligned}$$

where $\mathbf{Y} = (\mathbf{Y}_1, \dots, \mathbf{Y}_n)^T \in \mathbb{R}^{np}$, $\boldsymbol{\mu}(\boldsymbol{\theta}) = \mathbb{E}(\mathbf{Y}) \in \mathbb{R}^{np}$ and $\mathbf{C}(\boldsymbol{\theta}) = \text{Cov}(\mathbf{Y}) \in \mathbb{R}^{np \times np}$, even if \mathbf{Y} is not normally distributed (see Chapter 4 in Hannan and Deistler (1988) for details). For non-normal data we call the resulting estimator a quasi-maximum likelihood estimator (QMLE). The computation of the inverse $\mathbf{C}^{-1}(\boldsymbol{\theta})$ is expensive if the number of observations n is large. This computation can be avoided. We therefore write the log-likelihood in the *prediction error decomposition* (cf. Chapter 3.4 in Harvey (1990))

$$\log L(\boldsymbol{\theta}, \mathbf{Y}) = \sum_{t=1}^n \log p_{\boldsymbol{\theta}}(\mathbf{Y}_t | \mathbf{Y}_1^{t-1}),$$

where $p_{\boldsymbol{\theta}}(\mathbf{Y}_t | \mathbf{Y}_1^{t-1})$ denotes the density of \mathbf{Y}_t conditional on the information set $\mathbf{Y}_1^{t-1} := \{\mathbf{Y}_1, \dots, \mathbf{Y}_{t-1}\}$ at time $t-1$ and $p_{\boldsymbol{\theta}}(\mathbf{Y}_1 | \mathbf{Y}_1^0) = p_{\boldsymbol{\theta}}(\mathbf{Y}_1)$. Now it follows from the Kalman recursion under the assumption of normally distributed data \mathbf{Y} that the conditional distribution of \mathbf{Y}_t given \mathbf{Y}_1^{t-1} is again normal with mean

$$\mathbb{E}(\mathbf{Y}_t | \mathbf{Y}_1^{t-1}) = \boldsymbol{\mathcal{G}}_t(\boldsymbol{\theta}) \boldsymbol{\alpha}_{t|t-1}(\boldsymbol{\theta})$$

and covariance matrix $\boldsymbol{\Omega}_t(\boldsymbol{\theta})$ (cf. Chapter 4.2 in Durbin and Koopman (2001)). Hence the Gaussian log-likelihood is equal to

$$\log L(\boldsymbol{\theta}, \mathbf{Y}) = -\frac{np}{2} \log(2\pi) - \frac{1}{2} \sum_{t=1}^n (\log(\det(\boldsymbol{\Omega}_t(\boldsymbol{\theta}))) + \mathbf{v}_t^T(\boldsymbol{\theta}) \boldsymbol{\Omega}_t^{-1}(\boldsymbol{\theta}) \mathbf{v}_t(\boldsymbol{\theta})),$$

where $\mathbf{v}_t(\boldsymbol{\theta}) := \mathbf{Y}_t - \boldsymbol{\mathcal{G}}_t(\boldsymbol{\theta}) \boldsymbol{\alpha}_{t|t-1}(\boldsymbol{\theta})$, and can routinely be calculated by the Kalman filter. The problem of computing the inverse of a high-dimensional $np \times np$ matrix is therefore reduced to computing the inverse of a $p \times p$ matrix n times, which is an advantage since in most applications p is much smaller than n . In Chapter 6 we will apply the Kalman filter in a mixed effect model, where one component consist of a continuous time ARMA process defined in the next section.

1.4 Continuous time autoregressive moving average processes

In this section we want to introduce a time series model with continuous time parameter, which can be understood as the continuous time analogue of the well known ARMA model (see e.g. Chapter 3 in Brockwell and Davis (1991)). Members of this model class are therefore called *continuous time autoregressive moving average* processes, or shortly CARMA processes. The model class was found to be especially useful when modelling unequally spaced data as e.g. pointed out by Jones (1981) and Jones (1985) who used a continuous time autoregression (CAR). In Jones (1993) it is shown how to model the error process of a linear mixed effect model with irregularly spaced observation by a Gaussian CARMA process.

Originally CARMA processes have been defined with a Brownian motion as driving noise, cf. Chapter 3.7.7. in Priestley (1981). In recent years Brockwell (2001a, 2001b, 2004) has studied CARMA processes, which are driven by any Lévy process with finite k -th moment for some $k > 0$. This generalisation allows for heavy tailed marginal distributions and jumps in the sample path frequently observed in empirical data. Despite the fact that the model can be defined for Lévy processes with finite k -th moment, we will restrict ourselves to the case of Lévy processes with finite second moment, since this will be enough for our purpose.

1.4.1 Lévy driven CARMA(p, q) processes

Being the continuous time analogue of an ARMA process it is natural that the CARMA process is defined as the stationary solution of the continuous time analogue of the difference equation defining an ARMA process, i.e. a linear differential equation. Formally this is indeed the case. A second order Lévy driven CARMA(p, q) process $Y := (Y_t)_{t \geq 0}$ is defined as the stationary solution of the formal p -th order differential equation,

$$a(D)Y_t = b(D)DL_t, \quad t \geq 0, \quad (1.33)$$

in which D denotes differentiation with respect to t , $L := (L_t)_{t \geq 0}$ is a univariate Lévy process with $\mathbb{E}(L_1^2) < \infty$,

$$\begin{aligned} \text{autoregressive polynomial:} \quad a(z) &:= z^p + a_1 z^{p-1} + \cdots + a_p, \\ \text{moving-average polynomial:} \quad b(z) &:= 1 + b_1 z + \cdots + b_{p-1} z^{p-1}, \end{aligned}$$

and the coefficients b_j satisfy $b_q \neq 0$ and $b_j = 0$ for $q < j < p$. It is assumed that $a(z)$ and $b(z)$ have no common factors. Since in general the derivative of a Lévy process does not exist, one has to interpret (1.33) in a different way. And indeed (1.33) can be transformed by some algebraic manipulation into the equivalent state-space

representation:

$$\text{observations equation: } Y_t = \mathbf{b}^T \mathbf{X}_t, \quad (1.34)$$

and

$$\text{state equation: } d\mathbf{X}_t = \mathcal{A}\mathbf{X}_t dt + \mathbf{1}_p dL_t, \quad (1.35)$$

where

$$\mathcal{A} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ -a_p & -a_{p-1} & -a_{p-2} & \cdots & -a_1 \end{bmatrix}, \quad \mathbf{1}_p = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 1 \\ b_1 \\ \vdots \\ b_{p-2} \\ b_{p-1} \end{bmatrix}.$$

The state equation is therefore a system of linear stochastic differential equations.

Definition 1.4.1 *Let $p, q \in \mathbb{N}_0$ with $p > q$. Then we define a second order Lévy driven CARMA(p, q) process Y as the stationary solution of the continuous time state space model (1.34) and (1.35) with driving Lévy process L satisfying $\mathbb{E}(L_1^2) < \infty$.*

The state equation (1.35) is a system of linear stochastic differential equations driven by a one dimensional Lévy process. Thus it is of the same form as (1.30) and the solution can be found by applying Itô's formula, as shown in Example 1.2.2 (ii). The conditions for existence of a weakly (strictly) stationary solution are given in Proposition 1 (Proposition 2) in Brockwell and Marquardt (2005). We summarise them in the following Proposition.

Proposition 1.4.2 *If the real part of the roots $\lambda_1, \dots, \lambda_p$ of the autoregressive polynomial $a(z)$ is negative and \mathbf{X}_0 is independent of the driving Lévy process $(L_t)_{t \geq 0}$, with $\mathbb{E}(L_1^2) < \infty$, then the process*

$$Y_t = \mathbf{b}^T \mathbf{X}_t,$$

where

$$\mathbf{X}_t = e^{\mathcal{A}t} \mathbf{X}_0 + \int_0^t e^{\mathcal{A}(t-u)} \mathbf{1}_p dL_u,$$

i.e.

$$Y_t = \mathbf{b}^T e^{\mathcal{A}t} \mathbf{X}_0 + \int_0^t \mathbf{b}^T e^{\mathcal{A}(t-u)} \mathbf{1}_p dL_u, \quad (1.36)$$

for each $t \geq 0$, is a solution to the continuous time state space model (1.34) and (1.35). If \mathbf{X}_0 has the mean and covariance matrix of $\int_0^\infty e^{\mathcal{A}u} \mathbf{1}_p dL_u$, then Y is a

weakly stationary process. The mean and autocovariance function of the weakly stationary CARMA(p, q) process $(Y_t)_{t \geq 0}$ are

$$\mathbb{E}(Y_t) = -\mathbf{b}^T \mathcal{A}^{-1} \mathbf{1}_p \mathbb{E}(L_1) \quad (1.37)$$

and

$$\text{Cov}(Y_{t+h}, Y_t) = \text{Var}(L_1) \mathbf{b}^T e^{\mathcal{A}h} \Sigma \mathbf{b}, \quad (1.38)$$

where $\Sigma := \int_0^\infty e^{\mathcal{A}u} \mathbf{1}_p \mathbf{1}_p^T e^{\mathcal{A}^T u} du$ (see also Proposition A.19). In case \mathbf{X}_0 has the distribution of $\int_0^\infty e^{\mathcal{A}u} \mathbf{1}_p dL_u$, then Y is strictly stationary.

Remark 1.4.3 Let L be a zero mean Lévy process with finite second moment. Suppose we observe the CARMA(p, q) process (1.36) at time points $t_0 \leq t_1 \leq t_2 \leq \dots$, which may be unequally spaced. Then the state vectors \mathbf{X}_{t_i} and observations Y_{t_i} satisfy the discrete time state space model

$$\begin{aligned} Y_{t_i} &= \mathbf{b}^T \mathbf{X}_{t_i} \\ \mathbf{X}_{t_{i+1}} &= e^{\mathcal{A}(t_{i+1}-t_i)} \mathbf{X}_{t_i} + \mathbf{Z}_{t_i}, \quad i \in \mathbb{N}, \end{aligned}$$

where $(\mathbf{Z}_{t_i})_{i \in \mathbb{N}} \in \mathbb{R}^p$ is a sequence of independent random vectors with

$$\begin{aligned} \mathbb{E}(\mathbf{Z}_{t_i}) &= \mathbf{0} \\ \mathbb{E}(\mathbf{Z}_{t_i} \mathbf{Z}_{t_i}^T) &= \mathbb{E}(L_1)^2 \int_0^{t_{i+1}-t_i} e^{\mathcal{A}y} \mathbf{1}_p \mathbf{1}_p^T e^{\mathcal{A}^T y} dy. \end{aligned}$$

This is precisely the form of the discrete time state space model introduced in Chapter 1.3 without an observation noise.

The strictly stationary CARMA(p, q) process can also be defined on the negative real line by taking a two-sided Lévy process (see (1.17)) as driving process.

Definition 1.4.4 Let $p, q \in \mathbb{N}_0$ with $p > q$. If the real part of the roots $\lambda_1, \dots, \lambda_p$ of the autoregressive polynomial $a(z)$ is negative and L is a two-sided Lévy process, with $\mathbb{E}(L_1^2) < \infty$ then the second order Lévy driven CARMA(p, q) is the strictly stationary process

$$Y_t = \int_{-\infty}^{\infty} g(t-u) dL_u, \quad t \in \mathbb{R}, \quad (1.39)$$

with kernel function

$$g(t) := \mathbf{b}^T e^{\mathcal{A}t} \mathbf{1}_p \chi_{(0, \infty)}(t) \quad (1.40)$$

satisfying $g \in L^2(\mathbb{R}) \cap L^1(\mathbb{R})$.

From the representation (1.39) we see that Y is a *causal continuous time moving average process* (the class of (causal) continuous time moving average processes will be defined below). In the following we will always assume to work with a causal CARMA process.

Definition 1.4.5 *Let L be a square integrable Lévy process and $f \in L^2(\mathbb{R}) \cap L^1(\mathbb{R})$. Then the continuous time parameter process*

$$Z_t = \int_{-\infty}^{\infty} f(t-u) dL_u, \quad t \in \mathbb{R}, \quad (1.41)$$

is called a continuous time moving average process. If $f(t) = 0$ for all $t \leq 0$ the process Z is said to be causal, i.e. that Z_t depends only on the history $(L_s)_{s \leq t}$ of L for each $t \in \mathbb{R}$.

The process Z is well defined (see Section 4.3 in Applebaum (2004)), strictly stationary (see Theorem 4.3.16 in Applebaum (2004)) and by Theorem 2.7 in Rajput and Rosiński (1989) infinitely divisible with characteristic function at time $t \in \mathbb{R}$ given by

$$\mathbb{E}(e^{iuZ_t}) = \exp \left\{ iu\gamma_Z^t - \frac{1}{2}u^2\sigma_Z^{2t} + \int_{\mathbb{R}} (e^{iux} - 1 - iux\chi_{(-1,1)}(x)) \nu_Z^t(dx) \right\}, \quad u \in \mathbb{R},$$

where

$$\begin{aligned} \gamma_Z^t &:= \int_{\mathbb{R}} \gamma_L f(t-u) du + \\ &\quad \int_{\mathbb{R}} \int_{\mathbb{R}} xf(t-u)(\chi_{(-1,1)}(xf(t-u)) - \chi_{(-1,1)}(x)) \nu_L(dx) du, \\ \sigma_Z^{2t} &:= \sigma_L^2 \int_{\mathbb{R}} f^2(t-u) du, \\ \nu_Z^t(A) &:= \int_{\mathbb{R}} \int_{\mathbb{R}} \chi_A(f(t-u)x) \nu_L(dx) du, \quad A \in \mathcal{B}(\mathbb{R}), \end{aligned}$$

and $(\gamma_L, \sigma_L^2, \nu_L)$ is the characteristic triplet of L .

In particular the stationary distribution of the CARMA(p, q) process (1.39) is infinitely divisible. We will make use of this fact in Chapter 3 when we model the log-volatility process of a continuous time EGARCH process as a CARMA process.

We conclude this section with a result on the mixing behaviour of CARMA(p, q) processes (see Doukhan (1994) for a comprehensive treatment of mixing properties). In particular a strong mixing property is stated. Therefore we first recall the definition of strongly mixing for a stationary process.

Definition 1.4.6 For a stationary process $Z = (Z_s)_{s \geq 0}$ define the σ -algebras $\mathcal{F}_{[0,u]}^Z := \sigma((Z_s)_{s \in [0,u]})$ and $\mathcal{F}_{[u+t,\infty)}^Z := \sigma((Z_s)_{s \geq u+t})$ for all $u \geq 0$. Then Z is called strongly or α -mixing, if

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

as $t \rightarrow \infty$, for all $u \geq 0$.

The strong mixing property with exponential rate of the CARMA process is the subject of the next proposition, by which strongly mixing with exponential rate (exponentially α -mixing) means that $\alpha(t)$ decays to zero exponentially fast for $t \rightarrow \infty$.

Proposition 1.4.7 Let Y be defined by (1.34) and (1.35). Assume that $\mathbb{E}(L_1^2) < \infty$, the real part of the roots $\lambda_1, \dots, \lambda_p$ of the autoregressive polynomial $a(z)$ is negative and \mathbf{X}_0 has the same distribution as $\int_0^\infty e^{\mathbf{A}u} \mathbf{1}_p dM_u$, hence Y is strictly stationary. Then there exist constants $K > 0$ and $a > 0$ such that

$$\alpha_Y(t) \leq K \cdot e^{-at} \quad (1.42)$$

where $\alpha_Y(t)$ is the α -mixing coefficient of the CARMA(p, q) process Y .

Proof: The CARMA(p, q) process Y is equal to the first component of the p -dimensional Ornstein-Uhlenbeck process (henceforth referred to as OU process) $\mathbf{V} := (V^1, \dots, V^p)^T \in \mathbb{R}^p$ (see e.g. Section 4 of Brockwell (2001b)) where for fixed t

$$\mathbf{V}_t = e^{\mathbf{B}\mathbf{A}\mathbf{B}^{-1}(t-s)} \mathbf{V}_s + \int_s^t e^{\mathbf{A}(t-u)} \mathbf{B} \mathbf{1}_p dL_u \quad a.s., \quad (1.43)$$

with

$$\mathbf{B} = \begin{bmatrix} 1 & b_1 & b_2 & \cdots & b_{p-1} \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}.$$

Since L has finite second moment, \mathbf{V} also has finite second moment. Therefore the condition (4.5) in Masuda (2004) is satisfied. By Theorem 4.3 in Masuda (2004) \mathbf{V} is then exponentially α -mixing. Since every component of a multidimensional strongly mixing process is strongly mixing, Y is also exponentially α -mixing. \square

Remark 1.4.8 *A consequence of the strong mixing property of the CARMA(p, q) process (see e.g. Section 1.2.2 in Doukhan (1994)) is that*

$$|\text{Cov}(Y_{t+h}, Y_t)| \leq K \cdot e^{-ah}, \quad \forall h > 0. \quad (1.44)$$

In particular this means that the autocovariance function of the CARMA process will decay to zero at an exponential rate.

In case the roots $\lambda_1, \dots, \lambda_p$ are distinct we get from (2.12) in Brockwell (2004) the following representation of the autocovariance function

$$\text{Cov}(Y_{t+h}, Y_t) = \sum_{i=1}^p \frac{b(\lambda_i)b(-\lambda_i)}{a'(\lambda_i)a(-\lambda_i)} e^{\lambda_i|h|}, \quad h \in \mathbb{R}.$$

showing the exponential decay. Processes with an exponentially decaying autocovariance function will be called short memory processes.

1.4.2 Lévy driven FICARMA(p, d, q) processes

At the end of last section we have seen that the CARMA process is a short memory process. In different applications like finance (see e.g. Andersen and Bollerslev (1997)) or teletraffic (see Fay et al. (2006) and references therein) there is sometimes need for models explaining long range dependence in the data. Since there are different ways to characterise long range dependence or long memory, we first want to give the definition of a long memory process as we will use it before we continue.

Definition 1.4.9 *Let Z be a stationary stochastic process and $\gamma_Z(h) = \text{Cov}(Z_{t+h}, Z_t)$, $h \in \mathbb{R}$, be its autocovariance function. If there exists $0 < d < 0.5$ and a constant $c_Z > 0$ such that*

$$\lim_{h \rightarrow \infty} \frac{\gamma_Z(h)}{h^{2d-1}} = c_Z, \quad (1.45)$$

then Z is called a stationary process with long memory.

Thus we will say that a stationary process has long memory if the autocovariance function decays to zero at a hyperbolic rate.

For discrete time ARMA processes it was possible to extend the model to a long memory process resulting in the ARFIMA class of models (see e.g. Chapter 13.2 in Brockwell and Davis (1991)) and the same is true for the continuous time parameter case. Brockwell (2004) defined the class of fractionally integrated CARMA (FICARMA) processes. The idea behind this definition is to replace the kernel function g in (1.39) by its Riemann-Liouville fractional integral of order d . Let $0 < \alpha < 1$, then the Riemann-Liouville fractional integral I_+^α of order α of $\varphi \in L^p(\mathbb{R})$, $1 \leq p < 1/\alpha$, is defined by

$$(I_+^\alpha \varphi)(x) = \frac{1}{\Gamma(\alpha)} \int_{-\infty}^x \varphi(u)(x-u)^{\alpha-1} du.$$

The integral I_+^α is also called left-sided Riemann-Liouville fractional integral. For the definition of the right-sided one see Definition 2.1 in Samko et al. (1993), but we will only need I_+^α . The long memory kernel function g_d is then given by

$$g_d(t) = \int_0^t g(t-u) \frac{u^{d-1}}{\Gamma(d)} du. \quad (1.46)$$

Definition 1.4.10 *Let $0 < d < 0.5$. If the real part of the roots $\lambda_1, \dots, \lambda_p$ of the autoregressive polynomial $a(z)$ is negative and L is a two-sided Lévy process, with $\mathbb{E}(L_1^2) < \infty$ then the stationary FICARMA(p, d, q) process Y_d is defined as*

$$Y_{d,t} = \int_{-\infty}^{\infty} g_d(t-u) dL_u, \quad t \in \mathbb{R}, \quad (1.47)$$

with kernel function g_d defined in (1.46).

Remark 1.4.11 *From Theorem 5.3 in Samko et al. (1993) we get that $g_d \in L^2(\mathbb{R})$. Hence the stochastic integral (1.47) is indeed well-defined by Theorem 4.3.4 in Applebaum (2004).*

The long memory parameter d was taken between 0 and 0.5 to get a long memory process as defined in Definition 1.4.9, since

$$\lim_{h \rightarrow \infty} \frac{\text{Cov}(Y_{d,t+h}, Y_d)}{h^{2d-1}} = \frac{\mathbb{E}(L_1^2) \Gamma(1-2d)}{\Gamma(d) \Gamma(1-d)} \left(\frac{b(0)}{a(0)} \right)^2,$$

for $0 < d < 0.5$. For a proof of the long memory property of the FICARMA(p, d, q) process see Theorem 1.24 in Marquardt (2006a). The asymptotic behaviour of the autocovariance function is due to the behaviour of g_d . From (4.6) in Brockwell (2004) we know that the kernel function is asymptotically equivalent to $\frac{t^{d-1}}{\Gamma(d)} \cdot \frac{b(0)}{a(0)}$, which will be denoted by $g_d(t) \sim \frac{t^{d-1}}{\Gamma(d)} \cdot \frac{b(0)}{a(0)}$ as $t \rightarrow \infty$ and means that

$$\lim_{t \rightarrow \infty} g_d(t) \frac{\Gamma(d)}{t^{d-1}} \cdot \frac{a(0)}{b(0)} = 1.$$

In Chapter 5 we will show an application of the FICARMA model. The stationary log-volatility process of a continuous time fractionally integrated EGARCH process will be modeled by a FICARMA process.

1.5 Continuous time GARCH(1, 1) processes

We conclude this chapter by introducing the continuous time GARCH model of Klüppelberg et al. (2004). As we will see it is constructed in such a way to be the continuous time analogue of a discrete time GARCH(1, 1) process. The GARCH(1, 1) process is a model widely used by practitioners in the financial industry. It is defined as

$$Y_i = \sigma_i \epsilon_i \quad \text{with} \quad \sigma_i^2 = \omega_0 + \lambda Y_{i-1}^2 + \delta \sigma_{i-1}^2, \quad i \in \mathbb{N}, \quad (1.48)$$

where $\omega_0 > 0, \lambda, \delta \geq 0$ and $(\epsilon_i)_{i \in \mathbb{N}}$ is an i.i.d. innovation sequence (see also Bollerslev (1986)). This model captures some of the most prominent features in financial data, in particular in the volatility process. Empirical studies show that volatility changes randomly in time, has heavy or semi-heavy tails and clusters on high levels. These stylised features are modelled by the GARCH family as has been shown for the GARCH(1, 1) process in detail in Mikosch and Stărică (2000).

Approaches to create a continuous time GARCH model go back to Nelson (1990) and we refer to Drost and Werker (1996) for an overview. Such processes are diffusion limits to discrete time GARCH models, where, unfortunately, many of the above features of the GARCH process are wiped out in the limit; see Fasen et al. (2006). Since empirical work indicates upwards jumps in the volatility, a model driven by a Lévy process seems a natural approach. In Klüppelberg et al. (2004, 2006) such a model was suggested. They iterated the volatility equation in (1.48) to get

$$\begin{aligned} \sigma_i^2 &= \omega_0 \sum_{k=0}^{i-1} \prod_{j=k+1}^{i-1} (\delta + \lambda \epsilon_j^2) + \sigma_0^2 \prod_{j=0}^{i-1} (\delta + \lambda \epsilon_j^2) \\ &= \omega_0 \int_0^i \exp \left\{ \sum_{j=\lfloor u \rfloor + 1}^{i-1} \log(\delta + \lambda \epsilon_j^2) \right\} du + \sigma_0^2 \exp \left\{ \sum_{j=0}^{i-1} \log(\delta + \lambda \epsilon_j^2) \right\} \\ &= \left[\omega_0 \int_0^i \exp \left\{ - \sum_{j=0}^{\lfloor u \rfloor} \log(\delta + \lambda \epsilon_j^2) \right\} du + \sigma_0^2 \right] \exp \left\{ \sum_{j=0}^{i-1} \log(\delta + \lambda \epsilon_j^2) \right\} \\ &= \left[\omega_0 \int_0^i \exp \left\{ \eta(\lfloor u \rfloor + 1) - \sum_{j=0}^{\lfloor u \rfloor} \log(1 + \varphi \epsilon_j^2) \right\} du + \sigma_0^2 \right] \\ &\quad \times \exp \left\{ -\eta i + \sum_{j=0}^{i-1} \log(1 + \varphi \epsilon_j^2) \right\}, \end{aligned}$$

where $\lfloor u \rfloor$ denotes the integer part of $u \in \mathbb{R}$, $\eta := -\log(\delta)$ and $\varphi := \lambda/\delta$, then suggested to replace the noise variables ϵ_i by the jumps $\Delta L_t = L_t - L_{t-}$ of a Lévy process $L = (L_t)_{t \geq 0}$, which allowed them to define the volatility process σ^2 for all

$t \geq 0$ by

$$\sigma_t^2 := \left(\omega_0 \int_0^t e^{X_s} ds + \sigma_0^2 \right) e^{-X_t}, \quad (1.49)$$

where

$$X_t := \eta t - \sum_{0 < s \leq t} \log(1 + \varphi (\Delta L_s)^2), \quad t \geq 0.$$

The parameter space is given by $\omega_0, \eta > 0$ and $\varphi > 0$ and the process X will be referred to as the *auxiliary process*.

The COGARCH(1,1) process $G = (G_t)_{t \geq 0}$ is defined as the solution to the stochastic differential equation (SDE)

$$dG_t = \sigma_t dL_t. \quad (1.50)$$

We define $G_0 := 0$ and σ_0^2 is taken to be independent of L . In Theorem 2.2 in Brockwell et al. (2006) it is shown that the volatility process σ^2 can also be defined as the solution to the SDE

$$d\sigma_{t+}^2 = (\alpha\eta - \eta\sigma_t^2) dt + \varphi \sigma_t^2 d[L, L]_t^d, \quad (1.51)$$

where $\alpha > 0$ and $[L, L]_t^d = \sum_{0 \leq s \leq t} (\Delta L_s)^2$, $t \geq 0$, is the discrete part of the quadratic variation process $([L, L]_t)_{t \geq 0}$ of the Lévy process L (see also (A.2)). They showed that the solution to (1.51) coincides with (1.49) for $\omega_0 = \alpha\eta$. We will define the new parameter $\beta := \alpha\eta$ and work in the following with the SDE

$$d\sigma_{t+}^2 = (\beta - \eta\sigma_t^2) dt + \varphi \sigma_t^2 d[L, L]_t^d. \quad (1.52)$$

If our data consist of returns over time intervals of fixed length $r > 0$, we denote

$$G_t^{(r)} := G_t - G_{t-r} = \int_{(t-r, t]} \sigma_u dL_u, \quad t \geq r, \quad (1.53)$$

and $(G_{ri}^{(r)})_{i \in \mathbb{N}}$ describes an equidistant sequence of such non-overlapping returns. Hence $G^{(r)}$ corresponds to the discrete time GARCH process.

Here and in the next chapter the following terminology will be used:

G	<i>log-price process</i>
$G^{(r)}$	<i>log-return process</i>
σ^2	<i>volatility process.</i>

Calculating for the volatility the quantity corresponding to $G^{(r)}$ yields

$$\begin{aligned}\sigma_{ri}^{2(r)} &:= \sigma_{ri}^2 - \sigma_{r(i-1)}^2 = \int_{(r(i-1), ri]} \left((\beta - \eta \sigma_u^2) du + \varphi \sigma_u^2 d[L, L]_u^d \right) \\ &= \beta r - \eta \int_{(r(i-1), ri]} \sigma_u^2 du + \varphi \int_{(r(i-1), ri]} \sigma_u^2 d[L, L]_u^d,\end{aligned}\quad (1.54)$$

which is referred to as *actual volatility* in contrast to σ_t^2 , which is also called *instantaneous* or *spot volatility*. Whereas the process G is taken as being càdlàg, for the volatility process we assume càglàd sample paths. Further the volatility process is assumed to be stationary and latent. From Lemma 1 in Klüppelberg et al. (2006) we know that a strictly stationary version of σ^2 exists if $\mathbb{E}(|L_1|^{2s}) < \infty$ and $\Psi(s) \leq 0$ for some $s > 0$, where

$$\Psi(s) = -\eta s + \int_{\mathbb{R}} ((1 + \varphi x^2)^s - 1) \nu_L(dx), \quad s \geq 0. \quad (1.55)$$

is the Laplace exponent of the Laplace transform $\mathbb{E}(e^{-sX_t}) = e^{t\Psi(s)}$ of the auxiliary process X .

In contrast to classical stochastic volatility models, is the volatility σ^2 not independent of the process L , which drives the log-price process G . On the contrary, L drives both, the volatility and the log-price process. Note that G jumps at the same times as L does with jump size $\Delta G_t = \sigma_t \Delta L_t$, and that ΔL_t is independent of σ_t , since σ_t is left-continuous and hence does not include the jump ΔL_t .

One of the advantages of the COGARCH(1, 1) is that its second order structure is well-known. From Corollary 4.1 of Klüppelberg et al. (2004) we know for $k \in \mathbb{N}$ the moment $\mathbb{E}(\sigma^{2k})$ of the stationary volatility process, which exists if and only if $\mathbb{E}(L_1^{2k}) < \infty$ and $\Psi(k) < 0$. In particular, if $\mathbb{E}(L_1^4) < \infty$ and $\Psi(2) < 0$, then for $t, h \geq 0$

$$\mathbb{E}(\sigma_t^2) = \frac{\beta}{|\Psi(1)|} \quad \text{and} \quad \mathbb{E}(\sigma_t^4) = \frac{2\beta^2}{|\Psi(1)\Psi(2)|}, \quad (1.56)$$

$$\begin{aligned}\text{Cov}(\sigma_t^2, \sigma_{t+h}^2) &= \beta^2 \left(\frac{2}{|\Psi(1)\Psi(2)|} - \frac{1}{\Psi(1)^2} \right) e^{-h|\Psi(1)|} \\ &= \text{Var}(\sigma_t^2) e^{-h|\Psi(1)|}.\end{aligned}\quad (1.57)$$

In the following result we present the moments of $G_t^{(r)}$, which are independent of t by stationarity: expressions (1.58) and (1.60) have been already proved in Proposition 5.1 of Klüppelberg et al. (2004), however, under additional assumptions such as finite variation of L for (1.60). Here we shall give a different proof under less restrictive assumptions and also calculate the fourth moment of G .

Proposition 1.5.1 *Suppose that the Lévy process $(L_t)_{t \geq 0}$ has finite variance and zero mean, and that $\Psi(1) < 0$. Let $(\sigma_t^2)_{t \geq 0}$ be the stationary volatility process, so that $(G_t)_{t \geq 0}$ has stationary increments. Then $\mathbb{E}(G_t^2) < \infty$ for all $t \geq 0$, and for every $t, h \geq r > 0$ it holds*

$$\mathbb{E}(G_t^{(r)}) = 0, \quad \mathbb{E}(G_t^{(r)})^2 = \frac{\beta r}{|\Psi(1)|} \mathbb{E}(L_1^2), \quad \text{Cov}(G_t^{(r)}, G_{t+h}^{(r)}) = 0. \quad (1.58)$$

If further $\mathbb{E}(L_1^4) < \infty$ and $\Psi(2) < 0$, then $\mathbb{E}(G_t^4) < \infty$ for all $t \geq 0$ and, if additionally the Lévy measure ν_L of L is such that $\int_{\mathbb{R}} x^3 \nu_L(dx) = 0$, then it holds for every $t, h \geq r > 0$

$$\begin{aligned} & \mathbb{E}(G_t^{(r)})^4 \\ &= 6\mathbb{E}(L_1^2) \frac{\beta^2}{\Psi(1)^2} (2\eta\varphi^{-1} + 2\sigma_L^2 - \mathbb{E}(L_1^2)) \left(\frac{2}{|\Psi(2)|} - \frac{1}{|\Psi(1)|} \right) \left(r - \frac{1 - e^{-r|\Psi(1)|}}{|\Psi(1)|} \right) \\ & \quad + \frac{2\beta^2}{\varphi^2} \left(\frac{2}{|\Psi(2)|} - \frac{1}{|\Psi(1)|} \right) r + 3 \frac{\beta^2}{\Psi(1)^2} (\mathbb{E}(L_1^2))^2 r^2 \end{aligned} \quad (1.59)$$

and

$$\begin{aligned} \text{Cov}((G_t^{(r)})^2, (G_{t+h}^{(r)})^2) &= \frac{\mathbb{E}(L_1^2)\beta^2}{|\Psi(1)|^3} (2\eta\varphi^{-1} + 2\sigma_L^2 - \mathbb{E}(L_1^2)) \left(\frac{2}{|\Psi(2)|} - \frac{1}{|\Psi(1)|} \right) \\ & \quad \times \left(1 - e^{-r|\Psi(1)|} \right) \left(e^{r|\Psi(1)|} - 1 \right) e^{-h|\Psi(1)|} > 0. \end{aligned} \quad (1.60)$$

Proof: Since L has finite variance and zero mean, it is a square integrable martingale. Further, $\Psi(1) < 0$ implies $\mathbb{E}(\sigma_t^2) = \frac{\beta}{|\Psi(1)|} < \infty$ by (1.56), and it follows from equations (1.29) and (1.21) that

$$\mu := \mathbb{E}(G_t^2) \stackrel{(1.29)}{=} \mathbb{E}([G, G]_t) + 2\mathbb{E} \int_0^t G_{u-} \sigma_u dL_u \stackrel{(1.21)}{=} \mathbb{E}([G, G]_t).$$

since $G_{u-} \sigma_u$ is predictable. As

$$dG_t = \sigma_t dL_t \stackrel{(1.14)}{=} \sigma_t dB_t + \int_{\mathbb{R} - \{0\}} \sigma_t x \tilde{N}_L(dt, dx),$$

we get from (1.25)

$$[G, G]_t = \sigma_L^2 \int_0^t \sigma_u^2 du + \int_0^t \int_{\mathbb{R}} \sigma_u^2 x^2 N_L(du, dx) \stackrel{(1.23)}{=} \sigma_L^2 \int_0^t \sigma_u^2 du + \sum_{0 \leq u \leq t} \sigma_u^2 (\Delta L_u)^2.$$

Applying the compensation formula (see Proposition A.17) it then follows that

$$\begin{aligned} \mu &= \sigma_L^2 \int_0^t \mathbb{E}(\sigma_u^2) du + \mathbb{E} \left(\sum_{0 \leq u \leq t} \sigma_u^2 (\Delta L_u)^2 \right) \stackrel{(A.7)}{=} \int_0^t \mathbb{E}(\sigma_u^2) [\sigma_L^2 + \int_{\mathbb{R}} x^2 \nu_L(dx)] du \\ & \stackrel{(1.13)}{=} \mathbb{E}(L_1^2) \int_0^t \mathbb{E}(\sigma_u^2) du, \end{aligned}$$

since $\mathbb{E}(L_1) = 0$. Thus $\mathbb{E}(G_t^2)$ is finite and has the form specified in (1.58). The remaining equations in (1.58) are shown as in Proposition 5.1 of Klüppelberg et al. (2004).

Suppose that $\mathbb{E}(L_1^4) < \infty$ and $\Psi(2) < 0$. Then $\mathbb{E}(G_t^4)$ is finite by the Burkholder-Davis-Gundy inequality, cf. Protter (2004), p. 222, since

$$\mathbb{E}([G, G]_t^2) = \mathbb{E} \left(\left(\int_0^t \sigma_u^2 d(\sigma_L^2 u + \sum_{0 \leq u \leq t} (\Delta L_u)^2) \right)^2 \right) \stackrel{(A.3)}{=} \mathbb{E} \left(\left(\int_0^t \sigma_u^2 d[L, L]_u \right)^2 \right)$$

is finite as a consequence of $\mathbb{E}(\sigma_t^4) < \infty$ and $\mathbb{E}(L_1^4) < \infty$.

Now suppose additionally that $\int_{\mathbb{R}} x^3 \nu_L(dx) = 0$. To calculate the value of $\mathbb{E}(G_t^4)$, observe that by integration by parts (see equation (1.29)) and equation (A.6),

$$\begin{aligned} G_t^2 &= 2 \int_0^t G_{u-} dG_u + [G, G]_t = 2 \int_0^t G_{u-} \sigma_u dL_u + \int_0^t \sigma_u^2 d[L, L]_u, & (1.61) \\ G_t^4 &\stackrel{(1.29)}{=} 2 \int_0^t G_{u-}^2 dG_u^2 + [G^2, G^2]_t \\ &\stackrel{(1.61)}{=} 4 \int_0^t G_{u-}^3 \sigma_u dL_u + 2 \int_0^t G_{u-}^2 \sigma_u^2 d[L, L]_u \\ &\quad + \left[2 \int_0^t G_{u-} \sigma_u dL_u + \int_0^t \sigma_u^2 d[L, L]_u, 2 \int_0^t G_{u-} \sigma_u dL_u + \int_0^t \sigma_u^2 d[L, L]_u \right]_t \\ &\stackrel{(A.6)}{=} 4 \int_0^t G_{u-}^3 \sigma_u dL_u + 2 \int_0^t G_{u-}^2 \sigma_u^2 d[L, L]_u + 4 \int_0^t G_{u-}^2 \sigma_u^2 d[L, L]_u \\ &\quad + \int_0^t \sigma_u^4 d[[L, L], [L, L]]_u + 4 \int_0^t G_{u-} \sigma_u^3 d[[L, L], L]_u. \end{aligned}$$

Using (A.3)-(A.5) and the fact that the quadratic variation processes are of finite variation we can write

$$\begin{aligned} G_t^4 &= 4 \int_0^t G_{u-}^3 \sigma_u dL_u + 6 \left[\sigma_L^2 \int_0^t G_{u-}^2 \sigma_u^2 du + \sum_{0 \leq u \leq t} G_{u-}^2 \sigma_u^2 (\Delta L_u)^2 \right] \\ &\quad + \sum_{0 \leq u \leq t} \sigma_u^4 (\Delta L_u)^4 + 4 \sum_{0 \leq u \leq t} G_{u-} \sigma_u^3 (\Delta L_u)^3, & (1.62) \end{aligned}$$

for each $t \geq 0$. Applying again the compensation formula (A.7) for taking expectations in (1.62), the first and the last summand vanish due to the assumptions

$\mathbb{E}L_1 = 0$ and $\int_{\mathbb{R}} x^3 \nu_L(dx) = 0$, respectively, so that

$$\begin{aligned} \mathbb{E}(G_t^4) &= 6 \left[\sigma_L^2 \int_0^t \mathbb{E}(G_{u-}^2 \sigma_u^2) du + \int_{\mathbb{R}} x^2 \nu_L(dx) \int_0^t \mathbb{E}(G_{u-}^2 \sigma_u^2) du \right] \\ &\quad + \int_{\mathbb{R}} x^4 \nu_L(dx) \int_0^t \mathbb{E}(\sigma_u^4) du \\ &\stackrel{(1.13)}{=} 6\mathbb{E}(L_1^2) \int_0^t \mathbb{E}(G_{u-}^2 \sigma_u^2) du + \int_{\mathbb{R}} x^4 \nu_L(dx) \int_0^t \mathbb{E}(\sigma_u^4) du. \end{aligned} \quad (1.63)$$

The expression $\mathbb{E}(G_{u-}^2 \sigma_u^2)$ was already calculated in the proof of Proposition 5.1 in Klüppelberg et al. (2004), however, under additional assumptions which required in particular finite variation of L . The following calculations do not require these restrictions.

Let $Y_t := \int_0^t G_{u-} \sigma_u dL_u$, $t \geq 0$. Then $\mathbb{E}(Y_t) = 0$ for all $t \geq 0$, and integration by parts and substituting from (1.52) give

$$\begin{aligned} Y_t \sigma_{t+}^2 &= \int_0^t Y_{u-} d\sigma_{u+}^2 + \int_0^t \sigma_u^2 dY_u + [\sigma_+^2, Y]_t \\ &= \int_0^t Y_{u-} (\beta - \eta \sigma_u^2) ds + \int_0^t Y_{u-} \varphi \sigma_u^2 d[L, L]_u^d \\ &\quad + \int_0^t \sigma_u^3 G_{u-} dL_u + \left[\int_0^t (\beta - \eta \sigma_u^2) du + \int_0^t \varphi \sigma_u^2 d[L, L]_u^d, \int_0^t G_{u-} \sigma_u dL_u \right]_t \\ &\stackrel{(A.6)}{=} \int_0^t Y_{u-} (\beta - \eta \sigma_u^2) du + \int_0^t Y_{u-} \varphi \sigma_u^2 d[L, L]_u^d + \int_0^t \sigma_u^3 G_{u-} dL_u \\ &\quad + \int_0^t \varphi \sigma_u^3 G_{u-} d[[L, L]^d, L]. \end{aligned}$$

Taking expectations gives

$$\begin{aligned} \mathbb{E}(Y_t \sigma_{t+}^2) &\stackrel{(A.7)}{=} -\eta \int_0^t \mathbb{E}(Y_{u-} \sigma_u^2) du + \int_0^t \varphi \mathbb{E}(Y_{u-} \sigma_u^2) du \int_{\mathbb{R}} x^2 \nu_L(dx) \\ &\quad + \int_0^t \varphi \mathbb{E}(\sigma_u^2 G_{u-}) \int_{\mathbb{R}} x^3 \nu_L(dx) du \\ &\stackrel{(1.13)}{=} (\varphi(\mathbb{E}(L_1^2) - \sigma_L^2) - \eta) \int_0^t \mathbb{E}(Y_u \sigma_{u+}^2) du, \end{aligned}$$

where we used that $\mathbb{E}(Y_t) = 0$, $\int_{\mathbb{R}} x^3 \nu_L(dx) = 0$ and that $Y_{u-} \sigma_u^2 = Y_u \sigma_{u+}^2$ almost surely for fixed u , since $\Delta L_u = 0$ almost surely for fixed u (see e.g. Lemma 2.3.2 in Applebaum (2004)). Solving this integral equation and using that $Y_0 = 0$ implies

$\mathbb{E}(Y_0\sigma_{0+}^2) = 0$, it follows that $\mathbb{E}(Y_t\sigma_{t+}^2) = 0$ for all $t \geq 0$. Substituting

$$\begin{aligned} \int_0^t \sigma_u^2 d[L, L]_u &= \int_0^t \sigma_u^2 d[L, L]_u^c + \int_0^t \sigma_u^2 d[L, L]_u^d \\ &= \sigma_L^2 \int_0^t \sigma_u^2 du + \varphi^{-1} \left(\sigma_{t+}^2 - \sigma_0^2 - \int_0^t (\beta - \eta\sigma_u^2) du \right) \end{aligned}$$

from (1.52), equations (1.61) and (1.57) now give

$$\begin{aligned} \mathbb{E}(G_t^2\sigma_{t+}^2) &= 2\mathbb{E}(Y_t\sigma_{t+}^2) + \mathbb{E} \left(\sigma_{t+}^2 \int_0^t \sigma_u^2 d[L, L]_u \right) \\ &= \sigma_L^2 \int_0^t \mathbb{E}(\sigma_t^2\sigma_u^2) du + \varphi^{-1} \left(\mathbb{E}(\sigma_t^4) - \mathbb{E}(\sigma_t^2\sigma_0^2) - \beta\mathbb{E}(\sigma_t^2)t + \eta \int_0^t \mathbb{E}(\sigma_t^2\sigma_u^2) du \right) \\ &= (\sigma_L^2 + \varphi^{-1}\eta) \int_0^t [\text{Cov}(\sigma_t^2, \sigma_u^2) + (\mathbb{E}(\sigma_0^2))^2] du + \varphi^{-1} (\mathbb{E}(\sigma_t^4) - \mathbb{E}(\sigma_t^2\sigma_0^2) - \beta\mathbb{E}(\sigma_t^2)t) \\ &= (\sigma_L^2 + \varphi^{-1}\eta) \left[\text{Var}(\sigma_0^2)e^{-t|\Psi(1)|} \int_0^t e^{u|\Psi(1)|} du + t(\mathbb{E}(\sigma_0^2))^2 \right] \\ &\quad + \varphi^{-1} (\text{Var}(\sigma_0^2) - \text{Cov}(\sigma_t^2, \sigma_0^2) - \beta\mathbb{E}(\sigma_0^2)t) \\ &= (\sigma_L^2 + \varphi^{-1}\eta) \text{Var}(\sigma_0^2) \frac{1 - e^{-t|\Psi(1)|}}{|\Psi(1)|} + \varphi^{-1} \text{Var}(\sigma_0^2) (1 - e^{-t|\Psi(1)|}) \\ &\quad + ((\sigma_L^2 + \varphi^{-1}\eta)(\mathbb{E}(\sigma_0^2))^2 - \beta\varphi^{-1}\mathbb{E}(\sigma_0^2))t. \end{aligned} \tag{1.64}$$

Using (1.56), (1.57) and $\Psi(1) = -\eta + \varphi(\mathbb{E}(L_1^2) - \sigma_L^2)$ then leads to

$$\begin{aligned} \mathbb{E}(G_t^2\sigma_{t+}^2) &= \beta^2 \left(\frac{2}{|\Psi(1)\Psi(2)|} - \frac{1}{\Psi(1)^2} \right) \left((\sigma_L^2 + \varphi^{-1}\eta) \frac{1}{|\Psi(1)|} + \varphi^{-1} \right) (1 - e^{-t|\Psi(1)|}) \\ &\quad + \left((\sigma_L^2 + \varphi^{-1}\eta) \frac{\beta^2}{\Psi(1)^2} - \varphi^{-1} \frac{\beta^2}{\Psi(1)} \right) t \\ &= \frac{\beta^2}{\psi(1)^2} \left(\frac{2}{|\Psi(2)|} - \frac{1}{|\Psi(1)|} \right) (2\eta\varphi^{-1} + 2\sigma_L^2 - \mathbb{E}(L_1^2))(1 - e^{-t|\Psi(1)|}) \\ &\quad + \frac{\beta^2}{\Psi(1)^2} \mathbb{E}(L_1^2)t. \end{aligned}$$

If we substitute this into (1.63) and use (1.56) together with the fact that

$$\int_{\mathbb{R}} x^4 \nu_L(dx) = \frac{\Psi(2) - 2\Psi(1)}{\varphi^2} \tag{1.65}$$

by (1.55), we get

$$\begin{aligned}
& \mathbb{E}((G_t^{(r)})^4) \\
&= 6\mathbb{E}(L_1^2) \left[\int_0^r \frac{\beta^2}{\psi(1)^2} \left(\frac{2}{|\Psi(2)|} - \frac{1}{|\Psi(1)|} \right) (2\eta\varphi^{-1} + 2\sigma_L^2 - \mathbb{E}(L_1^2))(1 - e^{-u|\Psi(1)|}) du \right. \\
&\quad \left. + \int_0^r \frac{\beta^2}{\Psi(1)^2} \mathbb{E}(L_1^2) u du \right] + \frac{\Psi(2) - 2\Psi(1)}{\varphi^2} \int_0^r \frac{2\beta^2}{|\Psi(1)\Psi(2)|} du \\
&= 6\mathbb{E}(L_1^2) \left[\frac{\beta^2}{\psi(1)^2} \left(\frac{2}{|\Psi(2)|} - \frac{1}{|\Psi(1)|} \right) (2\eta\varphi^{-1} + 2\sigma_L^2 - \mathbb{E}(L_1^2)) \left(r + \frac{e^{-r|\Psi(1)|} - 1}{|\Psi(1)|} \right) \right. \\
&\quad \left. + \frac{\beta^2}{\Psi(1)^2} \mathbb{E}(L_1^2) \frac{r^2}{2} \right] + \frac{2\beta^2}{\varphi^2} \frac{\Psi(2) - \Psi(1)}{|\Psi(1)\Psi(2)|},
\end{aligned}$$

which implies (1.59). For the autocorrelation of the squared increments, observe that by equation (5.4) of Klüppelberg et al. (2004) we have

$$\text{Cov}((G_t^{(r)})^2, (G_{t+h}^{(r)})^2) = \left(\frac{e^{r|\Psi(1)|} - 1}{|\Psi(1)|} \right) \mathbb{E}(L_1^2) \text{Cov}(G_r^2, \sigma_r^2) e^{-h|\Psi(1)|} \quad (1.66)$$

(in Klüppelberg et al. (2004) this was stated under the additional assumption that L is a quadratic pure jump process (i.e. $\sigma_L^2 = 0$), but it can be seen that the proof given there holds true also for L having a Brownian motion component). This then implies (1.60) by (1.64), (1.58) and (1.56). \square

Lemma 1.5.2 *Under the conditions of Proposition 1.5.1 the process $((G_{ri}^{(r)})^2)_{i \in \mathbb{N}}$ has for each fixed $r > 0$ the autocorrelation structure of an ARMA(1,1) process.*

Proof: Denote by $\gamma(h) = \text{Cov}((G_{ri}^{(r)})^2, (G_{r(i+h)}^{(r)})^2)$, $h \in \mathbb{N}_0$, the autocovariance function and by $\rho(h) = \text{Corr}((G_{ri}^{(r)})^2, (G_{r(i+h)}^{(r)})^2)$, $h \in \mathbb{N}_0$, the autocorrelation function of the discrete time process $((G_{ri}^{(r)})^2)_{i \in \mathbb{N}}$ as defined in (1.53). Then from (1.60) it follows that

$$\frac{\rho(h)}{\rho(1)} = \frac{\gamma(h)}{\gamma(1)} = e^{-(h-1)r|\Psi(1)|}, \quad h \geq 1.$$

Moreover, for $h = 1$ we get

$$\rho(1) = \frac{\gamma(1)}{\text{Var}(G_r^2)}.$$

Recalling the autocorrelation function of an ARMA(1,1) process (see e.g. Brockwell and Davis (1991), Exercise 3.16), we identify $e^{-r|\Psi(1)|}$ as the autoregressive root ϕ . The moving average root θ can be determined by matching $\rho(1) = (1 + \phi\theta)(\phi + \theta)/(1 + \theta^2 + 2\phi\theta)$. \square

Chapter 2

Estimating the COGARCH(1, 1)

In this chapter we want to present a method of moment estimator of the parameters in the COGARCH(1, 1) model defined at the end of the last chapter. After considering the identifiability of the parameters, we will define our estimator and derive its asymptotic properties. The small sample properties will be analysed in a simulation study. There we will consider two different kinds of driving Lévy processes, a compound Poisson and a Variance Gamma process.

2.1 Identifiability of the model parameters

We aim at estimation of the model parameters (β, η, φ) from a sample of equally spaced log-returns, matching empirical autocorrelation function and moments to their theoretical counterparts given in Proposition 1.5.1. In our next result we show that the parameters are identifiable by this estimation procedure for driving Lévy processes L as in Proposition 1.5.1. We assume throughout that $\mathbb{E}(L_1) = 0$ and $\text{Var}(L_1) = 1$; furthermore, we assume that the variance σ_L^2 of the Brownian motion component in L is known. This last assumption is crucial for our analysis. Therefore we consider in our examples only pure jump Lévy processes. From an application point of view we suggest an infinite activity pure jump driving Lévy process. This seems reasonable in view of Madan and Seneta (1990) and Madan et al. (1998), where the Variance Gamma process is used to model the logarithm of stock prices. For the sake of simplicity we set $r = 1$.

Theorem 2.1.1 *Suppose $(L_t)_{t \geq 0}$ is a Lévy process such that $\mathbb{E}(L_1) = 0$, $\text{Var}(L_1) = 1$, the variance σ_L^2 of the Brownian motion component of L is known with $0 \leq \sigma_L^2 < \text{Var}(L_1) = 1$, $\mathbb{E}(L_1^4) < \infty$ and $\int_{\mathbb{R}} x^3 \nu_L(dx) = 0$. Assume also that $\Psi(2) < 0$, and denote by $(G_i^{(1)})_{i \in \mathbb{N}}$ the stationary increment process of the COGARCH(1,1) process*

with parameters $\beta, \eta, \varphi > 0$. Let $\mu, \gamma(0), k, p > 0$ be constants such that

$$\begin{aligned}\mathbb{E}((G_i^{(1)})^2) &= \mu, \\ \text{Var}((G_i^{(1)})^2) &= \gamma(0), \\ \rho(h) = \text{Corr}((G_i^{(1)})^2, (G_{i+h}^{(1)})^2) &= ke^{-hp}, \quad h \in \mathbb{N}.\end{aligned}$$

Define

$$M_1 := \gamma(0) - 2\mu^2 - 6\frac{1-p-e^{-p}}{(1-e^p)(1-e^{-p})}k\gamma(0), \quad (2.1)$$

$$M_2 := \frac{2k\gamma(0)p}{M_1(e^p-1)(1-e^{-p})}. \quad (2.2)$$

Then $M_1, M_2 > 0$, and the parameters β, η, φ are uniquely determined by $\mu, \gamma(0), k$ and p and are given by the formulas

$$\beta = p\mu, \quad (2.3)$$

$$\varphi = p\sqrt{1+M_2} - p, \quad (2.4)$$

$$\eta = p\sqrt{1+M_2}(1-\sigma_L^2) + p\sigma_L^2 = p + \varphi(1-\sigma_L^2). \quad (2.5)$$

Proof: Since $r = \mathbb{E}(L_1^2) = 1$, we obtain from Proposition 1.5.1

$$\mu = \frac{\beta}{|\Psi(1)|}, \quad (2.6)$$

$$\begin{aligned}\gamma(0) &= 6\frac{\beta^2}{|\Psi(1)|^3}(2\eta\varphi^{-1} + 2\sigma_L^2 - 1)\left(\frac{2}{|\Psi(2)|} - \frac{1}{|\Psi(1)|}\right)\left(|\Psi(1)| - 1 + e^{-|\Psi(1)|}\right) \\ &\quad + \frac{2\beta^2}{\varphi^2}\left(\frac{2}{|\Psi(2)|} - \frac{1}{|\Psi(1)|}\right) + \frac{2\beta^2}{\Psi(1)^2},\end{aligned} \quad (2.7)$$

$$=: \beta^2\hat{\gamma}(0)$$

$$p = |\Psi(1)|, \quad (2.8)$$

$$\begin{aligned}k &= \frac{\hat{\gamma}^{-1}(0)}{|\Psi(1)|^3}\left(\frac{2\eta}{\varphi} + 2\sigma_L^2 - 1\right)\left(\frac{2}{|\Psi(2)|} - \frac{1}{|\Psi(1)|}\right)\left(1 - e^{-|\Psi(1)|}\right) \\ &\quad \times \left(e^{|\Psi(1)|} - 1\right)\end{aligned} \quad (2.9)$$

Then (2.6) and (2.8) immediately give (2.3). Inserting (2.9) in (2.7) and using (2.6) and (2.8), we obtain

$$\gamma(0) = 6\frac{p-1+e^{-p}}{(1-e^{-p})(e^p-1)}k\gamma(0) + \frac{2\mu^2p^2}{\varphi^2}\left(\frac{2}{|\Psi(2)|} - \frac{1}{p}\right) + 2\mu^2.$$

By definition of M_1 and (1.65), we see that

$$M_1 = \frac{2\mu^2p^2}{\varphi^2}\left(\frac{2}{|\Psi(2)|} - \frac{1}{p}\right) = \frac{2\mu^2p^2}{\varphi^2}\frac{\varphi^2}{|\Psi(2)|p}\int_{\mathbb{R}^4}x^4\nu_L(dx) > 0,$$

so that

$$\frac{2}{|\Psi(2)|} - \frac{1}{p} = \frac{M_1 \varphi^2}{2\mu^2 p^2}.$$

Inserting this in (2.9) and using (2.3) gives

$$k\gamma(0) = \frac{2\eta\varphi^{-1} + 2\sigma_L^2 - 1}{p^3} \frac{M_1 \varphi^2}{2} (1 - e^{-p})(e^p - 1),$$

so that

$$0 < pM_2 = \frac{2k\gamma(0)p^2}{M_1(e^p - 1)(1 - e^{-p})} = \frac{2\eta\varphi^{-1} + 2\sigma_L^2 - 1}{p} \varphi^2 = \left(2 + \frac{\varphi}{p}\right) \varphi,$$

where we used

$$p = |\Psi(1)| = \eta - \varphi(\mathbb{E}(L_1^2) - \sigma_L^2) \quad (2.10)$$

from (1.55). Solving this quadratic equation for φ gives (2.4), which together with (2.10) implies (2.5). \square

We conclude from (2.3)–(2.5) that our model parameter vector (β, η, φ) is a continuous function of the first two moments $\mu, \gamma(0)$ and the parameters of the autocorrelation function p and k . Hence, by continuity, consistency of the moments will immediately imply consistency of the corresponding plug-in estimates for (β, η, φ) .

2.2 The estimation algorithm

The parameters are estimated under the following assumptions:

- (H1) We have equally spaced observations G_i , $i = 0, \dots, n$, giving log-return data $G_i^{(1)} = G_i - G_{i-1}$, $i = 1, \dots, n$.
- (H2) $\mathbb{E}(L_1) = 0$ and $\text{Var}(L_1) = 1$, i.e. σ^2 can be interpreted as the volatility.
- (H3) The variance σ_L^2 of the Brownian motion component of L is known and in $[0, 1)$.
- (H4) $\int_{\mathbb{R}} x^3 \nu_L(dx) = 0$, $\mathbb{E}(L_1^4) < \infty$ and $\Psi(2) < 0$.

Define the parameter vectors $\theta := (k, p)$ and $\vartheta := (\beta, \varphi, \eta)$, where k and p are as in Theorem 2.1.1.

Remark 2.2.1 In Theorem 2.1.1, under the chosen conditions, $\rho(h) > 0$ for all $h \in \mathbb{N}$. Furthermore, it was shown that M_1 and M_2 are strictly positive. However, this does not imply that the corresponding empirical estimates are strictly positive. As we shall prove in Theorem 2.3.9 the above estimators are strongly consistent. This means for almost all sufficiently large sample paths the empirical estimates will be strictly positive and all parameter estimates are well-defined.

Algorithm 2.2.2

(1) Calculate the moment estimator

$$\hat{\mu}_n := \frac{1}{n} \sum_{i=1}^n (G_i^{(1)})^2$$

of μ and for fixed $d \geq 2$ the empirical autocovariances

$$\hat{\gamma}_n := (\hat{\gamma}_n(0), \hat{\gamma}_n(1), \dots, \hat{\gamma}_n(d))^T$$

as

$$\hat{\gamma}_n(h) := \frac{1}{n} \sum_{i=1}^{n-h} \left((G_{i+h}^{(1)})^2 - \hat{\mu}_n \right) \left((G_i^{(1)})^2 - \hat{\mu}_n \right), \quad h = 0, \dots, d.$$

(2) Compute the empirical autocorrelations $\hat{\rho}_n := (\hat{\gamma}_n(1)/\hat{\gamma}_n(0), \dots, \hat{\gamma}_n(d)/\hat{\gamma}_n(0))^T$.

(3) For fixed $d \geq 2$ define the mapping $H : \mathbb{R}_+^{d+2} \rightarrow \mathbb{R}$ by

$$H(\hat{\rho}_n, \boldsymbol{\theta}) := \sum_{h=1}^d (\log(\hat{\rho}_n(h)) - \log k + ph)^2.$$

Compute the least squares estimator

$$\hat{\boldsymbol{\theta}}_n := \arg \min_{\boldsymbol{\theta} \in \mathbb{R}_+^2} H(\hat{\rho}_n, \boldsymbol{\theta}). \quad (2.11)$$

(4) Define the mapping $J : \mathbb{R}_+^4 \rightarrow [0, \infty)^3$ by

$$J(\mu, \gamma(0), \boldsymbol{\theta}) := (p\mu, p\sqrt{1+M_2} - p, p\sqrt{1+M_2}(1 - \sigma_L^2) + p\sigma_L^2) \quad (2.12)$$

if $p, M_2 > 0$ and $J(\mu, \gamma(0), \boldsymbol{\theta}) := (0, 0, 0)$ otherwise, where M_2 is defined as in (2.2). Compute the estimator

$$\hat{\boldsymbol{\vartheta}}_n = J(\hat{\mu}_n, \hat{\gamma}_n(0), \hat{\boldsymbol{\theta}}_n).$$

In part (3), alternatively, we could also have based the least squares estimation on the autocovariance function. It turned out, however, that the estimators chosen as above are considerably more accurate. The reason for this is that k is independent of β (see (2.9)) in contrast to $k_\gamma := \text{Cov}((G_i^{(1)})^2, (G_{i+1}^{(1)})^2)e^p$.

In addition to Remark 2.2.1 we emphasize that for a stationary model the parameter p has to be strictly positive. But if we compute the unrestricted minimum of $H(\widehat{\rho}_n, \boldsymbol{\theta})$ we get

$$\widehat{p}_n^* := -\frac{\sum_{h=1}^d \left(\log(\widehat{\rho}_n(h)) - \overline{\log(\widehat{\rho}_n)} \right) \left(h - \frac{d+1}{2} \right)}{\sum_{h=1}^d \left(h - \frac{d+1}{2} \right)^2} \quad (2.13)$$

$$\widehat{k}_n := \exp \left\{ \overline{\log(\widehat{\rho}_n)} + \frac{d+1}{2} \widehat{p}_n^* \right\}, \quad (2.14)$$

with $\overline{\log(\widehat{\rho}_n)} := \frac{1}{d} \sum_{h=1}^d \log(\widehat{\rho}_n(h))$ and \widehat{p}_n^* may be negative. As a remedy we define the estimator of p as

$$\widehat{p}_n := \max\{\widehat{p}_n^*, 0\} \quad (2.15)$$

and take $\widehat{p}_n = 0$ as an indication that the data is non-stationary.

Defining the mapping $S : \mathbb{R}_+^{d+1} \rightarrow \mathbb{R}_+^2$ by the equations (2.13)–(2.15) and noting that $\widehat{\rho}_n(h) = \widehat{\gamma}_n(h)/\widehat{\gamma}_n(0)$ presents the least squares estimator $\widehat{\boldsymbol{\theta}}_n := (\widehat{k}_n, \widehat{p}_n)$ as a function of $\widehat{\gamma}_n$:

$$\widehat{\boldsymbol{\theta}}_n = S(\widehat{\gamma}_n). \quad (2.16)$$

2.3 Asymptotic properties of the moment estimators

Strong mixing properties guarantee strong consistency and asymptotic normality of the empirical moments under suitable moment conditions. Remember that the definition of the strong mixing coefficient can also be found in the Chapter 1.4 (see Definition 1.4.6). We will further need the concept of $\tilde{\alpha}$ -mixing, which will now be defined.

Definition 2.3.1 For a stationary process $Y = (Y_s)_{s \geq 0}$ define the σ -algebras $\mathcal{F}_{[0,u]}^Y := \sigma((Y_s)_{s \in [0,u]})$ and $\mathcal{F}_{[u+t,\infty)}^Y := \sigma((Y_s)_{s \geq u+t})$ for all $u \geq 0$. Then $f \in b\mathcal{F}_{[u,\infty)}^Y$ means that f is a bounded $\mathcal{F}_{[u,\infty)}^Y$ -measurable random variable. Let $\|\cdot\|_{L^1(\mathbb{P})}$ and $\|\cdot\|_\infty$ be the L^1 -norm under \mathbb{P} and the supnorm, respectively. Then Y is called $\tilde{\alpha}$ -mixing, if

$$\begin{aligned} \tilde{\alpha}(t) &= \tilde{\alpha}(\mathcal{F}_{[0,u]}^Y, \mathcal{F}_{[u+t,\infty)}^Y) \\ &:= \sup\{\|\mathbb{E}(f|\mathcal{F}_{[0,u]}^Y) - \mathbb{E}(f)\|_{L^1(P)} : f \in b\mathcal{F}_{[u+t,\infty)}^Y, \|f\|_\infty \leq 1\} \rightarrow 0, \end{aligned}$$

as $t \rightarrow \infty$, for all $u \geq 0$.

Throughout the thesis all σ -algebras are assumed to include all \mathbb{P} -null sets. The following result shows that any α -mixing property is equivalent to the corresponding $\tilde{\alpha}$ -mixing property.

Lemma 2.3.2 $\alpha(\mathcal{F}_1, \mathcal{F}_2) \leq \tilde{\alpha}(\mathcal{F}_1, \mathcal{F}_2) \leq 6\alpha(\mathcal{F}_1, \mathcal{F}_2)$ holds for all $\mathcal{F}_1, \mathcal{F}_2 \subset \mathcal{F}$.

For the left-hand inequality assume $A^* \in \mathcal{F}_1$ and $B^* \in \mathcal{F}_2$ are such that $\alpha(\mathcal{F}_1, \mathcal{F}_2) = |P(A^* \cap B^*) - P(A^*)P(B^*)|$. Then take $f = \chi_{B^*}$ as the indicator of the set B^* , which implies $|P(A^* \cap B^*) - P(A^*)P(B^*)| \leq \|\mathbb{E}(\chi_{B^*}|A^*) - \mathbb{E}(\chi_{B^*})\|_{L^1(P)}$. See Lemma 3.5 in McLeish (1975) for the right-hand inequality.

Therefore the α -mixing property is equivalent to the $\tilde{\alpha}$ -mixing property. The following remark is the starting point of our analysis.

Remark 2.3.3 Let $\sigma^2 := (\sigma_t^2)_{t \geq 0}$ be the strictly stationary volatility process given by (1.49). Then σ^2 falls into the class of generalised Ornstein-Uhlenbeck processes (see Section 5 in Lindner and Maller (2005)). Consequently Proposition 3.1 of Fasen (2007) applies giving that σ^2 is exponentially β -mixing (see Definition A.16). This implies in particular that σ^2 is exponentially α -mixing, which means that $\alpha(t)$ decays to zero exponentially fast for $t \rightarrow \infty$.

In the following theorem we show that also the COGARCH(1,1) process satisfies a strong mixing condition.

Theorem 2.3.4 Suppose that $(L_t)_{t \geq 0}$ is such that $\mathbb{E}(L_1^4) < \infty$ and the parameters of the COGARCH(1,1) process satisfy $\Psi(2) < 0$. Let $\sigma^2 := (\sigma_t^2)_{t \geq 0}$ be the strictly stationary volatility process given as solution to (1.52). Then for every $r > 0$ the process $(G_{ir}^{(r)})_{i \in \mathbb{N}}$ is α -mixing with exponentially decreasing rate.

We want to show that the sequence $(G_{ir}^{(r)})_{n \in \mathbb{N}}$ is α -mixing with exponentially fast decreasing mixing coefficients. Recall from Lemma 2.3.2 that α -mixing is equivalent to the $\tilde{\alpha}$ -mixing. Therefore we will show that $(G_{ir}^{(r)})_{n \in \mathbb{N}}$ is $\tilde{\alpha}$ -mixing

Proof: We show that $(G_{ir}^{(r)})_{i \in \mathbb{N}}$ is $\tilde{\alpha}$ -mixing. Define the σ -algebra $\mathcal{F}_I^{dL} := \sigma(L_t - L_s : s, t \in I)$ for $I \subset \mathbb{R}$; i.e. generated from all increments of L over the interval I . Using similarly notation as in Definition 2.3.1 for the σ algebras generated by the log-return process $(G_{ir}^{(r)})_{i \in \mathbb{N}}$ and volatility process $(\sigma_t^2)_{t \geq 0}$, we can formulate the following inclusions

$$\mathcal{F}_{\{1, \dots, l\}}^{G^{(r)}} \subset \mathcal{F}_{[0, lr]}^{dL} \vee \mathcal{F}_{[0, lr]}^{\sigma^2} =: \mathcal{F}_{[0, lr]}^{dL, \sigma^2},$$

where \vee denotes the σ -algebra generated by the union of the two σ -algebras, and

$$\mathcal{F}_{\{k+l, k+l+1, \dots\}}^{G^{(r)}} \subset \mathcal{F}_{[(k+l-1)r, \infty)}^{dL} \vee \mathcal{F}_{[(k+l-1)r, \infty)}^{\sigma^2} =: \mathcal{F}_{[(k+l-1)r, \infty)}^{dL, \sigma^2}.$$

Using these relations we get

$$\begin{aligned}
\tilde{\alpha}_{G^{(r)}}(k) &:= \sup \left\{ \|\mathbb{E}(f|\mathcal{F}_{\{1,\dots,l\}}^{G^{(r)}}) - \mathbb{E}(f)\|_{L^1(P)} : f \in b\mathcal{F}_{\{k+l,k+l+1,\dots\}}^{G^{(r)}}, \|f\|_\infty \leq 1 \right\} \\
&\leq \sup \left\{ \|\mathbb{E}(f|\mathcal{F}_{[0,lr]}^{dL,\sigma^2}) - \mathbb{E}(f)\|_{L^1(P)} : f \in b\mathcal{F}_{[(k+l-1)r,\infty)}^{dL,\sigma^2}, \|f\|_\infty \leq 1 \right\} \\
&= \sup \left\{ \|\mathbb{E}(f|\mathcal{F}_{[0,lr]}^{dL,\sigma^2}) - \mathbb{E}(f)\|_{L^1(P)} : f \in b\mathcal{F}_{[(k+l-1)r,\infty)}^{dL} \right. \\
&\quad \left. \vee b\mathcal{F}_{\{(k+l-1)r\}}^{\sigma^2}, \|f\|_\infty \leq 1 \right\} \\
&= \sup \left\{ \|\mathbb{E}(f|\mathcal{F}_{[0,lr]}^{\sigma^2}) - \mathbb{E}(f)\|_{L^1(P)} : f \in b\mathcal{F}_{[(k+l-1)r,\infty)}^{dL} \vee b\mathcal{F}_{\{(k+l-1)r\}}^{\sigma^2}, \|f\|_\infty \leq 1 \right\} \\
&= \sup \left\{ \|\mathbb{E}(f|\mathcal{F}_{[0,lr]}^{\sigma^2}) - \mathbb{E}(f)\|_{L^1(P)} : f \in b\mathcal{F}_{[(k+l-1)r,\infty)}^{d[L,L]^d} \vee b\mathcal{F}_{\{(k+l-1)r\}}^{\sigma^2}, \|f\|_\infty \leq 1 \right\} \\
&= \tilde{\alpha}_{\sigma^2}((k-1)r).
\end{aligned}$$

The first equality (in the third line) holds since the driving process of σ^2 is the discrete part of the quadratic variation of L . $\mathcal{F}_{[0,lr]}^{dL}$ and $b\mathcal{F}_{[(k+l-1)r,\infty)}^{dL} \vee b\mathcal{F}_{\{(k+l-1)r\}}^{\sigma^2}$ are conditionally independent given $\mathcal{F}_{[0,lr]}^{\sigma^2}$, which is due to the Markov property of σ^2 (see Theorem 3.2 in Klüppelberg et al. (2004)) and the independence between $\mathcal{F}_{[0,lr]}^{dL}$ and $b\mathcal{F}_{[(k+l-1)r,\infty)}^{dL}$. This gives the second equality. The third one follows from the independence of $b\mathcal{F}_{[(k+l-1)r,\infty)}^{dL}$ and $b\mathcal{F}_{[(k+l-1)r,\infty)}^{d[L,L]^d}$ from $\mathcal{F}_{[0,lr]}^{\sigma^2}$ and again the Markov property of σ^2 , whereas the last one is due to the fact that $b\mathcal{F}_{[(k+l-1)r,\infty)}^{\sigma^2} = b\mathcal{F}_{[(k+l-1)r,\infty)}^{d[L,L]^d} \vee b\mathcal{F}_{\{(k+l-1)r\}}^{\sigma^2}$.

Consequently, $(G_{ir}^{(r)})_{i \in \mathbb{N}}$ inherits the mixing properties from the volatility process σ^2 . It has been shown in Proposition 3.1 in Fasen (2007) that σ^2 is α -mixing with exponential rate, i.e. there exist constants $K, a > 0$ such that

$$\hat{\alpha}_{G^{(r)}}(k) \leq \hat{\alpha}_{\sigma^2}((k-1)r) \leq 6\alpha_{\sigma^2}((k-1)r) \leq 6Ke^{-a(k-1)r},$$

implying that $(G_{ir}^{(r)})_{i \in \mathbb{N}}$ is α -mixing with exponential rate. \square

Remark 2.3.5 *We further like to mention that a different proof of the strong mixing property of $(G_{ir}^{(r)})_{i \in \mathbb{N}}$ can be found in the proof of Proposition 3.1 in Fasen (2007).*

Since we assumed in the above theorem that σ^2 is strictly stationary, the log-return process is also strictly stationary and together with the strong mixing property this implies that $(G_{ir}^{(r)})_{i \in \mathbb{N}}$ is ergodic. This enables us to apply Birkhoff's ergodic theorem to give strong consistency of the empirical moments and autocovariance function of $((G_i^{(1)})^2)_{i \in \mathbb{N}}$:

Corollary 2.3.6 *Under the same conditions as in Theorem 2.3.4 we obtain for $n \rightarrow \infty$*

$$\hat{\mu}_n \xrightarrow{\text{a.s.}} \mu, \quad \hat{\gamma}_n \xrightarrow{\text{a.s.}} \gamma. \quad (2.17)$$

Corollary 2.3.6 will imply strong consistency of the estimator $\widehat{\boldsymbol{\vartheta}}_n$, as stated in (2.21) below. To obtain asymptotic normality of the empirical estimates we want to apply a central limit theorem for strongly mixing processes.

Proposition 2.3.7 *Let the same conditions hold as in Theorem 2.3.4. Assume further*

(H5) *There exists a positive constant $\delta > 0$ such that $\mathbb{E}\left(G_1^{8+\delta}\right) < \infty$.*

Then as $n \rightarrow \infty$,

$$\sqrt{n} \left(\begin{bmatrix} \widehat{\mu}_n \\ \widehat{\boldsymbol{\gamma}}_n \end{bmatrix} - \begin{bmatrix} \mu \\ \boldsymbol{\gamma} \end{bmatrix} \right) \xrightarrow{d} N_{d+2}(\mathbf{0}, \boldsymbol{\Sigma}), \quad (2.18)$$

where the covariance $\boldsymbol{\Sigma}$ has components

$$\begin{aligned} \Sigma_{k+2, l+2} &= \text{Cov}((G_1^{(1)})^2(G_{1+k}^{(1)})^2, (G_1^{(1)})^2(G_{1+l}^{(1)})^2) \\ &\quad + 2 \sum_{j=1}^{\infty} \text{Cov}((G_1^{(1)})^2(G_{1+k}^{(1)})^2, (G_{1+j}^{(1)})^2(G_{1+l+j}^{(1)})^2) \end{aligned}$$

for $k, l = 0, \dots, d$,

$$\Sigma_{1, k+2} = \text{Cov}((G_1^{(1)})^2, (G_1^{(1)})^2(G_{1+k}^{(1)})^2) + 2 \sum_{j=1}^{\infty} \text{Cov}((G_1^{(1)})^2, (G_{1+j}^{(1)})^2(G_{1+k+j}^{(1)})^2)$$

for $k = 0, \dots, d$ and $\Sigma_{1,1} = \gamma(0) + 2 \sum_{h=1}^{\infty} k_{\gamma} e^{-ph}$.

Proof: We will first concentrate on the asymptotic behaviour of $(\widehat{\mu}_n, \boldsymbol{\gamma}_n^*)$, where $\boldsymbol{\gamma}_n^* = (\gamma_n^*(0), \dots, \gamma_n^*(d))$ and $\gamma_n^*(h) = \frac{1}{n} \sum_{i=1}^n [(G_{i+h}^{(1)})^2 - \mu][(G_i^{(1)})^2 - \mu]$, $h = 0, \dots, d$. Denote

$$\mathbf{Y}_i := ((G_i^{(1)})^2, [(G_i^{(1)})^2 - \mu]^2, [(G_i^{(1)})^2 - \mu][(G_{i+1}^{(1)})^2 - \mu], \dots, [(G_i^{(1)})^2 - \mu][(G_{i+d}^{(1)})^2 - \mu])^T.$$

For (2.18) to hold for $(\widehat{\mu}_n, \boldsymbol{\gamma}_n^*)$ in place of $(\widehat{\mu}_n, \widehat{\boldsymbol{\gamma}}_n)$, by the Cramér-Wold device, we have to show that as $n \rightarrow \infty$,

$$\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \boldsymbol{\lambda}^T \mathbf{Y}_i - \boldsymbol{\lambda}^T \begin{bmatrix} \mu \\ \boldsymbol{\gamma} \end{bmatrix} \right) \xrightarrow{d} N(0, \boldsymbol{\lambda}^T \boldsymbol{\Sigma} \boldsymbol{\lambda}), \quad (2.19)$$

for all vectors $\boldsymbol{\lambda} \in \mathbb{R}^{d+2}$ such that $\boldsymbol{\lambda}^T \boldsymbol{\Sigma} \boldsymbol{\lambda} > 0$. But as strong mixing is preserved under linear transformations as well as the rate, the sequence $(\boldsymbol{\lambda}^T \mathbf{Y}_i)_{i \in \mathbb{N}}$ is strongly mixing with exponentially decaying rate. Hence we get $\sum_{k=1}^{\infty} \{\alpha_{\boldsymbol{\lambda}^T \mathbf{Y}}(k)\}^c < \infty$ for every $c > 0$, and since $\mathbb{E}|Y_i|^{2+\delta} < \infty$ for some $\delta > 0$ by (H5), the central limit

theorem for strongly mixing processes is applicable (see Theorem 18.5.3 in Ibragimov and Linnik (1971)). Therefore, as $n \rightarrow \infty$,

$$\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \boldsymbol{\lambda}^T \mathbf{Y}_i - \boldsymbol{\lambda}^T \begin{bmatrix} \mu \\ \boldsymbol{\gamma} \end{bmatrix} \right) \xrightarrow{d} N(0, \sigma^2),$$

with

$$\sigma^2 := \text{Var}(\boldsymbol{\lambda}^T \mathbf{Y}_1) + 2 \sum_{i=1}^{\infty} \text{Cov}(\boldsymbol{\lambda}^T \mathbf{Y}_1, \boldsymbol{\lambda}^T \mathbf{Y}_{1+i}). \quad (2.20)$$

Evaluation of (2.20) and rearranging with respect to $\boldsymbol{\lambda}$ shows $\sigma^2 = \boldsymbol{\lambda}^T \boldsymbol{\Sigma} \boldsymbol{\lambda}$. Observing that $\sqrt{n} \left(n^{-1} \sum_{i=1}^n \boldsymbol{\lambda}^T \mathbf{Y}_i - \boldsymbol{\lambda}^T \begin{pmatrix} \hat{\mu}_n \\ \hat{\boldsymbol{\gamma}}_n \end{pmatrix} \right)$ converges in probability to zero as $n \rightarrow \infty$ for every $\boldsymbol{\lambda} \in \mathbb{R}^{d+2}$ such that $\boldsymbol{\lambda}^T \boldsymbol{\Sigma} \boldsymbol{\lambda} > 0$ (see e.g. the proof of Proposition 7.3.4. in Brockwell and Davis (1991)), it follows that $(\hat{\mu}_n, \hat{\boldsymbol{\gamma}}_n)$ has the same asymptotic behaviour as $(\hat{\mu}_n, \boldsymbol{\gamma}_n^*)$, giving (2.18). \square

Applying the delta method (see Theorem 3.1 in van der Vaart (1998)), we obtain:

Corollary 2.3.8 *Let the same conditions hold as in Proposition 2.3.7. Then as $n \rightarrow \infty$,*

$$\sqrt{n}(\hat{\boldsymbol{\rho}}_n - \boldsymbol{\rho}) \xrightarrow{d} N_d(\mathbf{0}, \boldsymbol{\Sigma}_\rho).$$

The following theorem gives asymptotic normality of our parameter estimates. The true parameter vector and the corresponding moments are from now on indicated by $\boldsymbol{\vartheta}_0$, μ_0 and $\boldsymbol{\gamma}_0$ respectively. We shall also denote by $\mathbb{P}_{\boldsymbol{\vartheta}_0}$ the probability with respect to the parameter vector $\boldsymbol{\vartheta}_0$.

Theorem 2.3.9 *Let the same conditions hold as in Theorem 2.3.4. Assume that (H1)–(H4) are satisfied. For $S(\boldsymbol{\gamma})$ as in (2.16), define the mapping $Q : \mathbb{R}^{d+2} \rightarrow \mathbb{R}^3$ by $(\mu, \boldsymbol{\gamma}^T) \mapsto Q((\mu, \boldsymbol{\gamma}^T)) := J(\mu, \boldsymbol{\gamma}(0), S(\boldsymbol{\gamma}))$. Then as $n \rightarrow \infty$,*

$$\hat{\boldsymbol{\vartheta}}_n \xrightarrow{\text{a.s.}} \boldsymbol{\vartheta}_0. \quad (2.21)$$

Assume additionally (H5). Then, under $\mathbb{P}_{\boldsymbol{\vartheta}_0}$, as $n \rightarrow \infty$,

$$\sqrt{n}(\hat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_0) \xrightarrow{d} \partial_{(\mu, \boldsymbol{\gamma})} Q((\mu_0, \boldsymbol{\gamma}_0)) N_{d+2}(\mathbf{0}, \boldsymbol{\Sigma}), \quad (2.22)$$

where $\boldsymbol{\Sigma}$ is as in Proposition 2.3.7.

Proof: Strong consistency of $\hat{\boldsymbol{\vartheta}}_n$ follows from (2.17) and the fact that the mapping Q is continuous in $(\mu, \boldsymbol{\gamma})$. Since $(\hat{\mu}_n, \hat{\boldsymbol{\gamma}}_n)$ is asymptotically normal and Q is differentiable at $(\mu_0, \boldsymbol{\gamma}_0)$, we can apply the delta method and the asymptotic normality of $\hat{\boldsymbol{\vartheta}}_n$ follows from (2.18). \square

2.4 Examples of COGARCH(1, 1) processes

Before we analyse the small sample properties of our estimator, two examples of COGARCH(1, 1) processes will be discussed. The main difference in the two examples is the activity of the driving Lévy process. In the first case it will be a compound Poisson process, while in the second one it is a Variance Gamma process. Thus we will consider Lévy processes with finite and infinite activity. An admissible choice of parameters satisfying the assumptions (H1)-(H5) is derived, which will be used in the simulation study. In the first example we also present an estimator for the jump rate λ of the driving compound Poisson process.

2.4.1 Compound Poisson COGARCH(1, 1)

This section is devoted to the compound Poisson COGARCH(1, 1) process, which corresponds to a compound Poisson driving process L given by

$$L_t = \sum_{k=1}^{N_t} Y_k, \quad t \geq 0,$$

where $N = (N_t)_{t \geq 0}$ is a Poisson process with jump rate $\lambda > 0$, and $(Y_k)_{k \in \mathbb{N}}$ are i.i.d. random variables, independent of N . We introduce a generic random variable Y with the same distribution function as the Y_k , denoted by F_Y . For this model (H3) is clearly satisfied, with $\sigma_L^2 = 0$. The Lévy measure of L has the representation $\nu_L(dx) = \lambda F_Y(dx)$. This allows us to calculate the Laplace exponent from (1.55) getting

$$\Psi(s) = -\eta s + \lambda \int_{\mathbb{R}} ((1 + \varphi y^2)^s - 1) F_Y(dy).$$

From this we obtain

$$\begin{aligned} \Psi(1) &= -\eta + \varphi \lambda \mathbb{E}(Y^2) \\ \Psi(2) &= -2\eta + 2\varphi \lambda \mathbb{E}(Y^2) + \varphi^2 \lambda \mathbb{E}(Y^4). \end{aligned}$$

Since Theorem 2.1.1 requires $\mathbb{E}(L_1) = 0$ and $\mathbb{V}\text{ar}(L_1) = \mathbb{E}(L_1^2) = 1$, we must have $\mathbb{E}(Y^2) = 1/\lambda$ yielding $p = |\Psi(1)| = \eta - \varphi$. The condition $\int_{\mathbb{R}} x^3 \nu_L(dx) = 0$ translates into $\mathbb{E}(Y^3) = 0$. Moreover, we obtain

$$\Psi(2) = 2(\varphi - \eta) + \varphi^2 \mathbb{E}(Y^4) / \mathbb{E}(Y^2) = -2p + \varphi^2 \mathbb{E}(Y^4) / \mathbb{E}(Y^2).$$

Then the condition $\Psi(2) < 0$ translates into $\varphi^2 < 2p / (\lambda \mathbb{E}(Y^4))$. Condition (H5) requires for G a finite moment of higher order than the eighth, which is the case if $\mathbb{E}(L_1^{8+\delta}) < \infty$ and the $(4 + \delta)$ -moment of the volatility is finite i.e. $\Psi(4 + \delta) < 0$. The $(8 + \delta)$ -moment of L will be finite if $\mathbb{E}(Y^{8+\delta}) < \infty$. The volatility will have a finite fourth moment if

$$\Psi(4) = 4(\varphi - \eta) + 6\lambda\varphi^2\mathbb{E}(Y^4) + 4\lambda\varphi^3\mathbb{E}(Y^6) + \lambda\varphi^4\mathbb{E}(Y^8)$$

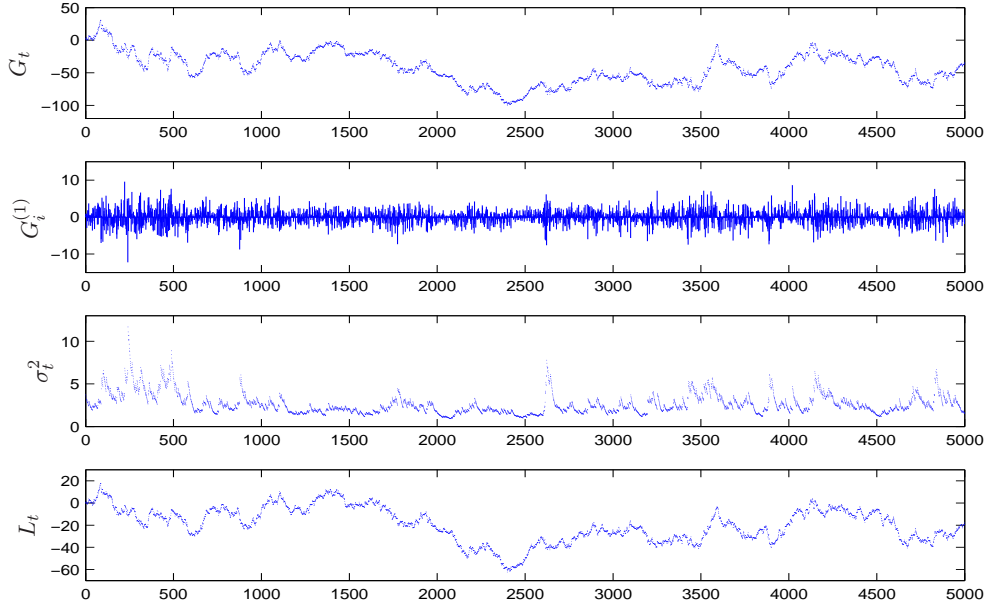


Figure 2.1: Simulated COGARCH(1, 1) process (G_t) with parameters $\beta = 0.04$, $\eta = 0.053$ and $\varphi = 0.038$ (*first*), log-return process ($G_i^{(1)}$) of order 1 (*second*), volatility process (σ_t^2) (*third*) and driving compound Poisson process (L_t) $_{0 \leq t \leq 5000}$ with jump rate $\lambda = 1$ and $N(0, 1)$ -distributed jumps (*last*).

is negative (this then implies $\Psi(4 + \delta) < 0$ for some $\delta > 0$ by continuity of Ψ , cf. Sato (1999), Lemma 26.4).

In Figure 2.1 we plotted simulated sample paths for the time interval $[0, 5000]$ of the the compound Poisson COGARCH(1,1) process G , the log-return process $G^{(1)}$, the volatility process σ^2 and the driving Lévy process L , respectively. All four sample paths have been simulated with the same random seed. As can be seen G looks similar to L , they only differ by the jump sizes. Also the volatility clustering, which is observed in real data, can be rediscovered in this simulation.

As jumps in a compound Poisson model with moderate frequency are rare, we should be able to estimate the jump rate λ from the discretised data $G_i^{(1)}$. This is shown in our next result. The analysis is based on $z(n)$, the number of intervals, where G does not change; i.e. $z(n) = \sum_{i=1}^n \chi_{\{0\}}(G_i^{(1)})$. This implies immediately that one needs a fine enough observation grid.

Proposition 2.4.1 *Let $(L_t)_{t \geq 0}$ be a compound Poisson process with continuous jump distribution F_Y and jump rate $\lambda > 0$. Then*

$$\widehat{\lambda}_n := -\log\left(\frac{z(n)}{n}\right) \xrightarrow{\text{a.s.}} \lambda, \quad n \rightarrow \infty,$$

and

$$\sqrt{n}(\widehat{\lambda}_n - \lambda) \xrightarrow{d} N(0, e^\lambda(1 - e^{-\lambda})), \quad n \rightarrow \infty. \quad (2.23)$$

Proof: Denote by S_i the number of jumps in the interval $(i-1, i]$. Then the S_i , $i = 1, \dots, n$, are i.i.d. Poisson distributed with parameter λ . Therefore, the indicator variables $\chi_{\{0\}}(S_i)$, $i = 1, \dots, n$, are also i.i.d. Since F_Y is continuous, we have

$$\chi_{\{0\}}(S_i) = \chi_{\{0\}}(G_i^{(1)}) \quad \text{a.s.}, \quad i = 1, \dots, n.$$

By the strong law of large numbers, we get

$$\frac{1}{n} \sum_{i=1}^n \chi_{\{0\}}(G_i^{(1)}) \xrightarrow{\text{a.s.}} \mathbb{E}(\chi_{\{0\}}(S_1)) = \mathbb{P}(S_1 = 0) = e^{-\lambda}, \quad n \rightarrow \infty,$$

and therefore

$$-\log\left(\frac{z(n)}{n}\right) \xrightarrow{\text{a.s.}} \lambda, \quad n \rightarrow \infty.$$

Moreover, as $\chi_{\{0\}}(G_i^{(1)})$, $i = 1, \dots, n$, are i.i.d., the central limit theorem applies, giving

$$\frac{z(n) - ne^{-\lambda}}{\sqrt{ne^{-\lambda}(1 - e^{-\lambda})}} \xrightarrow{d} N(0, 1), \quad n \rightarrow \infty.$$

Invoking the delta-method (see Theorem 3.1 in van der Vaart (1998)) to $-\log(\frac{z(n)}{n})$, using $\sqrt{e^{-\lambda}(1 - e^{-\lambda})/n} \rightarrow 0$ as $n \rightarrow \infty$ and the fact that $-\log(\cdot)$ is differentiable at $e^{-\lambda}$, we obtain (2.23). \square

Remark 2.4.2 The central limit result of Proposition 2.4.1 allows us to construct confidence intervals for the jump rate λ . Using (2.23) and

$$\frac{n}{z(n)} \left(1 - \frac{z(n)}{n}\right) \xrightarrow{\mathbb{P}} e^\lambda(1 - e^{-\lambda}), \quad n \rightarrow \infty,$$

we apply Slutsky's theorem to get

$$\frac{-\log\left(\frac{z(n)}{n}\right) - \lambda}{\sqrt{\frac{1}{z(n)}\left(1 - \frac{z(n)}{n}\right)}} \xrightarrow{d} N(0, 1). \quad (2.24)$$

Solving (2.24) with respect to λ , we get a $100(1 - \alpha)\%$ confidence interval for λ

$$\left[-\log\left(\frac{z(n)}{n}\right) - q_{1-\frac{\alpha}{2}}\sqrt{\frac{1}{z(n)} - \frac{1}{n}}, -\log\left(\frac{z(n)}{n}\right) + q_{1-\frac{\alpha}{2}}\sqrt{\frac{1}{z(n)} - \frac{1}{n}} \right],$$

where $q_{1-\frac{\alpha}{2}}$ is the $(1 - \frac{\alpha}{2})$ -quantil of the standard normal distribution.

2.4.2 Variance Gamma COGARCH(1, 1)

In this section we investigate the Variance Gamma COGARCH(1, 1) process. Again the driving Lévy process has no Brownian component. Consequently $\sigma_L^2 = 0$ and (H3) is satisfied. Assumption (H2) requires that the mean of L is zero and the variance is equal to one. Then it follows from (1.5) that the characteristic function at time $t \geq 0$ is given by

$$\mathbb{E}(e^{iuL_t}) = \left(1 + \frac{u^2}{2C}\right)^{-tC}.$$

The Lévy measure of L has the Lebesgue density

$$\nu_L(dx) = \frac{C}{|x|} \exp\left(- (2C)^{1/2} |x|\right) dx, \quad x \neq 0. \quad (2.25)$$

Inserting (2.25) into (1.55) we obtain

$$\Psi(1) = -\eta + \varphi \quad \text{and} \quad \Psi(2) = -2\eta + 2\varphi + 3\varphi^2 C^{-1}. \quad (2.26)$$

In Figure 2.2 we plotted simulated sample paths for the time interval $[0, 5000]$ of the the Variance Gamma COGARCH(1, 1) process G , the log-return process $G^{(1)}$, the volatility process σ^2 and the driving Lévy process L , respectively.

The first condition of (H4) is satisfied by symmetry. The only delicate point for choosing the parameters β , η and φ is the last condition of (H4). Econometric literature suggests that volatility is quite persistent. Hence the autocovariance function of $(G^{(1)})^2$ should not decrease too fast. From Proposition 1.5.1 we know that this is implied by $\Psi(1) < 0$ close to zero. Setting $\beta = 0.04$, $\eta = 0.053$ and $\varphi = 0.038$ gives $\Psi(1) = -0.015$ and $\Psi(2) = -0.0257$ which are satisfactory values. Condition (H5) requires for G a finite moment of higher order than the eighth, which is the case if $\mathbb{E}(L_1^{8+2\delta}) < \infty$ and the $(4 + \delta)$ -moment of the volatility is finite i.e. $\Psi(4 + \delta) < 0$. The VG process has finite moments of all orders for every $C > 0$, but for given η and φ the finiteness of $\mathbb{E}(\sigma_1^{8+2\delta})$ depends on C , since

$$\Psi(4) = -4\eta + 4\varphi + 18\varphi^2 C^{-1} + 120\varphi^3 C^{-2} + 630\varphi^4 C^{-3}$$

has to be strictly negative (this then implies $\Psi(4 + \delta) < 0$ for some $\delta > 0$ by continuity of Ψ , cf. Sato (1999), Lemma 26.4). Therefore we choose $C = 1$, resulting in $\Psi(4) = -0.0261$.

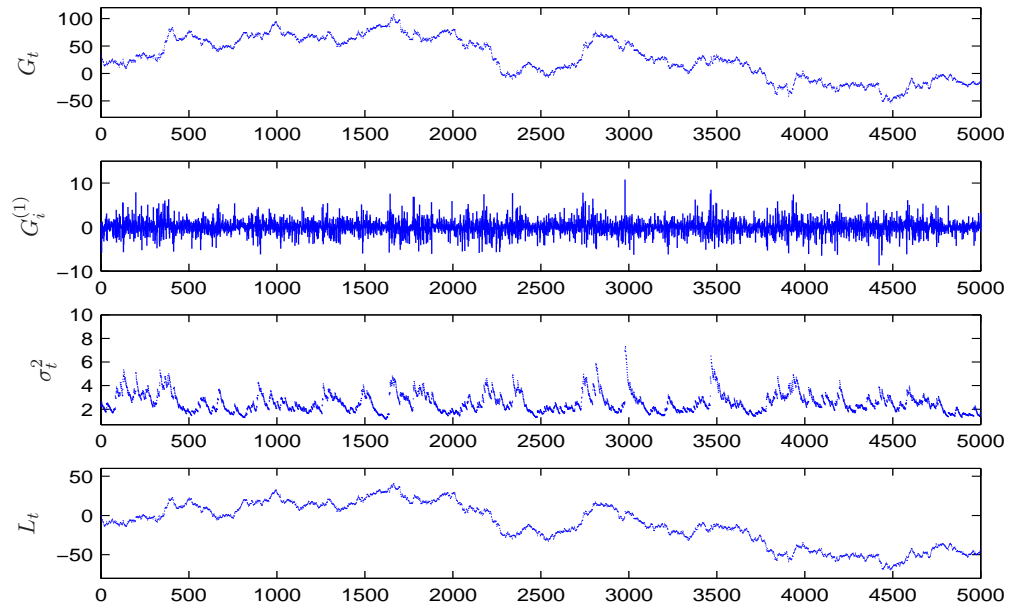


Figure 2.2: Simulated VG driven COGARCH(1,1) process $(G_t)_{0 \leq t \leq 5000}$ with parameters $\beta = 0.04, \eta = 0.053$ and $\varphi = 0.038$ (*first*), log-return process $(G_t^{(1)})$ of order 1 (*second*), the volatility process (σ_t^2) (*third*), the driving VG process (L_t) with parameter $C = 1$ (*last*).

2.5 Simulation study

We will perform the estimation procedure for two different sample sizes, namely 5000 and 20000. The estimates \hat{p}_n (2.13) and \hat{k}_n (2.14) are sensitive to the choice d of lags used and to outliers in the empirical autocorrelation function. Based on experience for linear models (recall Lemma 1.5.2), it seems reasonable to choose $d \approx \sqrt{n}$. Numerical experiments have indeed shown that d equal to 50 is sufficient for both our sample sizes. Moreover, we performed a robust linear regression (see e.g. Chapter 7 in Huber (2004)) to estimate the parameters, i.e. they are estimated by an iteratively reweighted least squares algorithm instead of ordinary least squares. The resulting estimates are not only less sensitive to outliers in the data, but also to the number of lags d taken into account.

Remark 2.5.1 *The choice of d in empirical applications is actually without any problems. An inspection of the empirical autocorrelation function immediately sug-*

gests a reasonable choice of d . But in a simulation study the parameters should be constant over all samples.

From (2.26) we know that $\Psi(1)$ is equal to $\Psi(1) = -\eta + \varphi$. Thus these two parameters give important characteristics of the model concerning stationarity and the rate p of decrease of the autocovariance and autocorrelation function. In case of $\hat{p}_n > 0$, which indicates that the data is stationary, it is also clear from (2.12), that the estimated parameters will always correspond to a stationary model, since $p > 0$ implies $\Psi(1) = -\eta + \varphi < 0$ and the same identity holds for the estimated parameters.

2.5.1 Estimation results for the compound Poisson COGARCH(1, 1)

In this section the behaviour of the moment estimators of Algorithm 2.2.2 is analysed in case of a driving compound Poisson process. All the conditions stated in Section 2.4.1 are satisfied if we choose standard normally distributed jumps implying a jump rate λ equal to one and model parameters $\beta = 0.04$, $\eta = 0.053$ and $\varphi = 0.038$. Observe that the chosen parameters imply a rather slow decay of the autocovariance function of $(G^{(1)})^2$, since $\Psi(1) = -0.015$ is close to zero. Further we get for these parameter values $\Psi(2) = -0.0257$ and $\Psi(4) = -0.0329$.

We first simulate 1 000 samples of $n = 5\,000$ equidistant observations of $G^{(1)}$. Table 2.1 summarises the outcome of our simulation study concerning the parameters β, η and φ . The empirical mean of all the estimated parameter values $\hat{\beta}_n, \hat{\eta}_n$ and $\hat{\varphi}_n$ is shown in the first line, with the empirical standard deviations in brackets. We also estimated mean square error (MSE) and mean absolute error (MAE), again with the corresponding standard deviation in brackets. Additionally the relative bias is reported. The corresponding results for a sample size of $n = 20\,000$ observations are reported in the last four lines of Table 2.1.

The estimator $\hat{\beta}$ seems to have the largest bias among all three estimators for both sample sizes. Comparing the results with respect to the two sample sizes we observe an increased accuracy, as expected due to the consistency of the estimators. Overall we got acceptable results except for the estimation of β based on 5 000 observations showing a bias of about 20 percent.

The estimation results concerning the jump rate λ and $\sigma_{Y_1}^2$, the variance of the jumps Y_k , are shown in Table 2.2 showing satisfactory performance. Again we calculated the empirical mean, MSE and MAE with corresponding empirical standard deviations.

2.5.2 Estimation results for the Variance Gamma COGARCH(1, 1)

Again we simulate first 1 000 samples of $n = 5\,000$ equidistant observations of $G^{(1)}$. Table 2.3 summarises the estimation results of our simulation study concerning the parameters β, η and φ .

n=5 000	$\hat{\beta}$	$\hat{\eta}$	$\hat{\varphi}$
Mean	0.03262 ($3 \cdot 10^{-4}$)	0.05212 ($5 \cdot 10^{-4}$)	0.03984 ($4 \cdot 10^{-4}$)
Rel. bias	-0.18461	-0.01649	0.04855
MSE	0.00017 ($6 \cdot 10^{-6}$)	0.00025 ($1 \cdot 10^{-5}$)	0.00017 ($9 \cdot 10^{-6}$)
MAE	0.01109 ($2 \cdot 10^{-4}$)	0.01282 ($3 \cdot 10^{-4}$)	0.01026 ($3 \cdot 10^{-4}$)
n=20 000	$\hat{\beta}$	$\hat{\eta}$	$\hat{\varphi}$
Mean	0.03942 ($3 \cdot 10^{-4}$)	0.05273 ($5 \cdot 10^{-4}$)	0.03787 ($4 \cdot 10^{-4}$)
Rel. bias	-0.01435	-0.00504	-0.00329
MSE	0.00008 ($4 \cdot 10^{-6}$)	0.00022 ($3 \cdot 10^{-6}$)	0.00016 ($3 \cdot 10^{-6}$)
MAE	0.00753 ($2 \cdot 10^{-4}$)	0.01015 ($3 \cdot 10^{-4}$)	0.00867 ($3 \cdot 10^{-4}$)

Table 2.1: Estimated mean, relative bias, MSE and MAE for $\hat{\beta}$, $\hat{\eta}$ and $\hat{\varphi}$ and corresponding estimated standard deviations in brackets. The true values are $\beta = 0.04$, $\eta = 0.053$ and $\varphi = 0.038$.

n=5 000	$\hat{\lambda}$	$\hat{\sigma}_{Y_1}^2$
Mean	1.00099 ($6 \cdot 10^{-4}$)	0.99935 ($6 \cdot 10^{-4}$)
MSE	0.00034 ($9 \cdot 10^{-6}$)	0.00034 ($9 \cdot 10^{-6}$)
MAE	0.01459 ($4 \cdot 10^{-4}$)	0.01456 ($4 \cdot 10^{-4}$)
n=20 000	$\hat{\lambda}$	$\hat{\sigma}_{Y_1}^2$
Mean	0.99999 ($3 \cdot 10^{-4}$)	1.00015 ($3 \cdot 10^{-4}$)
MSE	0.00008 ($3 \cdot 10^{-6}$)	0.00008 ($3 \cdot 10^{-6}$)
MAE	0.00707 ($2 \cdot 10^{-4}$)	0.00707 ($2 \cdot 10^{-4}$)

Table 2.2: Estimated mean, MSE and MAE for $\hat{\lambda}$ and $\hat{\sigma}_{Y_1}^2$ and corresponding estimated standard deviations in brackets. The true values are $c = 1$, and $\sigma_{Y_1}^2 = 1$.

The empirical mean of all the estimated parameter values $\hat{\beta}_n$, $\hat{\eta}_n$ and $\hat{\varphi}_n$ is shown in the first line, with the empirical standard deviations in brackets. As before we also estimated the relative bias, the MSE and the MAE. For the mean squared and absolute error the estimated standard deviation is again shown in brackets. The corresponding results for a sample size of $n = 20\,000$ observations are reported in the last four lines of Table 2.3.

The three estimators $\hat{\beta}_n$, $\hat{\eta}_n$ and $\hat{\varphi}_n$ show a similar power. Actually $\hat{\beta}_n$ seems to have the largest small sample variance. Contrarily to the compound Poisson case for $n = 5\,000$ the relative bias of $\hat{\beta}_n$ is the smallest. For $n = 20\,000$ estimated bias and variance of $\hat{\beta}_n$ are the largest among the three estimators. When one compares

n=5 000	$\hat{\beta}$	$\hat{\eta}$	$\hat{\varphi}$
Mean	0.04172 ($7 \cdot 10^{-4}$)	0.04897 ($7 \cdot 10^{-4}$)	0.03329 ($5 \cdot 10^{-4}$)
Rel. bias	0.04304	-0.07603	-0.12394
MSE	0.00053 ($3 \cdot 10^{-5}$)	0.00048 ($2 \cdot 10^{-5}$)	0.00023 ($9 \cdot 10^{-6}$)
MAE	0.01772 ($5 \cdot 10^{-4}$)	0.01724 ($4 \cdot 10^{-4}$)	0.01208 ($3 \cdot 10^{-4}$)
n=20 000	$\hat{\beta}$	$\hat{\eta}$	$\hat{\varphi}$
Mean	0.04309 ($4 \cdot 10^{-4}$)	0.05311 ($4 \cdot 10^{-4}$)	0.03689 ($3 \cdot 10^{-4}$)
Rel. bias	0.07735	0.00201	-0.02899
MSE	0.00019 ($9 \cdot 10^{-6}$)	0.00015 ($8 \cdot 10^{-6}$)	0.00007 ($4 \cdot 10^{-6}$)
MAE	0.01089 ($3 \cdot 10^{-4}$)	0.00954 ($2 \cdot 10^{-4}$)	0.00651 ($2 \cdot 10^{-4}$)

Table 2.3: Estimated mean, relative bias, MSE and MAE for $\hat{\beta}$, $\hat{\eta}$ and $\hat{\varphi}$ and corresponding estimated standard deviations in brackets. The true values are $\beta = 0.04$, $\eta = 0.053$ and $\varphi = 0.038$.

the estimates for the different sample sizes it can be seen that the MSE reduces for all three estimators, when the sample size is increased, and the reduction is roughly by a factor of four, which would correspond to the asymptotic properties of the estimators.

2.5.3 Estimation of the volatility σ_t^2

From an econometric point of view it is also of interest to estimate the volatility σ^2 . Therefore we want to present a recursive estimator of the volatility process given the estimated parameters $\hat{\vartheta}_n$. The estimator is then applied to the Variance Gamma example from the last section. Recall from (1.54) for $r = 1$,

$$\sigma_i^2 = \sigma_{i-1}^2 + \beta - \eta \int_{(i-1, i]} \sigma_s^2 ds + \varphi \sum_{i-1 < s \leq i} \sigma_s^2 (\Delta L_s)^2, \quad i \in \mathbb{N}. \quad (2.27)$$

Since σ_s is latent and ΔL_s is usually not observable, we have to approximate the integral and the sum on the right hand side. For the integral we use a simple Euler approximation

$$\int_{(i-1, i]} \sigma_s^2 ds \approx \sigma_{i-1}^2, \quad i \in \mathbb{N}.$$

As we observe G only at integer times we approximate

$$\sum_{i-1 < s \leq i} \sigma_s^2 (\Delta L_s)^2 \approx (G_i - G_{i-1})^2 = (G_i^{(1)})^2, \quad i \in \mathbb{N}.$$

An estimate of the volatility process $(\sigma_t^2)_{t \geq 0}$ can therefore be calculated recursively by

$$\hat{\sigma}_i^2 = \hat{\beta} + (1 - \hat{\eta})\hat{\sigma}_{i-1}^2 + \hat{\varphi}(G_i^{(1)})^2, \quad i \in \mathbb{N}. \quad (2.28)$$

Note that $\hat{\sigma}_i$ defines the conditional variance of a discrete time GARCH(1,1) model, which implies that we have to require $0 < \eta < 1$. The estimator (2.28) is plotted in Figure 2.3 together with the theoretical $(\sigma_t^2)_{t \geq 0}$ for one simulation.

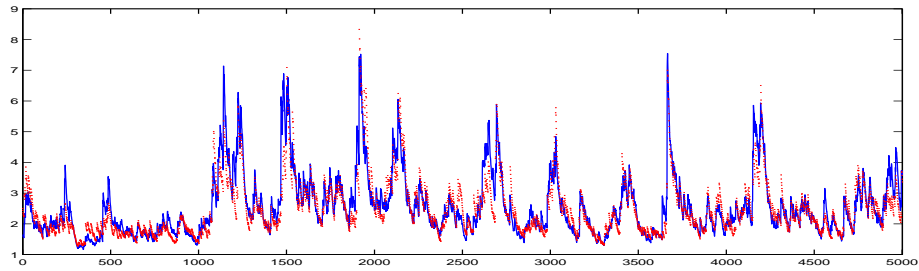


Figure 2.3: Sample paths of σ_t^2 (solid line) and $\hat{\sigma}_t^2$ (dotted line) of one simulation.

The goodness of fit of our estimation method is now investigated by a residual analysis. The estimated residuals are given by $G_i^{(1)}/\hat{\sigma}_{i-1}$ for $i = 1, \dots, n$. Since we assumed a symmetric jump distribution with zero mean, the residuals should be symmetric around zero and their mean should be close to zero. Furthermore, if the volatility has been estimated correctly, we expect the standard deviation to be close to one. Consequently, we estimated mean, MSE, MAE and the corresponding standard deviations for the mean, the standard deviation and the skewness of the residuals $G_i^{(1)}/\hat{\sigma}_{i-1}$ based on 1 000 simulations. The results for both sample sizes are reported in Table 2.4 and indicate a reasonable fit.

The correlation of the squared residuals was checked by performing a Ljung-Box test for each sample. For $n = 5\,000$ we computed the test statistic based on $70 \approx \sqrt{5\,000}$ lags and had to reject the null hypothesis of no correlation 140 times out of 1 000 simulations at the 0.05 level. Whereas for $n = 20\,000$ the test statistic was computed using $140 \approx \sqrt{20\,000}$ lags and the null hypothesis was rejected 137 times out of 1 000 simulations again at the 0.05 level.

2.6 Real data analysis

The COGARCH(1,1) model will be fitted to five minutes log-returns of three different stocks, which are General Motors (GM), Cisco and Intel. We have tick-by-tick

n=5 000	$\text{mean}(G_i^{(1)}/\hat{\sigma}_{i-1})$	$\text{std}(G_i^{(1)}/\hat{\sigma}_{i-1})$	$\text{skewness}(G_i^{(1)}/\hat{\sigma}_{i-1})$
Mean	0.00011 (0.00044)	1.00931 (0.00021)	-0.00152 (0.00428)
MSE	0.00019 ($9 \cdot 10^{-6}$)	0.00012 ($8 \cdot 10^{-6}$)	0.01838 (0.00098)
MAE	0.01110 (0.00027)	0.00945 (0.00020)	0.10671 (0.00264)
n=20 000	$\text{mean}(G_i^{(1)}/\hat{\sigma}_{i-1})$	$\text{std}(G_i^{(1)}/\hat{\sigma}_{i-1})$	$\text{skewness}(G_i^{(1)}/\hat{\sigma}_{i-1})$
Mean	0.00018 (0.00022)	0.01078 (0.00014)	-0.00285 (0.00205)
MSE	0.00005 ($2 \cdot 10^{-6}$)	0.00013 ($4 \cdot 10^{-6}$)	0.00422 (0.00019)
MAE	0.00577 (0.00013)	0.01079 (0.00014)	0.05119 (0.00126)

Table 2.4: Estimated mean, bias, MSE and MAE for the mean, standard deviation and skewness of the residuals with corresponding estimated standard deviations in brackets.

data of the *Trades and Quotes* database of the New York Stock Exchange (NYSE) and Nasdaq. The GM stock is from NYSE, whereas Cisco and Intel belong to Nasdaq. The data spans over 4 months starting in February 2002. We considered only the prices between 9.35am and 4pm to compute the five minutes log-returns based on previous tick interpolation. There were 83 trading days between the beginning of February and the end of May 2002. Hence each of the series has a total length of 6391 data points. This is part of a data set, which was analysed in Brodin and Klüppelberg (2006) with respect to the extreme dependence structure of the three stocks.

The effect of seasonality is common in high frequency data and also appears in the raw data. Therefore, the data was deseasonalised by a median filter, which is explained in Section 4.2 in Brodin and Klüppelberg (2006). The resulting time series are shown in Figure 2.4.

An application of Algorithm 2.2.2 produces moment estimates of β , η and φ under the assumption that the driving Lévy processes of each stock have no Brownian component. The results are shown in Table 2.5. To investigate the model fit, we

	$\hat{\beta}$	$\hat{\eta}$	$\hat{\varphi}$
GM	0.1091	0.1625	0.1357
Cisco	0.0621	0.0328	0.0126
Intel	0.0180	0.0396	0.0336

Table 2.5: $\hat{\beta}$, $\hat{\eta}$ and $\hat{\varphi}$ for the GM, Cisco and Intel data.

performed a Ljung-Box test for squared residuals of all three data sets. The test

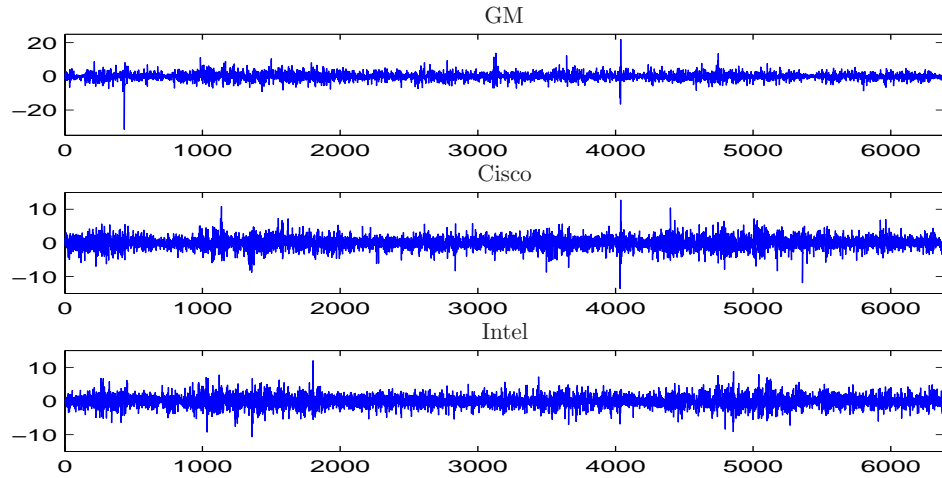


Figure 2.4: Deseasonalized 5 minutes log-returns of GM (*top*), Cisco (*middle*) and Intel (*bottom*).

statistics used 80 lags of the corresponding empirical autocorrelation function. The null hypothesis was not rejected for GM and Intel at the 0.05 level. For the GM squared residuals the p -value was 0.35, whereas for Intel it was only 0.27. The test statistic for the Cisco squared residuals was equal to 202.62, which led to a rejection of the null hypothesis, since the test had a critical value of 101.87 at the 0.05 level. This result is also obvious from Figure 2.5 where the empirical autocorrelation function of the squared residuals are plotted on the right, showing significant correlations of the Cisco residuals.

The estimated mean, standard deviation and skewness of the residuals are summarised in Table 2.6. The numbers show that the mean and variation of the residuals are according to our model, but that the residuals are significantly skewed. This skewness can also be seen in Figure 2.6 showing estimates of the log density for all three datasets.

	$\text{mean}(G_i^{(1)}/\hat{\sigma}_{i-1})$	$\text{std}(G_i^{(1)}/\hat{\sigma}_{i-1})$	$\text{skewness}(G_i^{(1)}/\hat{\sigma}_{i-1})$
GM	-0.0143	1.0785	-0.3714
Cisco	-0.0015	0.9832	-0.2082
Intel	-0.0002	1.0100	-0.0626

Table 2.6: Mean, standard deviation and skewness of the GM, Cisco and Intel residuals.

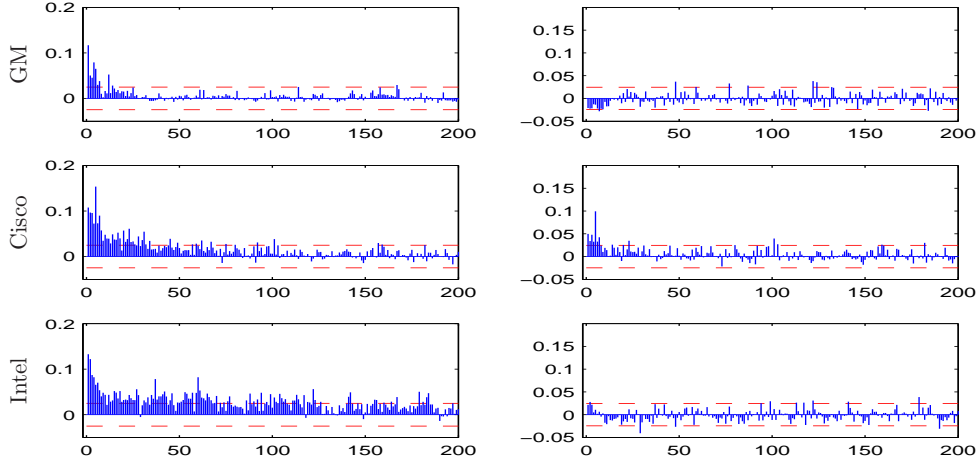


Figure 2.5: Empirical acf of the squared 5 minutes log-returns (*left*) and the squared residuals (*right*) of GM (*top*), Cisco (*middle*) and Intel (*bottom*).

It does not come as a surprise as it is a well-known fact that financial data are skewed. Although we will not deal with this problem in the framework of the COGARCH model, we want to discuss the assumptions, which prevent the modelling of skewness and also indicate some remedy to be worked out in detail in future work. This is not done in this thesis because we will introduce in the next chapter a new model, which takes this asymmetry into account directly.

The observed skewness indicates that the first condition of (H4) requiring that $\int_{\mathbb{R}} x^3 \nu_L(dx) = 0$ is violated. This introduces a bias into our estimates as in the calculation of $\mathbb{E}(G_t^4)$ the last term in (1.62) does not disappear.

Instead of including the term $\int_{\mathbb{R}} x^3 \nu_L(dx)$ in a statistical analysis we suggest to extend the model by an extra term in a similar fashion as Glosten et al. (1993) for the discrete GARCH process, to model the leverage in the market explicitly; thus taking care of the effect directly. Consequently, we extend the volatility model (1.52) for $\rho > 0$ to

$$d\sigma_{t+}^2 = (\beta - \eta\sigma_t^2) dt + \varphi\sigma_t^2 d[L, L]_t^{(d)} + \rho\sigma_t^2 dU_t, \quad (2.29)$$

where

$$U_t := \sum_{\substack{0 < s \leq t \\ \Delta L_s < 0}} (\Delta L_s)^2$$

and ρ is a positive constant. Then

$$\Delta\sigma_t^2 = \begin{cases} \varphi\sigma_t^2(\Delta L_t)^2, & \text{if } \Delta L_t > 0, \\ (\varphi + \rho)\sigma_t^2(\Delta L_t)^2, & \text{if } \Delta L_t < 0, \end{cases}$$

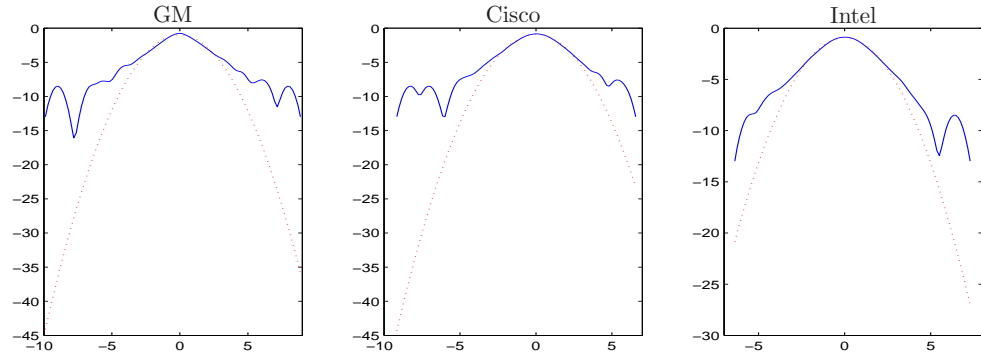


Figure 2.6: Kernel estimates of the log density (*solid line*) of the squared GM (*left*), Cisco (*middle*) and Intel (*right*) residuals together with a normal log density (*dashed line*) with mean and variance of the corresponding residual series.

so that a negative jump of L gives rise to a higher increase of the volatility than a positive jump of the same modulus does. Note that $(U_t)_{t \geq 0}$ is a subordinator, and so is

$$M_t := \varphi[L, L]_t^{(d)} + \rho U_t = \sum_{\substack{0 < s \leq t \\ \Delta L_s > 0}} \varphi(\Delta L_s)^2 + \sum_{\substack{0 < s \leq t \\ \Delta L_s < 0}} (\varphi + \rho)(\Delta L_s)^2,$$

so that (2.29) can be rewritten as

$$d\sigma_{t+}^2 = (\beta - \eta\sigma_t^2) dt + \sigma_t^2 dM_t,$$

an expression similar to (1.52).

Chapter 3

Exponential continuous time GARCH process

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 *stylised 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 modeled by a GARCH type process without further extensions. This finding led Nelson (1990) to introduce the exponential GARCH process, which is able to model this asymmetry in stock returns. The log-volatility of the EGARCH(p, q) process was modeled as an ARMA($q, p - 1$) process.

The availability of high frequency data, which increased enormously in the last years, is one reason to consider continuous time models with similar behaviour as discrete time GARCH models. The reason for this is of course that at the highest available frequency the observations of the price process occur at irregularly spaced time points and therefore it is natural to assume an underlying continuous time model. Different approaches have been taken to set up a continuous time model, which has the same features as discrete time GARCH processes. Recently Klüppelberg et al. (2004) developed a continuous time GARCH(1, 1) model (see also Chapter 1.5), shortly called COGARCH(1, 1). Their approach differs fundamentally from previous attempts, which could be summarised as diffusion approximations (see e.g. Nelson (1991)), by the fact that their model is driven by only one source of randomness (like discrete time GARCH) instead of two (as in the diffusion approximations). They replaced the noise process of discrete time GARCH by the jumps of a Lévy process. The COGARCH(1, 1) model was then extended by Brockwell et al. (2006) to a continuous time GARCH(p, q) process for general orders $p, q \in \mathbb{N}$, $q \geq p$, hence-

forth called COGARCH(p, q).

In this chapter a continuous time analogue of the EGARCH(p, q) model is introduced. The noise processes will also be modeled by the increments of a Lévy process. As in the discrete time case we describe the log-volatility process as a linear process, more precisely a continuous time ARMA($q, p - 1$) process.

3.1 The discrete time EGARCH process

Motivated by empirical evidence that stock returns are negatively correlated with changes in returns volatility Nelson (1990) defined the exponential GARCH process (EGARCH) to model this effect, which is called *leverage effect* (see also Chapter 3.4.2).

The process $(X_n)_{n \in \mathbb{Z}}$ of the form

$$X_n = \sigma_n \epsilon_n, \quad n \in \mathbb{Z}, \quad (3.1)$$

where $(\epsilon_n)_{n \in \mathbb{Z}}$ is an i.i.d. sequence with $\mathbb{E}(\epsilon_1) = 0$ and $\text{Var}(\epsilon_1) = 1$, is called an EGARCH process, if the volatility process $(\sigma_n^2)_{n \in \mathbb{Z}}$ satisfies

$$\log(\sigma_n^2) = \mu + \sum_{k=1}^{\infty} \beta_k f(\epsilon_{n-k}), \quad (3.2)$$

where $f : \mathbb{R} \rightarrow \mathbb{R}$ is some measurable real valued deterministic function, $\mu \in \mathbb{R}$ and $(\beta_k)_{k \in \mathbb{N}}$ are real coefficients such that

$$\mathbb{E}(|f(\epsilon_n)|) < \infty, \text{Var}(f(\epsilon_n)) < \infty \quad \text{and} \quad \sum_{k=1}^{\infty} |\beta_k| < \infty.$$

To achieve the asymmetric relation between the stock returns and the volatility, $f(\epsilon_n)$ must be a function of the magnitude and the sign of ϵ_n as noted by Nelson (1990). Therefore he proposed the following function:

$$f(\epsilon_n) := \theta \epsilon_n + \gamma[|\epsilon_n| - \mathbb{E}(|\epsilon_n|)], \quad (3.3)$$

with real coefficients θ and γ . We see that $f(\epsilon_n)$ is linear in ϵ_n and has slope $\theta + \gamma$ for positive shocks ϵ_n and slope $\theta - \gamma$ for negative ones. Therefore $f(\epsilon_n)$ allows the volatility process $(\sigma_n^2)_{n \in \mathbb{Z}}$ to respond asymmetrically to positive and negative jumps in the stock price.

Nelson (1990) also suggested a finite parameter model by modeling the log-volatility as an ARMA($q, p - 1$) process instead of an infinite moving average process. This leads to the EGARCH(p, q) model, which is defined in the following way.

Let $p, q \in \mathbb{N}, \mu, \alpha_1, \dots, \alpha_q, \beta_1, \dots, \beta_p \in \mathbb{R}$, suppose $\alpha_q \neq 0, \beta_p \neq 0$ and that the autoregressive polynomial

$$\alpha(z) := 1 - \alpha_1 z - \dots - \alpha_q z^q$$

and the moving average polynomial

$$\beta(z) := \beta_1 + \beta_2 z + \dots + \beta_p z^{p-1}$$

have no common zeros and that $\alpha(z) \neq 0$ on $\{z \in \mathbb{C} \mid |z| \leq 1\}$. Let $(\epsilon_n)_{n \in \mathbb{Z}}$ be an i.i.d. sequence with $\mathbb{E}(\epsilon_1) = 0$ and $\text{Var}(\epsilon_1) = 1$, and let $f(\cdot)$ be as in 3.3. Then $(X_n)_{n \in \mathbb{Z}}$, where $X_n = \sigma_n \epsilon_n$ and

$$\log(\sigma_n^2) = \mu + \sum_{k=1}^p \beta_k f(\epsilon_{n-k}) + \sum_{k=1}^q \alpha_k \log(\sigma_{n-k}^2)$$

is called an EGARCH(p,q) process.

3.2 Exponential COGARCH

The goal of this section is to construct a continuous time analogue of the discrete time EGARCH(p,q) process. Therefore we will use the idea of Klüppelberg et al. (2004) to replace the noise variables ϵ_n by the increments of a Lévy process $L = (L_t)_{t \geq 0}$. We consider univariate zero mean Lévy processes L with jumps $\Delta L_t := L_t - L_{t-}$ defined on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$, which is assumed to satisfy the usual conditions. Since a zero mean implies that $\gamma_L = -\int_{|x| \geq 1} x \nu_L(dx)$ the corresponding Lévy symbol is of the form

$$\psi_L(u) = -\sigma_L^2 \frac{u^2}{2} + \int_{\mathbb{R}} (e^{iux} - 1 - iux) \nu_L(dx),$$

and from (1.14) we know that the Lévy-Itô decomposition of L is

$$L_t = B_t + \int_{\mathbb{R} - \{0\}} x \tilde{N}_L(t, dx), \quad t \geq 0. \quad (3.4)$$

Now we define the *exponential continuous time GARCH(p,q) process* by specifying the log-volatility process as a continuous time ARMA(q,p-1) process, henceforth called CARMA(q,p-1) process (see Chapter 1.4 or Brockwell and Marquardt (2005) for details on CARMA processes). The driving noise process of the CARMA(q,p-1) process will be defined similarly to (3.3).

Definition 3.2.1 Let $L = (L_t)_{t \geq 0}$ be a zero mean Lévy process with Lévy measure ν_L such that $\int_{|x| \geq 1} x^2 \nu_L(dx) < \infty$. Then we define the exponential COGARCH(p, q) process G , shortly ECOGARCH(p, q), as the stochastic process satisfying,

$$dG_t := \sigma_t dL_t, \quad t > 0, \quad G_0 = 0, \quad (3.5)$$

where the log-volatility process $\log(\sigma^2) = (\log(\sigma_t^2))_{t \geq 0}$ is a CARMA($q, p-1$) process, $1 \leq p \leq q$, with mean $\mu \in \mathbb{R}$ and state space representation

$$\log(\sigma_t^2) := \mu + \mathbf{b}^T \mathbf{X}_{t-}, \quad t > 0, \quad \log(\sigma_0^2) = \mu + \mathbf{b}^T \mathbf{X}_0 \quad (3.6)$$

$$d\mathbf{X}_t = \mathcal{A}\mathbf{X}_t + \mathbf{1}_q dM_t, \quad t > 0 \quad (3.7)$$

where $\mathbf{X}_0 \in \mathbb{R}^q$ is independent of the driving Lévy process L and

$$M_t := \int_{\mathbb{R} - \{0\}} h(x) \tilde{N}_L(t, dx), \quad t > 0, \quad (3.8)$$

is a zero mean Lévy process (see Remark 3.2.2) with

$$h(x) := \theta x + \gamma|x|$$

and parameters $\theta, \gamma \in \mathbb{R}$. The $q \times q$ matrix \mathcal{A} and the vectors $\mathbf{b} \in \mathbb{R}^q$ and $\mathbf{1}_q \in \mathbb{R}^q$ are defined by

$$\mathcal{A} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ -a_q & -a_{q-1} & -a_{q-2} & \cdots & -a_1 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_{q-1} \\ b_q \end{bmatrix}, \quad \mathbf{1}_q = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}$$

with coefficients $a_1, \dots, a_q, b_1, \dots, b_p \in \mathbb{R}$, where $a_q \neq 0, b_p \neq 0$, and $b_{p+1} = \dots = b_q = 0$.

One has to observe that (3.6) implies independence between σ_t^2 and the jump ΔL_t at time t , i.e. σ^2 is left continuous. Returns over a time interval of length $r > 0$ are described by the increments of G

$$G_t^{(r)} := G_t - G_{t-r} = \int_{(t-r, t]} \sigma_s dL_s, \quad t \geq r > 0. \quad (3.9)$$

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 $(G_{ir}^{(r)})_{i \in \mathbb{N}}$.

In the rest of the chapter the following terminology will be used:

G	log-price process
$G^{(r)}$	log-return process
σ^2	volatility process
$\log(\sigma^2)$	log-volatility process.

Remark 3.2.2 (i) The process M defined by (3.8) is by construction a process with independent and stationary increments and by Theorem 4.3.4 in Applebaum (2004) well defined if

$$\int_{\mathbb{R}} |h(x)|^2 \nu_L(dx) < \infty. \quad (3.10)$$

Condition (3.10) is satisfied since ν_L is a Lévy measure and L has finite variance. By equation (2.9) of Applebaum (2004) the characteristic function of M at time $t \geq 0$ is given by

$$\begin{aligned} \mathbb{E}(e^{iuM_t}) &= \exp\left(t \int_{\mathbb{R}} [e^{iux} - 1 - iux] \nu_M(dx)\right) \\ &= \exp\left(t \left\{ iu\gamma_M + \int_{\mathbb{R}} [e^{iux} - 1 - iux\chi_{(-1,1)}(x)] \nu_M(dx) \right\}\right) \\ &=: \exp(t\psi_M(u)), \end{aligned}$$

where $\nu_M := \nu_L \circ h^{-1}$ is the Lévy measure of M and $\gamma_M := -\int_{|x|>1} x\nu_M(dx)$. The precise form of ν_M depends on the sign and size of θ and γ and is given in the following:

$$\nu_M((-\infty, -x]) = \begin{cases} \nu_L([-\frac{x}{\theta+\gamma}, \infty)) + \nu_L((-\infty, -\frac{x}{\theta-\gamma}]), & \gamma < \theta < -\gamma \\ \nu_L((-\infty, -\frac{x}{\theta-\gamma}]), & -\theta < \gamma < \theta \\ \nu_L([-\frac{x}{\theta+\gamma}, \infty)), & \theta < \gamma < -\theta \\ 0 & -\gamma < \theta < \gamma \end{cases}$$

and

$$\nu_M([x, \infty)) = \begin{cases} \nu_L([\frac{x}{\theta+\gamma}, \infty)) + \nu_L((-\infty, \frac{x}{\theta-\gamma}]), & -\gamma < \theta < \gamma \\ \nu_L((-\infty, \frac{x}{\theta-\gamma}]), & \theta < \gamma < -\theta \\ \nu_L([\frac{x}{\theta+\gamma}, \infty)), & -\theta < \gamma < \theta \\ 0 & \gamma < \theta < -\gamma \end{cases}$$

for each $x > 0$. One recognises that for $\theta + \gamma < 0 \vee \theta - \gamma > 0$ M is a spectrally negative Lévy process, i.e. M has only negative jumps, and for $\theta + \gamma > 0 \vee \theta - \gamma < 0$

M is a spectrally positive Lévy process. Therefore M has the characteristic triplet $(\gamma_M, 0, \nu_M)$.

(ii) The model can of course also be defined for a different choice of h , as long as condition (3.10) is satisfied.

(iii) In case the jump part of L is of finite variation M is a Lévy process of finite variation with Lévy-Itô decomposition

$$M_t := \sum_{0 \leq s \leq t} [\theta \Delta L_s + \gamma |\Delta L_s|] - Ct, \quad t > 0,$$

where $C := \gamma \int_{\mathbb{R}} |x| \nu_L(dx)$.

Proposition 3.2.3 Let σ^2 and G be as in Definition 3.2.1, with θ and γ not both equal to zero. If the eigenvalues of \mathcal{A} all have negative real parts and \mathbf{X}_0 has the same distribution as $\int_0^\infty e^{\mathcal{A}u} \mathbf{1}_q dM_u$, then $\log(\sigma^2)$ and σ^2 are strictly stationary.

Proof: The strict stationarity of $\log(\sigma^2)$ follows from Proposition 2 in Brockwell and Marquardt (2005), since it is a CARMA($q, p-1$) process. Since strict stationarity is invariant under continuous transformations, σ^2 also has this property. \square

Remark 3.2.4 The solution of the continuous time state space model (3.6) and (3.7) has the representation

$$\log(\sigma_{t+}^2) = \mu + \mathbf{b}^T e^{\mathcal{A}t} \mathbf{X}_0 + \int_0^t \mathbf{b}^T e^{\mathcal{A}(t-u)} \mathbf{1}_q dM_u, \quad t > 0.$$

Using the two-sided Lévy process L^* , defined in (1.17), instead of L in (3.8) we get an extension M^* of M . In the following we will write for simplicity L and M instead of L^* and M^* . In the strictly stationary case the log-volatility process can be defined on the whole real line

$$\log(\sigma_{t+}^2) = \mu + \int_{-\infty}^t g(t-u) dM_u, \quad t \in \mathbb{R}, \quad (3.11)$$

with kernel function

$$g(t) = \mathbf{b}^T e^{\mathcal{A}t} \mathbf{1}_q \chi_{(0, \infty)}(t) \quad (3.12)$$

(see also Chapter 1.4).

From (3.9) it follows directly that the increments $G_{\cdot}^{(r)} = \int_{(\cdot-r, \cdot]} \sigma_s dL_s$ of G are stationary if the volatility σ^2 is stationary, since the increments of L are stationary and independent by definition.

Corollary 3.2.5

If σ^2 is strictly stationary, then G has strictly stationary increments.

Remark 3.2.6 (i) If $q \geq p + 1$ the log-volatility process is $(q - p - 1)$ times differentiable, which follows from the state space representation (3.7) and (3.6) of $\log(\sigma^2)$, and hence the volatility process has continuous sample path. In particular the volatility will only contain jumps for $p = q$.

(ii) The volatility of the ECOGARCH(p, q) process is positive by definition. Therefore the parameters do not need to satisfy any constraints to assure positivity of the volatility. This is not the case for the COGARCH(p, q) model. For higher order COGARCH(p, q) processes these condition become quite difficult to check (see Theorem 5.1 in Brockwell et al. (2006)).

Example 3.2.7 As a first illustrative example we consider an ECOGARCH(1,1) process driven by a Lévy process L with Lévy symbol

$$\psi_L(u) = -\frac{u^2}{2} + \int_{\mathbb{R}} (e^{iux} - 1) \lambda \Phi_{0,1/\lambda}(dx),$$

where $\Phi_{0,1/\lambda}(\cdot)$ is the distribution function of a normal distribution with mean 0 and variance $1/\lambda$. This means that L is the sum of a standard Brownian motion W and the compound Poisson process $J_t = \sum_{k=1}^{N_t} Z_k$, $J_{-t} = \sum_{k=1}^{-N_{-t}} Z_{-k}$, $t \geq 0$, where $(N_t)_{t \in \mathbb{R}}$ is an independent Poisson process with intensity $\lambda > 0$ and jump times $(T_k)_{k \in \mathbb{Z}}$. The Poisson process N is also independent from the i.i.d. sequence of jump sizes $(Z_k)_{k \in \mathbb{Z}}$, with $Z_1 \sim N(0, 1/\lambda)$. The Lévy process M is in this case given by the following expression

$$M_t = \sum_{k=1}^{N_t} [\theta Z_k + \gamma |Z_k|] - Ct, \quad t > 0,$$

with $C = \gamma \int_{\mathbb{R}} |x| \lambda \Phi_{0,1/\lambda}(dx) = \sqrt{\frac{2\lambda}{\pi}} \gamma$. M_{-t} , $t \geq 0$ is defined analogously. If we just consider the case that $\theta < -\gamma < 0$ then the Lévy measure ν_M of M is defined by

$$\nu_M((-\infty, -x]) = \lambda \Phi_{0,1/\lambda}\left(\left[-\frac{x}{\theta + \gamma}, \infty\right)\right), \quad x > 0,$$

on the negative half real line and by

$$\nu_M([x, \infty)) = \lambda \Phi_{0,1/\lambda}\left(\left(-\infty, \frac{x}{\theta - \gamma}\right]\right), \quad x > 0,$$

on the positive half real line. In the top row of Figure 3.1 a simulated sample path of the compound Poisson process J , with $N(0, 1/2)$ distributed jumps, can be seen over three time scales. The corresponding Lévy process M , with parameters $\theta = -0.2$ and $\gamma = 0.1$, can be seen in the bottom row. Over all three time intervals one can recognise the desired asymmetry for this set of parameters. If J jumps up, then M jumps down and vice versa. If J does not move, then one observes the downwards

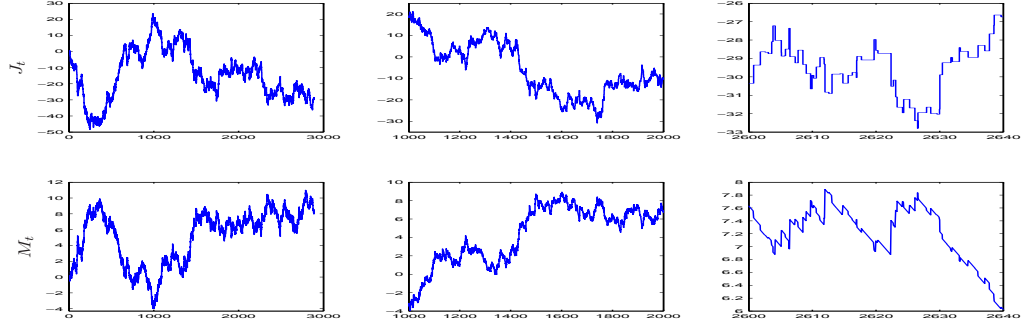


Figure 3.1: Simulated sample pathes of J (top row) and M (bottom row), with parameters $\theta = -0.2$ and $\gamma = 0.1$, over three different time scales.

drift of M , which can be seen on the right hand side of Figure 3.1.

The log-volatility process is then of the form

$$\begin{aligned} \log(\sigma_{t+}^2) &= \mu + \int_{-\infty}^t b_1 e^{-a_1(t-s)} dM_s \\ &= \mu + \sum_{\substack{k=-\infty \\ k \neq 0}}^{N_t} b_1 e^{-a_1(t-T_k)} [\theta Z_k + \gamma |Z_k|] - C \frac{b_1}{a_1}, \quad t > 0 \end{aligned}$$

and the log-price process is given by

$$G_t = \int_0^t \sigma_s dW_s + \sum_{k=1}^{N_t} \sigma_{T_k} Z_k, \quad t > 0, \quad G_0 = 0.$$

with jump times T_k , $k \in \mathbb{N}$.

The simulation of a sample path of the log-price process G and the log-volatility process $\log(\sigma^2)$ over a time interval $[0, T]$ is done in the following steps.

- (1) Choose observation times $0 = t_0 < t_1 < \dots < t_n \leq T$, possibly random.
- (2) Simulate the jump times (T_k) , $k = 1, \dots, n_T$, with $n_T := \max\{k \in \mathbb{N} : T_k \leq T\}$, of the compound Poisson process J .
- (3) Approximate the state process (3.7) of the log-volatility by a stochastic Euler scheme.

(4) Compute an approximation \widehat{G} via the recursion

$$\widehat{G}_{t_i} = \widehat{G}_{t_{i-1}} + \sigma_{t_{i-1}} \widetilde{W}_i + \sum_{k=N_{t_{i-1}}+1}^{N_{t_i}} \sqrt{\exp\{\mu + \mathbf{b}^T \widehat{\mathbf{X}}_{T_k-}\}} Z_k,$$

where $\widetilde{W}_i \sim N(0, t_i - t_{i-1})$ and $\widehat{\mathbf{X}}_{T_k-}$ is the Euler approximation without the jump ΔM_{T_k} .

In Figure 3.2 the results of the above simulation procedure are shown. The jump rate λ is now chosen to be $1/4$, which implies a variance of the jump sizes Z_i of 4. For exponentially distributed interarrival times $\Delta t_i := t_i - t_{i-1} \sim \text{expo}(1)$ the sample path of the log-price G , the return process $G^{(\Delta t)}$ and the volatility process σ^2 are displayed in the first three rows of Figure 3.2. The sample path of the driving Lévy process L is shown in the last row. From the plots of the return and volatility process we see the negative correlation between the two processes. We recognise increases in the volatility after large negative returns.

Mixing properties (see Doukhan (1994) for a comprehensive treatment of mixing properties) are useful for a number of applications. In particular for asymptotic statistics, since central limit theorems exist for mixing processes. Thus we will derive mixing properties of the strictly stationary volatility process and the return process over equidistant time intervals. We will need the concept of strong mixing (see Definition 1.4.6) and $\tilde{\alpha}$ -mixing (see Definition 2.3.1).

From Lemma 2.3.2 we know that any α -mixing property is equivalent to the corresponding $\tilde{\alpha}$ -mixing property. This fact will be used in Theorem 3.2.9 to show that $(G_{nr}^{(r)})_{n \in \mathbb{N}}$ is α -mixing. The strong mixing property with exponential rate of the log-volatility and volatility process is the subject of the next proposition. Thereby strongly mixing with exponential rate (exponentially α -mixing) means that $\alpha(t)$ decays to zero exponentially fast for $t \rightarrow \infty$.

Proposition 3.2.8 *Let $\log(\sigma^2)$ be defined by (3.6) and (3.7) with θ and γ not both equal to zero. Assume that $\mathbb{E}(L_1^2) < \infty$, the eigenvalues of A all have negative real parts and \mathbf{X}_0 has the same distribution as $\int_0^\infty e^{Au} \mathbf{1}_q dM_u$, so that $\log(\sigma^2)$ and σ^2 are strictly stationary. Then there exist constants $K > 0$ and $a > 0$ such that*

$$\alpha_{\log(\sigma^2)}(t) \leq K \cdot e^{-at} \quad \text{and} \quad \alpha_{\sigma^2}(t) \leq K \cdot e^{-at}, \quad \text{as } t \rightarrow \infty, \quad (3.13)$$

where $\alpha_{\log(\sigma^2)}(t)$ and $\alpha_{\sigma^2}(t)$ are the α -mixing coefficients of the log-volatility and volatility process, respectively.

Proof: The log-volatility process is a CARMA($q, p-1$) process and hence the mixing property follows from Proposition 1.4.7. The property of α -mixing is invariant under continuous transformations, which implies that σ^2 also has this property. \square

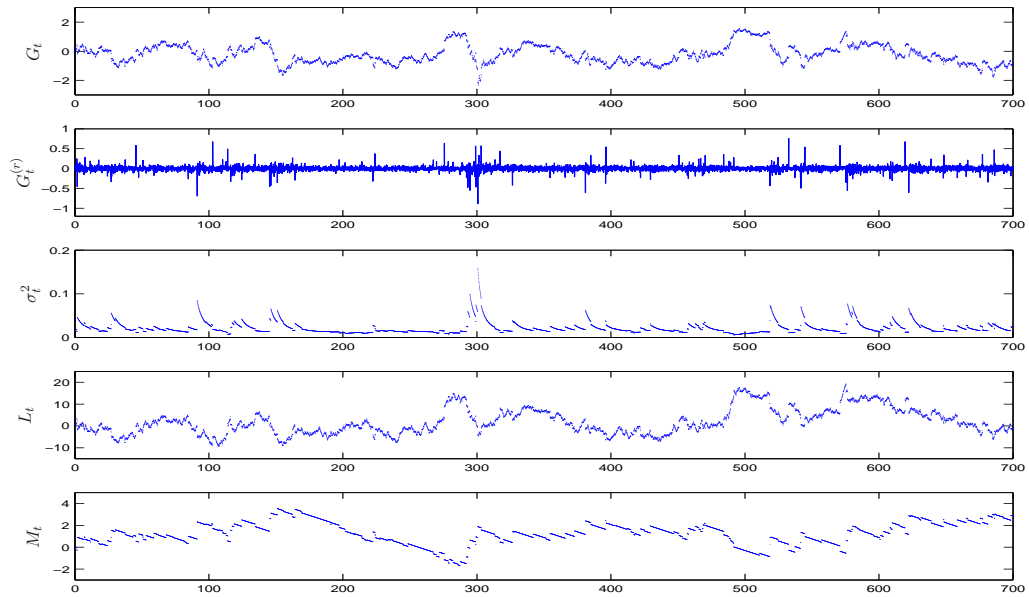


Figure 3.2: Observations of the log-price process G_t (top row), the return process $G_t^{(r)}$ (second row), the volatility process σ_t^2 (third row), with parameters $b_1 = 1$, $a_1 = 0.1$, $\mu = -4$, $\theta = -0.2$ and $\gamma = 0.1$ and the driving Lévy process L_t (last row) in the time interval $0, 700]$.

Theorem 3.2.9 *Assume that L is a Lévy process with finite variance. Let the volatility process σ^2 be strictly stationary and strongly mixing under \mathbb{P} . Then the discrete time process $(G_{ir}^{(r)})_{i \in \mathbb{N}}$,*

$$G_{ir}^{(r)} := G_{ir} - G_{(i-1)r} = \int_{((i-1)r, ir]} \sigma_s dL_s, i \in \mathbb{N},$$

is strongly mixing with geometric rate and since strict stationarity of σ^2 implies strict stationarity of $(G_{ir}^{(r)})_{i \in \mathbb{N}}$ also ergodic.

Proof: Consider the q -dimensional OU process \mathbf{V} defined in the proof of Proposition 1.4.7 corresponding to the log-volatility process. Equation (1.43) shows that \mathbf{V} is a Markov process. The process \mathbf{V} is also strictly stationary and $\tilde{\alpha}$ -mixing, since α -mixing and $\tilde{\alpha}$ -mixing are equivalent. The same is true for the q -dimensional process $\boldsymbol{\sigma}^2 := (\exp(V^1), \dots, \exp(V^q))$, since both mixing properties are invariant under

continuous transformations. Now define the discrete time q -dimensional process

$$\mathbf{G}_{ir}^{(r)} = \int_{((i-1)r, ir]} \sqrt{\sigma_s^2} dL_s, \quad i \in \mathbb{N}, \quad (3.14)$$

which is also strictly stationary under \mathbb{P} . Here we should mention that the integration of vectors in (3.14) is understood componentwise.

Define the σ -algebra $\mathcal{F}_I^{dL} := \sigma(L_t - L_s : s, t \in I)$ for $I \subset \mathbb{R}$. Further we denote the σ -algebra generated by the volatility process and the increments of L over the interval $[0, t]$ by $\mathcal{F}_{[0,t]}^{\sigma^2, dL} := \sigma((\sigma_s^2)_{s \in [0,t]}) \vee \mathcal{F}_{[0,t]}^{dL}$.

From $\mathcal{F}_{\{1,2,\dots,l\}}^{\mathbf{G}^{(r)}} \subset \mathcal{F}_{[0,lr]}^{\sigma^2, dL}$ and $\mathcal{F}_{\{k+l, k+l+1, \dots\}}^{\mathbf{G}^{(r)}} \subset \mathcal{F}_{[(k+l-1)r, \infty)}^{\sigma^2, dL}$ we get that

$$\begin{aligned} \tilde{\alpha}_{\mathbf{G}^{(r)}}(k) &= \sup \left\{ \|\mathbb{E}(f | \mathcal{F}_{\{1,2,\dots,l\}}^{\mathbf{G}^{(r)}}) - \mathbb{E}(f)\|_{L^1(P)} : \right. \\ &\quad \left. f \in b\mathcal{F}_{\{k+l, k+l+1, \dots\}}^{\mathbf{G}^{(r)}}, l \in \mathbb{N}, \|f\|_\infty \leq 1 \right\} \\ &\leq \sup \left\{ \|\mathbb{E}(f | \mathcal{F}_{[0,lr]}^{\sigma^2, dL}) - \mathbb{E}(f)\|_{L^1(P)} : f \in b\mathcal{F}_{[(k+l-1)r, \infty)}^{\sigma^2, dL}, l \in \mathbb{N}, \|f\|_\infty \leq 1 \right\} \\ &= \sup \left\{ \|\mathbb{E}(f | \mathcal{F}_{[0,lr]}^{\sigma^2, dL}) - \mathbb{E}(f)\|_{L^1(P)} : \right. \\ &\quad \left. f \in b\mathcal{F}_{(k+l-1)r}^{\sigma^2} \vee b\mathcal{F}_{[(k+l-1)r, \infty)}^{dL}, l \in \mathbb{N}, \|f\|_\infty \leq 1 \right\} \\ &= \sup \left\{ \|\mathbb{E}(f | \mathcal{F}_{[0,lr]}^{\sigma^2}) - \mathbb{E}(f)\|_{L^1(P)} : \right. \\ &\quad \left. f \in b\mathcal{F}_{(k+l-1)r}^{\sigma^2} \vee b\mathcal{F}_{[(k+l-1)r, \infty)}^{dL}, l \in \mathbb{N}, \|f\|_\infty \leq 1 \right\} \\ &= \sup \left\{ \|\mathbb{E}(f | \mathcal{F}_{[0,lr]}^{\sigma^2}) - \mathbb{E}(f)\|_{L^1(P)} : \right. \\ &\quad \left. f \in b\mathcal{F}_{(k+l-1)r}^{\sigma^2} \vee b\mathcal{F}_{[(k+l-1)r, \infty)}^{dM}, l \in \mathbb{N}, \|f\|_\infty \leq 1 \right\} \\ &= \tilde{\alpha}_{\sigma^2}((k-1)r) \end{aligned}$$

for any $k \in \mathbb{N}$. The second equality is due to the fact that the driving process of σ^2 is defined through the jumps of L . $\mathcal{F}_{[0,lr]}^{dL}$ and $b\mathcal{F}_{(k+l-1)r}^{\sigma^2} \vee b\mathcal{F}_{[(k+l-1)r, \infty)}^{dL}$ are conditionally independent given $\mathcal{F}_{[0,lr]}^{\sigma^2}$, which is due to the Markov property of σ^2 and the independence between $\mathcal{F}_{[0,lr]}^{dL}$ and $b\mathcal{F}_{[(k+l-1)r, \infty)}^{dL}$. This shows the third equality. The fourth one follows from the independence of $b\mathcal{F}_{[(k+l-1)r, \infty)}^{dL}$ and $b\mathcal{F}_{[(k+l-1)r, \infty)}^{dM}$ from $\mathcal{F}_{[0,lr]}^{\sigma^2}$ and again the Markov property of σ^2 , whereas the last one is due to the fact that M is the driving process of σ^2 .

Therefore $(\mathbf{G}_{nr}^{(r)})_{n \in \mathbb{N}}$ is $\tilde{\alpha}$ -mixing and by Lemma 2.3.2 also strongly mixing. The strict stationarity of $(\mathbf{G}_{nr}^{(r)})_{n \in \mathbb{N}}$ implies that it is also ergodic. Since both properties have to hold componentwise, we have shown that $(G_{nr}^{(r)})_{n \in \mathbb{N}}$ as the first component of $(\mathbf{G}_{nr}^{(r)})_{n \in \mathbb{N}}$ is strongly mixing and ergodic. From Lemma 2.3.2 and (3.13) we further

get that

$$\alpha_{G^{(r)}}(k) \leq 6\alpha_{\sigma^2}((k-1)r) \leq 6Ke^{-a(k-1)r},$$

as $k \rightarrow \infty$. Hence $(G_{nr}^{(r)})_{n \in \mathbb{N}}$ is strongly mixing with exponential rate and ergodicity follows from the strict stationarity. \square

3.3 Second order properties of the volatility process

In this section we derive moments and the autocovariance function of the volatility process σ^2 . Since it is a non-linear transformation of a CARMA($q, p-1$) process, we will first recall the moment structure and conditions for weak stationarity of a CARMA($q, p-1$) process.

Proposition 3.3.1 *If \mathbf{X}_0 has the same mean vector and covariance matrix as $\int_0^\infty e^{\mathcal{A}u} \mathbf{1}_q dM_u$, then $\log(\sigma^2)$ is weakly stationary. In the weakly stationary case the mean and autocovariance function of $\log(\sigma^2)$ are given by*

$$\mathbb{E}(\log(\sigma_t^2)) = \mu \quad \text{and} \quad \text{Cov}(\log(\sigma_{t+h}^2), \log(\sigma_t^2)) = \mathbb{E}(M_1^2) \mathbf{b}^T e^{\mathcal{A}h} \mathbf{\Sigma} \mathbf{b}, \quad h, t \geq 0,$$

where $\mathbf{\Sigma} := \int_0^\infty e^{\mathcal{A}s} \mathbf{1}_q \mathbf{1}_q^T e^{\mathcal{A}^T s} ds$.

The results follow immediately from the fact that the log-volatility process is a CARMA($q, p-1$) process, cf. Proposition 1.4.2.

The moments of the strictly stationary volatility process are exponential moments of the limit distribution of the log-volatility process. In Proposition 3.2.3 we gave conditions for the existence of a stationary distribution F of the log-volatility process. In the following proposition we want to further characterise this distribution.

Proposition 3.3.2 *Let $(\gamma_M, 0, \nu_M)$ be the characteristic triplet of the Lévy process M , where M is defined in (3.8), and F is the stationary distribution of the log-volatility process. Then F is infinitely divisible with characteristic triplet $(\gamma_\infty, 0, \nu_\infty)$, where*

$$\begin{aligned} \gamma_\infty &= \mu + \int_0^\infty g(s) \gamma_M ds + \int_0^\infty \int_{\mathbb{R}} g(s) x [\chi_{(-1,1)}(g(s)x) - \chi_{(-1,1)}(x)] \nu_M(dx) ds \\ \nu_\infty(B) &= \int_0^\infty \int_{\mathbb{R}} \chi_B(g(s)x) \nu_M(dx) ds, \quad B \in \mathcal{B}(\mathbb{R}), \end{aligned}$$

with $g(s) = \mathbf{b}^T e^{\mathbf{A}s} \mathbf{1}_q \chi_{(0,\infty)}(s)$. Let $\log(\sigma_\infty^2)$ be a random variable with distribution F , then the characteristic function of $\log(\sigma_\infty^2)$ can be written as

$$\begin{aligned} \mathbb{E}(e^{iu \log(\sigma_\infty^2)}) &= \exp\{\psi_\infty(u)\} \\ &= \exp\left\{iu\gamma_\infty + \int_{\mathbb{R}} (e^{iux} - 1 - iux\chi_{(-1,1)}(x)) \nu_\infty(dx)\right\} \\ &= \exp\left\{iu\mu + \int_0^\infty \psi_M(g(s)u) ds\right\}. \end{aligned}$$

Proof: In the strictly stationary case the log-volatility process is the continuous time moving average process (3.11). Since M has finite variance the kernel g and the driving Lévy process M satisfy the conditions in Theorem 2.7 in Rajput and Rosiński (1989). Therefore the stationary distribution F of the log-volatility process is infinitely divisible with characteristic triplet $(\gamma_\infty, 0, \nu_\infty)$. The third equality in the representation of the characteristic function follows from the fact that

$$\int_{\mathbb{R}} h(y) \nu_\infty(dy) = \int_0^\infty \int_{\mathbb{R}} h(g(s)x) \nu_M(dx) ds$$

for all h integrable with respect to ν_∞ (see e.g. Theorem 41 in Rocha-Arteaga and Sato (2003)). \square

Since the stationary distribution F of the log-volatility process is infinitely divisible, we can now apply Theorem 25.17 of Sato (1999) to calculate the exponential moments of $\log(\sigma_\infty^2)$, i.e. the moments of σ_∞^2 , in the next Proposition.

Proposition 3.3.3 *Let F be the stationary distribution of $\log(\sigma^2)$ with characteristic triplet $(\gamma_\infty, 0, \nu_\infty)$. Then the k -th moment of σ_t^2 is finite, if*

$$k \in K_\infty = \left\{s \in \mathbb{R} : \int_{|x|>1} e^{sx} \nu_\infty(dx) < \infty\right\}.$$

In this case

$$\Psi_\infty(k) := \gamma_\infty k + \int_{\mathbb{R}} \left(e^{kx} - 1 - kx\chi_{(-1,1)}(x)\right) \nu_\infty(dx), \quad (3.15)$$

is well defined and

$$\mathbb{E}(\sigma_t^{2k}) = e^{\Psi_\infty(k)}, \quad \forall t \geq 0. \quad (3.16)$$

Proof: The k -th exponential moment of a Lévy process $(X_t)_{t \geq 0}$ is computed in Theorem 25.17 of Sato (1999). Hence we can apply the Theorem for a Lévy process X with infinitely divisible distribution F at time one to get the k -th exponential moment of $\log(\sigma_t^2)$. It is then given by

$$\mathbb{E}(\exp(\log(\sigma_t^2))^k) = e^{\Psi_\infty(k)}, \quad \forall t \geq 0,$$

with $\Psi_\infty(k) = \gamma_\infty k + \int_{\mathbb{R}} (e^{kx} - 1 - kx\chi_{(-1,1)}(x)) \nu_\infty(dx)$ (see equation (25.11) in Sato (1999)). \square

Proposition 3.3.4 *Let $\log(\sigma_t^2)$ be the strictly stationary solution of (3.6) and (3.7). Assume that $\mathbb{E}(\sigma_t^4) < \infty$ for all $t \geq 0$. Let $\Psi_\infty^h(1)$ and $\Psi^h(1)$ be defined by (3.15) with kernel function g replaced by*

$$g_\infty^h(s) = \mathbf{b}^T (I_q + e^{\mathbf{A}h}) e^{\mathbf{A}s} \mathbf{1}_q \text{ and } g^h(s) = \mathbf{b}^T e^{\mathbf{A}s} \mathbf{1}_q \chi_{(0,h)}(s),$$

respectively. Then the autocovariance function of σ^2 is given by the following expression

$$\text{Cov}(\sigma_{t+h}^2, \sigma_t^2) = e^{\Psi_\infty^h(1)} e^{\Psi^h(1)} - e^{2\Psi_\infty(1)}, \quad h > 0, t \geq 0. \quad (3.17)$$

Proof: Let $\mathcal{F}_t^M = \sigma(M_s, -\infty < s \leq t)$ be the σ -algebra generated by the Lévy process M up to time t , then

$$\begin{aligned} & \mathbb{E}(\sigma_{t+h}^2 | \mathcal{F}_t^M) \\ &= \mathbb{E} \left(\exp \left\{ \int_{-\infty}^{t+h} g(t+h-s) dM_s \right\} \middle| \mathcal{F}_t^M \right) \\ &= \exp \left\{ \mu + \int_{-\infty}^t \mathbf{b}^T e^{\mathbf{A}h} e^{\mathbf{A}(t-s)} \mathbf{1}_q dM_s \right\} \mathbb{E} \left(\exp \left\{ \int_t^{t+h} g(t+h-s) dM_s \right\} \right). \end{aligned}$$

Therefore we get

$$\begin{aligned} \mathbb{E}(\sigma_{t+h}^2 \sigma_t^2) &= \mathbb{E}(\mathbb{E}(\sigma_{t+h}^2 \sigma_t^2 | \mathcal{F}_t^M)) = \mathbb{E}(\sigma_t^2 \mathbb{E}(\sigma_{t+h}^2 | \mathcal{F}_t^M)) \\ &= \mathbb{E} \left(\sigma_t^2 \exp \left\{ \mu + \int_{-\infty}^t \mathbf{b}^T e^{\mathbf{A}h} e^{\mathbf{A}(t-s)} \mathbf{1}_q dM_s \right\} \mathbb{E} \left(e^{\int_t^{t+h} g(t+h-s) dM_s} \right) \right) \\ &= \mathbb{E} \left(\exp \left\{ 2\mu + \int_{-\infty}^t \mathbf{b}^T (I_q + e^{\mathbf{A}h}) e^{\mathbf{A}(t-s)} \mathbf{1}_q dM_s \right\} \right) \mathbb{E} \left(e^{\int_0^h g(s) dM_s} \right) \\ &= \mathbb{E} \left(\exp \left\{ \mu + \int_0^\infty \mathbf{b}^T (I_q + e^{\mathbf{A}h}) e^{\mathbf{A}s} \mathbf{1}_q dM_s \right\} \right) \\ & \quad \times \mathbb{E} \left(\exp \left\{ \mu + \int_0^\infty \mathbf{b}^T e^{\mathbf{A}s} \mathbf{1}_q \chi_{(0,h)}(s) dM_s \right\} \right) \\ &= e^{\Psi_\infty^h(1)} e^{\Psi^h(1)}, \end{aligned}$$

where the last equality follows from (3.16) when we substitute the kernel g in (3.11) by $g_\infty^h(s)$ and g^h , respectively. This together with (3.16) yields (3.17). \square

Remark 3.3.5 *In Proposition 3.2.8 we have seen that the volatility process is strongly mixing with exponential rate. A consequence of this property (see e.g. Chapter 1.2.2 in Doukhan (1994)) is that*

$$|\text{Cov}(\sigma_{t+h}^2, \sigma_t^2)| \leq K_\sigma \cdot e^{-ah}, \quad \forall h > 0, \quad (3.18)$$

for some constant $K_\sigma > 0$. In particular this means that the autocovariance function of the volatility process will decay to zero at an exponential rate, thus σ^2 is a short memory process (cf. Remark 1.4.8).

3.4 Second order properties of the return process

In this section we will first derive the moment structure of the *return process*

$$G_t^{(r)} := G_t - G_{t-r} = \int_{(t-r, t]} \sigma_s dL_s, \quad t \geq r > 0.$$

We will only consider the case of a strictly stationary volatility process. Afterwards we discuss the *leverage effect* for our model.

3.4.1 Moments and autocovariance function of the return process

Proposition 3.4.1 *Let L be a Lévy process with $\mathbb{E}(L_1) = 0$ and $\mathbb{E}(L_1^2) < \infty$. Assume that the volatility process σ^2 is strictly stationary with finite mean. Then $\mathbb{E}(G_t^2) < \infty$ for all $t \geq 0$, and for every $t, h \geq r > 0$ it holds*

$$\mathbb{E}G_t^{(r)} = 0 \quad (3.19)$$

$$\mathbb{E}(G_t^{(r)})^2 = e^{\Psi_\infty(1)r} \mathbb{E}(L_1^2) \quad (3.20)$$

$$\mathbb{Cov}(G_t^{(r)}, G_{t+h}^{(r)}) = 0. \quad (3.21)$$

If further $\mathbb{E}(L_1^4) < \infty$ and the volatility process has finite second moment, then $\mathbb{E}(G_t^4) < \infty$ for all $t \geq 0$ and for every $t, h \geq r > 0$ we have

$$\mathbb{Cov}((G_t^{(r)})^2, (G_{t+h}^{(r)})^2) = \mathbb{E}(L_1^2) \int_h^{h+r} \mathbb{Cov}(G_r^2, \sigma_s^2) ds. \quad (3.22)$$

Proof: The proof of (3.19) - (3.21) is analogously to the proof of Proposition 5.1 in Klüppelberg et al. (2004), but we also present it here for completeness. Since G is a square integrable martingale, with mean zero, (3.19) follows immediately and (3.21) follows from the isometry (1.22). Observe that by Itô's product formula (see Example 1.2.2) and $[G, G]_t = \int_0^t \sigma_s^2 d[L, L]_s$ we get

$$G_t^2 = 2 \int_0^t G_{s-} \sigma_s dL_s + \int_0^t \sigma_s^2 d[L, L]_s, \quad t \geq 0, \quad (3.23)$$

where $[L, L]_t = [L, L]_t^c + \sum_{0 \leq s \leq t} (\Delta L_s)^2 = \sigma_L^2 t + \sum_{0 \leq s \leq t} (\Delta L_s)^2$. Application of the compensation formula (see Proposition A.17) now yields

$$\begin{aligned} \mathbb{E}(G_r^2) &= 2\mathbb{E} \int_0^r G_{s-} \sigma_s dL_s + \mathbb{E} \int_0^r \sigma_s^2 d[L, L]_s \\ &= \int_0^r \mathbb{E}(\sigma_s^2) \sigma_L^2 ds + \mathbb{E} \sum_{0 \leq s \leq r} \sigma_s^2 (\Delta L_s)^2 \\ &= e^{\Psi_\infty(1)r} (\sigma_L^2 + \int_{\mathbb{R}} x^2 \nu_L(dx)) \end{aligned}$$

which is finite under the assumptions $\mathbb{E}(L_1^2) < \infty$ and $1 \in K_\infty$ and equal to (3.20).

The martingale property and square integrability of G also implies

$$\mathbb{E}((G_r^{(r)})^2(G_{h+r}^{(r)})^2) = \mathbb{E}(G_r^2(G_{h+r} - G_h)^2) = \mathbb{E}(G_r^2(G_{h+r}^2 - G_h^2)).$$

Using this result, (3.23) and the compensation formula we get

$$\begin{aligned} \mathbb{E}((G_r^{(r)})^2(G_{h+r}^{(r)})^2) &= \mathbb{E}\left(2 \int_h^{h+r} G_r^2 G_{s-} \sigma_s dL_s + \int_h^{h+r} G_r^2 \sigma_s^2 d[L, L]_s\right) \\ &= \mathbb{E}\left(\int_h^{h+r} G_r^2 \sigma_s^2 d[L, L]_s\right) \\ &= \int_h^{h+r} \mathbb{E}(G_r^2 \sigma_s^2) \sigma_L^2 ds + \int_h^{h+r} \mathbb{E}(G_r^2 \sigma_s^2) ds \int_{\mathbb{R}} x^2 \nu_L(dx) \\ &= \mathbb{E}(L_1^2) \int_h^{h+r} \mathbb{E}(G_r^2 \sigma_s^2) ds \end{aligned}$$

Hence the covariance is equal to

$$\begin{aligned} \text{Cov}((G_r^{(r)})^2(G_{h+r}^{(r)})^2) &= \mathbb{E}((G_0^{(r)})^2(G_h^{(r)})^2) - (\mathbb{E}(G_t^{(r)})^2)^2 \\ &= \mathbb{E}(L_1^2) \int_h^{h+r} (\text{Cov}(G_r^2, \sigma_s^2) + \mathbb{E}(G_r^2)\mathbb{E}(\sigma_s^2)) ds - (\mathbb{E}(G_t^{(r)})^2)^2 \\ &= \mathbb{E}(L_1^2) \int_h^{h+r} \text{Cov}(G_r^2, \sigma_s^2) ds. \end{aligned}$$

The covariance is finite if $\mathbb{E}(G_t^4) < \infty$, $\forall t \geq 0$, and this follows with $\mathbb{E}(L_1^4) < \infty$ and $2 \in K_\infty$ analogously as in Proposition 1.5.1. \square

Example 3.4.2 *Let us consider again Example 3.2.7. From 50 000 equidistant observations of the simulated log-price we computed the empirical autocorrelation function of the returns and squared returns. In Figure 3.3 the first 40 lags of both empirical autocorrelation functions are shown. One recognises the GARCH like behaviour of zero correlation of the returns and significant correlation of the squared returns.*

In the next Proposition we want to consider the special case that $p = q = 1$. Under a further assumption on the mapping h , which has not to be the standard choice of Definition 3.2.1, and the Lévy measure ν_L we can express the covariance of the squared returns in terms of the covariance of the *actual volatility*

$$\sigma^2(h) := \int_0^{h+r} \sigma_s^2 ds - \int_0^h \sigma_s^2 ds, \quad h \geq r.$$

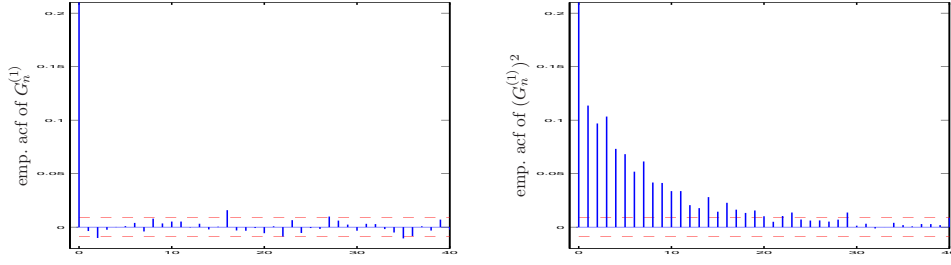


Figure 3.3: The first 40 lags of the empirical autocorrelation function of the return (*left*) and squared return (*right*) process.

Proposition 3.4.3 *Let $p = q = 1$ and assume that the mapping h satisfies (3.10) and additionally $\int_{\mathbb{R}} x \{ \exp(b_1 e^{-a_1 t} h(x)) - 1 \} \nu_L(dx) = 0$ for all $t \geq 0$. Then under the same conditions as in Proposition 3.4.1 we get*

$$\text{Cov}((G_t^{(r)})^2, (G_{t+h}^{(r)})^2) = (\mathbb{E}(L_1^2))^2 \text{Cov}(\sigma^2(h), \sigma^2(0)).$$

Proof: From the proof of Proposition 3.4.1 we know that

$$\begin{aligned} \text{Cov}((G_t^{(r)})^2, (G_{t+h}^{(r)})^2) &= \mathbb{E}(L_1^2) \int_h^{h+r} \mathbb{E}(G_r^2 \sigma_s^2) ds - (r \mathbb{E}(L_1^2) \mathbb{E}(\sigma_1^2))^2 \\ &= \mathbb{E}(L_1^2) \mathbb{E} \left(\int_h^{h+r} \left\{ 2 \int_0^r G_{u-\sigma_u} dL_u + \int_0^r \sigma_u^2 d[L, L]_u \right\} \sigma_s^2 ds \right) - (r \mathbb{E}(L_1^2) \mathbb{E}(\sigma_1^2))^2. \end{aligned}$$

Hence the result follows if we can show that $\mathbb{E}(\int_0^r G_{u-\sigma_u} \sigma_s^2 dL_u) = 0$, for all $s > r$, since

$$\mathbb{E} \left(\int_h^{h+r} \int_0^r \sigma_u^2 \sigma_s^2 d[L, L]_u ds \right) = (\mathbb{E}(L_1^2)) \mathbb{E}(\sigma^2(h) \sigma^2(0)).$$

Define $Y_t := \int_0^t G_{u-\sigma_u} dL_u$, $t \geq 0$, and $\tilde{\sigma}_{s,t}^2 := \exp(b_1 e^{-a_1(s-t)} X_t)$ for all $t \in [0, s)$, where X is the state process (3.7). Then

$$\mathbb{E}(Y_r \sigma_s^2) = \mathbb{E}(Y_r \tilde{\sigma}_{s,r+}^2) \mathbb{E}(\exp(\int_r^s b_1 e^{-a_1(s-u)} dM_u))$$

and we have to show $\mathbb{E}(Y_r \tilde{\sigma}_{s,r+}^2) = 0$. An application of Itô's formula (see Theorem 1.2.1) and substituting from (3.7) yields

$$\begin{aligned} \tilde{\sigma}_{s,r+}^2 &= \tilde{\sigma}_{s,0}^2 + \int_0^r \int_{\mathbb{R}-\{0\}} \tilde{\sigma}_{s,u}^2 b_1 e^{-a_1(s-u)} h(x) \tilde{N}_L(du, dx) \\ &\quad + \int_0^r \int_{\mathbb{R}-\{0\}} \tilde{\sigma}_{s,u}^2 \left\{ \exp(b_1 e^{-a_1(s-u)} h(x)) - 1 - b_1 e^{-a_1(s-u)} h(x) \right\} N_L(du, dx). \end{aligned}$$

Using this representation, integration by parts and taking expectation gives

$$\begin{aligned} \mathbb{E}(Y_r \tilde{\sigma}_{s,r+}^2) &= \\ & \int_0^r \int_{\mathbb{R}} \mathbb{E}(Y_u \tilde{\sigma}_{s,u}^2) \left\{ \exp(b_1 e^{-a_1(s-u)} h(x)) - 1 - b_1 e^{-a_1(s-u)} h(x) \right\} \nu_L(dx) du \\ & + \int_0^r \int_{\mathbb{R}} \mathbb{E}(\tilde{\sigma}_{s,u}^2 G_{u-\sigma_u}) x \left\{ \exp(b_1 e^{-a_1(s-u)} h(x)) - 1 \right\} \nu_L(dx) du \\ & = \int_0^r \int_{\mathbb{R}} \mathbb{E}(Y_u \tilde{\sigma}_{s,u+}^2) \left\{ \exp(b_1 e^{-a_1(s-u)} h(x)) - 1 - b_1 e^{-a_1(s-u)} h(x) \right\} \nu_L(dx) du, \end{aligned}$$

where we used $\int_{\mathbb{R}} x \left\{ \exp(b_1 e^{-a_1 t} h(x)) - 1 \right\} \nu_L(dx) = 0$ and $Y_u \tilde{\sigma}_{s,u}^2 = Y_u \tilde{\sigma}_{s,u+}^2$ almost surely for fixed u . Solving this linear integral equation with initial condition $\mathbb{E}(Y_0 \tilde{\sigma}_{s,0+}^2) = 0$, it follows that $\mathbb{E}(Y_r \tilde{\sigma}_{s,r+}^2) = 0$ for all $r \geq 0$ is the only solution. \square

Remark 3.4.4 In Theorem 3.2.9 we have seen that $(G_{nr}^{(r)})_{n \in \mathbb{N}}$ is strongly mixing with exponential rate. A consequence of this property (see e.g. Chapter 1.2.2 in Doukhan (1994)) is that there exists a constant $K_G > 0$ such that

$$|\text{Cov}((G_{(n+h)r}^{(r)})^2, (G_{nr}^{(r)})^2)| \leq K_G \cdot e^{-ah}, \quad \forall h > 0. \quad (3.24)$$

In particular this means that the autocovariance function of the squared returns will decay to zero at an exponential rate, thus $(G^{(r)})^2$ is a short memory process (cf. Remark 1.4.8).

3.4.2 Leverage effect

In empirical return data researchers have found evidence (see e.g. Section 1 in Nelson (1990)) that current returns are negatively correlated with future volatility. This means that a negative shock increases the future volatility more than a positive one or increases it while a positive one even decreases the volatility. This phenomenon is called *leverage effect* in the literature.

If we take a look at the shocks of the state process \mathbf{X} in the ECOGARCH(p, q) model

$$\Delta M_t = \begin{cases} (\theta + \gamma) \Delta L_t, & \Delta L_t \geq 0 \\ (\theta - \gamma) \Delta L_t, & \Delta L_t < 0 \end{cases},$$

we see that:

- (i) a positive shock in the return data increases \mathbf{X} less (more) than a negative one for $-\gamma < \theta < 0$ ($0 < \theta < \gamma$), while a negative one decreases it for $\theta > |\gamma|$.
- (ii) a positive shock in the return data decreases \mathbf{X} less (more) than a negative one for $0 < \theta < -\gamma$ ($\gamma < \theta < 0$), while a negative one increases it for $\theta < -|\gamma|$.

If we compare this to the COGARCH(p, q) process, we see that in the COGARCH model the innovations of the volatility process at time t are given by the squared innovations of the log-price process (see Section 2 of Brockwell et al. (2006)). Hence the volatility process of the COGARCH model reacts in the same way to positive and negative shocks. But remember that this can be changed as e.g. explained in Chapter 2.6. We will consider now an *instantaneous leverage effect*, which is defined as

$$\text{Cov}(\Delta G_t, \sigma_{t+}^2 \mid |\Delta L_t| > \epsilon)$$

being negative. Intuitively it is clear that this correlation can be different from zero only if the sample path of σ^2 can have jumps. But from Remark 3.2.6 (i) we know that this is the case only if $p = q$. The reason is that for $p < q$ the parameter b_q will be zero and therefore the jump ΔL_t at time t just contributes to the $(q - 1)$ th derivative of the state process \mathbf{X} , but is not taken into account for the log-volatility at that time point. Thus we will expect an instantaneous leverage effect only for the ECOGARCH(p, p) models. This will be shown in the next proposition, in particular we will show that the sign of the correlation is equal to the sign of θb_q . This result is similar to the discrete time case (see Proposition 2.9 in Surgailis and Viano (2002)).

Proposition 3.4.5 *Assume that the distribution of the jumps of L is symmetric, i.e. for all $\epsilon > 0$,*

$$\mathbb{P}(\Delta L_t \in dx \mid |\Delta L_t| > \epsilon) = \mathbb{P}(\Delta L_t \in -dx \mid |\Delta L_t| > \epsilon), \quad t \geq 0.$$

Conditionally on the event that $|\Delta L_t| > \epsilon$, the sign of $\text{Cov}(\Delta G_t, \sigma_{t+}^2)$ is equal to the sign of θb_q .

Proof: Since the distribution of the jumps of L is symmetric we get

$$\begin{aligned} & \text{Cov}(\Delta G_t, \sigma_{t+}^2 \mid |\Delta L_t| > \epsilon) \\ &= \mathbb{E}(\Delta G_t \sigma_{t+}^2 \mid |\Delta L_t| > \epsilon) - \mathbb{E}(\Delta G_t \mid |\Delta L_t| > \epsilon) \mathbb{E}(\sigma_{t+}^2 \mid |\Delta L_t| > \epsilon) \\ &= \mathbb{E}(\Delta G_t \sigma_{t+}^2 \mid |\Delta L_t| > \epsilon) - \mathbb{E}(\sigma_t) \mathbb{E}(\Delta L_t \mid |\Delta L_t| > \epsilon) \mathbb{E}(\sigma_{t+}^2 \mid |\Delta L_t| > \epsilon) \\ &= \mathbb{E}(\Delta G_t \sigma_{t+}^2 \mid |\Delta L_t| > \epsilon) - \mathbb{E}(\sigma_t) \left(\int_{x>\epsilon} x \mathbb{P}(\Delta L_t \in dx \mid |\Delta L_t| > \epsilon) \right. \\ & \quad \left. - \int_{x>\epsilon} x \mathbb{P}(\Delta L_t \in dx \mid |\Delta L_t| > \epsilon) \right) \mathbb{E}(\sigma_{t+}^2 \mid |\Delta L_t| > \epsilon) \\ &= \mathbb{E} \left(\Delta G_t e^{\log(\sigma_t^2) + b_q \Delta M_t} \mid |\Delta L_t| > \epsilon \right) \\ &= \mathbb{E} \left(\sigma_t^3 \Delta L_t \exp\{b_q(\theta \Delta L_t + \gamma |\Delta L_t|)\} \mid |\Delta L_t| > \epsilon \right) \end{aligned}$$

Since ΔL_t is independent of σ_t^3 we get

$$\begin{aligned} & \mathbb{Cov}(\Delta G_t, \sigma_{t+}^2 \mid |\Delta L_t| > \epsilon) \\ &= \mathbb{E}(\sigma_t^3) \mathbb{E}(\Delta L_t \exp\{b_q(\theta \Delta L_t + \gamma |\Delta L_t|)\} \mid |\Delta L_t| > \epsilon) \\ &= \mathbb{E}(\sigma_t^3) \int_{x>\epsilon} x \exp(b_q \gamma x) (\exp(\theta b_q x) - \exp(-\theta b_q x)) P(\Delta L_t \in dx \mid |\Delta L_t| > \epsilon). \end{aligned}$$

From $\text{sgn}(\exp(\theta b_q x) - \exp(-\theta b_q x)) = \text{sgn}(\theta b_q)$ for all $x > \epsilon$ the desired result follows.

□

Chapter 4

QMLE of compound Poisson ECOGARCH(1, 1)

A first step in estimating the ECOGARCH process will be taken in this chapter. We will restrict ourselves to the compound Poisson ECOGARCH(1,1) process and introduce a quasi maximum likelihood estimator. It will be derived under the assumption of full observations of the sample path, i.e. observation of every jump. Since the jump points are a series of unequally spaced time points it is clear that the estimation can and in fact should be done for irregularly spaced data.

4.1 Quasi MLE of compound Poisson ECOGARCH(1, 1)

We consider an ECOGARCH(1,1) process driven by a compound Poisson process L with Lévy symbol

$$\psi_L(u) = \int_{\mathbb{R}} (e^{iux} - 1) \lambda F_{0,1/\lambda}(dx),$$

where $F_{0,1/\lambda}(\cdot)$ is a distribution function with mean 0 and variance $1/\lambda$. Hence the mean and variance of L are independent of λ and given by $\mathbb{E}(L_t) = 0$ and $\text{Var}(L_t) = t$, respectively. This means L has representation $L_t = \sum_{k=1}^{J_t} Z_k$, $t > 0$, $L_0 = 0$, where $(J_t)_{t \geq 0}$ is an independent Poisson process with intensity $\lambda > 0$ and jump times $(t_k)_{k \in \mathbb{N}}$. The Poisson process J is also independent from the i.i.d. sequence of jump sizes $(Z_k)_{k \in \mathbb{N}}$, with distribution function $F_{0,1/\lambda}$. The Lévy process M , defined in (3.8), is in this case also a compound Poisson process and given by the following expression

$$M_t = \sum_{k=1}^{J_t} [\theta Z_k + \gamma |Z_k|] - \gamma \lambda K t, \quad t > 0,$$

with $K = \int_{\mathbb{R}} |x| F_{0,1/\lambda}(dx)$. The log-volatility process at time t is then of the form

$$\begin{aligned} \log(\sigma_t^2) &= \mu + b_1 X_{t-} = \mu + b_1 e^{-a_1 t} X_0 + \int_0^{t-} b_1 e^{-a_1(t-s)} dM_s \\ &= \mu + b_1 e^{-a_1 t} X_0 + \sum_{k=1}^{J_{t-}} b_1 e^{-a_1(t-t_k)} [\theta Z_k + \gamma |Z_k|] \\ &\quad - \gamma \lambda K \frac{b_1}{a_1} (1 - e^{-a_1 t}), \quad t > 0 \end{aligned}$$

and from (3.5) it follows that the log-price process is given by

$$G_t = \sum_{k=1}^{J_t} \sigma_{t_k} Z_k, \quad t > 0, \quad G_0 = 0,$$

with jump times t_k , $k \in \mathbb{N}$. Note that $\log(\sigma_t^2)$ does not include a possible jump $\Delta L_t = Z_{J_t}$ at time t , while the log-price G_t includes such a jump.

We assume now that we observe G at n consecutive jump times $0 = t_0 < t_1 < \dots < t_n < T < t_{n+1}$, $n \in \mathbb{N}$, over the time interval $[0, T]$. The state process X has then the following autoregressive representation

$$\begin{aligned} b_1 X_{t_i} &= b_1 e^{-a_1 \Delta t_i} X_{t_{i-1}} + \sum_{k=J_{t_{i-1}}+1}^{J_{t_i}} b_1 e^{-a_1(t_i-t_k)} [\theta Z_k + \gamma |Z_k|] \\ &\quad - \gamma \lambda \int_{t_{i-1}}^{t_i} b_1 e^{-a_1(t_i-s)} K ds \\ &= b_1 e^{-a_1 \Delta t_i} X_{t_{i-1}} + b_1 \theta Z_i + b_1 \gamma \left(|Z_i| - \frac{\lambda K}{a_1} (1 - e^{-a_1 \Delta t_i}) \right), \end{aligned} \quad (4.1)$$

where $\Delta t_i := t_i - t_{i-1}$, $i = 1, \dots, n$. Here we used $J_{t_{i-1}} + 1 = J_{t_i} = i$. This implies that the left-hand limit is

$$b_1 X_{t_i-} = b_1 e^{-a_1 \Delta t_i} X_{t_{i-1}} - b_1 \gamma \frac{\lambda K}{a_1} (1 - e^{-a_1 \Delta t_i}),$$

since X_{t_i-} does not include the jump at time t_i and its corresponding jump size Z_i . Thus the log-volatility process at the jump times $0 = t_0 < t_1 < \dots < t_n$ is given by

$$\log(\sigma_{t_i}^2) = \mu + b_1 e^{-a_1 \Delta t_i} X_{t_{i-1}} - b_1 \gamma \frac{\lambda K}{a_1} (1 - e^{-a_1 \Delta t_i}). \quad (4.2)$$

Since b_1 serves only as a scaling coefficient it will be set equal to one for identifiability reasons from now on. The observations of the log-price process are given by

$$G_{t_i} = \sum_{k=1}^{J_{t_i}} \sigma_{t_k} Z_k = G_{t_{i-1}} + \sigma_{t_i} Z_i. \quad (4.3)$$

Hence the return at time t_i is equal to $G_{t_i}^{\Delta t_i} = \sigma_{t_i} Z_i$. Observe that σ_{t_i} is independent of Z_i for each $i = 1, \dots, n$.

The parameter estimation is done in two steps. The rate λ of the Poisson process J can be estimated given only the jump times t_i , therefore this is done in a first step. Since we observe the total number n of jumps for the Poisson process J over T intervals of length one the MLE of λ is given by

$$\hat{\lambda}_n := \frac{n}{T}.$$

To estimate the remaining parameters $\boldsymbol{\vartheta} := (a_1, \theta, \gamma, \mu)$ we use similar ideas as in the discrete time case to solve the parameter estimation problem. Quasi maximum likelihood estimation in discrete time conditionally heteroscedastic time series models is e.g. explained in Straumann (2005).

Consider the following decomposition of the conditional log-likelihood given the initial value X_0

$$\log \rho_{\boldsymbol{\vartheta}}(G_{t_1}^{\Delta t_1}, \dots, G_{t_n}^{\Delta t_n} | X_0) = \sum_{i=1}^n \log \rho_{\boldsymbol{\vartheta}}(G_{t_i}^{\Delta t_i} | G_{t_{i-1}}^{\Delta t_{i-1}}, \dots, G_{t_1}^{\Delta t_1}, X_0),$$

where we assume that $G_{t_i}^{\Delta t_i}$ given $G_{t_{i-1}}^{\Delta t_{i-1}}, \dots, G_{t_1}^{\Delta t_1}, X_0$ is conditionally normal distributed with mean zero and variance $\sigma_{t_i}^2/\lambda$. This implies that the conditional log-likelihood has the representation

$$\log \rho_{\boldsymbol{\vartheta}}(G_{t_1}^{\Delta t_1}, \dots, G_{t_n}^{\Delta t_n} | X_0) = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \sum_{i=1}^n \left(\log(\sigma_{t_i}^2/\lambda) - \frac{(G_{t_i}^{\Delta t_i})^2}{\sigma_{t_i}^2/\lambda} \right). \quad (4.4)$$

Since the volatility is unobservable (4.4) can not be evaluated numerically. Therefore we need an approximation of the state process X , which together with (4.2) gives estimates of $\sigma_{t_1}^2, \dots, \sigma_{t_n}^2$. Given parameters $\boldsymbol{\vartheta}$ and λ , an approximation of the recursion (4.1) is used to compute estimates of the state process X by

$$\hat{X}_{t_i}(\boldsymbol{\vartheta}, \lambda) = e^{-a_1 \Delta t_i} \hat{X}_{t_{i-1}}(\boldsymbol{\vartheta}, \lambda) + \theta \frac{G_{t_i}^{\Delta t_i}}{\hat{\sigma}_{t_i}(\boldsymbol{\vartheta}, \lambda)} + \gamma \left(\frac{|G_{t_i}^{\Delta t_i}|}{\hat{\sigma}_{t_i}(\boldsymbol{\vartheta}, \lambda)} - \lambda \hat{K} \Delta t_i \right), \quad (4.5)$$

$i = 1, \dots, n$, where $\widehat{K} := \sqrt{\frac{2}{\pi\lambda}} = \mathbb{E}(|W|)$, $W \sim N(0, 1/\lambda)$.

Here we used $(1 - e^{-z}) \approx z$ for z small. The recursion needs a starting value \widehat{X}_0 , which will be the mean value zero of the stationary distribution of X . (4.5) can be understood as the log-volatility description in a discrete time EGARCH(1, 1) model for irregularly spaced data. We work with this approximation since in our numerical experiences it provided better results than (4.1). This is due to the independence of a_1 from the part compensating the absolute jumps, which is approximated by the Gaussianity assumption. If we would use (4.1) the optimisation with respect to a_1 tries to account for the approximation, which results in convergence problems and biased estimates of the autoregressive parameter e^{-a_1} .

Recursion (4.5) together with expression (4.2) provides then estimates of the volatility given by

$$\widehat{\sigma}_{t_i}^2(\boldsymbol{\vartheta}, \lambda) := \exp(\mu + e^{-a_1 \Delta t_i} \widehat{X}_{t_{i-1}}(\boldsymbol{\vartheta}, \lambda) - \gamma \lambda \widehat{K} \Delta t_i), \quad i = 1, \dots, n.$$

Based on the approximation of the volatility we define the quasi log-likelihood function for $\boldsymbol{\vartheta}$ given the data $\mathbf{G}^\Delta := (G_{t_1}^{\Delta t_1}, \dots, G_{t_n}^{\Delta t_n})$ and the MLE $\widehat{\lambda}_n$ by

$$L(\boldsymbol{\vartheta} | \mathbf{G}^\Delta, \widehat{\lambda}_n) := -\frac{1}{2} \sum_{i=1}^n \log(\widehat{\sigma}_{t_i}^2(\boldsymbol{\vartheta}, \widehat{\lambda}_n)) - \frac{1}{2} \sum_{i=1}^n \frac{(G_{t_i}^{\Delta t_i})^2}{\widehat{\sigma}_{t_i}^2(\boldsymbol{\vartheta}, \widehat{\lambda}_n) / \widehat{\lambda}_n}. \quad (4.6)$$

Observe that $L(\boldsymbol{\vartheta} | \mathbf{G}^\Delta, \widehat{\lambda}_n)$ does not contain the constant $\frac{n}{2} \log(\widehat{\lambda}_n / 2\pi)$, since it does not affect the optimisation with respect to $\boldsymbol{\vartheta}$.

Maximising the log-likelihood function (4.6) with respect to $\boldsymbol{\vartheta}$ over the parameter space $\Theta := \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}^2$ yields QML estimates

$$\widehat{\boldsymbol{\vartheta}}_n := \arg \max_{\boldsymbol{\vartheta} \in \Theta} L(\boldsymbol{\vartheta} | \mathbf{G}^\Delta, \widehat{\lambda}_n) \quad (4.7)$$

of $\boldsymbol{\vartheta}$. As a byproduct we get a parametric estimator of the volatility. If we first determine the QMLE $\widehat{\boldsymbol{\vartheta}}_n$ in (4.7) then we can substitute $\widehat{\boldsymbol{\vartheta}}_n$ into (4.5) and get estimates

$$\widehat{\sigma}_{t_i}^2(\widehat{\boldsymbol{\vartheta}}_n, \widehat{\lambda}_n) := \exp(\widehat{\mu}_n + e^{-\widehat{a}_1^n \Delta t_i} \widehat{X}_{t_{i-1}}(\widehat{\boldsymbol{\vartheta}}_n, \widehat{\lambda}_n) - \widehat{\gamma}_n \widehat{\lambda}_n \widehat{K} \Delta t_i) \quad (4.8)$$

of the volatility at the jump times t_1, \dots, t_n based on $\widehat{\boldsymbol{\vartheta}}_n = (\widehat{a}_1^n, \widehat{\theta}_n, \widehat{\gamma}_n, \widehat{\mu}_n)$ and $\widehat{\lambda}_n$.

4.2 Simulation study

The small sample behaviour of the QMLE introduced in the last section will be investigated in a simulation study. In all of the following simulation cases we will consider a compound Poisson ECOGARCH(1, 1) observed at all jump times t_i over

the time interval $[0, 1500]$. The estimates will be computed for 1000 independent replications in each case. Since we are able to estimate the volatility by (4.8), an analysis of the fitted innovations will also be done in one of the simulation examples.

In all of the following cases we have taken the parameter a_1 equal to 0.1, the intensity λ equal to 2 and the mean μ of the log-volatility process will be equal to -3 . This implies that we expect per replication 3000 observations. The leverage parameter θ and γ will vary over the examples. In most of the cases θ will be negative and γ positive, i.e. we model the leverage effect as observed in stock price data. If $-\gamma < \theta < 0$, this corresponds to the case where a positive shock in the return data increases the log-volatility process less than a negative one of the same magnitude. For $\theta < -\gamma < 0$ a positive shock decreases the log-volatility, whereas a negative one increases it. The last example will illustrate the case where a positive shock in the log-price process increases the log-volatility process more than a negative one. For the distribution of the innovations Z_i we will consider two different cases.

4.2.1 Leverage case with Gaussian jump distribution

First the innovations Z_i are normally distributed with mean 0 and variance $1/\lambda$. We computed the empirical mean ($\widehat{\text{mean}}$), relative bias ($\widehat{\text{rbias}}$), and mean squared error ($\widehat{\text{MSE}}$) for all parameter estimates based on 1000 independent replications. The corresponding results are summarised in Table 4.1. In the leverage case we observe a satisfying performance of the QMLE. The relative bias of \widehat{a}_1 , $\widehat{\theta}$ and $\widehat{\gamma}$ varies between -0.0361 and 0.0211 . over the different parameter sets. For the mean $\widehat{\mu}$ of the log-volatility process a larger relative bias is observed. It also increases for larger values of γ . Moreover we seem to underestimate μ consistently, shown by a negative bias in each case. The quality of the separately estimated $\widehat{\lambda}$ remains of course unchanged over all parameter settings.

The goodness of fit of our estimation method is further investigated by an analysis of the fitted innovations for the case, where $\theta = -0.3$ and $\gamma = 0.4$. The fitted innovations are given by $\widehat{Z}_i := G_{t_i}^{\Delta t_i} / \widehat{\sigma}_{t_i}$, $i = 1, \dots, n$. Since our innovations were normally distributed with mean zero and standard deviation equal to $1/\sqrt{2}$, we expect the average $\frac{1}{n} \sum_{i=1}^n \widehat{Z}_i$ of the fitted innovations close to zero, their empirical standard deviation close to $1/\sqrt{2}$ and their empirical skewness close to zero. For all three quantities we computed $\widehat{\text{mean}}$ and $\widehat{\text{MSE}}$. The results are reported in Table 4.2 and indicate a good fit.

Under the assumption of a correctly estimated volatility the fitted innovations are a white noise series, in particular the innovations and also the squared innovations should be uncorrelated. The correlation of the squared innovations was checked by

Gaussian	\hat{a}_1	$\hat{\theta}$	$\hat{\gamma}$	$\hat{\mu}$	$\hat{\lambda}$
ϑ, λ	0.1000	-0.1000	0.2000	-3.0000	2.0000
$\widehat{\text{mean}}$	0.1021 ($5 \cdot 10^{-4}$)	-0.1007 ($5 \cdot 10^{-4}$)	0.1927 ($6 \cdot 10^{-4}$)	-2.9058 ($3 \cdot 10^{-3}$)	2.0008 ($1 \cdot 10^{-3}$)
$\widehat{\text{rbias}}$	0.0211	0.0077	-0.0361	-0.0313	0.0004
$\widehat{\text{MSE}}$	0.0003 ($1 \cdot 10^{-5}$)	0.0003 ($1 \cdot 10^{-5}$)	0.0004 ($2 \cdot 10^{-5}$)	0.0161 ($6 \cdot 10^{-4}$)	0.0014 ($6 \cdot 10^{-5}$)
ϑ, λ	0.1000	-0.1800	0.2000	-3.0000	2.0000
$\widehat{\text{mean}}$	0.0979 ($7 \cdot 10^{-4}$)	-0.1797 ($1 \cdot 10^{-3}$)	0.1938 ($6 \cdot 10^{-4}$)	-2.8980 ($4 \cdot 10^{-3}$)	2.0001 ($1 \cdot 10^{-3}$)
$\widehat{\text{rbias}}$	-0.0202	-0.0016	-0.0310	-0.0339	$2 \cdot 10^{-5}$
$\widehat{\text{MSE}}$	0.0005 ($6 \cdot 10^{-5}$)	0.0014 ($4 \cdot 10^{-4}$)	0.0004 ($2 \cdot 10^{-5}$)	0.0274 ($4 \cdot 10^{-3}$)	0.0012 ($6 \cdot 10^{-5}$)
ϑ, λ	0.1000	-0.3000	0.2000	-3.0000	2.0000
$\widehat{\text{mean}}$	0.1004 ($3 \cdot 10^{-4}$)	-0.3027 ($1 \cdot 10^{-3}$)	0.1930 ($6 \cdot 10^{-4}$)	-2.8982 ($4 \cdot 10^{-3}$)	2.0000 ($1 \cdot 10^{-3}$)
$\widehat{\text{rbias}}$	0.0038	0.0091	-0.0349	-0.0339	$1 \cdot 10^{-5}$
$\widehat{\text{MSE}}$	0.0001 ($1 \cdot 10^{-5}$)	0.0013 ($5 \cdot 10^{-4}$)	0.0004 ($5 \cdot 10^{-4}$)	0.0290 ($1 \cdot 10^{-2}$)	0.0012 ($6 \cdot 10^{-5}$)
ϑ, λ	0.1000	-0.1000	0.4000	-3.0000	2.0000
$\widehat{\text{mean}}$	0.1017 ($4 \cdot 10^{-4}$)	-0.0996 ($6 \cdot 10^{-4}$)	0.3877 ($7 \cdot 10^{-4}$)	-2.8028 ($4 \cdot 10^{-3}$)	2.0001 ($1 \cdot 10^{-3}$)
$\widehat{\text{rbias}}$	0.0170	-0.0037	-0.0307	-0.0657	0.0005
$\widehat{\text{MSE}}$	0.0001 ($7 \cdot 10^{-6}$)	0.0004 ($2 \cdot 10^{-5}$)	0.0006 ($3 \cdot 10^{-5}$)	0.0562 ($2 \cdot 10^{-3}$)	0.0013 ($6 \cdot 10^{-5}$)
ϑ, λ	0.1000	-0.1800	0.4000	-3.0000	2.0000
$\widehat{\text{mean}}$	0.1007 ($4 \cdot 10^{-4}$)	-0.1792 ($7 \cdot 10^{-4}$)	0.3864 ($7 \cdot 10^{-4}$)	-2.8122 ($4 \cdot 10^{-3}$)	1.9985 ($1 \cdot 10^{-3}$)
$\widehat{\text{rbias}}$	0.0071	-0.0044	-0.0338	-0.0625	-0.0007
$\widehat{\text{MSE}}$	0.0001 ($6 \cdot 10^{-6}$)	0.0005 ($2 \cdot 10^{-5}$)	0.0007 ($3 \cdot 10^{-5}$)	0.0526 ($2 \cdot 10^{-3}$)	0.0013 ($5 \cdot 10^{-5}$)
ϑ, λ	0.1000	-0.3000	0.4000	-3.0000	2.0000
$\widehat{\text{mean}}$	0.1010 ($3 \cdot 10^{-4}$)	-0.3004 ($7 \cdot 10^{-4}$)	0.3880 ($7 \cdot 10^{-4}$)	-2.7998 ($4 \cdot 10^{-3}$)	2.0023 ($1 \cdot 10^{-3}$)
$\widehat{\text{rbias}}$	0.0102	0.0014	-0.0298	-0.0667	0.0011
$\widehat{\text{MSE}}$	0.0001 ($4 \cdot 10^{-6}$)	0.0005 ($2 \cdot 10^{-5}$)	0.0006 ($2 \cdot 10^{-5}$)	0.0595 ($2 \cdot 10^{-3}$)	0.0014 ($6 \cdot 10^{-5}$)

Table 4.1: Estimated mean and MSE for $\hat{a}_1, \hat{\theta}, \hat{\gamma}, \hat{\mu}$ and $\hat{\lambda}$ with corresponding estimated standard deviations in brackets based on 1000 replications with normally distributed jump sizes. In the third row of each case the relative bias is shown.

Gaussian	mean(\widehat{Z}_i)	std(\widehat{Z}_i)	skewness(\widehat{Z}_i)
$\widehat{\text{mean}}$	-0.00004 ($4 \cdot 10^{-4}$)	0.70743 ($2 \cdot 10^{-4}$)	0.00062 ($1 \cdot 10^{-3}$)
$\widehat{\text{MSE}}$	0.00017 ($8 \cdot 10^{-6}$)	0.00004 ($2 \cdot 10^{-6}$)	0.00209 ($9 \cdot 10^{-5}$)
student- t	mean(\widehat{Z}_i)	std(\widehat{Z}_i)	skewness(\widehat{Z}_i)
$\widehat{\text{mean}}$	-0.00007 ($4 \cdot 10^{-4}$)	0.70763 ($2 \cdot 10^{-4}$)	-0.00295 ($6 \cdot 10^{-3}$)
$\widehat{\text{MSE}}$	0.00016 ($7 \cdot 10^{-6}$)	0.00006 ($7 \cdot 10^{-6}$)	0.03846 ($5 \cdot 10^{-3}$)

Table 4.2: Estimated mean and MSE for the mean, standard deviation and skewness of the residuals with corresponding estimated standard deviations in brackets for normally (*top*) and student- t (*bottom*) distributed jumps based on 1 000 replications.

performing a Ljung-Box test. The test statistic is given by

$$Q = n(n+2) \sum_{k=1}^m \frac{\widehat{\rho}_{\widehat{Z}^2}(k)^2}{n-k},$$

where $\widehat{\rho}_{\widehat{Z}^2}(k)$ is the empirical autocorrelation function of the the squared innovations, and asymptotically χ^2 -distributed with m degrees of freedom under the null hypothesis of no correlation. The number of lags m taken into account to compute the statistic was set equal to \sqrt{n} (cf. Section 9.4 in (Brockwell and Davis 1991)). The null hypothesis of no correlation was rejected 49 times out of 1 000 simulations at the 0.05 level. The empirical mean of the 1 000 p -values was equal to 0.52, which shows that a majority of the test statistics has a rather large p -value confirming the hypothesis of no correlation.

4.2.2 Leverage case with student- t jump distribution

In all the previous examples the jump distribution was Gaussian. Now we want to compute the QMLE under the assumption of sampling innovations from a student- t distribution. We will consider a t -distribution with 6 degrees of freedom. Since we assume $\mathbb{E}(Z_1) = 0$ and $\text{Var}(Z_1) = 1/\lambda$, we have to scale the innovations in an appropriate way. The intensity λ will again be equal to 2 and μ remains -3 . The parameters a_1, θ and γ take on the same values as in the Gaussian case. The results are reported in Table 4.3.

The quality of the estimators for a_1, θ, γ and μ is reduced due to the model misspecification. The relative bias and MSE have increased for almost all parameter settings. Concerning μ the relative bias for example has doubled compared to results for normally distributed jumps. But overall the results are still satisfying. Indicating

student- t	\hat{a}_1	$\hat{\theta}$	$\hat{\gamma}$	$\hat{\mu}$	$\hat{\lambda}$
ϑ, λ	0.1000	-0.1000	0.2000	-3.0000	2.0000
$\widehat{\text{mean}}$	0.0978	-0.0995	0.1917	-2.7748	2.0020
	($1 \cdot 10^{-3}$)	($1 \cdot 10^{-3}$)	($9 \cdot 10^{-4}$)	($6 \cdot 10^{-3}$)	($1 \cdot 10^{-3}$)
$\widehat{\text{rbias}}$	-0.0213	-0.0046	-0.0415	-0.0750	0.0010
$\widehat{\text{MSE}}$	0.0012	0.0017	0.0010	0.0873	0.0014
	($8 \cdot 10^{-5}$)	($6 \cdot 10^{-4}$)	($5 \cdot 10^{-5}$)	($9 \cdot 10^{-3}$)	($6 \cdot 10^{-5}$)
ϑ, λ	0.1000	-0.1800	0.2000	-3.0000	2.0000
$\widehat{\text{mean}}$	0.1007	-0.1829	0.1934	-2.7776	2.0001
	($8 \cdot 10^{-4}$)	($1 \cdot 10^{-3}$)	($9 \cdot 10^{-4}$)	($5 \cdot 10^{-4}$)	($1 \cdot 10^{-3}$)
$\widehat{\text{rbias}}$	0.0077	0.0165	-0.0331	-0.0741	$4 \cdot 10^{-5}$
$\widehat{\text{MSE}}$	0.0007	0.0020	0.0008	0.0757	0.0013
	($6 \cdot 10^{-5}$)	($6 \cdot 10^{-4}$)	($5 \cdot 10^{-5}$)	($5 \cdot 10^{-3}$)	($6 \cdot 10^{-5}$)
ϑ, λ	0.1000	-0.3000	0.2000	-3.0000	2.0000
$\widehat{\text{mean}}$	0.1013	-0.3034	0.1926	2.0012	-2.7738
	($5 \cdot 10^{-4}$)	($1 \cdot 10^{-3}$)	($9 \cdot 10^{-4}$)	($1 \cdot 10^{-3}$)	($4 \cdot 10^{-4}$)
$\widehat{\text{rbias}}$	0.0131	0.0115	-0.0369	0.0006	-0.0754
$\widehat{\text{MSE}}$	0.0002	0.0015	0.0008	0.0014	0.0668
	($2 \cdot 10^{-5}$)	($5 \cdot 10^{-4}$)	($6 \cdot 10^{-5}$)	($6 \cdot 10^{-5}$)	($3 \cdot 10^{-3}$)
ϑ, λ	0.1000	-0.1000	0.4000	-3.0000	2.0000
$\widehat{\text{mean}}$	0.1022	-0.1010	0.3884	-2.5405	2.0025
	($5 \cdot 10^{-4}$)	($1 \cdot 10^{-3}$)	($1 \cdot 10^{-3}$)	($6 \cdot 10^{-3}$)	($1 \cdot 10^{-3}$)
$\widehat{\text{rbias}}$	0.0221	0.0101	-0.0290	-0.1531	0.0012
$\widehat{\text{MSE}}$	0.0003	0.0009	0.0012	0.2441	0.0014
	($2 \cdot 10^{-5}$)	($4 \cdot 10^{-5}$)	($6 \cdot 10^{-5}$)	($6 \cdot 10^{-3}$)	($6 \cdot 10^{-5}$)
ϑ, λ	0.1000	-0.1800	0.4000	-3.0000	2.0000
$\widehat{\text{mean}}$	0.1015	-0.1819	0.3861	-2.5518	2.0004
	($5 \cdot 10^{-4}$)	($1 \cdot 10^{-3}$)	($1 \cdot 10^{-3}$)	($6 \cdot 10^{-3}$)	($1 \cdot 10^{-3}$)
$\widehat{\text{rbias}}$	0.0151	0.0109	-0.0347	-0.1493	0.0002
$\widehat{\text{MSE}}$	0.0003	0.0011	0.0013	0.2346	0.0013
	($1 \cdot 10^{-5}$)	($2 \cdot 10^{-4}$)	($5 \cdot 10^{-5}$)	($5 \cdot 10^{-3}$)	($6 \cdot 10^{-5}$)
ϑ, λ	0.1000	-0.3000	0.4000	-3.0000	2.0000
$\widehat{\text{mean}}$	0.1020	-0.3030	0.3876	-2.5481	1.9997
	($4 \cdot 10^{-4}$)	($1 \cdot 10^{-3}$)	($1 \cdot 10^{-3}$)	($6 \cdot 10^{-3}$)	($1 \cdot 10^{-3}$)
$\widehat{\text{rbias}}$	0.0199	0.0099	-0.0309	-0.1506	-0.0002
$\widehat{\text{MSE}}$	0.0013	0.0009	0.0010	0.0789	0.0014
	($8 \cdot 10^{-5}$)	($4 \cdot 10^{-4}$)	($5 \cdot 10^{-5}$)	($4 \cdot 10^{-3}$)	($6 \cdot 10^{-5}$)

Table 4.3: Estimated mean and MSE for $\hat{a}_1, \hat{\theta}, \hat{\gamma}, \hat{\mu}$ and $\hat{\lambda}$ with corresponding estimated standard deviations in brackets based on 1 000 replications with t -distributed jump sizes. In the third row of each case the relative bias is shown.

that the QMLE provides reasonable values even if the true distribution of the returns is much heavier tailed than the assumed one.

Analogously to the Gaussian case we investigated the goodness of fit by an analysis of the fitted innovations for the case, where $\theta = -0.3$ and $\gamma = 0.4$. Since we scaled the innovations such that they have mean zero and a standard deviation equal to $1/\sqrt{2}$, we expect the empirical mean of the estimated innovations close to zero, their empirical standard deviation close to $1/\sqrt{2}$ and their empirical skewness close to zero. For all three quantities we computed $\widehat{\text{mean}}$ and $\widehat{\text{MSE}}$. The results are reported in Table 4.2 and indicate a reasonable fit. The null hypothesis of the Ljung-Box test was rejected 84 times out of 1 000 simulations at the 0.05 level. The empirical mean of the 1 000 p -values was 0.59.

4.2.3 Non-leverage case

So far we have only considered the leverage case. In this last example we will have the following relation: $0 < \theta < \gamma$. This means that a positive shock in the return data increases the log-volatility process more than a negative one. For normally distributed jumps the results in the non-leverage case are shown in the upper section of Table 4.4.

Gaussian	\widehat{a}_1	$\widehat{\theta}$	$\widehat{\gamma}$	$\widehat{\mu}$	$\widehat{\lambda}$
ϑ, λ	0.1000	0.1000	0.2000	-3.0000	2.0000
$\widehat{\text{mean}}$	0.0938 ($1 \cdot 10^{-3}$)	0.0980 ($1 \cdot 10^{-3}$)	0.1911 ($8 \cdot 10^{-4}$)	-2.8977 ($7 \cdot 10^{-3}$)	2.0024 ($1 \cdot 10^{-3}$)
$\widehat{\text{rbias}}$	-0.0614	-0.0201	-0.0447	-0.0341	0.0012
$\widehat{\text{MSE}}$	0.0012 ($9 \cdot 10^{-5}$)	0.0015 ($5 \cdot 10^{-4}$)	0.0007 ($1 \cdot 10^{-4}$)	0.0595 ($1 \cdot 10^{-2}$)	0.0014 ($6 \cdot 10^{-5}$)
student- t	\widehat{a}_1	$\widehat{\theta}$	$\widehat{\gamma}$	$\widehat{\mu}$	$\widehat{\lambda}$
ϑ, λ	0.1000	0.1000	0.2000	-3.0000	2.0000
$\widehat{\text{mean}}$	0.0976 ($1 \cdot 10^{-3}$)	0.0988 ($1 \cdot 10^{-3}$)	0.1913 ($1 \cdot 10^{-3}$)	-2.7873 ($6 \cdot 10^{-3}$)	2.0031 ($1 \cdot 10^{-3}$)
$\widehat{\text{rbias}}$	-0.0235	-0.0117	-0.0432	-0.0709	0.0015
$\widehat{\text{MSE}}$	0.0013 ($8 \cdot 10^{-5}$)	0.0009 ($4 \cdot 10^{-5}$)	0.0009 ($5 \cdot 10^{-5}$)	0.0789 ($4 \cdot 10^{-3}$)	0.0013 ($6 \cdot 10^{-5}$)

Table 4.4: Estimated mean and MSE for $\widehat{a}_1, \widehat{\theta}, \widehat{\gamma}, \widehat{\lambda}$ and $\widehat{\mu}$ with corresponding estimated standard deviations in brackets based on 1 000 replications with normally distributed (*top*) and t -distributed (*bottom*) jump sizes. In the third row of each case the relative bias is shown.

One can observe an increased relative bias for a_1, θ and γ compared to the leverage case. But the results are still acceptable. For t -distributed innovations the results are shown in the lower section of Table 4.4. The conclusions which can be made are similarly to the Gaussian case. It is interesting however that the leverage parameter θ is estimated more accurately for t -distributed jumps. It can be concluded that the influence of the sign of the leverage parameter θ is not so important for the QMLE.

4.3 Prediction

The aim of this section is to show how prediction can be done in this framework. In particular we will construct a one-step ahead prediction of the volatility process and also derive the prediction density in that case. Since future innovations of the log-price process are positive or negative with probability 0.5, it is not possible to define a sensible one-step ahead prediction of the log-price process. But we can construct a prediction interval for the next observation.

4.3.1 One-step ahead prediction of the volatility

In this section we want to show how to compute a prediction $P_r(\sigma_{t_{n+1}+}^2)$ of the volatility $\sigma_{t_{n+1}+}^2$, which means the volatility right after the next jump, conditional on the information $\widehat{X}_{t_n}, \widehat{\vartheta}_n, \widehat{\lambda}_n$ and Δt_{n+1} . Remember that the parameters $\widehat{\vartheta}_n$ and $\widehat{\lambda}_n$ are estimated based on the observations G_{t_1}, \dots, G_{t_n} . Definitely we will need as a first step a prediction of the log-volatility process at time $t_{n+1}+$. Since we estimated the parameters by maximising a Gaussian likelihood it seems reasonable to simulate the future jump \widehat{Z}_{n+1} as a normally distributed random variable with mean 0 and variance $1/\widehat{\lambda}_n$. If we substitute in the recursion (4.5) for $i = n + 1$ the fitted innovations $G_{t_{n+1}}^{\Delta t_{n+1}}/\widehat{\sigma}_{t_{n+1}}(\vartheta, \lambda)$ by \widehat{Z}_{n+1} , we get

$$\widehat{X}_{t_{n+1}}(\vartheta, \lambda) = e^{-a_1 \Delta t_{n+1}} \widehat{X}_{t_n}(\vartheta, \lambda) + \theta \widehat{Z}_{n+1} + \gamma \left(|\widehat{Z}_{n+1}| - \lambda \widehat{K} \Delta t_{n+1} \right).$$

Hence we will get a prediction of the log-volatility process at time $t_{n+1}+$, which is denoted by $P_r(\log(\sigma_{t_{n+1}+}^2))$, by the following equation

$$P_r(\log(\sigma_{t_{n+1}+}^2)) := \widehat{\mu}_n + e^{-\widehat{a}_1 \Delta t_{n+1}} \widehat{X}_{t_n}(\widehat{\vartheta}_n, \widehat{\lambda}_n) + \widehat{\theta}_n \widehat{Z}_{n+1} + \widehat{\gamma}_n \left(|\widehat{Z}_{n+1}| - \widehat{\lambda}_n \widehat{K} \Delta t_{n+1} \right).$$

The one-step ahead prediction of the volatility process at time $t_{n+1}+$, which is denoted by $P_r(\sigma_{t_{n+1}+}^2)$, is then defined by applying the exponential function to the prediction $P_r(\log(\sigma_{t_{n+1}+}^2))$, i.e.

$$P_r(\sigma_{t_{n+1}+}^2) := \exp(P_r(\log(\sigma_{t_{n+1}+}^2))). \quad (4.9)$$

Given \widehat{X}_{t_n} , $\widehat{\boldsymbol{\vartheta}}_n$, $\widehat{\lambda}_n$ and Δt_{n+1} $P_r(\sigma_{t_{n+1}+}^2)$ is a monotone function of \widehat{Z}_{n+1} . Hence we can in addition easily derive a prediction density of $\sigma_{t_{n+1}+}^2$, which is denoted by $p(s|\widehat{X}_{t_n}, \widehat{\boldsymbol{\vartheta}}_n, \widehat{\lambda}_n, \Delta t_{n+1})$. The form of the density depends on the sign and size of $\widehat{\theta}_n$ and $\widehat{\gamma}_n$. If e.g. $\widehat{\theta}_n < -\widehat{\gamma}_n < 0$, then the prediction density is given by

$$\begin{aligned}
p(s|\widehat{X}_{t_n}, \widehat{\boldsymbol{\vartheta}}_n, \widehat{\lambda}_n, \Delta t_{n+1}) = & \\
& \frac{1}{s} \left[\chi_{(-\infty, \varphi(\widehat{X}_{t_n}, \Delta t_{n+1}))}(\log(s) - \widehat{\mu}_n) \left| \frac{1}{\widehat{\theta}_n + \widehat{\gamma}_n} \right| \right. \\
& \quad \times \sqrt{\frac{\widehat{\lambda}_n}{2\pi}} \exp \left\{ -\frac{\widehat{\lambda}_n}{2} \left(\frac{\log(s) - \widehat{\mu}_n - \varphi(\widehat{X}_{t_n}, \Delta t_{n+1})}{\widehat{\theta}_n + \widehat{\gamma}_n} \right)^2 \right\} \\
& + \chi_{(\varphi(\widehat{X}_{t_n}, \Delta t_{n+1}), \infty)}(\log(s) - \widehat{\mu}_n) \left| \frac{1}{\widehat{\theta}_n - \widehat{\gamma}_n} \right| \\
& \quad \times \sqrt{\frac{\widehat{\lambda}_n}{2\pi}} \exp \left\{ -\frac{\widehat{\lambda}_n}{2} \left(\frac{\log(s) - \widehat{\mu}_n - \varphi(\widehat{X}_{t_n}, \Delta t_{n+1})}{\widehat{\theta}_n - \widehat{\gamma}_n} \right)^2 \right\} \left. \right], \quad (4.10)
\end{aligned}$$

for each $s > 0$, where

$$\varphi(\widehat{X}_{t_n}, \Delta t_{n+1}) := e^{-\widehat{a}_1 \Delta t_{n+1}} \widehat{X}_{t_n} - \widehat{\gamma}_n \widehat{\lambda}_n \widehat{K} \Delta t_{n+1}.$$

Next we want to illustrate how the shape of the prediction density for $\sigma_{t_{n+1}+}^2$ depends on the estimated parameters. Therefore we consider again 1 000 samples of a compound Poisson ECOGARCH(1, 1) process with parameters $\mu = 0$, $a_1 = 0.1$, $\theta = -0.3$, $\gamma = 0.2$ and $\lambda = 2$. The sample paths are over the time interval $[0, 1 000]$. First we estimate the parameters, as explained in Section 4.1, based on the first $n = 1900$ observations. To compute a prediction density for $\sigma_{t_{n+1}+}^2$ by (4.10) we will further need \widehat{X}_{t_n} and Δt_{n+1} . Since both of them will be different for each sample, we set them equal to 1 and 0.5, respectively, to make prediction densities comparable. For each sample the prediction density was then computed over the same grid on the interval $[1.5, 4.7]$. Everything was done for two different jump distributions. In the first case the jumps Z_i were normally distributed with mean 0 and variance $1/\lambda$, whereas in the second example scaled t -distributed with 6 degrees of freedom and same mean and variance as in the normal case. In the first row of Figure 4.1 one can see on the left hand side six replications of prediction densities $p(\cdot|1, \widehat{\boldsymbol{\vartheta}}_n, \widehat{\lambda}_n, 0.5)$ for normally distributed jumps and on the right hand side for the t -distributed ones.

In both cases we also computed a mean prediction density

$$\widehat{p}(s|1, \widehat{\boldsymbol{\vartheta}}, \widehat{\lambda}_n, 0.5) := \frac{1}{N} \sum_{i=1}^N p^i(s|1, \widehat{\boldsymbol{\vartheta}}_n^i, \widehat{\lambda}_n^i, 0.5),$$

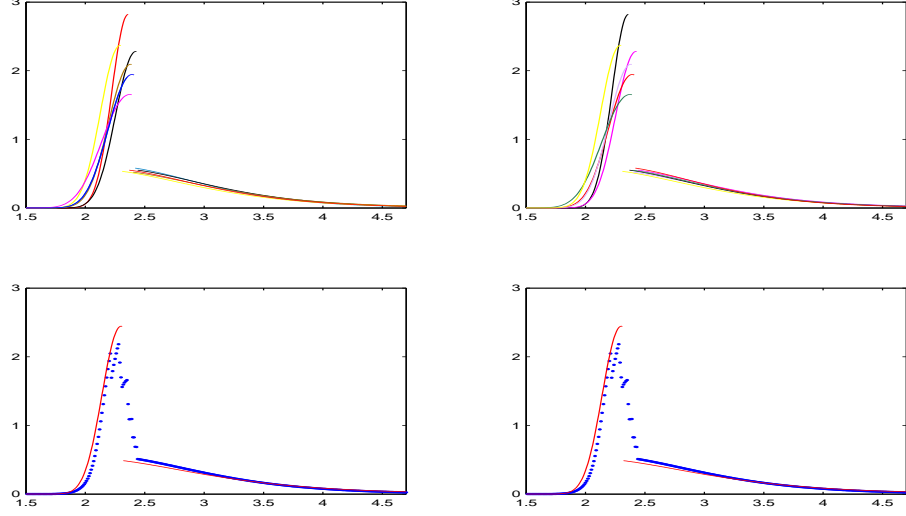


Figure 4.1: Prediction densities $p^i(\cdot|1, \widehat{\boldsymbol{\vartheta}}_n^i, \widehat{\lambda}_n^i, 0.5)$ for the first 6 replication (*top row*) together with mean prediction density $\widehat{p}(s|1, \widehat{\boldsymbol{\vartheta}}, \widehat{\lambda}_n, 0.5)$ and true (*solid line*) prediction density $p(\cdot|1, \boldsymbol{\vartheta}_n, \lambda_n, 0.5)$ (*bottom row*) for normally (*left*) and t-distributed (*right*) jumps.

for s lying on the grid and $N = 1000$. The results are shown in the second row of Figure 4.1 together with the true prediction density $p(\cdot|1, \boldsymbol{\vartheta}, \lambda, 0.5)$. One can observe that despite the fact that we have a larger bias in the estimates $\widehat{\boldsymbol{\vartheta}}_n$ for t -distributed jumps compared to normally distributed ones, the prediction densities show similar behaviour. Especially for the mean prediction density there can be no big difference observed.

For assessing the quality of the forecasts we will consider a scoring rule, which assigns a numerical score based on the prediction density and the value actually observed. We will work with the logarithmic score, for other scoring rules see Gneiting and Raftery (2006). Assume we have N observations of the volatility $\sigma_{t_{n+1}^+}^2, \dots, \sigma_{t_{n+1}^+}^2$. For each of them we compute a prediction density

$$p(\cdot|\widehat{X}_{t_n^i}, \widehat{\boldsymbol{\vartheta}}_n^i, \widehat{\lambda}_n^i, \Delta t_{n+1}^i), \quad i = 1, \dots, N,$$

which will be evaluated at the observation point. The logarithmic score is then defined as

$$LS \left(p(\cdot|\widehat{X}_{t_n^i}, \widehat{\boldsymbol{\vartheta}}_n^i, \widehat{\lambda}_n^i, \Delta t_{n+1}^i), \sigma_{t_{n+1}^+}^2 \right) := \log \left(p(\sigma_{t_{n+1}^+}^2|\widehat{X}_{t_n^i}, \widehat{\boldsymbol{\vartheta}}_n^i, \widehat{\lambda}_n^i, \Delta t_{n+1}^i) \right).$$

Observe that the higher LS is, the better the prediction will be.

To associate a numerical score to all N samples we further calculate the mean logarithmic score $MLS \left(p(\cdot | \widehat{X}_{t_n}, \widehat{\boldsymbol{\vartheta}}_n, \widehat{\lambda}_n, \Delta t_{n+1}), \sigma_{t_{n+1}+}^2 \right)$ defined as

$$MLS := \frac{1}{N} \sum_{i=1}^N \log \left(p(\sigma_{t_{n+1}+}^2 | \widehat{X}_{t_n}^i, \widehat{\boldsymbol{\vartheta}}_n^i, \widehat{\lambda}_n^i, \Delta t_{n+1}^i) \right).$$

For our previous considered simulation examples we get the following mean logarithmic scores over $N = 1000$ replications

$$\begin{aligned} \text{Gaussian:} & \quad -0.0716 \\ \text{student-t:} & \quad -0.8285. \end{aligned}$$

This result is not surprising, since we assumed normally distributed jumps to compute the prediction density. Thus we would expect a better performance of the prediction, if the jumps are really normally distributed. One also has to mention that the logarithmic score is rather sensitive to outliers which produce large negative values.

Given the prediction density we are able to define a second one-step ahead prediction. More precisely we take the mode of the corresponding prediction distribution as prediction of $\sigma_{t_{n+1}+}^2$, defined by

$$P_m(\sigma_{t_{n+1}+}^2) := \operatorname{argmax} p(\cdot | \widehat{X}_{t_n}, \widehat{\boldsymbol{\vartheta}}_n, \widehat{\lambda}_n, \Delta t_{n+1}).$$

Both predictions P_r and P_m are compared for the two simulation cases by computing the empirical relative bias

$$\widehat{\text{rbias}}(P_k) := \frac{1}{N} \sum_{i=1}^N \left(\frac{P_k(\sigma_{t_{n+1}+}^2)}{\sigma_{t_{n+1}+}^2} - 1 \right)$$

and the empirical relative mean squared error

$$\widehat{\text{RMSE}}(P_k) := \frac{1}{N} \sum_{i=1}^N \left(\frac{P_k(\sigma_{t_{n+1}+}^2)}{\sigma_{t_{n+1}+}^2} - 1 \right)^2,$$

for $k = r, m$. Remember that we have $N = 1000$ replications in each case. The results were the following:

	$\widehat{\text{rbias}}(P_r)$	$\widehat{\text{RMSE}}(P_r)$	$\widehat{\text{rbias}}(P_m)$	$\widehat{\text{RMSE}}(P_m)$
Gaussian	0.0364	0.1301	-0.0895	0.0428
student-t	0.0399	0.1383	-0.0974	0.0517

Concluding one can say that despite the reduced performance of the QMLE in the non-normal case, the one-step ahead prediction ability is acceptable compared to the normal case. The results for the recursive prediction P_r are actually very similar. For the mode prediction greater differences can be observed. This is also rather obvious, since it predicts the mode of a distribution, which is more suitable if the innovations Z are not so heavy tailed. Concerning the RMSE the mode prediction P_m is preferable compared to P_r in both cases, but one has to deal with a larger relative bias of P_m . In Figure 4.2 we plotted for 50 of the 1000 samples the volatility $\sigma_{t_{n+1}+}^2$ (\diamond) together with the two predictions $P_r(\sigma_{t_{n+1}+}^2)$ (*) and $P_m(\sigma_{t_{n+1}+}^2)$ (x). The top row corresponds to the case of normal and the bottom one to the case of non-normal jumps.

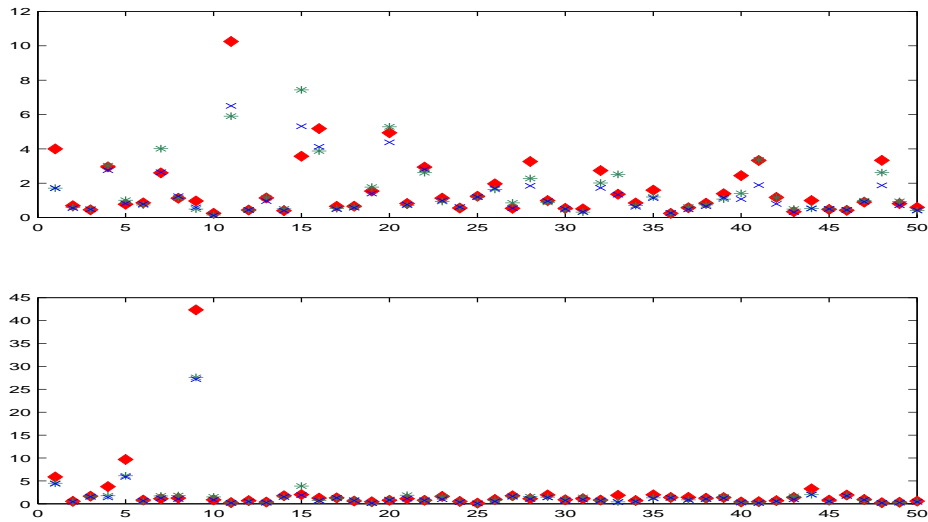


Figure 4.2: Future volatilities $\sigma_{t_{n+1}+}^2$ (\diamond) together with their predictions $P_r(\sigma_{t_{n+1}+}^2)$ (*) and $P_m(\sigma_{t_{n+1}+}^2)$ (x) for normally (*top panel*) and t-distributed (*bottom panel*) jumps.

Finally we want to analyse how the accuracy of the prediction density depends on the number of observations. For the 1 000 sample paths in each case, we estimated the parameters based on all observations G_{t_1}, \dots, G_{t_n} , the last 1 500 $G_{t_{n-1499}}, \dots, G_{t_n}$, the last 1 000 $G_{t_{n-999}}, \dots, G_{t_n}$ and finally the last 500 $G_{t_{n-499}}, \dots, G_{t_n}$. For all four scenarios we computed the prediction density and logarithmic score, which was then averaged over all 1 000 samples. The resulting MLS's are:

Number of observations	n	1 500	1 000	500
MLS (Gaussian)	-0.0716	-0.1426	-0.8002	-3.5185
MLS (student- t)	-0.8285	-0.5776	-3.5121	-5.1267

Remember that the average number of observations over the 1 000 samples is 2 000. Comparing the MLS's one can recognise that at least 1 500 observations should be taken into account to estimate the parameters. If not the prediction quality will be reduced considerably. In our real data example considered in Section 4.4, 1 500 observations correspond to roughly three quarters of a trading day.

4.3.2 Prediction interval for the log-price

If we consider again the log-price process G defined in (4.3) one observes that the future return $G_{t_{i+1}}^{\Delta t_{i+1}}$ has variance $\sigma_{t_{i+1}}^2/\lambda$ under the assumption $Z_{i+1} \sim N(0, 1/\lambda)$. From (4.2) it follows that given X_{t_i} and Δt_{i+1} the volatility $\sigma_{t_{i+1}}^2$ is known up to parameters $\boldsymbol{\vartheta}$ and λ and can be calculated by (4.8). This enables us to construct a 95% prediction interval for the one-step ahead log-price $G_{t_{i+1}}$. It is given by the following expression

$$PI(G_{t_{i+1}}) := \left(G_{t_i} - 1.96 \cdot \sqrt{\widehat{\sigma}_{t_{i+1}}^2(\widehat{\boldsymbol{\vartheta}}_i, \widehat{\lambda}_i)/\widehat{\lambda}_i}, G_{t_i} + 1.96 \cdot \sqrt{\widehat{\sigma}_{t_{i+1}}^2(\widehat{\boldsymbol{\vartheta}}_i, \widehat{\lambda}_i)/\widehat{\lambda}_i} \right).$$

For a simulated log-price process G , with observation times t_1, \dots, t_{n+1} and normally distributed jumps, the empirical quality of the prediction interval was tested. The parameters of the process are set equal to $\boldsymbol{\vartheta} = (-3, 0.1, -0.3, 0.4)$ and $\lambda = 2$. Starting with 100 observations, we reestimated the model for each new observation to get $\widehat{\boldsymbol{\vartheta}}_i, \widehat{\lambda}_i$ and computed in each step the volatility $\sigma_{t_{i+1}}^2(\widehat{\boldsymbol{\vartheta}}_i, \widehat{\lambda}_i)$, $i = 100, \dots, n$. Then we calculated the prediction interval $PI(G_{t_{i+1}})$ and counted over three time intervals $(t_{100}, t_{k_1}]$, $(t_{k_1}, t_{k_2}]$ and $(t_{k_2}, t_{n+1}]$, for $100 < k_1 < k_2 < n + 1$, containing the same number of observations G_{t_i} how many of them are observed within $PI(G_{t_i})$. Notice that the length of the intervals differs for each sample. The results are the following:

	$t_i \in (t_{100}, t_{k_1}]$	$t_i \in (t_{k_1}, t_{k_2}]$	$t_i \in (t_{k_2}, t_{n+1}]$
# $G_{t_i} \in PI(G_{t_i})$ (%)	575 (0.9200)	597 (0.9552)	603 (0.9633)

Even for the first interval the results are already satisfactory but the number of observations lying in the prediction interval is still less than the expected 95%. But already for the second interval this is the case. On the left hand side of Figure 4.3 we plotted the observations $G_{t_{201}}, \dots, G_{t_{300}}$ and on the right hand side the last 50 observations $G_{t_{n-99}}, \dots, G_{t_n}$ together with the corresponding prediction interval.

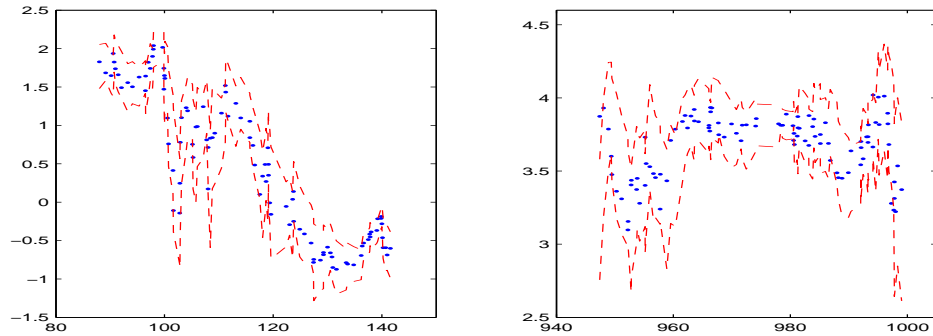


Figure 4.3: Simulated compound Poisson ECOGARCH(1, 1) observations with prediction intervals (*dotted line*).

4.4 Analysis of General Motors stock prices

In this section we will fit the compound Poisson ECOGARCH(1, 1) model to tick-by-tick data for the General Motors (GM) stock. The data spans one week starting from 4th of May 2002 until the 10th of May. The data will be analysed on a daily basis. Therefore no strong seasonality effect will be present. On the other hand we have to take into account a market microstructure noise on this fine level. This will be done by considering mid quotes, which are the average of the last bid and ask quote just before the trade, as our price data. In particular this means, if we have observation points t_1, \dots, t_n , then the log-price G_{t_i} is given by

$$G_{t_i} = \frac{1}{2} (\log(b_{t_i-}) + \log(a_{t_i-})) \cdot 1000, \quad i = 1, \dots, n,$$

where b_{t_i-} (a_{t_i-}) denotes the last bid (ask) quote just before time t_i . This will reduce the effect of bid-ask bounces. We also multiplied the log-price by 1000 not to run into numerical difficulties due to very small values of the volatility process. If equal transaction times occurred, the corresponding trades are combined to a single trade. Further we have omitted the first and last 5 minutes on each trading day due to possible irregularities during that time. Hence we only consider the trading between 9.35 and 15.55. The resulting log-price series together with the corresponding returns are given in Figure 4.4. Observe that we have transformed the observation time such that one unit corresponds to 30 seconds in calendar time.

The parameters are then estimated as explained in Section 4.1. The results are reported in Table 4.5. The parameter values suggest that we have a leverage effect, which is the case if $\hat{\theta} < 0$, on all days except of May 8th.

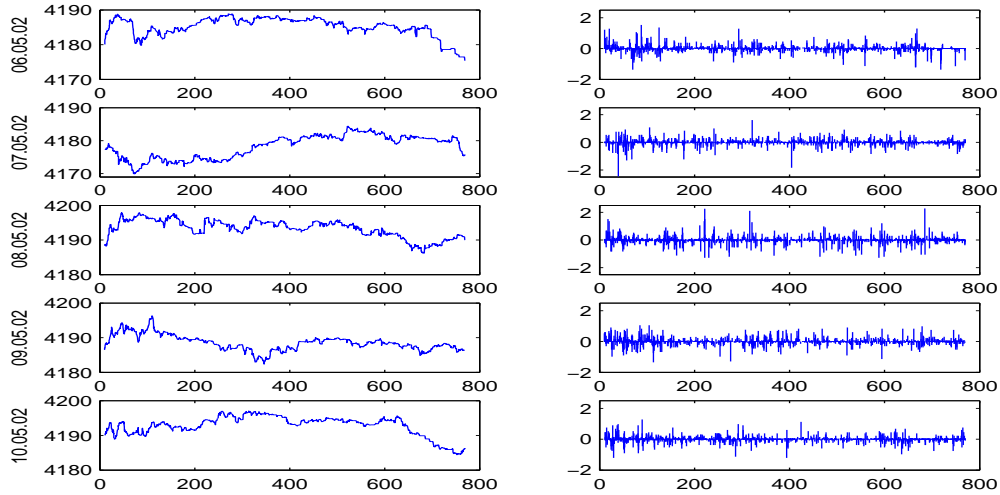


Figure 4.4: Tick-by-tick log-prices G_{t_i} (left) together with the corresponding returns (right) of GM over the time span 06.05.02-10.05.02.

For these dates a positive shock to the log-price will increase the log-volatility less than a negative one. On May the 8th a positive shock increases the log-volatility process more than a negative one. The estimated γ for May the 7th is significantly smaller than on the other days. Therefore we will expect no large jumps in the estimated volatility.

Given the parameter estimates $(\hat{\vartheta}_n, \hat{\lambda}_n)$ we are able, due to equation (4.8), to estimate the volatility for the five days. The results are plotted on the left hand side in Figure 4.5.

Date	\hat{a}_1	$\hat{\theta}$	$\hat{\gamma}$	$\hat{\mu}$	$\hat{\lambda}$
06.05.02	0.2139	-0.0960	0.1701	-3.3392	2.4757
07.05.02	0.2859	-0.0604	0.0667	-3.0909	2.3579
08.05.02	0.5618	0.0691	0.1436	-2.8646	2.7004
09.05.02	0.2533	-0.0477	0.1816	-3.2837	2.4484
10.05.02	0.5972	-0.0230	0.2745	-3.5036	2.5888

Table 4.5: Estimated parameters for the GM data over the time span 06.05.02-10.05.02.

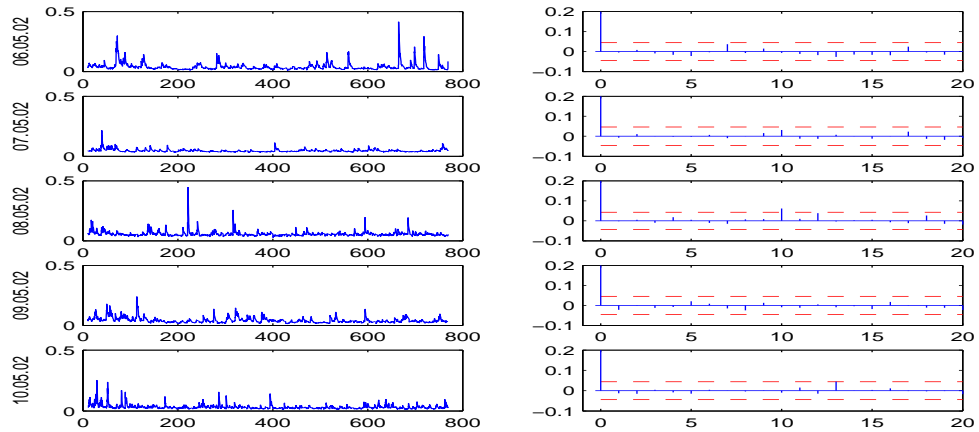


Figure 4.5: Estimated volatility $\sigma_{t_i}^2$ of GM tick-by-tick data (*left*) and estimated autocorrelation function of the squared residuals (*right*) over the time span 06.05.02-10.05.02.

Analogously to the simulation the remaining correlation in the squared fitted innovations was checked by performing a Ljung-Box test. For all five days the hypothesis of no correlation could not be rejected, indicating a suitable fit of the data. This is also confirmed by the empirical autocorrelation function of the squared residuals $\widehat{Z}_i = G_{t_i}^{\Delta t_i} / \widehat{\sigma}_{t_i}$, which can be seen on the right hand side of Figure 4.5.

As explained in Section 4.3.1 we are able to compute a prediction density $p(s|\widehat{X}_{t_n}, \widehat{\vartheta}_n, \widehat{\lambda}_n, \Delta t_{n+1})$ for the one-step ahead volatility $\sigma_{t_{n+1}+}^2$. This is also done for the volatility of the last observation on each of the five days. The densities are shown in Figure 4.6. One observes that the support of the density for May 7th is smaller than for the other days, due to the fact that the log-price process on May 7th was not that volatile. Notice also that up to now we have seen in Figure 4.1 only the shape of the prediction density in the leverage case where a negative jump in the price increases the log-volatility process, while a positive one decreases it.

Here we have for the 8th of May an example for the non-leverage case and on the other days we see prediction densities for the leverage case where a positive shock increases the log-volatility process less than a negative one. In the bottom panel we plotted the corresponding logarithmic prediction densities.

In Section 4.3.2 we introduced a prediction interval $PI(G_{t_{i+1}})$ for the one-step ahead log-price process $G_{t_{i+1}}$. For each of the five days we started with 100 observations to estimate the model and then reestimated it for each new observation up to G_{t_n} . At each time point t_i we calculated the prediction interval $PI(G_{t_{i+1}})$ for

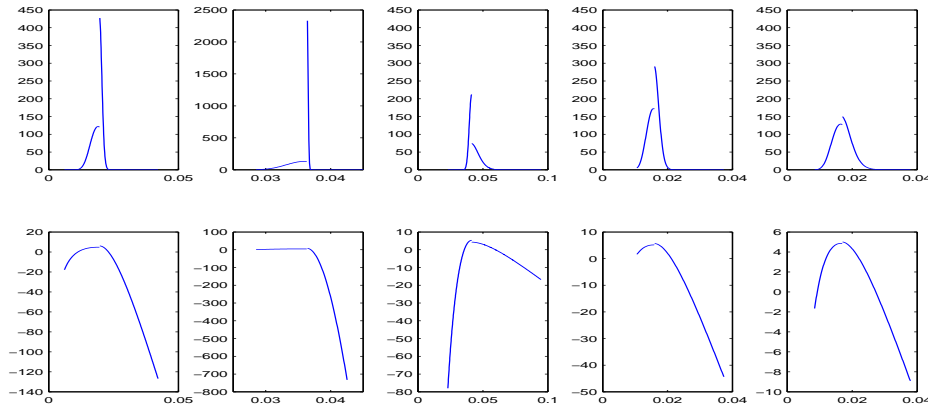


Figure 4.6: Estimated prediction density (*top*) and prediction log density (*bottom*) of the volatility for the last trade on the days 06.05.02 to 10.05.02.

$i = 100, \dots, n$, and then counted how many observations $G_{t_{i+1}}$ were actually observed within $PI(G_{t_{i+1}})$ separately for the time intervals $T_1 = (9.35, 12.00)$, $T_2 = [12.00, 14.00)$ and $T_3 = [14.00, 15.55]$. Let M_k be the number of observation in the time interval T_k , $k = 1, 2, 3$. In Table 4.6 we report $J_k := \#\{G_{t_i} \in PI(G_{t_i}); t_i \in T_k\}$ the number of observations lying in the prediction interval, the percentage J_k/M_k and the average prediction interval length

$$\widehat{PI}_k = \frac{1}{M_k} \sum_{t_j \in T_k} |PI(G_{t_j})| \quad k = 1, 2, 3.$$

Date	J_1 (J_1/M_1); \widehat{PI}_1	J_2 (J_2/M_2); \widehat{PI}_2	J_3 (J_3/M_3); \widehat{PI}_3
06.05.02	600 (0.8982); 3.5602	402 (0.8072); 0.3636	565 (0.8856); 0.3357
07.05.02	593 (0.9324); 4.9761	402 (0.8816); 0.6593	567 (0.9116); 0.5920
08.05.02	728 (0.8656); 0.5790	393 (0.8433); 0.4939	593 (0.8838); 0.5391
09.05.02	539 (0.8242); 1053.6	430 (0.7847); 0.3742	497 (0.8569); 0.3686
10.05.02	599 (0.8139); 0.5194	436 (0.8651); 0.3574	564 (0.8650); 0.3467

Table 4.6: The number (percentage) of observations G_{t_i} lying in the prediction interval $PI(G_{t_i})$ on the days 06.05.02 to 10.05.02.

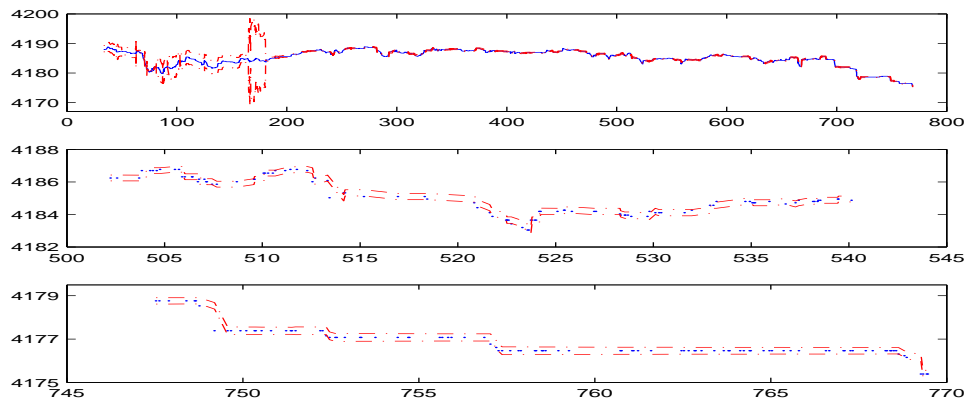


Figure 4.7: Log-price process on May 6th with corresponding prediction intervals over the whole trading day (*top panel*), the last 100 observations in T_2 (*middle panel*) and the last 100 observations in T_3 (*bottom panel*).

The first thing, which stands out are the large average interval length during the first third of the day on May the 6th, 7th and especially the 9th. For the other two days they are comparatively small. In Figure 4.7 we plotted the log-price process on May 6th together with the corresponding prediction intervals.

There one can observe that the interval can actually grow over time and then become smaller again. Hence the percentage numbers for the time interval T_1 can not be taken into account since the intervals are too wide. For the other two time periods J_k/M_k varies between 78% and 91%. If the data were normally distributed, we would have expected about 95% of the observations within the prediction interval. This suggests that the data is heavier tailed than the normal distribution.

Chapter 5

Fractionally integrated ECOGARCH process

In Chapter 3 we have introduced an EGARCH like process in continuous time. We have seen that the ECOGARCH process is a short memory process (see Remark 3.4.4), which is inappropriate in some financial applications. Andersen and Bollerslev (1997) have found empirical evidence for the existence of long-run volatility persistence by analysing high-frequency foreign exchange data. Baillie et al. (1996), Bollerslev and Mikkelsen (1996) and Baillie (1996) provide an overview over long memory processes in econometrics. These findings have to be treated carefully since certain empirical evidence, like a slow decay of the empirical autocorrelation function, could also be due to non-stationarity of the data. This was e.g. shown by Mikosch and Stărică (2000) for a long time series of S&P 500 log-returns. In the following this problem will not be our subject.

In the discrete time GARCH framework there are various models with long range dependence in the volatility process. Among these are the IGARCH(p, q) process of Engle and Bollerslev (1986), the FIGARCH(p, d, q) process proposed by Baillie et al. (1996) or the fractionally integrated EGARCH(p, d, q) process of Bollerslev and Mikkelsen (1996). The FIGARCH process has to be treated carefully since the existence of a stationary version is not clear; see Section 4 in Mikosch and Stărică (2000) and Remark 3.2 in Kazakevičius and Leipus (2003). The FIEGARCH(p, d, q) process is a modification of the EGARCH model of Nelson (1990) in the sense that the log-volatility process is modeled by a fractionally integrated ARMA($q, d, p - 1$) process instead of a short memory ARMA process. This long memory effect introduced in the log-volatility process propagates to the volatility and the squared return process. This was shown by Surgailis and Viano (2002).

In the continuous time setting there are also various approaches to incorporate a long memory effect. Comte and Renault (1998) defined a continuous time stochastic volatility (SV) model by specifying the log-volatility process as an OU process

driven by a fractional Brownian motion. Brockwell and Marquardt (2005) proposed to model the stochastic volatility as a non-negative fractionally integrated CARMA process. Another non-Gaussian continuous time SV model with long memory was introduced by Anh et al. (2002), where they define their model via the Green function solution of a fractional differential equation driven by a Lévy process. Since this shows considerable interest in continuous time models with long memory in the volatility process, the aim of this chapter is to extend the ECOGARCH(p, q) in such a way.

5.1 Fractionally integrated exponential COGARCH

In this section we want to construct a continuous time analogue of the discrete time fractionally integrated EGARCH(p, d, q) process, which is defined in the following way:

Let $p, q \in \mathbb{N}, \mu, \alpha_1, \dots, \alpha_q, \beta_1, \dots, \beta_p \in \mathbb{R}$, suppose $\alpha_q \neq 0, \beta_p \neq 0$ and that the autoregressive polynomial

$$\alpha(z) := 1 - \alpha_1 z - \dots - \alpha_q z^q$$

and the moving average polynomial

$$\beta(z) := \beta_1 + \beta_2 z + \dots + \beta_p z^{p-1}$$

have no common zeros and that $\alpha(z) \neq 0$ on $\{z \in \mathbb{C} \mid |z| \leq 1\}$. Let $(\epsilon_n)_{n \in \mathbb{Z}}$ be an i.i.d. sequence with $\mathbb{E}(\epsilon_1) = 0$ and $\text{Var}(\epsilon_1) = 1$ and $-0.5 < d < 0.5$. Define the measurable function $f : \mathbb{R} \rightarrow \mathbb{R}$ by

$$f(x) := \theta x + \gamma[|x| - \mathbb{E}(|x|)], \quad x \in \mathbb{R}, \tag{5.1}$$

with real coefficients θ and γ . Then we call $(X_n)_{n \in \mathbb{Z}}$, where $X_n = \sigma_n \epsilon_n$, a FIEGARCH(p, d, q) process if

$$\log(\sigma_n^2) = \mu + \alpha(B)^{-1}(1 - B)^{-d}(1 + \beta(B))f(\epsilon_{n-1}),$$

where B is the backward shift operator, $BX_n = X_{n-1}$.

We will define the process using the idea of Klüppelberg et al. (2004) to replace the innovations ϵ_n of the discrete time model by the jumps of a Lévy process $L = (L_t)_{t \geq 0}$.

We consider zero mean Lévy processes L with jumps $\Delta L_t := L_t - L_{t-}$ defined on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$, which is assumed to satisfy the usual conditions. Since $\mathbb{E}(L_t) = t(\gamma_L + \int_{|x| > 1} x \nu_L(dx))$, a zero mean implies that $\gamma_L = - \int_{|x| > 1} x \nu_L(dx)$ and hence the corresponding Lévy symbol is of the form

$$\psi_L(u) = -\sigma_L^2 \frac{u^2}{2} + \int_{\mathbb{R}} (e^{iux} - 1 - iux) \nu_L(dx),$$

and from (1.14) we know that the Lévy-Itô decomposition of L is

$$L_t = B_t + \int_{\mathbb{R}-\{0\}} x \tilde{N}_L(t, dx), \quad t \geq 0, \quad (5.2)$$

where B is a Brownian motion with variance σ_L^2 and $\tilde{N}_L(t, dx) = N_L(t, dx) - t\nu_L(dx)$, $t \geq 0$, is the compensated random measure associated to the Poisson random measure

$$N_L(t, A) = \#\{0 \leq s < t; \Delta L_s \in A\} = \sum_{0 \leq s \leq t} \chi_A(\Delta L_s), \quad A \in \mathcal{B}(\mathbb{R} - \{0\}),$$

on $\mathbb{R}_+ \times \mathbb{R} - \{0\}$, which is independent of B .

Now we are able to define the *fractionally integrated exponential continuous time GARCH* (p, d, q) process, shortly called FIECOGARCH(p, d, q). The stationary log-volatility process will be modeled by a fractionally integrated continuous time ARMA($q, d, p - 1$) process, henceforth called FICARMA($q, d, p - 1$) process (see Chapter 1.4.2 for details on FICARMA processes). The driving noise process of the log-volatility process will be defined similarly to (5.1).

Definition 5.1.1 Let $L = (L_t)_{t \geq 0}$ be a Lévy process with $\mathbb{E}(L_1) = 0$, $\text{Var}(L_1) = 1$ and Lévy measure ν_L and let the $q \times q$ matrix \mathcal{A} and vectors $\mathbf{b} \in \mathbb{R}^q$ and $\mathbf{1}_q \in \mathbb{R}^q$ be defined by

$$\mathcal{A} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ -a_q & -a_{q-1} & -a_{q-2} & \cdots & -a_1 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_{q-1} \\ b_q \end{bmatrix}, \quad \mathbf{1}_q = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}$$

with coefficients $a_1, \dots, a_q, b_1, \dots, b_p \in \mathbb{R}$, where $a_q \neq 0, b_p \neq 0$, and $b_{p+1} = \dots = b_q = 0$. Then for $0 < d < 0.5$ we define the fractionally integrated exponential COGARCH(p, d, q) process G_d as the stochastic process satisfying,

$$dG_{d,t} = \sigma_{d,t} dL_t, \quad t > 0, \quad G_0 = 0, \quad (5.3)$$

where the log-volatility process is given by

$$\log(\sigma_{d,t+}^2) = \mu + \int_{-\infty}^t g_d(t-u) dM_u, \quad t > 0, \quad (5.4)$$

with mean $\mu \in \mathbb{R}$ and initial value $\log(\sigma_{d,0}^2)$ independent of the driving Lévy process L . The process

$$M_t := \int_{\mathbb{R}-\{0\}} h(x) \tilde{N}_L(t, dx), \quad t > 0,$$

is a zero mean Lévy process (see Remark 5.1.2) with

$$h(x) := \theta x + \gamma|x|$$

and parameters $\theta, \gamma \in \mathbb{R}$. The kernel function

$$g_d(t) = \int_0^t g(t-u) \frac{u^{d-1}}{\Gamma(d)} du, \quad 0 < d < 0.5, \quad (5.5)$$

is the Riemann-Liouville fractional integral of order d (see Definition 2.1 in Samko et al. (1993)) of the kernel function $g(t) = \mathbf{b}^T e^{-\mathbf{A}t} \mathbf{1}_q \chi_{(0, \infty)}(t)$.

One has to observe that (5.4) implies predictability of σ^2 . Returns over a time interval of length $r > 0$ are described by the increments of G_d

$$G_{d,t}^{(r)} := G_{d,t} - G_{d,t-r} = \int_{(t-r,t]} \sigma_{d,s} dL_s, \quad t \geq r > 0.$$

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 $(G_{nr}^{(r)})_{n \in \mathbb{N}}$.

In the rest of the chapter the following terminology will be used:

G_d	(log-)price process
$G_d^{(r)}$	(log-)return process
σ_d^2	volatility process
$\log(\sigma_d^2)$	log-volatility process.

Remark 5.1.2 (i) If the real part of the eigenvalues of A is negative, we get from Theorem 5.3 in Samko et al. (1993) that $g_d \in L^2(\mathbb{R})$. Hence the log-volatility process (5.4) is indeed well-defined and stationary, since then

$$\int_{\mathbb{R}} \int_{\mathbb{R}} |g_d(t-s)x|^2 \nu_L(dx) ds < \infty, \quad \forall t \geq 0,$$

and we can apply Theorem 4.3.4 and 4.3.16 in Applebaum (2004) from which the assumptions follow.

In this case the log-volatility process is a zero mean FICARMA(p, d, q) process plus a constant mean μ . In the following we denote the stationary distribution, which is also the limit distribution of $\log(\sigma_{d,t}^2)$ as $t \rightarrow \infty$, by F_d .

(ii) The driving process M is the same as in the ECOGARCH model and was already discussed in Remark 3.2.2. Analogously to the ECOGARCH case the mapping h given in Definition 5.1.1 can be replaced by any other mapping satisfying the integrability condition (3.10).

Alternatively the log-volatility process can be defined in terms of the fractional Lévy process M_d associated with M . We recall the definition of a fractional Lévy process from Marquardt (2006b).

Remark 5.1.3 *Let $M = (M_t)_{t \in \mathbb{R}}$ be a Lévy process on \mathbb{R} with $\mathbb{E}M_1 = 0$, $\mathbb{E}M_1^2 < \infty$ and without Brownian component. For the fractional integration parameter $0 < d < 0.5$ the stochastic process*

$$M_{d,t} = \frac{1}{\Gamma(d+1)} \int_{\mathbb{R}} [(t-s)_+^d - (-s)_+^d] dM_s, \quad t \in \mathbb{R}, \quad (5.6)$$

is called a fractional Lévy process.

The strictly stationary log-volatility process (5.4) is then equal to

$$\mu + \int_{-\infty}^t \mathcal{D}_+^d g_d(t-u) dM_{d,u}, \quad t > 0, \quad (5.7)$$

in the L^2 -sense, where $\mathcal{D}_+^d g_d(x) = \frac{1}{\Gamma(1-d)} \frac{d}{dx} \int_{-\infty}^x \frac{g_d(u)}{(x-u)^d} du$ is the Riemann-Liouville fractional derivative of g_d of order d (see Definition 2.2 in Samko et al. (1993)). Since $g \in L^1(\mathbb{R})$ we get from Theorem 2.4 in Samko et al. (1993) that $\mathcal{D}_+^d g_d = g$. The proof of the equivalence of (5.4) and (5.7) can be found in Marquardt (2006b), Theorem 6.5.

If the Lévy process M is of finite activity, i.e. $\nu_M(\mathbb{R}) < \infty$, then the corresponding fractional Lévy process M_d is of finite variation. In this case the integral in (5.7) can be defined as a Riemann-Stieltjes integral. In case where M is not of finite activity the corresponding fractional Lévy process is not a semimartingale, but for a deterministic integrand the integral with respect to M_d can be defined in the L^2 -sense (we refer to Section 5 of Marquardt (2006b) for details). We do not restrict the driving Lévy process to be of finite activity but we only deal with deterministic integrands and hence this turns out to be sufficient for our purpose.

The log-volatility process (5.7) is now the solution of the continuous time state space model

$$\log(\sigma_{d,t}^2) = \mu + \mathbf{b}^T \mathbf{X}_{d,t-}, \quad t > 0, \quad \log(\sigma_{d,0}^2) = \mu + \mathbf{b}^T \mathbf{X}_{d,0} \quad (5.8)$$

$$d\mathbf{X}_{d,t} = \mathcal{A}\mathbf{X}_{d,t} dt + \mathbf{1}_q dM_{d,t}, \quad t > 0, \quad (5.9)$$

where $\mathbf{X}_{d,0}$ is independent of $(M_{d,t})_{t \geq 0}$ and \mathcal{A} , \mathbf{b} and $\mathbf{1}_q$ are defined in Definition 5.1.1. The state space representation of the log-volatility process is also advantageous for the purpose of simulating the log-price process G_d . The simulation procedure is the following:

- (1) Choose simulation times $0 = t_0 < t_1 < \dots < t_n \leq T$, possibly random.
- (2) Generate increments $M_{d,t_{i+1}} - M_{d,t_i}$, $i = 0, \dots, n-1$, of the driving fractional Lévy process.
- (3) Approximate the state process (5.9) of the log-volatility by a stochastic Euler scheme.
- (4) Compute

$$\log(\hat{\sigma}_{d,t_i}^2) = \mu + \mathbf{b}^T \hat{\mathbf{X}}_{d,t_{i-1}}$$

for $i = 1, \dots, n$.

- (5) Compute an approximation \hat{G}_d by a stochastic Euler scheme:

$$\hat{G}_{d,t_i} = \hat{G}_{d,t_{i-1}} + \hat{\sigma}_{d,t_{i-1}} W_i + \hat{\sigma}_{d,t_{i-1}} J_i,$$

where $W_i \sim N(0, t_i - t_{i-1})$ and J_i is an increment of the jump part of L over the time interval $[t_{i-1}, t_i]$.

Since the fractional Lévy process M_d at time t is an integral with respect to the driving Lévy process M it can be approximated by the corresponding Riemann sums. This approximation is explained in Chapter 2.4.3 in Marquardt (2006a).

Defined in this way $\log(\sigma_d^2)$ is not strictly stationary by definition. The conditions for stationarity of $\log(\sigma_d^2)$, the volatility process σ_d^2 and the return process $G_d^{(r)}$ are summarised in the next proposition. The autocovariance function of the log-volatility process and its asymptotic behaviour is also stated.

Proposition 5.1.4 *Let $\log(\sigma_d^2)$ be defined by (5.8) and (5.9) and G_d as in Definition 5.1.1. If the eigenvalues of \mathcal{A} all have negative real parts and $X_{d,0}$ has the same distribution as $\int_0^\infty e^{\mathcal{A}s} \mathbf{1}_q dM_{d,s}$, then $\log(\sigma_d^2)$ and σ_d^2 are strictly stationary and G_d has strictly stationary increments. The log-volatility process is weakly stationary if $X_{d,0}$ has the same mean and covariance matrix as $\int_0^\infty e^{\mathcal{A}s} \mathbf{1}_q dM_{d,s}$. Let $t > 0, h \geq 0$, then the weakly stationary $\log(\sigma_d^2)$ has autocovariance function*

$$\text{Cov}(\log(\sigma_{d,t+h}^2), \log(\sigma_{d,t}^2)) = \mathbb{E}(M_1^2) \int_{\mathbb{R}} g_d(u+h) g_d(u) du, \quad (5.10)$$

$$\sim C_1 h^{2d-1}, \quad \text{as } h \rightarrow \infty, \quad (5.11)$$

where $C_1 := \frac{\Gamma(1-2d)}{\Gamma(d)\Gamma(1-d)} \mathbb{E}M_1^2 \left(\int_{\mathbb{R}} g(s) ds \right)^2$.

The strict stationarity of $\log(\sigma_d^2)$, σ_d^2 and the increments of G_d follows from the same reasoning as in the short memory case (see Proposition 3.2.3 and Corollary 3.2.5). The proof of (5.10) and (5.11) is given in Marquardt (2006b), Theorem 6.7 and 6.6.

Remark 5.1.5 *The asymptotic behaviour of the autocovariance function of a FI-CARMA process was derived by Brockwell (2004). The result depends on the asymptotic behaviour of the kernel function g_d*

$$g_d(s) \sim \left(\int_{\mathbb{R}} g(x) dx \right) s^{d-1}, \quad \text{for } s \rightarrow \infty, \quad (5.12)$$

which was shown in Brockwell (2004), Section 4. In the following the constant in (5.12) will be denoted by C_2 .

5.2 Second order properties of the volatility process

Second order properties are now derived under the assumption that the log-volatility process is strictly stationary. The stationary distribution F_d (see also Remark 5.1.2) is infinitely divisible, which follows from Theorem 2.7 in Rajput and Rosiński (1989) and the characteristic function of $\log(\sigma_{d,t+}^2)$ is given by

$$\mathbb{E}(e^{iu \log(\sigma_{d,t+}^2)}) = e^{iu\mu} \exp \left\{ \int_0^\infty \psi_M(g_d(s)u) ds \right\}.$$

This implies that the Lévy symbol has the form

$$\begin{aligned} & \log(\mathbb{E}(e^{iu \log(\sigma_{d,t+}^2)})) \\ &= iu \left(\mu + \int_0^\infty g_d(s) \gamma_M ds + \int_0^\infty \int_{\mathbb{R}} g_d(s)x (\chi_{(-1,1)}(g_d(s)x) - \chi_{(-1,1)}(x)) \nu_M(dx) ds \right) \\ & \quad + \int_0^\infty \int_{\mathbb{R}} (e^{iug_d(s)x} - 1 - iug_d(s)x \chi_{(-1,1)}(g_d(s)x)) \nu_M(dx) ds. \end{aligned}$$

The stationary distribution F_d of $\log(\sigma_d^2)$ is therefore specified by the characteristic triplet $(\gamma_{d,\infty}, 0, \nu_{d,\infty})$, where

$$\begin{aligned} \gamma_{d,\infty} &= \mu + \int_0^\infty g_d(s) \gamma_M ds \\ & \quad + \int_0^\infty \int_{\mathbb{R}} g_d(s)x (\chi_{(-1,1)}(g_d(s)x) - \chi_{(-1,1)}(x)) \nu_M(dx) ds \quad (5.13) \end{aligned}$$

$$\nu_{d,\infty}(B) = \int_0^\infty \int_{\mathbb{R}} \chi_B(g_d(s)x) \nu_M(dx) ds, \quad B \in \mathcal{B}(\mathbb{R}). \quad (5.14)$$

The second order behaviour is now summarised in the following proposition.

Proposition 5.2.1 *Let $\log(\sigma_d^2)$ be strictly stationary with marginal distribution F_d , where F_d is infinitely divisible with characteristic triplet $(\gamma_{d,\infty}, 0, \nu_{d,\infty})$. The k -th moment of $\sigma_{d,t}^2$ is finite, if*

$$k \in K_{d,\infty} = \left\{ s \in \mathbb{R} : \int_{|x|>1} e^{sx} \nu_{d,\infty}(dx) < \infty \right\}.$$

In this case

$$\Psi_{d,\infty}(k) := \gamma_{d,\infty}k + \int_{\mathbb{R}} \left(e^{kx} - 1 - kx\chi_{(-1,1)}(x) \right) \nu_{d,\infty}(dx), \quad (5.15)$$

is well defined and

$$\mathbb{E}(\sigma_{d,t}^{2k}) = e^{\Psi_{d,\infty}(k)}, \quad \forall t \geq 0. \quad (5.16)$$

Assume that $\mathbb{E}(\sigma_{d,t}^4) < \infty$. Let $\Psi_{d,\infty}^h(k)$ and $\Psi_d^h(k)$ be defined by (5.15) with kernel function g_d replaced by

$$g_{d,\infty}^h(s) := g_d(s) + g_d(s+h) \quad \text{and} \quad g_d^h(s) := g_d(s)\chi_{(0,h)}(s)$$

respectively. Then the autocovariance function of σ_d^2 is given by

$$\text{Cov}(\sigma_{d,t+h}^2, \sigma_{d,t}^2) = e^{\Psi_{d,\infty}^h(1)} e^{\Psi_d^h(1)} - e^{2\Psi_{d,\infty}(1)}. \quad (5.17)$$

If we replace the kernel functions in the proof of Proposition 3.3.3 and 3.3.4 with the kernel functions $g_d, g_{d,\infty}^h$ and g_d^h , then the result follows.

Next we will show that the long memory property introduced in the log-volatility process implies also a long memory effect in the volatility process. The proof is based on a result for the FICARMA(p, d, q) process which can be found in Lemma 1.23 in Marquardt (2006a).

Theorem 5.2.2 *Let $\log(\sigma_d^2)$ be the strictly stationary long memory process (5.4) with long memory parameter $0 < d < 0.5$ and assume that $2 \in K_{d,\infty}$. Then $\mathbb{E}(\sigma_{d,t}^4) < \infty, \quad \forall t \geq 0$, and*

$$\text{Cov}(\sigma_{d,t+h}^2, \sigma_{d,t}^2) \sim e^{2\Psi_{d,\infty}(1)} C_1 h^{2d-1}, \quad \text{as } h \rightarrow \infty,$$

where $C_1 = \frac{\Gamma(1-2d)}{\Gamma(d)\Gamma(1-d)} \mathbb{E}(M_1^2) \left(\int_{\mathbb{R}} g(s) ds \right)^2$.

Proof: From equation (5.17) it follows that

$$\begin{aligned} & \text{Cov}(\sigma_{d,t+h}^2, \sigma_{d,t}^2) \\ &= e^{\Psi_{d,\infty}^h(1)} e^{\Psi_d^h(1)} - e^{2\Psi_{d,\infty}(1)} = e^{2\Psi_{d,\infty}(1)} (e^{\Psi_{d,\infty}^h(1) + \Psi_d^h(1) - 2\Psi_{d,\infty}(1)} - 1) \\ &= e^{2\Psi_{d,\infty}(1)} \left[\Psi_{d,\infty}^h(1) + \Psi_d^h(1) - 2\Psi_{d,\infty}(1) + O\left((\Psi_{d,\infty}^h(1) + \Psi_d^h(1) - 2\Psi_{d,\infty}(1))^2 \right) \right]. \end{aligned}$$

If we can show that $\Psi_{d,\infty}^h(1) + \Psi_d^h(1) - 2\Psi_{d,\infty}(1) \sim C_1 h^{2d-1}$, as $h \rightarrow \infty$, the result follows. Consider therefore

$$\begin{aligned}
& \Psi_{d,\infty}^h(1) - \Psi_{d,\infty}(1) + \Psi_d^h(1) - \Psi_{d,\infty}(1) \\
&= \int_0^\infty \int_{\mathbb{R}} \left\{ e^{g_{d,\infty}^h(s)x} - 1 - g_{d,\infty}^h(s)x - \left[e^{g_d(s)x} - 1 - g_d(s)x \right] \right\} \nu_M(dx) ds \\
&\quad + \int_0^\infty \int_{\mathbb{R}} \left\{ e^{g_d^h(s)x} - 1 - g_d^h(s)x - \left[e^{g_d(s)x} - 1 - g_d(s)x \right] \right\} \nu_M(dx) ds \\
&= \int_0^\infty \int_{\mathbb{R}} \left\{ e^{g_{d,\infty}^h(s)x} - e^{g_d(s)x} + e^{g_d^h(s)x} - e^{g_d(s)x} \right\} \nu_M(dx) ds \\
&= \int_0^\infty \int_{\mathbb{R}} e^{g_d(s)x} \left\{ e^{g_d(s+h)x} - 1 \right\} \nu_M(dx) ds - \int_h^\infty \int_{\mathbb{R}} \left\{ 1 - e^{g_d(s)x} \right\} \nu_M(dx) ds.
\end{aligned}$$

Series expansion of the exponential function yields

$$\begin{aligned}
& \Psi_{d,\infty}^h(1) - \Psi_{d,\infty}(1) + \Psi_d^h(1) - \Psi_{d,\infty}(1) \\
&= \int_0^\infty \int_{\mathbb{R}} \left[\sum_{k=1}^\infty \frac{(g_d(s+h)x)^k}{k!} + g_d(s)x \sum_{k=1}^\infty \frac{(g_d(s+h)x)^k}{k!} \right. \\
&\quad \left. + \sum_{m=2}^\infty \frac{(g_d(s)x)^m}{m!} \sum_{k=1}^\infty \frac{(g_d(s+h)x)^k}{k!} - \sum_{k=1}^\infty \frac{(g_d(s+h)x)^k}{k!} \right] \nu_M(dx) ds \\
&= \int_0^\infty \int_{\mathbb{R}} \left[x g_d(s) \sum_{k=1}^\infty \frac{(g_d(s+h)x)^k}{k!} \right. \\
&\quad \left. + \sum_{m=2}^\infty \frac{(g_d(s)x)^m}{m!} \sum_{k=1}^\infty \frac{(g_d(s+h)x)^k}{k!} \right] \nu_M(dx) ds \\
&= \int_0^\infty \int_{\mathbb{R}} \left[x^2 g_d(s) g_d(s+h) + g_d(s)x \sum_{k=2}^\infty \frac{(g_d(s+h)x)^k}{k!} \right. \\
&\quad \left. + g_d(s+h)x \sum_{m=2}^\infty \frac{(g_d(s)x)^m}{m!} + \sum_{m=2}^\infty \frac{(g_d(s)x)^m}{m!} \sum_{k=2}^\infty \frac{(g_d(s+h)x)^k}{k!} \right] \nu_M(dx) ds \\
&= \int_0^\infty \int_{\mathbb{R}} \left[g_d(s+h)x \sum_{m=1}^\infty \frac{(g_d(s)x)^m}{m!} \right. \\
&\quad \left. + \sum_{m=1}^\infty \frac{(g_d(s)x)^m}{m!} \sum_{k=2}^\infty \frac{(g_d(s+h)x)^k}{k!} \right] \nu_M(dx) ds.
\end{aligned}$$

Define $M_j := \int_{\mathbb{R}} x^j \nu_M(dx)$, $j \in \mathbb{N}$. Since $\int_{|x|>1} e^x \nu_{d,\infty}(dx) < \infty$ implies that $\int_{\mathbb{R}} |x|^k \nu_M(dx) < \infty$, $k \geq 2$, we get that all moments M_j , $j \geq 2$, of the Lévy measure

ν_M are finite. Consider now the integral

$$I_1(h) := \int_0^\infty g_d(s+h)g_d(s) \left[M_2 + \sum_{m=2}^\infty \frac{(g_d(s))^{m-1}}{m!} M_{m+1} \right] ds.$$

We want to show that

$$I_1(h) \sim M_2 \int_0^\infty G_d(s+h)G_d(s)ds =: I_G(h), \quad \text{as } h \rightarrow \infty, \quad (5.18)$$

with $G_d(s) := C_2 s^{d-1}$, since $I_G(h) \sim C_1 h^{2d-1}$, $h \rightarrow \infty$. We show first, that

$$\begin{aligned} g_d(s+h+h^{d/2})g_d(s+h^{d/2}) \left[M_2 + \sum_{m=2}^\infty \frac{(g_d(s+h^{d/2}))^{m-1}}{m!} M_{m+1} \right] \\ \sim M_2 G_d(s+h+h^{d/2})G_d(s+h^{d/2}), \end{aligned}$$

if $h \rightarrow \infty$. Consider therefore the limit

$$\begin{aligned} \lim_{s \geq 0, h \rightarrow \infty} \frac{g_d(s+h+h^{d/2})g_d(s+h^{d/2}) \left[M_2 + \sum_{m=2}^\infty \frac{(g_d(s+h^{d/2}))^{m-1}}{m!} M_{m+1} \right]}{M_2 G_d(s+h+h^{d/2})G_d(s+h^{d/2})} \\ = 1 + \lim_{s \geq 0, h \rightarrow \infty} M_2^{-1} \sum_{m=2}^\infty \frac{(g_d(s+h^{d/2}))^{m-1}}{m!} M_{m+1}, \end{aligned}$$

which is equal to 1 because of (5.12),

$$\begin{aligned} \left| g_d(s+h^{d/2})^{-1} \sum_{m=2}^\infty \frac{(g_d(s+h^{d/2}))^m}{m!} M_{m+1} \right| \\ \leq M^* |g_d(s+h^{d/2})|^{-1} \left(e^{|g_d(s+h^{d/2})|} - |g_d(s+h^{d/2})| - 1 \right), \end{aligned}$$

with $M^* := \sup_{j \in \mathbb{N}} |M_j| < \infty$, and $\lim_{x \geq 0, x \rightarrow 0} x^{-1}(e^x - x - 1) = 0$. From Lemma 1.22 in Marquardt (2006a) it follows that

$$\tilde{I}_1(h) \sim \tilde{I}_G(h), \quad \text{for } h \rightarrow \infty, \quad (5.19)$$

where

$$\tilde{I}_1(h) := \int_{h^{d/2}}^\infty g_d(s+h)g_d(s) \left[M_2 + \sum_{m=2}^\infty \frac{(g_d(s))^{m-1}}{m!} M_{m+1} \right] ds$$

and

$$\tilde{I}_G(h) := M_2 \int_{h^{d/2}}^\infty G_d(s+h)G_d(s)ds.$$

Now (5.18) follows if we can show

$$\frac{|I_1(h) - I_G(h)|}{|I_G(h)|} \leq \frac{|I_1(h) - \tilde{I}_1(h)|}{|I_G(h)|} + \frac{|\tilde{I}_1(h) - \tilde{I}_G(h)|}{|I_G(h)|} + \frac{|\tilde{I}_G(h) - I_G(h)|}{|I_G(h)|} \rightarrow 0$$

for $h \rightarrow \infty$. This can be done in a similar way as in the proof of Lemma 1.23 in Marquardt (2006a). In particular, since $|I_G(h)| \geq |\tilde{I}_G(h)|$ it follows from (5.19) that

$$\frac{|\tilde{I}_1(h) - \tilde{I}_G(h)|}{|I_G(h)|} \leq \frac{|\tilde{I}_1(h) - \tilde{I}_G(h)|}{|\tilde{I}_G(h)|} \rightarrow 0.$$

For $d < 0.5$ we get $|I_G(h)| \geq |C_2|^2 \frac{h^{2d-1}}{1-2d}$ and for all $h \geq K$, K large enough, we have $|g_d(s+h)| \leq 2|C_2|h^{d-1}$. There exists also a constant $C_g > 0$ with $\sup_{s \geq 0} |g_d(s)| \leq C_g$. This yields for $h \geq K$

$$\begin{aligned} |I_1(h) - \tilde{I}_1(h)| &= \left| \int_0^{h^{d/2}} g_d(s+h)g_d(s) \left[M_2 + \sum_{m=2}^{\infty} \frac{(g_d(s))^{m-1}}{m!} M_{m+1} \right] ds \right| \\ &\leq \int_0^{h^{d/2}} 2|C_2|h^{d-1}C_g [M_2 + M^*C_g^{-1}(e^{C_g} - C_g - 1)] ds \\ &\leq 2|C_2|C_g [M_2 + M^*C_g^{-1}(e^{C_g} - C_g - 1)] h^{2d-1-d/2} \end{aligned}$$

and hence

$$\frac{|I_1(h) - \tilde{I}_1(h)|}{|I_G(h)|} \leq \frac{2C_g [M_2 + M^*C_g^{-1}(e^{C_g} - C_g - 1)] h^{2d-1-d/2}}{|C_2|^2 \frac{h^{2d-1}}{1-2d}} \rightarrow 0 \quad \text{for } h \rightarrow \infty.$$

Similarly we get

$$\frac{|\tilde{I}_G(h) - I_G(h)|}{|I_G(h)|} \leq \frac{C_g h^{2d-1-d/2}}{|C_2|^2 \frac{h^{2d-1}}{1-2d}} \quad \text{for } h \rightarrow \infty,$$

from which the result follows. Analogously we get with

$$\int_0^\infty g_d^k(s+h)g_d(s)ds \sim C_3 h^{(k+1)d-k}, \quad k \geq 2,$$

that

$$I_k(h) := \int_0^\infty \frac{1}{k!} g_d^k(s+h)g_d(s) \left[M_{k+1} + \sum_{m=2}^{\infty} \frac{(g_d(s))^{m-1}}{m!} M_{m+k} \right] ds = o(h^{2d-1}).$$

and hence it follows that

$$\Psi_{d,\infty}^h(1) + \Psi_d^h(1) - 2\Psi_{d,\infty}(1) \sim C_1 h^{2d-1}, \quad \text{for } h \rightarrow \infty,$$

which proves the assertion. \square

Example 5.2.3 *In this example we consider a fractionally integrated ECGARCH(1, 0.4, 1) process driven by a Lévy process L with Lévy symbol*

$$\psi_L(u) = -\frac{u^2}{2} + \int_{\mathbb{R}} (e^{iux} - 1) \lambda \Phi_{0,1/\lambda}(dx),$$

where $\Phi_{0,1/\lambda}(\cdot)$ is the distribution function of a normal distribution with mean 0 and variance $1/\lambda$. This means that L is the sum of a standard Brownian motion W and the compound Poisson process

$$J_t = \sum_{k=1}^{N_t} Z_k, \quad J_{-t} = \sum_{k=1}^{-N_{-t}} Z_{-k}, \quad t \geq 0,$$

where $(N_t)_{t \in \mathbb{R}}$ is an independent Poisson process with intensity $\lambda > 0$ and jump times $(T_k)_{k \in \mathbb{Z}}$. The Poisson process N is also independent from the i.i.d. sequence of jump sizes $(Z_k)_{k \in \mathbb{Z}}$, with $Z_1 \sim N(0, 1/\lambda)$. The Lévy process M is in this case given by the following expression

$$M_t = \sum_{k=1}^{N_t} [\theta Z_k + \gamma |Z_k|] - Ct, \quad t > 0,$$

with $C = \gamma \int_{\mathbb{R}} |x| \lambda \Phi_{0,1/\lambda}(dx) = \sqrt{\frac{2\lambda}{\pi}} \gamma$. M_{-t} , $t \geq 0$ is defined analogously. The parameter θ is equal to -0.15 and γ is equal to 0.1 . The stationary log-volatility process is of the form

$$\log(\sigma_{d,t+}^2) = \mu + \int_{-\infty}^t b_1 e^{-a_1(t-s)} dM_{d,s}, \quad t > 0,$$

where $\mu = -5$, $a_1 = 0.5$ and $b_1 = 1$. In Figure 5.1 we plotted 3 000 observations of the sample path of the simulated log-volatility $\log(\sigma_d^2)$ and volatility σ_d^2 process observed at 10 000 equidistant time points in the first row. The long memory parameter d in this example is 0.4. Hence we will expect a slow decay of the autocorrelation function (acf). This is indeed the case. The empirical autocorrelation function of the volatility σ_d^2 and log-volatility $\log(\sigma_d^2)$ process are shown in the bottom left panel of Figure 5.1 as a dashed and solid line, respectively. One observes that the empirical acf of both series show similar asymptotic behaviour as indicated from Theorem 5.2.2. In the bottom right panel the corresponding return process is shown.

Remark 5.2.4 *In the last Theorem we have shown, that the autocovariance function of the volatility process decays at a hyperbolic rate. For the discrete time EGARCH process this was shown by Surgailis and Viano (2002).*

In the continuous time setting Comte and Renault (1998) showed this effect for a long memory stochastic volatility model, where the log-volatility process was modeled

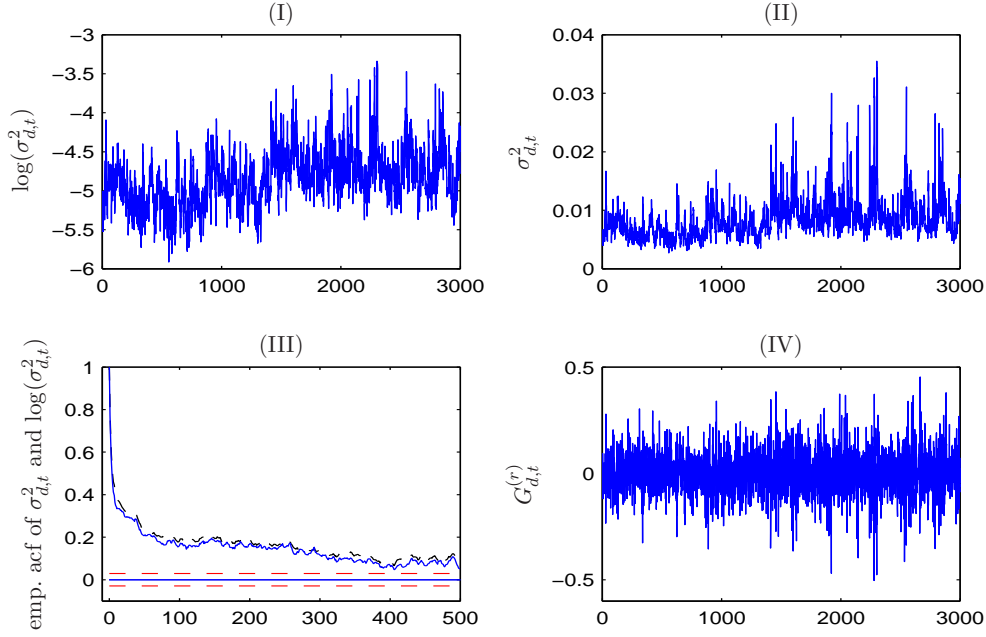


Figure 5.1: (I) The log-volatility process $\log(\sigma_d^2)$ and (II) volatility process σ_d^2 of a FIECOGARCH(1, 0.4, 1) with parameters $a_1 = 0.5, b_1 = 1, \mu = -5, \theta = -0.15, \gamma = 0.1$ and $d = 0.4$. (III) The empirical autocorrelation function of σ_d^2 (solid line) and $\log(\sigma_d^2)$ (dashed line). (IV) 3000 observations of the return process $G_d^{(1)}$. The jumps of the compound Poisson process are $N(0, 1/2)$ distributed.

as an OU process driven by a fractional Brownian motion. Hence our result can also be applied to a continuous-time stochastic volatility model, where the log-price process $Y = (Y_t)_{t \geq 0}$ satisfies

$$dY_t = \sigma_t dW_t, \quad t \geq 0, \quad (5.20)$$

with a Brownian motion W , and the log-volatility process $\log(\sigma^2)$ is described by a FICARMA(p, d, q), $p \geq q$, process, where the Lévy measure of the driving noise process has finite moments of all orders.

5.3 Second order properties of the return process

Second order properties are now derived under the assumption that the log-volatility process is strictly stationary. The structure of the price process G_d is the same as

that of an ECOGARCH(p, q) process. Therefore the result concerning the first and second moment, as well as the autocovariance function, of the return process is analogous to the result in Proposition 3.4.1.

Proposition 5.3.1 *Let L be a Lévy process with $\mathbb{E}(L_1) = 0$ and $\mathbb{E}(L_1^2) < \infty$. Assume that $\log(\sigma_d^2)$ is strictly stationary with marginal distribution F_d , where F_d is infinitely divisible with characteristic triplet $(\gamma_{d,\infty}, 0, \nu_{d,\infty})$ and $1 \in K_{d,\infty}$. Then $\mathbb{E}(G_{d,t}^2) < \infty$ for all $t \geq 0$, and for every $t, h \geq r > 0$ it holds*

$$\mathbb{E}G_{d,t}^{(r)} = 0 \quad (5.21)$$

$$\mathbb{E}(G_{d,t}^{(r)})^2 = e^{\Psi_{d,\infty}(1)r} \mathbb{E}(L_1^2) \quad (5.22)$$

$$\text{Cov}(G_{d,t}^{(r)}, G_{d,t+h}^{(r)}) = 0. \quad (5.23)$$

If further $\mathbb{E}(L_1^4) < \infty$ and the volatility process has finite second moment, then $\mathbb{E}(G_{d,t}^4) < \infty$ for all $t \geq 0$ and for every $t, h \geq r > 0$ it holds

$$\text{Cov}((G_{d,t}^{(r)})^2, (G_{d,t+h}^{(r)})^2) = \mathbb{E}(L_1^2) \int_h^{h+r} \text{Cov}(G_{d,r}^2, \sigma_s^2) ds. \quad (5.24)$$

Chapter 6

Mixed effect models for absolute log returns of ultra high frequency data

In the previous chapters we have considered models for the log-price process of financial data with stochastic volatility. Now we want to concentrate on providing estimates just for the volatility, but allowing for explanatory variables and accounting for market micro structures. For this we use ultra high frequency (uhf) financial data. The term uhf data was defined by Engle (2000). He calls financial data uhf data, if they consist of all transactions and quotes recorded during the trading day. The recorded transactions of course do not take place at regularly spaced time points, i.e. we have to analyse irregularly spaced time series. One way would be to sample it at a given frequency, but this results in a loss of information. Therefore we set up a model directly dealing with this irregular time spacing. Our object of interest will be the absolute log return, which is a proxy for the unobservable instantaneous standard deviation σ_{t_i} , where t_i is the time of the i -th trade, of the log price $S_{t_i} = \log(P_{t_i})$. By modeling the mean of the absolute log returns, we get a model-based estimate of the instantaneous standard deviation. This could then be used for example, as in Jungbacker and Koopman (2006), to estimate actual volatility of the interval $[t_i, t_j]$, $j > i$, given by

$$\sigma^{*2}(t_i, t_j) = \int_{t_i}^{t_j} \sigma_t^2 dt$$

based on all available information. Here it is important to account for microstructure noise, when dealing with ultra high frequencies. The problem of market microstructure noise at this frequency is for example explained in Ait-Sahalia et al. (2005). It is more common to account for microstructure effects on the return level, while we will account for these effects on the absolute log return scale. This is more appropriate

in the context of the regression set-up we follow for the absolute log returns. The absolute log-return $|S_{t_i} - S_{t_{i-1}}|$ will be modeled in this Chapter given the past information $\mathcal{G}_{t_{i-1}} = \sigma(S_{t_j}, d_{t_j}; j \leq i-1)$ and current duration $d_{t_i} = t_i - t_{i-1}$. Since the duration process is a stochastic process itself one also needs a model for this regularly spaced (measured in tick time) time series. A popular model for the durations given the past information, called *Autoregressive Conditional Duration* (ACD) model, has been proposed by Engle and Russell (1998). There are a number of modifications of the ACD model, which are described for example in Bauwens et al. (2004).

To cope with the problem of unequally spaced data, we will assume a continuous time parameter price process. The absolute log returns will be the response in a regression framework with the current duration as one of the explanatory variables and correlated residuals. They have the correlation structure of a continuous time ARMA process. The estimation of correlation for unequally spaced time series is problematic, since e.g. the sample autocorrelation function cannot be estimated directly. We compute the sample variogram, which is defined in terms of increments and therefore adequate for irregularly spaced observations. We have already said, that the absolute log return is viewed in this paper as a noisy measure of instantaneous volatility. It can be decomposed into a fixed effect, a random effect and a measurement error. The fixed effect describes the time dependent mean of the data, whereas the random effect specifies the correlation structure. Since the fixed effect is a function of time varying explanatory variables it allows for time of day effects (see for example Bauwens and Giot (2001)). The measurement error accounts for the market microstructure noise on this absolute return level. The presence of microstructure effects also allows us to assume the mean function to be a continuous variable, despite the fact that the prices are multiple of one cent of a US dollar. The return of irregularly spaced transaction data is also modeled as a continuous variable for example in Meddahi et al. (1998), whereas Russell and Engle (2005) or Liesenfeld and Pohlmeier (2003) assume that it takes on only countably many values. The influence of the explanatory variables will be modeled in a parametric way, which allows us to compute predictions based on past information and current duration in a very easy way. By using the mean squared error as scoring rule, we are able to quantify the loss in predictive power, when duration is not used as an explanatory variable. Here we would like to mention that initially we are interested in detecting certain dependencies between the response and the explanatory variables. In a further step one could think about additionally applying an ACD model to compute predictions in real applications. Visualisation of the explanatory variable effect on the absolute log returns on a daily basis is also possible. Renault and Werker (2006) studied the instantaneous causality effect from transaction durations to price volatility and found significant empirical evidence for it. There are also further regression models with measures of volatility as response. Corsi (2004), Andersen et al. (2006) and Ghysels et al. (2002) have set-up different kinds of linear regression models with for example realized volatility (see Barndorff-Nielsen and Shephard (2002)) as response.

An overview over these three models can be found in Forsberg and Ghysels (2004). As we have already mentioned, Jungbacker and Koopman (2006) estimated actual volatility of ultra-high frequency data in a model-based approach. They considered a state space model for the return process, which is defined for every second. This leads to a missing values problem. We also used a state space approach, but rather prefer to work with time dependent matrices, to account for the irregular time spacing, than to deal with a large number of missing values per day.

6.1 A mixed effect regression model for irregularly spaced data

The main characteristic of the data we deal with is that we have observations at irregularly spaced time points. Therefore we think it is natural to assume that these observations are observations from a continuous time model. The volatility of high frequency data is often modeled as a continuous time linear process. Examples are the stochastic volatility model in Barndorff-Nielsen and Shephard (2001) or Jungbacker and Koopman (2006) or the continuous time GARCH(1,1) model of Klüppelberg, Lindner, and Maller (2004) (see also Chapter 2). Since the absolute log return is a measure of the instantaneous standard deviation, we will model them in such a way that they have the autocorrelation structure of a continuous time linear process. To be precise, we assume the autocorrelation structure of a continuous time ARMA(p, q) process as defined in Chapter 1.4.

6.1.1 Regression mean specification

Ultra high frequency data exhibit some time of day effects (see for example Bauwens and Giot (2001)), which result in a non-stationary time series. We try to explain these effects as being influenced by explanatory variables, which have time of day dependent values. In our set-up this explanatory variable information is used to model the mean of the data,

$$\mu_{t_i} := \mathbb{E}(|r_{t_i}|),$$

with

$$|r_{t_i}| := |\log(P_{t_i}) - \log(P_{t_{i-1}})| \cdot 100, \quad i = 1, \dots, n, \quad (6.1)$$

where P_{t_i} is the stock price observed at time t_i , as in a typical regression set-up. There will be no assumption made about a stock price model, except that we assume, that it is a continuous time process. To assure positivity of the mean we will use a log-link, i.e.

$$\log(\mu_{t_i}) := \mathbf{x}_{t_i}^T \boldsymbol{\beta}, \quad i = 1, \dots, n, \quad (6.2)$$

with $\mathbf{x}_{t_i}^T \in \mathbb{R}^{1 \times s+1}$ the i -th row of the design matrix

$$\mathbf{X} = \begin{pmatrix} \mathbf{x}_{t_1}^T \\ \vdots \\ \mathbf{x}_{t_n}^T \end{pmatrix} \in \mathbb{R}^{n \times s+1}$$

and parameter vector $\boldsymbol{\beta}^T := (\beta_0, \dots, \beta_s)^T \in \mathbb{R}^{s+1 \times 1}$. As can be seen from (6.2), a parametric approach is taken. The specific structure of the design matrix will be discussed in the applications. Potential explanatory variables are

$$\begin{aligned} b_{t_i} &:= \text{the last bid-ask spread before time } t_i \\ d_{t_i} &:= \text{the duration } t_i - t_{i-1} \\ v_{t_i} &:= \text{the volume of the the last trade before time } t_i. \end{aligned}$$

The choice of explanatory variables will be discussed in the applications. The explanatory variable d_{t_i} is unknown before time t_i and has therefore to be estimated, by some *autoregressive conditional duration model*, if the model is used for prediction.

6.1.2 Correlated residuals

As we have said in the beginning we model the absolute log returns as an autocorrelated process. The question is if autocorrelation is really present in this uhf data. The answer to this question is part of the analysis. The problem with empirically estimating the autocorrelation in uhf data is the irregular time structure. Therefore the empirical autocorrelation function cannot be computed. One way out is to consider the variogram (it will be introduced and discussed in Appendix B), which is defined for irregularly spaced data. But the variogram is also not defined for $(|r_{t_i}|)$, because the mean of the increments is not a linear function of the time lag, i.e. $\mathbb{E}(|r_t| - |r_s|) \neq C \cdot (t - s)$, which has to be the case. The variogram is however defined, when we consider the residuals

$$\varepsilon_{t_i} := |r_{t_i}| - \mu_{t_i}, \quad i = 1, \dots, n, \quad (6.3)$$

with $\mathbb{E}(\varepsilon_{t_i}) = 0$ and $\text{Var}(\varepsilon_{t_i}) =: \sigma_\varepsilon^2$. The ε_{t_i} are autocorrelated because of the following assumption

$$\varepsilon_{t_i} =: Y_{t_i} + \tilde{\varepsilon}_{t_i}, \quad i = 1, \dots, n, \quad (6.4)$$

where Y is a CARMA(p,q) process, which is defined in Chapter 1.4, and $(\tilde{\varepsilon}_{t_i})$ is an i.i.d. sequence and uncorrelated with (Y_{t_i}) . To motivate (6.4) think of (Y_{t_i}) as the random effect of the absolute log returns, which describes their correlation structure. The mean, as we have already said, will be accounted for by μ_{t_i} . But since we will not observe $\mu_{t_i} + Y_{t_i}$ due to some microstructure noise, like for example the fixed

tick size of the log returns, we will make some measurement error $\tilde{\varepsilon}_{t_i}$. To assure that Y is non-negative, the driving Lévy process L of the CARMA(p,q) process Y has to be non-decreasing and the kernel of Y has to be non-negative. By substituting (6.4) into (6.3) we get

$$\tilde{\varepsilon}_{t_i} = |r_{t_i}| - \mu_{t_i} - Y_{t_i},$$

which leads to

$$\mathbb{E}(\tilde{\varepsilon}_{t_i}) = -\mathbb{E}(Y_{t_i}) = \mathbf{b}^T \mathcal{A}^{-1} \mathbf{1}_q \mathbb{E}(L_1).$$

The variance of ε_{t_i} decomposes into

$$\begin{aligned} \sigma_\varepsilon^2 &= \text{Var}(Y_{t_i}) + \text{Var}(\tilde{\varepsilon}_{t_i}), \\ &=: \text{Var}(L_1) \mathbf{b}^T \Sigma \mathbf{b} + \sigma_{\tilde{\varepsilon}}^2, \end{aligned}$$

and the autocovariance function of (ε_{t_i}) is equal to that of (Y_{t_i}) , i.e.

$$\text{Cov}(\varepsilon_{t_i}, \varepsilon_{t_{i-1}}) = \text{Var}(L_1) \mathbf{b}^T e^{\mathcal{A}(t_i - t_{i-1})} \Sigma \mathbf{b}.$$

6.1.3 A generalised regression model with CARMA(p,q) random effects

The above considerations have led us to the model

$$|r_{t_i}| = \exp(\mathbf{x}_{t_i}^T \boldsymbol{\beta}) + Y_{t_i} + \tilde{\varepsilon}_{t_i}, \quad i = 1, \dots, n. \quad (6.5)$$

In (6.5) we will understand $\exp(\mathbf{x}_{t_i}^T \boldsymbol{\beta})$ as some fixed effect, Y_{t_i} as some random effect and $\tilde{\varepsilon}_{t_i}$ as a measurement error. The parameters which have to be estimated are

$$\boldsymbol{\theta} := (a_1, \dots, a_p, b_1, \dots, b_q, \sigma^2, \beta_0, \dots, \beta_s, \sigma_{\tilde{\varepsilon}}^2),$$

with $\sigma^2 := \text{Var}(L_1)$. This is done by an iterated estimation algorithm, which will be described in the next section.

6.2 Parameter Estimation

The actual parameter estimation can be done in two ways. The first one (henceforth called direct approach) works directly on the linear regression model approximation to model (6.5), which will be introduced in the following, and the second one (henceforth called state space approach) on the associated state space model with application of the Kalman filter. Both estimation procedures will be explained in Chapter 6.2.1 and 6.2.2, respectively. But first we start by describing the general estimation algorithm. Therefore consider equation (6.3) in vector notation

$$|\mathbf{r}| = \boldsymbol{\mu} + \boldsymbol{\varepsilon}, \quad (6.6)$$

with $|\mathbf{r}| = (|r_{t_1}|, \dots, |r_{t_n}|)^T$, $\boldsymbol{\mu}$ and $\boldsymbol{\varepsilon}$ similarly. Since we chose the logarithm as link function, we have the relationship

$$\log(\boldsymbol{\mu}) = \boldsymbol{\mathcal{X}}\boldsymbol{\beta} =: \boldsymbol{\eta}. \quad (6.7)$$

The covariance matrix of $\boldsymbol{\varepsilon}$ will be denoted by

$$\boldsymbol{\mathcal{V}}(\boldsymbol{\xi}) = \text{Cov}(\mathbf{Y}) + \sigma_{\boldsymbol{\varepsilon}}^2 \boldsymbol{\mathcal{I}}_n,$$

with $\boldsymbol{\xi} := (a_1, \dots, a_p, b_1, \dots, b_q, \sigma^2, \sigma_{\boldsymbol{\varepsilon}}^2)$ and $\mathbf{Y} = (Y_{t_1}, \dots, Y_{t_n})^T$. Equation (6.6) is just a nonlinear regression model with correlated errors. Therefore the parameters can be estimated by maximizing

$$G(\boldsymbol{\theta}, |\mathbf{r}|) := -(|\mathbf{r}| - \boldsymbol{\mu})^T \boldsymbol{\mathcal{V}}(\boldsymbol{\xi})^{-1} (|\mathbf{r}| - \boldsymbol{\mu}). \quad (6.8)$$

Applying the Fisher scoring algorithm to maximize (6.8) leads to an iterative generalised least squares problem. The linear model, occurring in each iteration step, can be constructed as in generalised linear models (McCullagh and Nelder (1983) p.40) by applying the link function $g(\cdot) := \log(\cdot)$ to the data $|\mathbf{r}|$ and linearise to the first order. The estimation algorithm, which can also be found e.g. in Schall (1991), is described in the following.

General Estimation Algorithm:

- (1) Linearise $\mathbf{g}(|\mathbf{r}|) := (g(|r_{t_1}|), \dots, g(|r_{t_n}|))^T$ to the first order

$$g(|\mathbf{r}|) = g(\boldsymbol{\mu}) + \partial_{\boldsymbol{\mu}} g(\boldsymbol{\mu})(|\mathbf{r}| - \boldsymbol{\mu}),$$

where $\partial_{\boldsymbol{\mu}} g(\boldsymbol{\mu}) = \text{diag}(\frac{\partial}{\partial \mu_{t_1}} g(\mu_{t_1}), \dots, \frac{\partial}{\partial \mu_{t_n}} g(\mu_{t_n}))$, and define the new dependent variable

$$\begin{aligned} \mathbf{z} &:= g(\boldsymbol{\mu}) + \partial_{\boldsymbol{\mu}} g(\boldsymbol{\mu})(|\mathbf{r}| - \boldsymbol{\mu}) \\ &= \boldsymbol{\eta} + \partial_{\boldsymbol{\mu}} g(\boldsymbol{\mu})\boldsymbol{\varepsilon} \\ &= \boldsymbol{\eta} + \mathbf{e}, \end{aligned}$$

where $\mathbf{e} := \partial_{\boldsymbol{\mu}} \boldsymbol{\eta} \boldsymbol{\varepsilon}$. Now we have a linear regression model with correlated errors

$$\mathbf{z} = \boldsymbol{\mathcal{X}}\boldsymbol{\beta} + \mathbf{e}, \quad (6.9)$$

where $\mathbb{E}(\mathbf{z}) = \boldsymbol{\mathcal{X}}\boldsymbol{\beta}$ and $\text{Cov}(\mathbf{e}) = \partial_{\boldsymbol{\mu}} \boldsymbol{\eta} \boldsymbol{\mathcal{V}}(\boldsymbol{\xi}) \partial_{\boldsymbol{\mu}} \boldsymbol{\eta}^T$.

- (2) To get starting values $\hat{\boldsymbol{\eta}}^0, \hat{\mathbf{z}}^0$ we fit a generalised linear model to (6.6) assuming uncorrelated errors, i.e. $\text{Cov}(\boldsymbol{\varepsilon}) = \sigma_{\boldsymbol{\varepsilon}}^2 \boldsymbol{\mathcal{I}}_n$.
- (3) Start Iteration $k = 1$

- (4) The parameters $\boldsymbol{\beta}$ and $\boldsymbol{\xi}$ in (6.9) are then estimated in the direct or state space approach giving parameter estimates

$$\hat{\boldsymbol{\beta}}^k \text{ and } \hat{\boldsymbol{\xi}}^k,$$

respectively.

- (5) Construct new estimates of $\boldsymbol{\eta}$, i.e. define

$$\hat{\boldsymbol{\eta}}^k := \boldsymbol{\mathcal{X}} \hat{\boldsymbol{\beta}}^k.$$

Check if

$$\|\hat{\boldsymbol{\eta}}^k - \hat{\boldsymbol{\eta}}^{k-1}\| < TOL$$

is satisfied. If not set

$$\begin{aligned} \hat{\boldsymbol{\mu}}^k &:= g^{-1}(\hat{\boldsymbol{\eta}}^k) \\ \hat{\mathbf{z}}^k &:= \hat{\boldsymbol{\eta}}^k + \left(\partial_{\boldsymbol{\mu}} \hat{\boldsymbol{\eta}}^k \Big|_{\boldsymbol{\mu}=\hat{\boldsymbol{\mu}}^k} \right) (|\mathbf{r}| - \hat{\boldsymbol{\mu}}^k) \end{aligned}$$

$k = k + 1$ and go to (4).

Both estimation approaches will perform *quasi maximum likelihood* (QML) estimation (see for example Chapter 2 in White (1994)) of the parameters, which requires only the knowledge of the first two moments of the model for the data. In particular the quasi maximum likelihood estimate (QMLE) $\hat{\boldsymbol{\theta}}_n$ of an arbitrary parameter vector $\boldsymbol{\theta}$ is defined, in this case, to maximise the QML-estimation criterion

$$Q_n(\boldsymbol{\theta}, \mathbf{z}) := -\frac{1}{n} \left[\log(|\boldsymbol{\Lambda}(\boldsymbol{\xi})|) + (\mathbf{z} - \boldsymbol{\mathcal{X}}\boldsymbol{\beta})^T \boldsymbol{\Lambda}(\boldsymbol{\xi})^{-1} (\mathbf{z} - \boldsymbol{\mathcal{X}}\boldsymbol{\beta}) \right] \quad (6.10)$$

where

$$\boldsymbol{\Lambda}(\boldsymbol{\xi}) := \partial_{\boldsymbol{\mu}} \boldsymbol{\eta} \boldsymbol{\mathcal{V}}(\boldsymbol{\xi}) \partial_{\boldsymbol{\mu}} \boldsymbol{\eta}^T.$$

Therefore

$$\hat{\boldsymbol{\theta}}_n := \operatorname{argmax}_{\boldsymbol{\theta} \in \Theta} Q_n(\boldsymbol{\theta}, \mathbf{z}), \quad (6.11)$$

where $\Theta := \tilde{\Theta} \times \mathbb{R}_+ \times \mathbb{R}^{s+1} \times \mathbb{R}_+$, with

$$\tilde{\Theta} := \{(a_1, \dots, a_p, b_1, \dots, b_q) \mid a(z) \neq 0 \text{ if } \operatorname{Re}(z) \geq 0; b(z) \neq 0 \text{ if } \operatorname{Re}(z) > 0 : \\ \text{the kernel of } Y \text{ is non-negative} \}.$$

Conditions for the kernel of Y to be non-negative are given in Tsai and Chan (2005).

6.2.1 Direct approach

The estimation of parameters in (6.9) is a generalised least squares problem. It can be solved in the following way. Since $\mathbf{\Lambda}(\boldsymbol{\xi})$ is positive definite there exists a positive definite lower triangular matrix $\mathbf{K}(\boldsymbol{\xi})$ with ones on the leading diagonal, and a positive definite diagonal matrix $\mathcal{F}(\boldsymbol{\xi})$, such that

$$\mathbf{\Lambda}(\boldsymbol{\xi})^{-1} := \mathbf{K}(\boldsymbol{\xi})^T \mathcal{F}(\boldsymbol{\xi})^{-1} \mathbf{K}(\boldsymbol{\xi}).$$

If we transform the data

$$\mathbf{z}^*(\boldsymbol{\xi}) := \mathbf{K}(\boldsymbol{\xi})\mathbf{z}, \quad \mathcal{X}^*(\boldsymbol{\xi}) := \mathbf{K}(\boldsymbol{\xi})\mathcal{X}, \quad \mathbf{e}^*(\boldsymbol{\xi}) := \mathbf{K}(\boldsymbol{\xi})\mathbf{e},$$

we get the heteroscedastic regression model

$$\mathbf{z}^*(\boldsymbol{\xi}) = \mathcal{X}^*(\boldsymbol{\xi})\boldsymbol{\beta} + \mathbf{e}^*(\boldsymbol{\xi}) \quad \text{with} \quad \text{Cov}(\mathbf{e}^*) = \mathcal{F}(\boldsymbol{\xi}). \quad (6.12)$$

If we assume that $\boldsymbol{\xi}$ is known and fixed, we get the generalised least squares estimate of $\boldsymbol{\beta}$ by solving an ordinary least-squares problem:

$$\begin{aligned} \widehat{\boldsymbol{\beta}}_n(\boldsymbol{\xi}) &= [(\mathcal{F}(\boldsymbol{\xi})^{-1/2} \mathcal{X}^*(\boldsymbol{\xi}))^T \mathcal{F}(\boldsymbol{\xi})^{-1/2} \mathcal{X}^*(\boldsymbol{\xi})]^{-1} (\mathcal{F}(\boldsymbol{\xi})^{-1/2} \mathcal{X}^*(\boldsymbol{\xi}))^T \mathcal{F}(\boldsymbol{\xi})^{-1/2} \mathbf{z}^*(\boldsymbol{\xi}) \\ &= [\mathcal{X}^T \mathbf{\Lambda}^{-1}(\boldsymbol{\xi}) \mathcal{X}]^{-1} \mathcal{X}^T \mathbf{\Lambda}^{-1}(\boldsymbol{\xi}) \mathbf{z}. \end{aligned} \quad (6.13)$$

Replacing $\boldsymbol{\beta}$ in (6.10) by the above estimate on gets the reduced QML-estimation criterion

$$Q_n(\boldsymbol{\xi}, \mathbf{z}) := \frac{1}{n} \sum_{i=1}^n \left[-\log(F_{t_i}(\boldsymbol{\xi})) - \frac{v_{t_i}^2(\boldsymbol{\xi})}{F_{t_i}(\boldsymbol{\xi})} \right], \quad (6.14)$$

with $v_{t_i}(\boldsymbol{\xi}) = z_{t_i}^*(\boldsymbol{\xi}) - x_{t_i}^{*T}(\boldsymbol{\xi}) \widehat{\boldsymbol{\beta}}_n(\boldsymbol{\xi})$ and $F_{t_i}(\boldsymbol{\xi}) = (\mathcal{F}(\boldsymbol{\xi}))_{i,i}$. QMLE of the parameters are therefore obtained by first maximizing (6.14) with respect to $\boldsymbol{\xi}$ to get $\widehat{\boldsymbol{\xi}}_n$. Afterwards one replaces $\boldsymbol{\xi}$ in $\widehat{\boldsymbol{\beta}}_n(\boldsymbol{\xi})$ by $\widehat{\boldsymbol{\xi}}_n$ to get the generalised least squares estimate of $\boldsymbol{\beta}$.

Remark 6.2.1 *The estimation of the parameters in the direct approach includes the computation of the inverse of $\mathbf{\Lambda}(\boldsymbol{\xi})$. In the application, which we have in mind, the dimension of $\mathbf{\Lambda}(\boldsymbol{\xi})$ varies from 2000 to 3000. $\mathbf{\Lambda}(\boldsymbol{\xi})^{-1}$ will also be a full matrix in comparison to regularly spaced observation, where $\mathbf{\Lambda}(\boldsymbol{\xi})^{-1}$ will be sparse (see Jones (1993) for details). Computationally it is not efficient to compute this inverse, and therefore we reformulate (6.9) as a state space model and apply the Kalman filter to compute (6.14). The idea to rewrite a regression model in state space form is explained for example in Chapter 6.2 in Durbin and Koopman (2001) and Chapter 5 in Jones (1993).*

6.2.2 State space approach

Consider again the linear regression model with correlated errors

$$\mathbf{z} = \mathbf{X}\boldsymbol{\beta} + \partial_{\boldsymbol{\mu}}\boldsymbol{\eta}\boldsymbol{\varepsilon}.$$

Since $\boldsymbol{\varepsilon} = \mathbf{Y} + \tilde{\boldsymbol{\varepsilon}}$, where $Y_{t_i} = \mathbf{b}^T \mathbf{X}_{t_i}$ is a CARMA(p,q) process, and $\partial_{\boldsymbol{\mu}}\boldsymbol{\eta} = \text{diag}(1/\mu_{t_1}, \dots, 1/\mu_{t_n})$, because of the log-link, we get the following state space representation of (6.9).

(1) Observation equation:

$$z_{t_i} = \mathbf{x}_{t_i}^T \boldsymbol{\beta} + \mathbf{G}_{t_i} \boldsymbol{\alpha}_{t_i} + \frac{1}{\mu_{t_i}} \tilde{\boldsymbol{\varepsilon}}_{t_i}, \quad (6.15)$$

where

$$\mathbf{G}_{t_i} := \frac{1}{\mu_{t_i}} \mathbf{b}^T \text{ and } \boldsymbol{\alpha}_{t_i} := \mathbf{X}_{t_i}.$$

with $\mathbf{x}_{t_i}^T$ the i -th row of $\mathbf{X} \in \mathbb{R}^{n \times s+1}$.

(2) State equation:

$$\boldsymbol{\alpha}_{t_{i+1}} = \mathcal{T}_{t_i} \boldsymbol{\alpha}_{t_i} + \boldsymbol{\zeta}_{t_i}, \quad (6.16)$$

where

$$\mathcal{T}_{t_i} := e^{\mathcal{A}(t_{i+1}-t_i)} \text{ and } \boldsymbol{\zeta}_{t_i} := \int_{t_i}^{t_{i+1}} e^{\mathcal{A}(t_{i+1}-u)} \mathbf{1}_p dL_u.$$

One standard assumption for state-space models is the zero mean of the noise processes. This assumption is not satisfied in (6.15) and (6.16). But we can construct a second state-space model defining observations ($z_{t_i}^*$, which are identical to z_{t_i} in $L^2(\mathbb{R})$). Since the QML-estimation criterion (6.10) only depends on the first two moments of the observations we will get the same QMLE for both models. For ease of notation we denote the observations in the second model also by \mathbf{z} . Because of the assumption $\mathbb{E}(\tilde{\boldsymbol{\varepsilon}}_{t_i}) = -\mathbb{E}(Y_{t_i})$, a zero mean CARMA(p,q) process $(Y_t^*)_{t \geq 0} = (\mathbf{b}^T \mathbf{X}_t^*)_{t \geq 0}$, with $\text{Cov}(Y_t^*, Y_s^*) = \text{Cov}(Y_t, Y_s)$, together with an i.i.d. noise sequence $(\tilde{\boldsymbol{\varepsilon}}_{t_i}^*)$, with $\mathbb{E}(\tilde{\boldsymbol{\varepsilon}}_{t_i}^*) = 0$, $\text{Var}(\tilde{\boldsymbol{\varepsilon}}_{t_i}^*) = \sigma_{\tilde{\boldsymbol{\varepsilon}}}^2$ and uncorrelated with Y^* , will lead to the same first and second order structure of z_{t_i} . Let L^* be the driving Lévy process of Y^* , with $\mathbb{E}(L_1^*) = 0$ and $\text{Var}(L_1^*) = \text{Var}(L_1)$. Then we get the state-space model:

(1) Observation equation:

$$z_{t_i} = \mathbf{x}_{t_i}^T \boldsymbol{\beta} + \mathbf{G}_{t_i} \boldsymbol{\alpha}_{t_i}^* + \frac{1}{\mu_{t_i}} \tilde{\boldsymbol{\varepsilon}}_{t_i}^*, \quad (6.17)$$

where

$$\mathbf{G}_{t_i} = \frac{1}{\mu_{t_i}} \mathbf{b}^T \text{ and } \boldsymbol{\alpha}_{t_i}^* := \mathbf{X}_{t_i}^*.$$

with $\mathbf{x}_{t_i}^T$ the i -th row of $\mathbf{X} \in \mathbb{R}^{n \times s+1}$.

(2) State equation:

$$\boldsymbol{\alpha}_{t_{i+1}}^* = \boldsymbol{T}_{t_i} \boldsymbol{\alpha}_{t_i}^* + \boldsymbol{\zeta}_{t_i}^*, \quad (6.18)$$

where

$$\boldsymbol{T}_{t_i} = e^{\boldsymbol{A}(t_{i+1}-t_i)} \text{ and } \boldsymbol{\zeta}_{t_i}^* := \int_{t_i}^{t_{i+1}} e^{\boldsymbol{A}(t_{i+1}-u)} \mathbf{1}_p dL_u^*.$$

An augmented Kalman filter (see e.g. Chapter 5.7 in Durbin and Koopman (2001)) will be applied to (6.17) and (6.18). The idea of this filter is to apply the Kalman filter with observation matrix \mathbf{G}_{t_i} and state matrix \boldsymbol{T}_{t_i} to the variables $z_{t_i}, x_{t_i,1}^T, \dots, x_{t_i,s+1}^T$ consecutively. $x_{t_i,k}^T$ is the k -th element of the row vector $\mathbf{x}_{t_i}^T$. For each of the variables $x_{t_i,1}^T, \dots, x_{t_i,s+1}^T$ a new state vector $\boldsymbol{\alpha}_{t_i}^k$, $k = 1, \dots, s+1$ is taken, but the variance elements in the Kalman filter are the same as for z_{t_i} . The Kalman filter computes best linear predictions $\widehat{z}_{t_i}, \widehat{x}_{t_i,1}^T, \dots, \widehat{x}_{t_i,s+1}^T$ based on all past observations $\{z_{t_j}, x_{t_j,1}^T, \dots, x_{t_j,s+1}^T; 1 \leq j < i\}$. In each step of the filter we store the one-step forecast errors $z_{t_i}^*(\boldsymbol{\xi}) := z_{t_i} - \widehat{z}_{t_i}, x_{t_i,1}^{*T}(\boldsymbol{\xi}) := x_{t_i,1}^T - \widehat{x}_{t_i,1}^T, \dots, x_{t_i,s+1}^{*T}(\boldsymbol{\xi}) := x_{t_i,s+1}^T - \widehat{x}_{t_i,s+1}^T$. These forecast errors can then be used to calculate the generalised least square estimates $\widehat{\boldsymbol{\beta}}$, given by

$$\widehat{\boldsymbol{\beta}}(\boldsymbol{\xi}) := \left(\sum_{i=1}^n \mathbf{x}_{t_i}^{*T}(\boldsymbol{\xi}) F_{t_i}^{-1}(\boldsymbol{\xi}) \mathbf{x}_{t_i}^{*T}(\boldsymbol{\xi}) \right)^{-1} \sum_{i=1}^n \mathbf{x}_{t_i}^{*T}(\boldsymbol{\xi}) F_{t_i}^{-1}(\boldsymbol{\xi}) z_{t_i}^*(\boldsymbol{\xi}), \quad (6.19)$$

where $\mathbf{x}_{t_i}^{*T}(\boldsymbol{\xi}) := (x_{t_i,1}^{*T}(\boldsymbol{\xi}), \dots, x_{t_i,s+1}^{*T}(\boldsymbol{\xi}))$ and $F_{t_i}(\boldsymbol{\xi}) := \text{Var}(z_{t_i}^*(\boldsymbol{\xi}) - \mathbf{x}_{t_i}^{*T}(\boldsymbol{\xi})\boldsymbol{\beta})$. To see that (6.19) is equal to (6.13) one has to recall that

$$\boldsymbol{\Lambda}^{-1}(\boldsymbol{\xi}) = \boldsymbol{\mathcal{K}}^T(\boldsymbol{\xi}) \boldsymbol{\mathcal{F}}^{-1}(\boldsymbol{\xi}) \boldsymbol{\mathcal{K}}(\boldsymbol{\xi}). \quad (6.20)$$

Inserting (6.20) into (6.13) yields

$$\widehat{\boldsymbol{\beta}}(\boldsymbol{\xi}) = [(\boldsymbol{\mathcal{K}}(\boldsymbol{\xi}) \boldsymbol{\mathcal{X}})^T \boldsymbol{\mathcal{F}}^{-1}(\boldsymbol{\xi}) \boldsymbol{\mathcal{K}}(\boldsymbol{\xi}) \boldsymbol{\mathcal{X}}]^{-1} (\boldsymbol{\mathcal{K}}(\boldsymbol{\xi}) \boldsymbol{\mathcal{X}})^T \boldsymbol{\mathcal{F}}^{-1}(\boldsymbol{\xi}) \boldsymbol{\mathcal{K}}(\boldsymbol{\xi}) \mathbf{z}.$$

Since the Kalman filter performs the Cholesky decomposition (6.20) (see Chapter 3.4 in Harvey (1990)), we see that applying the Kalman filter is equivalent to the multiplication by the matrix $\boldsymbol{\mathcal{K}}(\boldsymbol{\xi})$. For more details on the augmented Kalman filter see Chapter 5.7 Durbin and Koopman (2001) or Chapter 3.4 in Harvey (1990).

The procedure to estimate the parameters is then exactly the same as in the direct approach. First $\boldsymbol{\xi}$ is estimated by maximizing

$$\begin{aligned} Q_n(\boldsymbol{\xi}, \mathbf{z}) &= \frac{1}{n} \sum_{i=1}^n \left[-\log(F_{t_i}(\boldsymbol{\xi})) - \frac{(v_{t_i}^*(\boldsymbol{\xi}) - \mathbf{x}_{t_i}^{*T}(\boldsymbol{\xi})\widehat{\boldsymbol{\beta}}(\boldsymbol{\xi}))^2}{F_{t_i}(\boldsymbol{\xi})} \right] \\ &= \frac{1}{n} \sum_{i=1}^n \left[-\log(F_{t_i}(\boldsymbol{\xi})) - \frac{v_{t_i}^2(\boldsymbol{\xi})}{F_{t_i}(\boldsymbol{\xi})} \right] \end{aligned}$$

with respect to ξ . This estimate is denoted by $\hat{\xi}_n$. Afterwards ξ in (6.19) is replaced by $\hat{\xi}_n$ to get the generalised least squares estimate of β .

6.3 Simulation results

The performance of the QML estimator using the state space approach is going to be analysed in a small simulation study. The parameters are estimated in two set-ups. One with regularly spaced observations and the other with irregularly spaced ones. For the regularly spaced observations we created 2000 equidistant time points in the interval $(0, 400)$. In case of irregularly sampling the durations are exponentially distributed, with a mean value of 0.2, to assure that time points are also in the interval $(0, 400)$.

In each of the 100 simulations the sample size was 2000. As an explanatory variable we took real bid ask spreads from the IBM stock. The regression coefficient β was taken equal to 0.3. We did not include an intercept in the regression. The correlation was simulated by a CARMA(1,0) process with parameter $a = 0.8$. As driving Lévy process L we chose a compound Poisson process (see also Example 1.1.6) with exponentially distributed jumps (X_k) i.i.d. $\text{expo}(100)$. The jump rate of the Poisson process N was taken equal to 3. The mean and variance of L_1 are then 0.0375 and $\sigma^2 = 0.0006$, respectively. The choice of the parameter values was motivated by similar parameter values obtained in the application presented later. The measurement noise $\tilde{\varepsilon}$ was simulated as a Gaussian i.i.d. noise with mean -0.0375 and variance $\sigma_{\tilde{\varepsilon}}^2 = 0.0001$, respectively.

For the resulting estimates we computed estimates of mean, bias, mean absolute error (MAE), mean squared error (MSE) and the estimated standard errors of these estimates. The results can be seen in Table 6.1 and 6.2 showing satisfying performance for both settings.

6.4 Application

The data, which we will use, comes from the *Trades and Quotes* (TAQ) database of the New York Stock Exchange (NYSE). We will work with IBM trade data from September 30, 2002 up to October 31, 2002. The NYSE market opens 9.30 am and closes at 4.00 pm. Tradings outside these official trading hours have been deleted. Since we want to concentrate on real price changes we also excluded all zero returns and the corresponding explanatory variables. We also eliminated all multiple trades. Trades for the same transaction price were treated as a single trade by adding up the volumes. Different transaction prices were averaged and the volumes totalled. The resulting data set consists of transaction, bid and ask prices (all measured in cents of US dollars), transaction times (measured in seconds) and volumes (measured in the number of shares) realised over the specified time period. No further data

	\hat{a}	$\hat{\beta}$	$\hat{\sigma}^2$	$\hat{\sigma}_{\varepsilon}^2$
True value	0.8000	$3.0000 \cdot 10^{-1}$	$6.0000 \cdot 10^{-4}$	$1.0000 \cdot 10^{-4}$
Mean	0.8122 (0.0095)	$2.9881 \cdot 10^{-1}$ $(1.1341 \cdot 10^{-1})$	$6.1091 \cdot 10^{-4}$ $(6.9019 \cdot 10^{-6})$	$0.9939 \cdot 10^{-4}$ $(7.8563 \cdot 10^{-7})$
Rel. bias	0.0152	$-3.9676 \cdot 10^{-3}$	$1.8193 \cdot 10^{-2}$	$-6.0439 \cdot 10^{-3}$
MAE	0.0781 (0.0056)	$8.8016 \cdot 10^{-3}$ $(7.1971 \cdot 10^{-4})$	$5.7419 \cdot 10^{-5}$ $(3.9417 \cdot 10^{-6})$	$6.2597 \cdot 10^{-6}$ $(4.7446 \cdot 10^{-7})$
MSE	0.0092 (0.0013)	$1.2875 \cdot 10^{-4}$ $(1.8454 \cdot 10^{-5})$	$4.8352 \cdot 10^{-9}$ $(6.0478 \cdot 10^{-10})$	$6.1470 \cdot 10^{-11}$ $(7.7510 \cdot 10^{-12})$

Table 6.1: Mean, relative bias, mean absolute error (MAE) and mean squared error (MSE) for \hat{a} , $\hat{\beta}$, $\hat{\sigma}^2$ and $\hat{\sigma}_{\varepsilon}^2$ together with their estimated standard errors in parentheses in case of *regularly spaced observations*.

	\hat{a}	$\hat{\beta}$	$\hat{\sigma}^2$	$\hat{\sigma}_{\varepsilon}^2$
True value	0.8000	$3.0000 \cdot 10^{-1}$	$6.0000 \cdot 10^{-4}$	$1.0000 \cdot 10^{-4}$
Mean	0.8015 (0.0092)	$2.9844 \cdot 10^{-1}$ $(9.4843 \cdot 10^{-4})$	$6.0974 \cdot 10^{-4}$ $(6.9191 \cdot 10^{-6})$	$9.8657 \cdot 10^{-5}$ $(5.4509 \cdot 10^{-7})$
Rel. bias	0.0019	$-5.1803 \cdot 10^{-3}$	$1.6248 \cdot 10^{-2}$	$-1.3423 \cdot 10^{-2}$
MAE	0.0696 (0.0059)	$8.1259 \cdot 10^{-3}$ $(5.0689 \cdot 10^{-4})$	$5.5842 \cdot 10^{-5}$ $(4.1634 \cdot 10^{-6})$	$4.4295 \cdot 10^{-6}$ $(3.4225 \cdot 10^{-7})$
MSE	0.0082 (0.0014)	$9.1468 \cdot 10^{-5}$ $(1.1264 \cdot 10^{-5})$	$4.8344 \cdot 10^{-9}$ $(6.8064 \cdot 10^{-10})$	$3.1220 \cdot 10^{-11}$ $(4.2670 \cdot 10^{-12})$

Table 6.2: Mean, relative bias, mean absolute error (MAE) and mean squared error (MSE) for \hat{a} , $\hat{\beta}$, $\hat{\sigma}^2$ and $\hat{\sigma}_{\varepsilon}^2$ together with their estimated standard errors in parentheses in case of *irregularly spaced observations*.

manipulations have been carried out. Sample absolute log returns of six trading days have been plotted in Figure 6.1.

In Chapter 6.1.1 we have said, that a parametric approach is used. But up to now we have not specified the parametric set-up. To get an idea how the absolute log return may depend on the explanatory variables, we perform some kind of explorative data analysis by fitting a *Generalized Additive Model* (see Hastie and Tibshirani (1990)) with uncorrelated errors to the data. The functional relationship displayed by the model, will then be used to set up a parametric model. The aim of the analysis in this section is to fit our model to the data. Then to check if the fitted correlation structure can be justified and investigate the predictive power of the explanatory variables. The one step ahead predictions of the absolute log return for October 14th

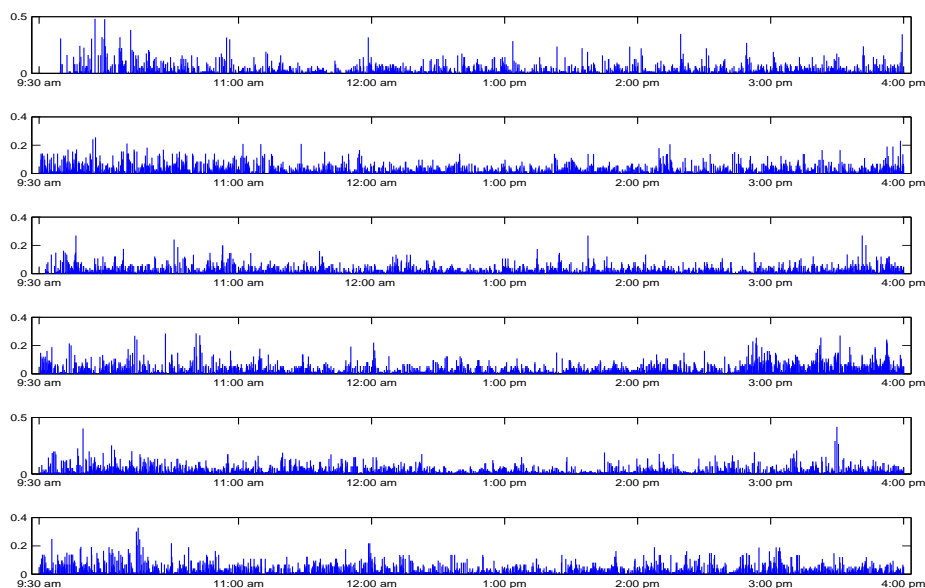


Figure 6.1: Absolute log returns of the 11th (*first row*), 18th (*second row*), 22nd (*third row*), 23rd (*fourth row*), 24th (*fifth row*) and 25th (*last row*) of October 2002. The time is measured in real time.

until October 31st, 2002, will be computed using the information corresponding to each of the following four set-ups:

- (1) the last day
- (2) the last three days
- (3) the last day and the same day one week ago
- (4) the same day one and two weeks ago.

The different forecasts are then compared using the mean squared error as criterion. Exploratory we will present the estimation results for the days needed to predict October 25th, 2002.

6.4.1 Explorative data analysis

Initially we chose only the bid-ask spread and the duration as explanatory variables. The influence of the volume will be analysed in a further study. Therefore the

generalised additive model under consideration is the following one

$$\log(\mu_{t_i}) = s_1(b_{t_i}) + s_2(d_{t_i}),$$

where $s_i()$, $i = 1, 2$, are smoothing splines and b_{t_i} (bid-ask spread) and d_{t_i} (durations) are the explanatory variables. This model is fitted using the Splus function `gam()` under the assumption of uncorrelated errors. The results of this estimation procedure can be seen in Figure 6.2.

For the bid-ask spread as well as the duration one can recognise a relatively smooth functional relationship. We decided, that a polynomial of third order has enough flexibility to model both explanatory variables. This led us to consider a model with design matrix \mathcal{X} , where

$$\mathbf{x}_{t_i}^T \boldsymbol{\beta} := \beta_0 + \beta_1 b_{t_i} + \beta_2 b_{t_i}^2 + \beta_3 b_{t_i}^3 + \beta_4 d_{t_i} + \beta_5 d_{t_i}^2 + \beta_6 d_{t_i}^3,$$

with bid-ask spread b_{t_i} and duration d_{t_i} .

6.4.2 Estimation results

The application of the augmented Kalman filter, which was described in Chapter 6.2.2, and the quasi maximum likelihood estimation of the remaining parameters resulted in the parameter estimates, which can be seen in Table 6.3. The coefficients $\hat{\beta}_k$, $k = 4, 5, 6$, correspond to durations measured in one-hundredth of a second, whereas the time was measured in seconds. The plots of the absolute log returns together with their fitted mean values are shown in Figure 6.3 demonstrating no obvious lack of fit.

The regression coefficients lead to estimates of the two polynomials

$$p_b(b_{t_i}) := \beta_0 + \beta_1 b_{t_i} + \beta_2 b_{t_i}^2 + \beta_3 b_{t_i}^3 \quad (6.21)$$

$$p_d(d_{t_i}) := \beta_4 d_{t_i} + \beta_5 d_{t_i}^2 + \beta_6 d_{t_i}^3. \quad (6.22)$$

The estimated polynomials of the m -th day are denoted by

$$\hat{p}_b^m(x) := \hat{\beta}_0^m(\mathbf{b}^m, \mathbf{d}^m) + \hat{\beta}_1^m(\mathbf{b}^m, \mathbf{d}^m)x + \hat{\beta}_2^m(\mathbf{b}^m, \mathbf{d}^m)x^2 + \hat{\beta}_3^m(\mathbf{b}^m, \mathbf{d}^m)x^3$$

and

$$\hat{p}_d^m(x) := \hat{\beta}_4^m(\mathbf{b}^m, \mathbf{d}^m)x + \hat{\beta}_5^m(\mathbf{b}^m, \mathbf{d}^m)x^2 + \hat{\beta}_6^m(\mathbf{b}^m, \mathbf{d}^m)x^3$$

and the observations on the m -th day by

$$\mathbf{b}^m := (b_{t_1}^m, \dots, b_{t_{n_m}}^m) \quad \text{and} \quad \mathbf{d}^m := (d_{t_1}^m, \dots, d_{t_{n_m}}^m)$$

where n_m is the number of observations on day m . These estimated polynomials are shown in Figure 6.2.

Day	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$
October 11, 2002	-4.2726	18.0106	-48.9357	49.1082	1.8313
October 18, 2002	-4.4576	18.0729	-27.0261	-42.8318	2.6261
October 22, 2002	-4.6144	24.0957	-113.1130	253.7693	2.1601
October 23, 2002	-4.3120	17.2861	-38.7623	22.5341	1.3981
October 24, 2002	-4.4129	15.7375	-27.6543	16.4028	2.8124
October 25, 2002	-4.6366	26.6262	-117.5710	228.4430	1.8190
Day	$\hat{\beta}_5$	$\hat{\beta}_6$	\hat{a}	$\hat{\sigma}^2$	$\hat{\sigma}_\varepsilon^2$
October 11, 2002	-2.1036	0.7714	0.3942	$1.1 \cdot 10^{-3}$	$2.6 \cdot 10^{-09}$
October 18, 2002	-4.7253	2.6941	0.5942	$7.4 \cdot 10^{-4}$	$9.1 \cdot 10^{-13}$
October 22, 2002	-3.8395	3.2206	0.9886	$2.1 \cdot 10^{-4}$	$4.1 \cdot 10^{-04}$
October 23, 2002	-0.2093	-0.7194	0.7301	$1.3 \cdot 10^{-3}$	$2.1 \cdot 10^{-10}$
October 24, 2002	-3.4322	-0.0021	0.5253	$7.1 \cdot 10^{-4}$	$4.0 \cdot 10^{-08}$
October 25, 2002	-1.4558	0.2407	0.8991	$9.8 \cdot 10^{-4}$	$1.4 \cdot 10^{-04}$

Table 6.3: QMLE based on the augmented Kalman filter.

6.4.3 Analysis of the correlation structure

In the end we want to take a look at the sample variograms of the residuals, and see if the assumed correlation structure can be justified. The variogram is defined in Appendix B, where we also present four examples of sample variograms of simulated CARMA(p,q) processes. Figure 6.3 contains the sample variograms and variograms of the estimated models for all six residual processes.

The rough structure of the sample variogram is due to the irregularly spaced observations, because the irregular spacing leads to greater changes in the number of observations for consecutive lags. For October 11, 2002 the estimated model proposes stronger correlation than the sample variogram, but despite this fact, the shape of the sample variogram and the variogram based on the estimated model is quite similar. The reason for this might be a numerical imprecision or a misspecified correlation structure, which has to be further analysed. The other days show less correlation in the residuals, which can be seen by the faster increasing variograms. The sample variograms represent the proposed structure of the model variogram quite well. Only for the first few lags we see consistently smaller values of the sample variogram $\hat{\gamma}(h)$ compared to the model variogram $\gamma(h)$. This may be due to the fact that $\gamma(h) \rightarrow \sigma_\varepsilon^2$ but $\hat{\gamma}(h) \rightarrow 0$ as $h \rightarrow 0$ (see also the appendix). This effect is known in the geostatistics literature as a *nugget effect* and appears because of the superposition of independent noise on an underlying process. The nugget effect can be seen on all six days. Therefore one could try to fit CARMA processes of higher

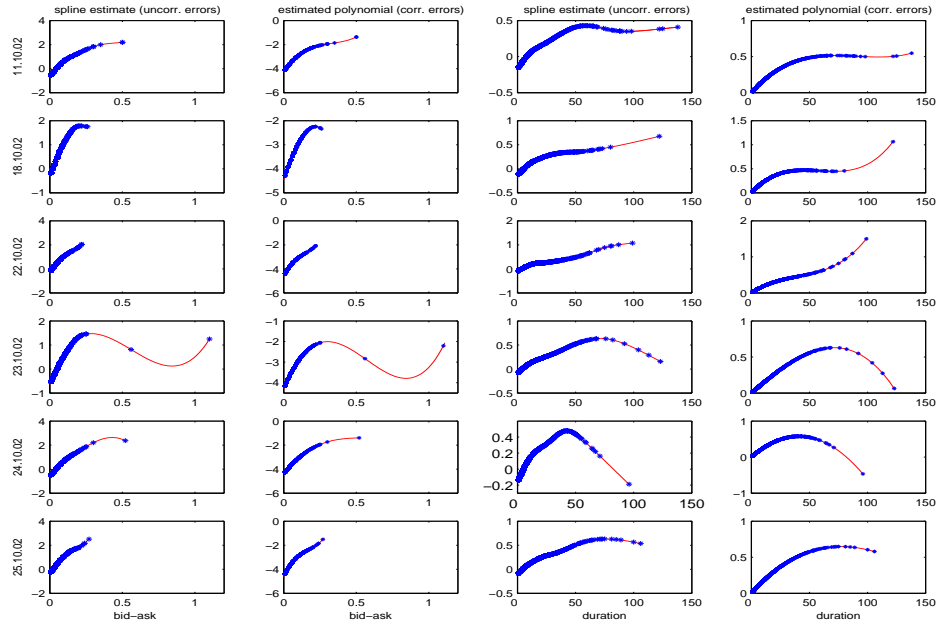


Figure 6.2: Smoothing spline estimates and estimated bid-ask and duration polynomials $\hat{p}_b^m(\cdot)$ and $\hat{p}_d^m(\cdot)$ for the days 11th (*first row*), 18th (*second row*), 22nd (*third row*), 23rd (*fourth row*), 24th (*fifth row*) and 25th (*last row*) of October 2002. The marks represent the observed values of the explanatory variables.

order to the data on October 11th to see, if the fit could be improved. For the remaining days the proposed correlation could be justified.

6.4.4 Prediction

Since we have shown how to estimate the polynomials, we want to explain now how to predict the mean of the absolute log return of the next trading day. Imagine that we have estimates for $m = 1, \dots, M$ days. Using these $2M$ polynomials we construct two *mean piecewise polynomials* by averaging over the observed data points

$$\overline{p}_b^M(x) := \frac{1}{|M^b(x)|} \sum_{m \in M^b(x)} \hat{p}_b^m(x) \quad (6.23)$$

$$\overline{p}_d^M(x) := \frac{1}{|M^d(x)|} \sum_{m \in M^d(x)} \hat{p}_d^m(x), \quad (6.24)$$

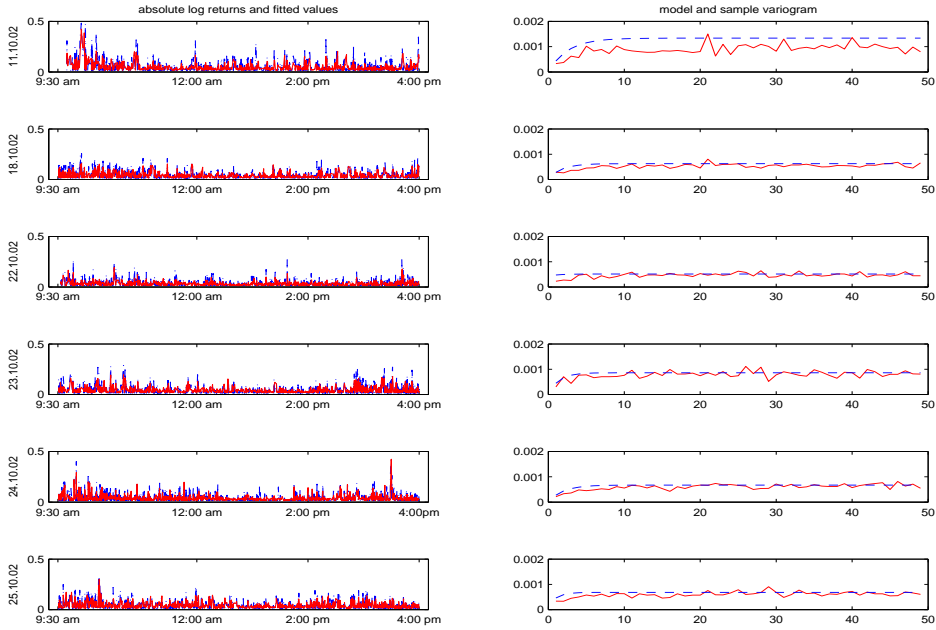


Figure 6.3: Left column: Absolute log returns (*dashed line*) together with the fitted values (*solid line*) for the days 11th (*top row*), 18th (*second row*), 22nd (*third row*), 23rd (*fourth row*), 24th (*fifth row*) and 25th (*bottom row*) of October 2002. Right column: Model (*dashed line*) and sample variogram of the residuals ε_{t_i} (*solid line*) for the days 11th (*top row*), 18th (*second row*), 22nd (*third row*), 23rd (*fourth row*), 24th (*fifth row*) and 25th (*bottom row*) of October 2002.

where

$$M^b(x) := \{m \in \{1, \dots, M\} \mid x \in [0, \max_i b_{t_i}^m]\}$$

$$|M^b(x)| := \text{card } M^b(x)$$

and

$$M^d(x) := \{m \in \{1, \dots, M\} \mid x \in [0, \max_i d_{t_i}^m]\}$$

$$|M^d(x)| := \text{card } M^d(x).$$

A smoothed version of these two piecewise polynomials for day $M + 1$ we get by fitting two smoothing splines at $p_b^M(\cdot)$ and $p_d^M(\cdot)$ over the intervals $[0, \max_m b_{t_{nm}}^m]$

and $[0, \max_m d_{t_{nm}}^m]$. The smoothing splines $\overline{p}_b(\cdot)$ and $\overline{p}_d(\cdot)$ minimise

$$\sum_{i=1}^n \left(\overline{p}_b^M(x_{t_i}^b) - \overline{p}_b(x_{t_i}^b) \right)^2 + \lambda_b \int_0^{T_b} \left[\frac{\partial^2 \overline{p}_b(x)}{\partial^2 x} \right]^2 dx, \quad x_{t_i}^b \in [0, \max_m b_{t_{nm}}^m], \quad (6.25)$$

and

$$\sum_{i=1}^n \left(\overline{p}_d^M(x_{t_i}^d) - \overline{p}_d(x_{t_i}^d) \right)^2 + \lambda_d \int_0^{T_d} \left[\frac{\partial^2 \overline{p}_d(x)}{\partial^2 x} \right]^2 dx, \quad x_{t_i}^d \in [0, \max_m d_{t_{nm}}^m] \quad (6.26)$$

respectively, where $\lambda_b, \lambda_d > 0$ are smoothing parameters, $T_b := \max_m b_{t_{nm}}^m$ and T_d similarly. λ_b and λ_d are maximum likelihood estimates. Maximum likelihood estimation of smoothing parameters for spline smoothing is explained in Chapter 3.11 in Durbin and Koopman (2001).

The predicted mean values of the absolute log returns $|\widehat{r}_{t_i}|$ of the $M + 1$ -th day are then defined like this

$$P(|r_{t_i}^{M+1}|) := \exp(\overline{p}_b(b_{t_i}^{M+1}) + \overline{p}_d(d_{t_i}^{M+1})). \quad (6.27)$$

Remark 6.4.1 Observe that d_{t_i} is unknown up to time t_i . Since we mainly want to investigate the dependence on the explanatory variables, we will assume in a first step, that the durations are known. In a second step an ACD model could be fitted to the durations, to get forecasts also for the durations.

6.4.5 Prediction results

As we mentioned at the beginning of this section, the one step ahead predictions of the absolute log return for the days October 14th-31st, 2002, will be computed using the data of:

- (i) the last day
- (ii) the last three days
- (iii) the last day and the same day one week ago
- (iv) the same day one and two weeks ago.

Performing the steps described in Chapter 6.4.4 produced for each day the smoothing spline estimates $\overline{p}_b^k(\cdot)$, $k = 1, \dots, 4$ and $\overline{p}_d^k(\cdot)$, $k = 1, \dots, 4$. In the first prediction set-up (i) the smoothing splines are equal to the estimated polynomials for the last day, since we have only one polynomial observation in each case. For the 25th of October, the smoothing splines together with the mean piecewise polynomials are shown in Figure 6.4. The absolute log returns together with corresponding predictions can also be seen.

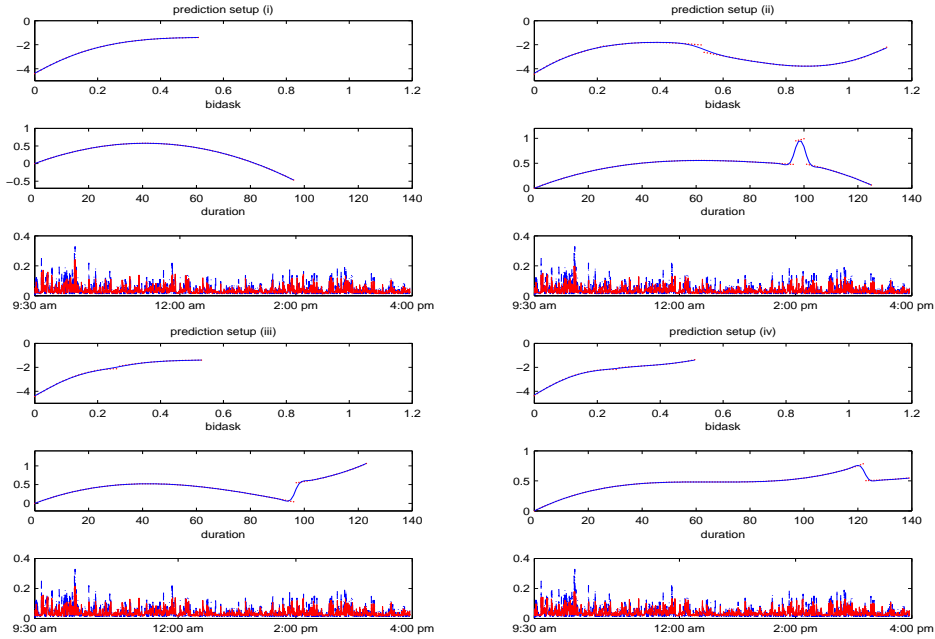


Figure 6.4: Smoothing spline (*solid line*) and mean piecewise polynomials (*dashed line*) in rows 1,2,4 and 5, absolute log returns (*dashed line*) and mean value predictions (*solid line*) in rows 3 and 6 for the prediction set-up (i) (*top left 3 panels*), (ii) (*top right 3 panels*), (iii) (*bottom left 3 panels*) and (iv) (*bottom right 3 panels*).

The different forecasts are now compared using the mean squared errors

$$MSE^{k,j} := \frac{1}{n_j} \sum_{i=1}^{n_j} (|r_{t_i}^j| - P^k(|r_{t_i}^j|))^2, \quad k = 1, \dots, 4, j \in \mathcal{I},$$

where

$$P^k(|r_{t_i}^j|) := \exp(\overline{p_b}^k(b_{t_i}^j) + \overline{p_d}^k(d_{t_i}^j)), \quad k = 1, \dots, 4, j \in \mathcal{I},$$

and \mathcal{I} is the index set of the sample including October 14th to 31st, as criterion. These MSE are shown in Table 5.4. In parentheses one can see the rank of the prediction within each day. For October 14th the random effect could not be described by a CARMA(1,0) process. Therefore we fitted a CARMA(2,1) process to the data. To compare the different prediction set-ups we calculated average ranks over the days. For this data the best strategy would be to use the information of the last three days for prediction. set-up (iii) is the second best strategy and set-up (i) and

Day	set-up (i)	set-up (ii)	set-up (iii)	set-up (iv)
October 14	$1.387 \cdot 10^{-3}$ (3)	$1.356 \cdot 10^{-3}$ (1)	$1.368 \cdot 10^{-3}$ (2)	$1.389 \cdot 10^{-3}$ (4)
October 15	$6.825 \cdot 10^{-4}$ (2)	$6.830 \cdot 10^{-4}$ (3)	$6.804 \cdot 10^{-4}$ (1)	$7.776 \cdot 10^{-4}$ (4)
October 16	$9.541 \cdot 10^{-4}$ (1)	$9.639 \cdot 10^{-4}$ (2)	$9.771 \cdot 10^{-4}$ (3)	$1.013 \cdot 10^{-3}$ (4)
October 17	$4.538 \cdot 10^{-4}$ (1)	$5.157 \cdot 10^{-4}$ (2)	$5.661 \cdot 10^{-4}$ (3)	$8.871 \cdot 10^{-4}$ (4)
October 18	$6.453 \cdot 10^{-4}$ (3)	$6.203 \cdot 10^{-4}$ (1)	$6.211 \cdot 10^{-4}$ (2)	$6.485 \cdot 10^{-4}$ (4)
October 21	$5.598 \cdot 10^{-4}$ (1)	$5.942 \cdot 10^{-4}$ (3)	$5.665 \cdot 10^{-4}$ (2)	$8.365 \cdot 10^{-4}$ (4)
October 22	$5.261 \cdot 10^{-4}$ (3)	$5.254 \cdot 10^{-4}$ (2)	$5.228 \cdot 10^{-4}$ (1)	$5.952 \cdot 10^{-4}$ (4)
October 23	$9.311 \cdot 10^{-4}$ (2)	$8.705 \cdot 10^{-4}$ (1)	$3.846 \cdot 10^{-3}$ (4)	$1.475 \cdot 10^{-3}$ (3)
October 24	$7.644 \cdot 10^{-4}$ (4)	$7.581 \cdot 10^{-4}$ (3)	$7.537 \cdot 10^{-4}$ (2)	$7.378 \cdot 10^{-4}$ (1)
October 25	$7.054 \cdot 10^{-4}$ (4)	$7.011 \cdot 10^{-4}$ (3)	$6.932 \cdot 10^{-4}$ (2)	$6.928 \cdot 10^{-4}$ (1)
October 28	$1.048 \cdot 10^{-3}$ (4)	$8.157 \cdot 10^{-4}$ (2)	$8.048 \cdot 10^{-4}$ (1)	$8.746 \cdot 10^{-4}$ (3)
October 29	$8.529 \cdot 10^{-4}$ (2)	$8.421 \cdot 10^{-4}$ (1)	$8.561 \cdot 10^{-4}$ (3)	$8.828 \cdot 10^{-4}$ (4)
October 30	$2.657 \cdot 10^{-3}$ (4)	$1.829 \cdot 10^{-3}$ (3)	$1.226 \cdot 10^{-3}$ (1)	$1.323 \cdot 10^{-3}$ (2)
October 31	$5.463 \cdot 10^{-4}$ (1)	$5.558 \cdot 10^{-4}$ (2)	$5.700 \cdot 10^{-4}$ (3)	$6.821 \cdot 10^{-4}$ (4)
average rank	(2.50)	(2.07)	(2.14)	(3.28)

Table 6.4: MSE of the one step ahead predictions on the next trading day for the set-up (i), (ii), (iii) and (iv) together with the corresponding rank in parentheses.

(iv) are third and fourth. This presents a method which allows to empirically investigate the performance of different prediction strategies. The predictive power of the duration can be seen, when we recompute the predictions for the set-up with the smallest MSE without using the duration. We observed an increase in the MSE between 5 and 20 percent. For October 25th the resulting predictions are shown in Figure 6.5. The mean squared error in this case is equal to $8.1874 \cdot 10^{-4}$, showing a significant increase of about 18 percent.

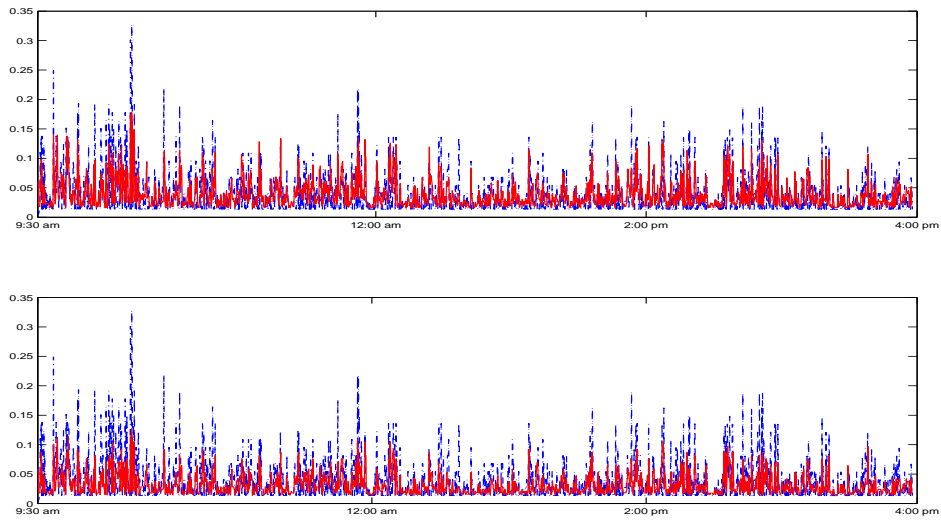


Figure 6.5: Absolute log returns on October 25th (*dashed line*) and mean value predictions (*solid line*) for prediction set-up (iv) using bid ask and duration (*top*) and using only bid ask (*bottom*).

Conclusion and outlook

Analysing ultra high frequency financial data is nowadays mostly based on a continuous time model. In this thesis we addressed several new approaches. First of all the exponential continuous time GARCH(p, q) process was introduced and its probabilistic properties studied to some extent. It was shown that the model can explain a leverage effect in the way that a jump in the log-price process can be negatively correlated with future volatility for certain parameters. For the simplest case, the compound Poisson ECOGARCH(1, 1) process, we also considered inference in this new model. Future work in this direction has to be done to develop estimators applicable for more general driving Lévy processes.

To adjust for long range dependence in the volatility, which is sometimes observed in financial data, we introduced the fractionally integrated ECOGARH(p, d, q) process. Due to the long memory property, which we showed for the volatility process, the estimation of the model will be more complicated. Inference of this model should probably be based on Whittle type estimators and use the linear structure of the log-volatility process. In both models we have not taken into account any explanatory variables, which are provided to us by the market. Finding suitable extensions to allow for such information should also be one point in future research.

The mixed effect model we introduced at the end of the thesis takes into account such explanatory variables, but models a proxy, the absolute log-return, of the instantaneous volatility. We analysed the impact of this exogenous information on the absolute log-return on a daily basis. Significant and time varying influence of the bid-ask spread and the duration was found. Since the duration is unknown prior to the trade, one should consider in future work an autoregressive condition duration model for the unknown duration. Also the influence of further explanatory variables has to be analysed.

Multivariate approaches to model a portfolio of stocks have not been considered in this thesis. Though the interest in multivariate continuous time models with stochastic volatility has increased in recent time, see e.g. Barndorff-Nielsen and Shephard (2001), Barndorff-Nielsen and Stelzer (2007) or Pigorsch and Stelzer (2007). The research on extending the ECOGARCH(p, q) process to the multivariate case is joint work with Robert Stelzer and a first multivariate model is currently under investigation.

Appendix A

Basics on stochastic processes and related topics

A.1 Definitions and review of basic results

We review some basic facts about the theory of stochastic processes and related topics. For further details we refer to Protter (2004), Karatzas and Shreve (1991) and Applebaum (2004).

Definition A.1 *Let \mathbb{T} be a parameter space. A stochastic process is a family of random variables $(\mathbf{X}_t)_{t \in \mathbb{T}}$ defined on a probability space (Ω, \mathcal{F}, P) .*

In the following we will only consider d -dimensional continuous time stochastic processes. In that case the parameter space \mathbb{T} is identified with \mathbb{R}_+ .

Definition A.2 *Let \mathcal{F} and \mathcal{G} be two σ -algebras. Then the product σ -algebra formed from \mathcal{F} and \mathcal{G} is defined by $\mathcal{F} \otimes \mathcal{G} := \sigma(A \times B; A \in \mathcal{F}, B \in \mathcal{G})$.*

Definition A.3

- (1) *A random variable defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is called \mathcal{F} -measurable if $\mathbf{X}^{-1}(\mathcal{B}(\mathbb{R}^d)) \subset \mathcal{F}$. If \mathbf{X} is a bounded \mathcal{F} -measurable random variable we denote it by $\mathbf{X} \in b\mathcal{F}$.*
- (2) *A d -dimensional stochastic process \mathbf{X} defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is called measurable if, for every $A \in \mathcal{B}(\mathbb{R}^d)$, the set $\{(t, \omega); \mathbf{X}_t(\omega) \in A\}$ belongs to the product σ -algebra $\mathcal{B}([0, \infty)) \otimes \mathcal{F}$.*

If we equip the sample space (Ω, \mathcal{F}) with a *filtration*, i.e. a nondecreasing family $\{\mathcal{F}_t; t \geq 0\}$ of sub- σ -algebras of \mathcal{F} : $\mathcal{F}_s \subseteq \mathcal{F}_t \subseteq \mathcal{F}$ for $0 \leq s < t < \infty$, then an adapted stochastic process can be defined as follows.

Definition A.4 The stochastic process \mathbf{X} is adapted to the filtration $\{\mathcal{F}_t\}$, if for every $t \geq 0$, \mathbf{X}_t is an \mathcal{F}_t -measurable random variable.

A probability space $(\Omega, \mathcal{F}, \mathbb{P})$ equipped with a filtration $\{\mathcal{F}_t\}$ will be called a *filtered probability space* and denoted by $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$.

Definition A.5 Let $E \in \mathcal{B}(\mathbb{R}^m)$. Then the smallest σ -algebra \mathcal{P} on $\mathbb{R}_+ \times E \times \Omega$, with respect to which all mappings $\mathbf{F} : \mathbb{R}_+ \times E \times \Omega \rightarrow \mathbb{R}^d$ satisfying:

(1) for each $t \in \mathbb{R}_+$ the mapping $(x, \omega) \mapsto \mathbf{F}(t, x, \omega)$ is $\mathcal{B}(E) \otimes \mathcal{F}_t$ -measurable

(2) for each $x \in E, \omega \in \Omega$ the mapping $t \mapsto \mathbf{F}(t, x, \omega)$ is left-continuous

are measurable, is called the *predictable σ -algebra*. A mapping measurable with respect to \mathcal{P} is said to be *predictable*.

Notice that we would call a mapping defined on $\mathbb{R}_+ \times \Omega$ also a stochastic process.

Definition A.6 Consider a measurable space (Ω, \mathcal{F}) equipped with a filtration $\{\mathcal{F}_t\}$.

(1) A random time T is a \mathcal{F} -measurable random variable, with values in $[0, \infty]$. A random time T is a *stopping time* of the filtration, if

$$\{T \leq t\} \in \mathcal{F}_t, \quad \forall t \in \mathbb{R}_+.$$

(2) Let \mathbf{X} be stochastic process and T a stopping time. The *stopped process* \mathbf{X}^T is defined by $\mathbf{X}_s^T(\omega) = \mathbf{X}_{T \wedge s}(\omega)$.

Definition A.7

(1) Let us define $\mathcal{F}_{t-} := \sigma(\bigcup_{s < t} \mathcal{F}_s)$ to be the σ -algebra of events strictly prior to $t > 0$, $\mathcal{F}_{t+} := \bigcap_{\epsilon > 0} \mathcal{F}_{t+\epsilon}$ to be the σ -algebra of events immediately after $t \geq 0$ and $\mathcal{F}_{0-} := \mathcal{F}_0$. Then the filtration $\{\mathcal{F}_t\}$ is right- (left-) continuous if $\mathcal{F}_t = \mathcal{F}_{t+}$ ($\mathcal{F}_t = \mathcal{F}_{t-}$) holds for every $t \geq 0$.

(2) A filtration $\{\mathcal{F}_t\}$ is said to satisfy the *usual conditions* if it is right-continuous and \mathcal{F}_0 contains all the \mathbb{P} -negligible events (events which occur with probability zero) in \mathcal{F} .

Definition A.8 A stochastic process \mathbf{X} is said to be *càdlàg (càglàd)* if it a.s. has sample paths which are right- (left-) continuous, with left- (right-) hand limits.

Definition A.9 The stochastic process \mathbf{X} is called (left-, right-) continuous in probability, if

$$\lim_{(h \downarrow 0, h \uparrow 0), h \rightarrow 0} \mathbb{P}(\|\mathbf{X}_{t+h} - \mathbf{X}_t\| > \epsilon) = 0, \quad \forall (t > 0, t \geq 0) t \geq 0 \text{ and } \epsilon > 0.$$

Definition A.10 Two stochastic processes \mathbf{X} and \mathbf{Y} are modifications if $\mathbf{X}_t = \mathbf{Y}_t$ a.s., for each t .

Definition A.11 Let (S, \mathcal{A}) be a measurable space and $(\Omega, \mathcal{F}, \mathbb{P})$ a probability space. Then a random measure M on (S, \mathcal{A}) is defined as a collection of random variables $\{M(B), B \in \mathcal{A}\}$ on $(\Omega, \mathcal{F}, \mathbb{P})$ such that:

(R1) $M(\emptyset) = 0$

(R2)

$$M\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} M(A_n), \quad a.s.,$$

for any given sequence $(A_n)_{n \in \mathbb{N}}$ of mutually disjoint sets in \mathcal{A} .

(R3) for each disjoint family B_1, \dots, B_n of sets in \mathcal{A} , the random variables $M(B_1), \dots, M(B_n)$ are independent.

Definition A.12

(1) Let \mathbf{X} be an adapted process on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$ such that $\mathbb{E}(\|\mathbf{X}_t\|) < \infty$ for all $t \geq 0$. Then we say that \mathbf{X} is a martingale if, for all $0 \leq s < t < \infty$,

$$\mathbb{E}(\mathbf{X}_t | \mathcal{F}_s) = \mathbf{X}_s \quad a.s..$$

(2) Let U be a topological space and M a random measure on $S = \mathbb{R}_+ \times U$. For each $A \in \mathcal{B}(U)$, define the process $M_A := (M([0, t] \times A))_{t \geq 0}$. Then M is called a martingale-valued random measure if there exists $V \in \mathcal{B}(U)$ such that M_A is a martingale whenever $\bar{A} \cap V = \emptyset$.

Definition A.13 Let $\pi = \{a = t_0 < t_1 < \dots < t_n = b\}$ be a finite partition of the interval $[a, b] \subset \mathbb{R}$ and $f : [a, b] \rightarrow \mathbb{R}^m$ a càdlàg mapping. Then the variation $var_{\pi}(f)$ of f over the partition π is defined as

$$var_{\pi}(f) := \sum_{i=1}^n \|f(t_i) - f(t_{i-1})\|.$$

If $\sup_{\pi} var_{\pi}(f) < \infty$, then f is said to be of finite variation on $[a, b]$. If f is defined on \mathbb{R} , then f has finite variation if it has finite variation on every compact interval.

Remark A.14 A stochastic process X is said to be of finite variation if the trajectories $(X_t(\omega))_{t \geq 0}$ are of finite variation for almost all $\omega \in \Omega$.

Definition A.15 Let X and Y be univariate Lévy-type stochastic integrals as defined in (1.24), i.e.

$$dX_t = G(t)dt + F(t)dB_t + \int_{|x|<1} H(t, x)\tilde{N}_L(dt, dx) + \int_{|x|\geq 1} K(t, x)N_L(dt, dx),$$

where $|G|^{1/2}, F \in \mathcal{H}_2(T)$, $H \in \mathcal{H}_2(T, B_1 - \{0\})$ and K is predictable. For each $t \geq 0$ let $(\pi_n)_{n \in \mathbb{N}}$, where $\pi_n = \{0 = t_0^{(n)} < t_1^{(n)} < \dots < t_{k_n}^{(n)} = t\}$, be a sequence of finite partitions of $[0, t]$ with mesh $m_n := \max_{1 \leq i \leq k_n} |t_i^{(n)} - t_{i-1}^{(n)}|$ tending to zero for $n \rightarrow \infty$. Then we define

$$[X, Y]_t := \lim_{n \rightarrow \infty} \sum_{i=1}^{k_n} \left(X_{t_i^{(n)}} - X_{t_{i-1}^{(n)}} \right) \left(Y_{t_i^{(n)}} - Y_{t_{i-1}^{(n)}} \right), \quad (\text{A.1})$$

where the limit is taken in probability, i.e.

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(\left| \sum_{i=1}^{k_n} \left(X_{t_i^{(n)}} - X_{t_{i-1}^{(n)}} \right) \left(Y_{t_i^{(n)}} - Y_{t_{i-1}^{(n)}} \right) - [X, Y]_t \right| > \epsilon \right) = 0$$

for all $\epsilon > 0$. $[X, Y]_t$ will be called the quadratic variation of X and Y at time t and $([X, Y]_t)_{t \geq 0}$ the quadratic variation process of X and Y .

From Corollary II.1 in Protter (2004) we know that $[X, Y]$ has sample paths of finite variation on every compact interval. From the polarization identity (see e.g. Chapter II.6 in Protter (2004))

$$[X, Y] = \frac{1}{2}([X + Y, X + Y] - ([X, X] + [Y, Y]))$$

one can see that $[X, Y]$ is the difference of two non-decreasing processes, hence its sample path is of finite variation (cf. Corollary II.1 in Protter (2004)). Then we know by Lemma 21.8 (iv) in Sato (1999) that the sample path can be decomposed into a continuous $[X, Y]^c$ and pure jump $[X, Y]^d$ part. Therefore we can write

$$\begin{aligned} [X, Y]_t &= [X, Y]_t^c + [X, Y]_t^d = [X, Y]_t^c + \sum_{0 \leq u \leq t} \Delta[X, Y]_u^d \\ &= [X, Y]_t^c + \sum_{0 \leq u \leq t} \Delta X_u \Delta Y_u. \end{aligned} \quad (\text{A.2})$$

In particular the quadratic variation of a Lévy process L with Lévy-Itô decomposition (1.11) is given by

$$\begin{aligned} [L, L]_t &= [L, L]_t^c + [L, L]_t^d = [L, L]_t^c + \sum_{0 \leq u \leq t} (\Delta L_u)^2 \\ &= \sigma_L^2 t + \int_0^t \int_{\mathbb{R} - \{0\}} x^2 N_L(du, dx). \end{aligned} \quad (\text{A.3})$$

In Chapter 1.5 we need also the following quadratic variation processes:

$$\begin{aligned} [L, [L, L]]_t &= [L, [L, L]]_t^c + \sum_{0 \leq u \leq t} \Delta L_u \Delta [L, L]_u = [L^c, [L, L]^c]_t + \sum_{0 \leq u \leq t} (\Delta L_u)^3 \\ &= [\gamma_L \cdot + B, \sigma_L^2 \cdot]_t + \sum_{0 \leq u \leq t} (\Delta L_u)^3 = \sum_{0 \leq u \leq t} (\Delta L_u)^3, \end{aligned} \quad (\text{A.4})$$

where the last equality follows from the formal relations $dt \cdot dt = dt \cdot dB_t = 0$.

$$\begin{aligned} [[L, L], [L, L]]_t &= [[L, L], [L, L]]_t^c + \sum_{0 \leq u \leq t} \Delta [L, L]_u \Delta [L, L]_u \\ &= [\sigma_L^2 \cdot, \sigma_L^2 \cdot]_t + \sum_{0 \leq u \leq t} (\Delta L_u)^4 = \sum_{0 \leq u \leq t} (\Delta L_u)^4. \end{aligned} \quad (\text{A.5})$$

The quadratic variation of two Lévy-type integrals can be expressed as a stochastic integral with respect to the quadratic variation of the driving Lévy processes. Assume therefore $H, K \in \mathcal{H}_2(\mathbb{R})$ and X and Y are two Lévy processes. Then

$$\left[\int_0^\cdot H_u dX_u, \int_0^\cdot K_u dY_u \right]_t = \int_0^t H_u K_u d[X, Y]_u, \quad (\text{A.6})$$

for each $t \geq 0$. For a proof of this result see Theorem II.29 in Protter (2004).

In Chapter 2.3 we use the β -mixing concept which is defined below.

Definition A.16 For a stationary process $Z = (Z_s)_{s \geq 0}$ define the σ -algebras $\mathcal{F}_{[0, u]}^Z := \sigma((Z_s)_{s \in [0, u]})$ and $\mathcal{F}_{[u+t, \infty)}^Z := \sigma((Z_s)_{s \geq u+t})$ for all $u \geq 0$. Then Z is called β -mixing, if

$$\begin{aligned} \beta(t) &= \beta(\mathcal{F}_{[0, u]}^Z, \mathcal{F}_{[u+t, \infty)}^Z) \\ &:= \mathbb{E}(\sup\{|\mathbb{P}(B | \mathcal{F}_{[0, u]}^Z) - \mathbb{P}(B)| : B \in \mathcal{F}_{[u+t, \infty)}^Z\}) \rightarrow 0, \end{aligned}$$

as $t \rightarrow \infty$, for all $u \geq 0$.

From Proposition 1.1 in Doukhan (1994) we know that β -mixing implies α -mixing, since

$$2\alpha(\mathcal{F}_{[0, u]}^Z, \mathcal{F}_{[u+t, \infty)}^Z) \leq \beta(\mathcal{F}_{[0, u]}^Z, \mathcal{F}_{[u+t, \infty)}^Z).$$

A.2 Auxiliary results

A useful tool to calculate certain expression when dealing with Lévy processes is the *compensation formula* (see e.g. Section 0.5 in Bertoin (1996)).

Proposition A.17 (Compensation formula) *Let L be a Lévy process with jumps taking values in $\mathbb{R} - \{0\}$ and $H \in \mathcal{P}_2(T, \mathbb{R} - \{0\})$. Then we have*

$$\mathbb{E} \left(\sum_{0 \leq u \leq T} H(u, \Delta L_u) \right) = \mathbb{E} \left(\int_0^T \int_{\mathbb{R}} H(u, x) \nu_L(dx) du \right). \quad (\text{A.7})$$

Lemma A.18 (Applebaum (2004) Lemma 4.2.2) *Let $F \in \mathcal{H}_2(T, E)$ be a simple mapping, i.e.*

$$F(t, x) = \sum_{i=1}^m \sum_{j=1}^k F_{i,j} \chi_{(t_i, t_{i+1}]}(t) \chi_{A_j}(x),$$

for each $t \in [0, T]$, $x \in E$ and where each $F_{i,j} \in b\mathcal{F}_{t_i}$, and M a martingale-valued random measure satisfying (M1)-(M3). The stochastic integral of F with respect to M is defined as

$$I_T(F) := \sum_{i=1}^m \sum_{j=1}^k F_{i,j} [M(t_{i+1}, A_j) - M(t_i, A_j)],$$

for each $T > 0$. Then we have

$$\mathbb{E}(I_T(F)) = 0, \quad (\text{A.8})$$

$$\mathbb{E}(I_T(F)^2) = \int_0^T \int_E \mathbb{E}(|F(t, x)|^2) \mu(dx) dt. \quad (\text{A.9})$$

Proof: Due to (M2) is $M(t_{i+1}, A_j) - M(t_i, A_j)$ independent of \mathcal{F}_{t_i} and hence

$$\mathbb{E}(I_T(F)) = \sum_{i=1}^m \sum_{j=1}^k \mathbb{E}(F_{i,j}) \mathbb{E}([M(t_{i+1}, A_j) - M(t_i, A_j)]) = 0.$$

The second moment is equal to

$$\begin{aligned} & \mathbb{E}(I_T(F)^2) \\ &= \sum_{i,j=1}^{m,k} \sum_{h,l=1}^{m,k} \mathbb{E}(F_{i,j} [M(t_{i+1}, A_j) - M(t_i, A_j)] F_{h,l} [M(t_{h+1}, A_l) - M(t_h, A_l)]) \\ &= \sum_{i,j=1}^{m,k} \sum_{l=1}^k \sum_{h < i} \mathbb{E}(F_{i,j} [M(t_{i+1}, A_j) - M(t_i, A_j)] F_{h,l} [M(t_{h+1}, A_l) - M(t_h, A_l)]) \\ & \quad + \sum_{i,j=1}^{m,k} \sum_{l=1}^k \mathbb{E}(F_{i,j} [M(t_{i+1}, A_j) - M(t_i, A_j)] F_{i,l} [M(t_{i+1}, A_l) - M(t_i, A_l)]) \\ & \quad + \sum_{i,j=1}^{m,k} \sum_{l=1}^k \sum_{h > i} \mathbb{E}(F_{i,j} [M(t_{i+1}, A_j) - M(t_i, A_j)] F_{h,l} [M(t_{h+1}, A_l) - M(t_h, A_l)]) \end{aligned}$$

The first and third term are zero due to the independent increment property (M2) of M , since

$$\begin{aligned} & \sum_{i,j=1}^{m,k} \sum_{l=1}^k \sum_{h<i} \mathbb{E}(F_{i,j}[M(t_{i+1}, A_j) - M(t_i, A_j)]F_{h,l}[M(t_{h+1}, A_l) - M(t_h, A_l)]) \\ &= \sum_{i,j=1}^{m,k} \sum_{l=1}^k \sum_{h<i} \mathbb{E}(F_{i,j}F_{h,l}[M(t_{h+1}, A_l) - M(t_h, A_l)])\mathbb{E}([M(t_{i+1}, A_j) - M(t_i, A_j)]) \\ &= 0, \end{aligned}$$

and similarly for the third term. Again using (M2) and (R3) in Definition A.11 we get

$$\begin{aligned} & \sum_{i,j=1}^{m,k} \sum_{l=1}^k \mathbb{E}(F_{i,j}[M(t_{i+1}, A_j) - M(t_i, A_j)]F_{i,l}[M(t_{i+1}, A_l) - M(t_i, A_l)]) \\ &\stackrel{(M2)}{=} \sum_{i,j=1}^{m,k} \sum_{l=1}^k \mathbb{E}(F_{i,j}F_{i,l})\mathbb{E}([M(t_{i+1}, A_j) - M(t_i, A_j)][M(t_{i+1}, A_l) - M(t_i, A_l)]) \\ &= \sum_{i,j=1}^{m,k} \sum_{\substack{l=1 \\ l \neq j}}^k \mathbb{E}(F_{i,j}F_{i,l})\mathbb{E}([M(t_{i+1}, A_j) - M(t_i, A_j)][M(t_{i+1}, A_l) - M(t_i, A_l)]) \\ &\quad + \sum_{i,j=1}^{m,k} \mathbb{E}(F_{i,j}^2)\mathbb{E}([M(t_{i+1}, A_j) - M(t_i, A_j)]^2) \\ &\stackrel{(R3)}{=} \sum_{i,j=1}^{m,k} \mathbb{E}(F_{i,j}^2)\mathbb{E}([M(t_{i+1}, A_j) - M(t_i, A_j)]^2) \end{aligned}$$

Using the martingale property of M and (M3) we finally get

$$\begin{aligned} & \sum_{i,j=1}^{m,k} \mathbb{E}(F_{i,j}^2)\mathbb{E}([M(t_{i+1}, A_j) - M(t_i, A_j)]^2) \\ &= \sum_{i,j=1}^{m,k} \mathbb{E}(F_{i,j}^2) [\mathbb{E}(M(t_{i+1}, A_j)^2) - 2\mathbb{E}(\mathbb{E}(M(t_{i+1}, A_j)M(t_i, A_j)|\mathcal{F}_t)) + \mathbb{E}(M(t_i, A_j)^2)] \\ &= \sum_{i,j=1}^{m,k} \mathbb{E}(F_{i,j}^2) [\mathbb{E}(M(t_{i+1}, A_j)^2) - \mathbb{E}(M(t_i, A_j)^2)] \\ &= \sum_{i,j=1}^{m,k} \mathbb{E}(F_{i,j}^2)(t_{i+1} - t_i)\mu(A_j) = \int_0^T \int_E \mathbb{E}(|F(t, x)|^2)\mu(dx)dt. \end{aligned}$$

□

Proposition A.19 *Let Y be the stationary CARMA(p, q) process*

$$Y_t = \mathbf{b}^T e^{\mathcal{A}t} \mathbf{X}_0 + \int_0^t \mathbf{b}^T e^{\mathcal{A}(t-u)} \mathbf{1}_p dL_u.$$

Then the mean and autocovariance function of Y are

$$\mathbb{E}(Y_t) = -\mathbf{b}^T \mathcal{A}^{-1} \mathbf{1}_p \mathbb{E}(L_1) \quad \text{and} \quad \text{Cov}(Y_{t+h}, Y_t) = \text{Var}(L_1) \mathbf{b}^T e^{\mathcal{A}h} \Sigma \mathbf{b},$$

where $\Sigma := \int_0^\infty e^{\mathcal{A}u} \mathbf{1}_p \mathbf{1}_p^T e^{\mathcal{A}^T u} du$.

Proof: The stationary solution of the state equation (1.35) is given by

$$\begin{aligned} \mathbf{X}_t &= e^{\mathcal{A}t} \mathbf{X}_0 + \int_0^t e^{\mathcal{A}(t-u)} \mathbf{1}_p dL_u \\ &= e^{\mathcal{A}t} \mathbf{X}_0 + \int_0^t e^{\mathcal{A}(t-u)} \mathbf{1}_p \left(dB_u + \gamma_L du + \int_{|x|<1} x \tilde{N}_L(du, dx) \right. \\ &\quad \left. + \int_{|x|\geq 1} x N_L(du, dx) \right). \end{aligned}$$

Hence we get

$$\begin{aligned} \mathbb{E}(\mathbf{X}_t) &= e^{\mathcal{A}t} \mathbb{E}(\mathbf{X}_0) + \int_0^t e^{\mathcal{A}(t-u)} \mathbf{1}_p \left(\gamma_L + \int_{|x|\geq 1} x \nu_L(dx) \right) du \\ &= e^{\mathcal{A}t} \mathbb{E}(\mathbf{X}_0) + (e^{\mathcal{A}t} - \mathcal{I}_p) \mathcal{A}^{-1} \mathbf{1}_p \mathbb{E}(L_1). \end{aligned}$$

Since \mathbf{X} is stationary, $\mathbb{E}(\mathbf{X}_t)$ has to be independent of t . Thus $\mathbb{E}(\mathbf{X}_0)$ has to be $-\mathcal{A}^{-1} \mathbf{1}_p \mathbb{E}(L_1)$, implying

$$\mathbb{E}(Y_t) = -\mathbf{b}^T \mathcal{A}^{-1} \mathbf{1}_p \mathbb{E}(L_1).$$

For computing the autocovariance $\text{Cov}(\mathbf{X}_{t+h}, \mathbf{X}_t)$, $h \geq 0$, consider the zero mean state process $\tilde{\mathbf{X}}_t := \mathbf{X}_t - \mathbb{E}(\mathbf{X}_t) = e^{\mathcal{A}t} \tilde{\mathbf{X}}_0 + \int_0^t e^{\mathcal{A}(t-u)} \mathbf{1}_p d\tilde{L}_u$, with $\tilde{L}_t := L_t - \mathbb{E}(L_t)$

for each $t \geq 0$. Then

$$\begin{aligned}
& \text{Cov}(\mathbf{X}_{t+h}, \mathbf{X}_t) \\
&= \mathbb{E} \left[\left(e^{\mathcal{A}(t+h)} \tilde{\mathbf{X}}_0 + \int_0^{t+h} e^{\mathcal{A}(t+h-u)} \mathbf{1}_p d\tilde{L}_u \right) \left(e^{\mathcal{A}t} \tilde{\mathbf{X}}_0 + \int_0^t e^{\mathcal{A}(t-u)} \mathbf{1}_p d\tilde{L}_u \right)^T \right] \\
&= e^{\mathcal{A}(t+h)} \mathbb{E}(\tilde{\mathbf{X}}_0 \tilde{\mathbf{X}}_0^T) e^{\mathcal{A}^T t} + e^{\mathcal{A}(t+h)} \mathbb{E} \left[\tilde{\mathbf{X}}_0 \left(\int_0^t e^{\mathcal{A}(t-u)} \mathbf{1}_p d\tilde{L}_u \right)^T \right] \\
&\quad + \mathbb{E} \left[\int_0^{t+h} e^{\mathcal{A}(t+h-u)} \mathbf{1}_p d\tilde{L}_u \tilde{\mathbf{X}}_0^T e^{\mathcal{A}^T t} \right] \\
&\quad + \mathbb{E} \left[\int_0^{t+h} e^{\mathcal{A}(t+h-u)} \mathbf{1}_p (dB_u + \int_{\mathbb{R}-\{0\}} x \tilde{N}_L(du, dx)) \right. \\
&\quad \quad \left. \times \left(\int_0^t e^{\mathcal{A}(t-u)} \mathbf{1}_p (dB_u + \int_{\mathbb{R}-\{0\}} x \tilde{N}_L(du, dx)) \right)^T \right] \\
&= e^{\mathcal{A}(t+h)} \left[\text{Cov}(\mathbf{X}_0, \mathbf{X}_0) + \int_0^{t+h} e^{-\mathcal{A}u} \mathbf{1}_p \mathbf{1}_p^T e^{-\mathcal{A}^T u} (\sigma_L^2 + \int_{\mathbb{R}} x^2 \nu_L(dx)) du \right] e^{\mathcal{A}^T t}
\end{aligned}$$

One can show that $\text{Cov}(\mathbf{X}_t, \mathbf{X}_t)$ solves the linear differential equation

$$\frac{d}{dt} \boldsymbol{\nu}_t = \mathcal{A} \boldsymbol{\nu}_t + \boldsymbol{\nu}_t \mathcal{A}^T + \text{Var}(L_1) \mathbf{1}_p \mathbf{1}_p^T$$

(cf. (6.13)' in Karatzas and Shreve (1991)). But since \mathbf{X} is stationary, $\text{Cov}(\mathbf{X}_t, \mathbf{X}_t)$ has to be a constant solution, i.e. $\text{Cov}(\mathbf{X}_t, \mathbf{X}_t)$ has to solve the algebraic matrix equation

$$\mathcal{A} \boldsymbol{\nu} + \boldsymbol{\nu} \mathcal{A}^T = -\text{Var}(L_1) \mathbf{1}_p \mathbf{1}_p^T. \quad (\text{A.10})$$

Substituting the right hand side of (A.10) into the expression for $\boldsymbol{\Sigma}$

$$\begin{aligned}
\boldsymbol{\Sigma} &= \int_0^\infty e^{\mathcal{A}u} \mathbf{1}_p \mathbf{1}_p^T e^{\mathcal{A}^T u} du = -\text{Var}(L_1)^{-1} \int_0^\infty e^{\mathcal{A}u} (\mathcal{A} \boldsymbol{\nu} + \boldsymbol{\nu} \mathcal{A}^T) e^{\mathcal{A}^T u} du \\
&= -\text{Var}(L_1)^{-1} \int_0^\infty \frac{d}{du} (e^{\mathcal{A}u} \boldsymbol{\nu} e^{\mathcal{A}^T u}) du = -\text{Var}(L_1)^{-1} e^{\mathcal{A}u} \boldsymbol{\nu} e^{\mathcal{A}^T u} \Big|_0^\infty \\
&= \text{Var}(L_1)^{-1} \boldsymbol{\nu}.
\end{aligned}$$

shows that $\text{Var}(L_1) \boldsymbol{\Sigma}$ is a solution to (A.10) and thus $\text{Cov}(\mathbf{X}_t, \mathbf{X}_t) = \text{Var}(L_1) \boldsymbol{\Sigma}$. From

$$\text{Cov}(\mathbf{X}_t, \mathbf{X}_t) = e^{\mathcal{A}t} \left[\text{Cov}(\mathbf{X}_0, \mathbf{X}_0) + \text{Var}(L_1) \int_0^t e^{-\mathcal{A}u} \mathbf{1}_p \mathbf{1}_p^T e^{-\mathcal{A}^T u} du \right] e^{\mathcal{A}^T t}$$

it then follows that

$$\int_0^t e^{-\mathcal{A}u} \mathbf{1}_p \mathbf{1}_p^T e^{-\mathcal{A}^T u} du = e^{-\mathcal{A}t} \Sigma e^{-\mathcal{A}^T t} - \Sigma.$$

Thus one gets for the autocovariance

$$\begin{aligned} \mathbb{C}\text{ov}(\mathbf{X}_{t+h}, \mathbf{X}_t) &= e^{\mathcal{A}(t+h)} \left[\text{Var}(L_1) \left(\Sigma + e^{-\mathcal{A}t} \Sigma e^{-\mathcal{A}^T t} - \Sigma \right) \right] e^{\mathcal{A}^T t} \\ &= \text{Var}(L_1) e^{\mathcal{A}h} \Sigma, \end{aligned}$$

which implies

$$\mathbb{C}\text{ov}(Y_{t+h}, Y_t) = \text{Var}(L_1) \mathbf{b}^T e^{\mathcal{A}h} \Sigma \mathbf{b}.$$

□

Appendix B

Variogram for irregularly spaced time series

The variogram is mainly used in geostatistics. Applications for time series data are rare, despite the fact that it has the advantage to be defined for irregularly spaced and even non-stationary time series in comparison to the autocovariance function (see Haslett (1997)).

Definition B.1 (*variogram*) Let $(Z_t)_{0 \leq t < \infty}$ be a process, such that

$$\mathbb{E}(Z_{t+h} - Z_t) = Ch,$$

with a constant C , and

$$\text{Var}(Z_{t+h} - Z_t) =: 2\gamma(h), \tag{B.1}$$

where $\gamma(h)$ is a conditionally negative definite function. Then $\gamma(h)$ is called the variogram.

Remark B.2 The requirement that $\gamma(h)$ be conditionally negative definite means that $\text{Var}(\sum_i a_i Y_{t_i})$ (which is equal to $-\sum_{i,j} a_i a_j \gamma(t_i - t_j)$ when $\sum_i a_i = 0$) be non-negative definite when $\sum_i a_i = 0$.

For observations Z_{t_1}, \dots, Z_{t_n} , with $C = 0$, the variogram can be estimated through the *sample variogram*

$$\hat{\gamma}(h) := \frac{1}{2}(n - |N_h|)^{-1} \sum_{(i,j) \in I_h} (Z_{t_i} - Z_{t_j})^2, \tag{B.2}$$

where $N_h := \{(i, j), i, j \in \{1, \dots, n\} \mid |t_i - t_j| = h\}$.

To compare the sample variogram of the residuals $(\hat{\varepsilon}_{t_i})$ in (6.5) with the theoretical one, we have to compute the variogram of (ε_{t_i}) . It is given by the following expression

$$\gamma_\varepsilon(h) = \text{Var}(L_1) \mathbf{b}^T (I_p - e^{Ah}) \Sigma \mathbf{b} + \sigma_\varepsilon^2. \tag{B.3}$$

Example B.3 As an example consider a Lévy driven CARMA(p,q) process (Y_t) . Here the driving Lévy process (L_t) is chosen to be a compound Poisson process with exponentially distributed jumps, i.e.

$$L_t = \sum_{k=1}^{N_t} X_k,$$

where (X_k) i.i.d. with density $f(x) = 100e^{-100x}$ and $N_t \sim \text{Pois}(15t)$. The simulated sample path has 2000 equidistant observations. The variogram $\gamma(h)$ and sample variogram $\hat{\gamma}(h)$ for the following parameter sets:

- (1) $p = 1, q = 0, a(z) = z + 0.1, b(z) = 1$
- (2) $p = 2, q = 1, a(z) = z^2 + 0.9z + 0.5, b(z) = 1 + z$
- (3) $p = 2, q = 1, a(z) = z^2 + 0.09z + 0.5, b(z) = 1 + z$
- (4) $p = 3, q = 2, a(z) = z^3 + 1.1z^2 + 2.8174z + 0.2717, b(z) = 1 + 5z + z^2$.

are shown in Figure B.1. They are all computed for a maximal lag of 30. Figure B.1 shows the flexibility of the CARMA(p,q) process to model a wide variety of correlation structures, represented by a slowly, fast increasing or oscillating variogram.

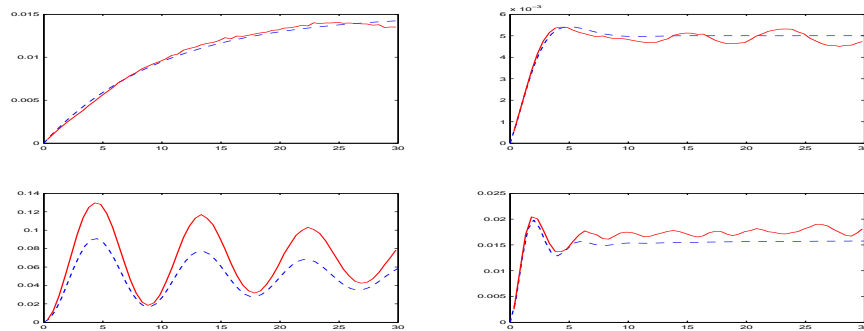


Figure B.1: The variogram $\gamma(h)$ (dashed line) and sample variogram $\hat{\gamma}(h)$ (solid line) for the following processes: (i) CAR(1) with $a(z) = z + 0.1$ and $b(z) = 1$ (top left), (ii) CARMA(2,1) with $a(z) = z^2 + 0.9z + 0.5$ and $b(z) = 1 + z$ (top right), (iii) CARMA(2,1) with $a(z) = z^2 + 0.09z + 0.5$ and $b(z) = 1 + z$ (bottom left), (iv) CARMA(3,2) with $a(z) = z^3 + 1.1z^2 + 2.8174z + 0.2717$ and $b(z) = 1 + 5z + z^2$ (bottom right)

References

- Aït-Sahalia, Y., P. Mykland, and L. Zhang (2005). How often to sample a continuous-time process in the presence of market microstructure noise. *Review of Financial Studies* **18**, 351–416.
- Andersen, T. and T. Bollerslev (1997). Heterogeneous information arrivals and return volatility dynamics: Uncovering the long-run in high frequency returns. *J. Finance* **52**, 975–1005.
- Andersen, T., T. Bollerslev, and F. Diebold (2006). Roughing it up: Including jump components in the measurement, modeling and forecasting of return volatility. *Review of Economics and Statistics* *forthcoming*.
- Anh, V. V., C. C. Heyde, and N. N. Leonenko (2002). Dynamic models of long-memory processes driven by Lévy noise. *J. Appl. Probab.* **39**(4), 730–747.
- Applebaum, D. (2004). *Lévy Processes and Stochastic Calculus*, Volume 93 of *Cambridge Studies in Advanced Mathematics*. Cambridge: Cambridge University Press.
- Baillie, R. T. (1996). Long memory processes and fractional integration in econometrics. *J. Econometrics* **73**(1), 5–59.
- Baillie, R. T., T. Bollerslev, and H. O. Mikkelsen (1996). Fractionally integrated generalized autoregressive conditional heteroskedasticity. *J. Econometrics* **74**(1), 3–30.
- Barndorff-Nielsen, O. E. and N. Shephard (2001). Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial economics. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **63**(2), 167–241.
- Barndorff-Nielsen, O. E. and N. Shephard (2002). Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **64**(2), 253–280.
- Barndorff-Nielsen, O. E. and R. Stelzer (2007). Positive-definite matrix processes of finite variation. *Probability and Mathematical Statistics* **27**(1). To appear.
- Bauwens, L. and P. Giot (2001). *Econometric Modelling of Stock Market Intraday Activity*. Dordrecht: Kluwer Academic.

- Bauwens, L., P. Giot, J. Grammig, and D. Veredas (2004). A comparison of financial duration models through density forecasts. *Int. J. Forecasting* **20**, 589–609.
- Bertoin, J. (1996). *Lévy Processes*, Volume 121 of *Cambridge Tracts in Mathematics*. Cambridge: Cambridge University Press.
- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *J. Econometrics* **31**(3), 307–327.
- Bollerslev, T. and H. Mikkelsen (1996). Modeling and pricing long memory in stock market volatility. *J. Econometrics* **73**, 151–184.
- Brockwell, P., E. Chandraa, and A. Lindner (2006). Continuous time GARCH processes. *Ann. Appl. Probab.* **16**(2), 790–826.
- Brockwell, P. J. (2001a). Continuous-time ARMA processes. In *Stochastic processes: theory and methods*, Volume 19 of *Handbook of Statist.*, pp. 249–276. Amsterdam: North-Holland.
- Brockwell, P. J. (2001b). Lévy-driven CARMA processes. *Ann. Inst. Statist. Math.* **53**(1), 113–124. Nonlinear non-Gaussian models and related filtering methods (Tokyo, 2000).
- Brockwell, P. J. (2004). Representations of continuous-time ARMA processes. *J. Appl. Probab.* **41A**, 375–382. Stochastic methods and their applications.
- Brockwell, P. J. and R. A. Davis (1991). *Time Series: Theory and Methods* (Second ed.). Springer Series in Statistics. New York: Springer-Verlag.
- Brockwell, P. J. and T. Marquardt (2005). Lévy-driven and fractionally integrated ARMA processes with continuous time parameter. *Statist. Sinica* **15**(2), 477–494.
- Brodin, E. and C. Klüppelberg (2006). Modelling, estimation and visualization of multivariate dependence for high-frequency data. Submitted.
- Comte, F. and E. Renault (1998). Long memory in continuous-time stochastic volatility models. *Math. Finance* **8**(4), 291–323.
- Cont, R. (2001). Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance* **1**, 223–236.
- Cont, R. and P. Tankov (2004). *Financial Modelling with Jump Processes*. Chapman & Hall/CRC Financial Mathematics Series. Chapman & Hall/CRC, Boca Raton, FL.
- Corsi, F. (2004). A simple long memory model of realized volatility. Available at SSRN: <http://ssrn.com/abstract=626064>.
- Doukhan, P. (1994). *Mixing*, Volume 85 of *Lecture Notes in Statistics*. New York: Springer-Verlag. Properties and examples.

- Drost, F. C. and B. J. M. Werker (1996). Closing the GARCH gap: continuous time GARCH modeling. *J. Econometrics* **74**(1), 31–57.
- Durbin, J. and S. J. Koopman (2001). *Time Series Analysis by State Space Methods*, Volume 24 of *Oxford Statistical Science Series*. Oxford: Oxford University Press.
- Engle, R. (2000). The econometrics of ultra high frequency data. *Econometrica* **68**, 1–22.
- Engle, R. and T. Bollerslev (1986). Modeling the persistence of conditional variances. *Econometric Reviews* **5**, 1–50.
- Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* **50**(4), 987–1007.
- Engle, R. F. and J. R. Russell (1998). Autoregressive conditional duration: a new model for irregularly spaced transaction data. *Econometrica* **66**(5), 1127–1162.
- Fasen, V. (2007). Asymptotic results for sample autocovariance functions and extremes of integrated generalized Ornstein-Uhlenbeck processes. Submitted.
- Fasen, V., C. Klüppelberg, and A. Lindner (2006). Extremal behavior of stochastic volatility models. In *Stochastic Finance*, pp. 107–155. Berlin: Springer.
- Fay, G., B. Gonzalez-Arevalo, T. Mikosch, and G. Samorodnitsky (2006). Modeling teletraffic arrivals by a Poisson cluster process. *Queueing Systems fourthcoming*.
- Forsberg, L. and E. Ghysels (2004). Why do absolute returns predict volatility so well? Available at <http://www.unc.edu/eghysels/>.
- Ghysels, E., P. Santa-Clara, and V. R. (2002). The MIDAS touch: Mixed data sampling regressions. Available at <http://www.unc.edu/eghysels/>.
- Glosten, L., R. Jagannathan, and D. Runkle (1993). On the relation between the expected value and the volatility of the nominal excess on stocks. *J. Finance* **43**, 1770–1801.
- Gneiting, T. and A. Raftery (2006). Strictly proper scoring rules, prediction and estimation. *JASA*. Forthcoming.
- Hannan, E. J. and M. Deistler (1988). *The Statistical Theory of Linear Systems*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. New York: John Wiley & Sons Inc.
- Harvey, A. (1990). *Forecasting structural time series models and the Kalman filter*. Cambridge: Cambridge University Press.
- Haslett, J. (1997). On the sample variogram and the sample autocovariance for non-stationary time series. *The Statistician* **46**, 475–485.

- Hastie, T. J. and R. J. Tibshirani (1990). *Generalized Additive Models*, Volume 43 of *Monographs on Statistics and Applied Probability*. London: Chapman and Hall Ltd.
- Haug, S. and C. Czado (2006a). An exponential continuous time GARCH process. Submitted.
- Haug, S. and C. Czado (2006b). A fractionally integrated ECOGARCH process. Submitted.
- Haug, S. and C. Czado (2006c). Mixed effect models for absolute log returns of ultra high frequency data. *Appl. Stochastic Models Bus. Ind.* **22**, 243–267.
- Haug, S., C. Klüppelberg, A. Lindner, and M. Zapp (2007). Method of moment estimation in the COGARCH(1, 1) model. *The Econometrics Journal*. Forthcoming.
- Huber, P. J. (2004). *Robust Statistics*. New York: John Wiley & Sons Inc. Wiley Series in Probability and Mathematical Statistics.
- Ibragimov, I. A. and Y. V. Linnik (1971). *Independent and Stationary Sequences of Random Variables*. Wolters-Noordhoff Publishing, Groningen. With a supplementary chapter by I. A. Ibragimov and V. V. Petrov, Translation from the Russian edited by J. F. C. Kingman.
- Jones, R. H. (1981). Fitting a continuous time autoregression to discrete data. In *Applied time series analysis II.*, pp. 651–682. New York: Academic Press.
- Jones, R. H. (1985). Time series analysis with unequally spaced data. In *Time Series in the Time Domain*, Volume 5 of *Handbook of Statist.*, pp. 157–177. Amsterdam: North-Holland.
- Jones, R. H. (1993). *Longitudinal Data with Serial Correlation: A State-space Approach*, Volume 47 of *Monographs on Statistics and Applied Probability*. London: Chapman & Hall.
- Jungbacker, B. and S. Koopman (2006). Model-based measurement of actual volatility in high-frequency data. In *Econometric Analysis of Financial and Economic Time Series*, Volume 20, Part A of *Advances in Econometrics*. New York: JAI Press.
- Karatzas, I. and S. E. Shreve (1991). *Brownian Motion and Stochastic Calculus* (Second ed.), Volume 113 of *Graduate Texts in Mathematics*. New York: Springer-Verlag.
- Kazakevičius, V. and R. Leipus (2003). A new theorem on the existence of invariant distributions with applications to ARCH processes. *J. Appl. Probab.* **40**(1), 147–162.

- Klüppelberg, C., A. Lindner, and R. Maller (2004). A continuous-time GARCH process driven by a Lévy process: stationarity and second-order behaviour. *J. Appl. Probab.* **41**(3), 601–622.
- Klüppelberg, C., A. Lindner, and R. Maller (2006). Continuous time volatility modelling: COGARCH versus Ornstein-Uhlenbeck models. In *From stochastic calculus to mathematical finance*, pp. 393–419. Berlin: Springer.
- Kyprianou, A. (2006). *Introductory Lectures on Fluctuations of Lévy Processes with Applications*. Universitext. Springer-Verlag.
- Liesenfeld, R. and W. Pohlmeier (2003). A dynamic integer count data model for financial transaction prices. Available at <http://econometrics.wiwi.uni-konstanz.de/prof/papers>.
- Lindner, A. and R. Maller (2005). Lévy integrals and the stationarity of generalised Ornstein-Uhlenbeck processes. *Stochastic Process. Appl.* **115**(10), 1701–1722.
- Madan, D. B., P. Carr, and E. Chang (1998). The Variance Gamma process and option pricing. *European Finance Review* **2**, 79–105.
- Madan, D. B. and E. Seneta (1990). The Variance Gamma (V.G.) model for share market returns. *J. Business* **63**, 511–524.
- Marquardt, T. (2006a). Fractional Lévy processes, CARMA processes and related topics. PhD thesis, Centre of Mathematical Sciences, Munich University of Technology.
- Marquardt, T. (2006b). Fractional Lévy processes with an application to long memory moving average processes. *Bernoulli* **12**(6), 1009–1126.
- Masuda, H. (2004). On multidimensional Ornstein-Uhlenbeck processes driven by a general Lévy process. *Bernoulli* **10**(1), 97–120.
- McCullagh, P. and J. A. Nelder (1983). *Generalized Linear Models*. Monographs on Statistics and Applied Probability. London: Chapman & Hall.
- McLeish, D. L. (1975). Invariance principles for dependent variables. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **32**(3), 165–178.
- Meddahi, N., E. Renault, and B. Werker (1998). Modelling high-frequency data in continuous time. Available at <http://www.cireq.umontreal.ca/personnel/meddahi.html>.
- Mikosch, T. and C. Stărică (2000). Limit theory for the sample autocorrelations and extremes of a GARCH (1, 1) process. *Ann. Statist.* **28**(5), 1427–1451.
- Nelson, D. B. (1990). ARCH models as diffusion approximations. *J. Econometrics* **45**(1-2), 7–38.

- Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: a new approach. *Econometrica* **59**(2), 347–370.
- Pigorsch, C. and R. Stelzer (2007). A multivariate generalization of the Ornstein-Uhlenbeck stochastic volatility model. Submitted.
- Priestley, M. B. (1981). *Spectral Analysis and Time Series. Vol. 1*. London: Academic Press Inc. [Harcourt Brace Jovanovich Publishers]. Univariate series, Probability and Mathematical Statistics.
- Protter, P. E. (2004). *Stochastic Integration and Differential Equations* (Second ed.), Volume 21 of *Applications of Mathematics (New York)*. Berlin: Springer-Verlag. Stochastic Modelling and Applied Probability.
- Rajput, B. S. and J. Rosiński (1989). Spectral representations of infinitely divisible processes. *Probab. Theory Related Fields* **82**(3), 451–487.
- Renault, E. and B. Werker (2006). Causality effects in return volatility measures with random times. Tilburg University, available at: <http://center.uvt.nl/staff/werker/preprints/>.
- Rocha-Arteaga, A. and K. Sato (2003). *Topics in Infinitely Divisible Distributions and Lévy Processes*, Volume 17 of *Aportaciones Matemáticas: Investigación [Mathematical Contributions: Research]*. México: Sociedad Matemática Mexicana.
- Russell, J. R. and R. F. Engle (2005). A discrete-state continuous-time model of financial transactions prices and times: the autoregressive conditional multinomial—autoregressive conditional duration model. *J. Bus. Econom. Statist.* **23**(2), 166–180.
- Samko, S. G., A. A. Kilbas, and O. I. Marichev (1993). *Fractional Integrals and Derivatives*. Yverdon: Gordon and Breach Science Publishers. Theory and applications, Edited and with a foreword by S. M. Nikol'skiĭ, Translated from the 1987 Russian original, Revised by the authors.
- Sato, K. (1999). *Lévy Processes and Infinitely Divisible Distributions*, Volume 68 of *Cambridge Studies in Advanced Mathematics*. Cambridge: Cambridge University Press. Translated from the 1990 Japanese original, Revised by the author.
- Sato, K. and M. Yamazato (1983). Stationary processes of Ornstein-Uhlenbeck type. In *Probability theory and mathematical statistics (Tbilisi, 1982)*, Volume 1021 of *Lecture Notes in Math.*, pp. 541–551. Berlin: Springer.
- Schall, R. (1991). Estimation in generalised linear models with random effects. *Biometrika* **78**, 719–727.
- Schoutens, W. (2003). *Lévy processes in Finance: Pricing Financial Derivatives*. Wiley.

- Straumann, D. (2005). *Estimation in Conditionally Heteroscedastic Time Series Models*, Volume 181 of *Lecture Notes in Statistics*. Berlin, Heidelberg: Springer-Verlag.
- Surgailis, D. and M.-C. Viano (2002). Long memory properties and covariance structure of the EGARCH model. *ESAIM Probab. Statist.* **6**, 311–329 (electronic). New directions in time series analysis (Luminy, 2001).
- Tsai, H. and K. S. Chan (2005). A note on non-negative continuous time processes. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **67**(4), 589–597.
- van der Vaart, A. W. (1998). *Asymptotic Statistics*, Volume 3 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge: Cambridge University Press.
- Wang, Y. (2002). Asymptotic nonequivalence of GARCH models and diffusions. *Ann. Statist.* **30**(3), 754–783. Dedicated to the memory of Lucien Le Cam.
- White, H. (1994). *Estimation, Inference and Specification Analysis*, Volume 22 of *Econometric Society Monographs*. Cambridge: Cambridge University Press.

Index

- Leb , iv
- ΔL_t , 8
- $\mathcal{H}_2(T, E)$, 19
- $\mathcal{P}(T, E)$, 20
- χ_A , iv
- δ_x , iv
- $\mathbb{M}_{d,m}(\mathbb{R})$, iv
- $b\mathcal{F}$, 140

- activity
 - (in-)finite, 10
- adapted process, 141
- auxiliary process, 35

- Brownian motion, 8

- càdlàg, 141
- càglàd, 141
- CARMA(p, q), 28
 - α -mixing, 31
- causal, 30
- characteristic triplet, 7
- COGARCH(1, 1), 35
 - $G, G^{(r)}, \sigma^2$, 35
 - α -mixing, 48
 - compound Poisson, 52
 - Variance Gamma, 55
- compensation formula, 145
- compound Poisson process, 10
- continuous in probability, 141

- Dirac measure, iv
- duration, 119

- ECOGARCH(p, q), 68
 - $G, G^{(r)}, \sigma^2$, 69
 - α -mixing, 74
- EGARCH(p, q), 67
- exponential matrix, 23

- FICARMA(p, d, q) process, 33
- FIECOGARCH(p, d, q), 106
 - $G_d, G_d^{(r)}, \sigma_d^2$, 107
- FIEGARCH(p, d, q), 105
- filtration, 140
 - right-(left-)continuous, 141
- fractional integral, 33

- GARCH(1, 1), 34

- infinitely divisible, 7
- integration by parts, 22
- isometry, 19
- Itô formula, 22

- Kalman filter, 26
 - augmented, 127
- kernel function, 30
 - long memory, 33

- Lévy measure, 7
- Lévy process, 6
 - jump, 8
 - two-sided, 17
- Lévy symbol, 7
- Lévy-Itô decomposition, 15
- Lévy-Khinchine formula, 7
- leverage effect, 63, 66, 82
- long memory, 32
 - FIECOGARCH(p, d, q), 111

- mapping, 18
 - simple, 19
- martingale, 142
- martingale property, 19
- measurable, 140
- mixed effect model, 122
 - prediction, 135
- mixing
 - $\tilde{\alpha}$ -, 47
 - α - (strong), 31
 - β -, 144
 - exponentially α -, 31
- modification, 142
- moment estimator, 46
 - asymptotic normality, 51
- moving average, 30
 - infinitely divisible, 30

- observation equation, 24

- Poisson integral, 14
- Poisson process, 9
- Poisson random measure, 13
 - compensated, 13
- predictable
 - σ - algebra, 141
 - mapping, 141
- product σ -algebra, 140

- quadratic variation, 143

- random measure, 142
 - martingale-valued, 18, 142

- short memory, 32
- state equation, 24
- state space model, 24
- stochastic integral, 20
 - Lévy-type, 21
 - simple mapping, 19
- stopping time, 141

- usual conditions, 141

- Variance Gamma process, 10
- variation, 142
 - finite, 17, 142
- variogram, 132, 150

Abbreviations

a.s.	almost surely
ARMA	autoregressive moving average
CARMA	continuous time ARMA
COGARCH	continuous time generalised autoregressive conditionally heteroscedastic
ECOGARCH	exponential COGARCH
EGARCH	exponential generalised autoregressive conditionally heteroscedastic
FICARMA	fractionally integrated CARMA
FIECOGARCH	fractionally integrated ECOGARCH
FIEGARCH	fractionally integrated EGARCH
GARCH	generalised autoregressive conditionally heteroscedastic
i.i.d.	independent identically distributed
OU	Ornstein Uhlenbeck
SDE	stochastic differential equation
SV	stochastic volatility

Symbols

$\mathbb{N}, \mathbb{N}_0, \mathbb{Z}$	$\{1, 2, \dots\}, \{0, 1, 2, \dots\}, \{\dots, -1, 0, 1, \dots\}$
\mathbb{R}, \mathbb{R}_+	$(-\infty, \infty), [0, \infty)$
$\mathcal{B}(\mathbb{R})$	Borel σ -algebra over \mathbb{R}
$\mathbb{M}_{d,m}(\mathbb{R})$	space of all real $d \times m$ matrices
\mathcal{A}^T	transpose of matrix \mathcal{A}
\mathcal{I}_d	identity in $\mathbb{M}_{d,d}(\mathbb{R})$
\log, \exp	natural logarithm, exponential function
\mathbb{P}, \mathbb{E}	probability, expectation
Var, Cov	variance, covariance
$\xrightarrow{d}, \xrightarrow{\mathbb{P}}, \xrightarrow{\text{a.s.}}$	convergence in distribution, probability, almost surely
$\stackrel{d}{=}$	equality in distribution
$\partial_i f$	i -th partial derivative of f
δ_x	Dirac measure at x
$\text{Leb}(A)$	Lebesgue measure of a set A
χ_A	indicator function of the set A
A^c, \bar{A}	complement, closure of a set A
L^p	space of p -integrable functions
$(\gamma_L, \sigma_L^2, \nu_L)$	characteristic triplet of a Lévy process L

Kurzfassung: In dieser Arbeit werden verschiedene zeitstetige Ansätze zur Modellierung hochfrequenter Finanzdaten behandelt. Neben dem Schätzen des bereits existierenden zeitstetigen GARCH(1,1) (COGARCH) Prozesses, werden drei neue Modelle vorgeschlagen. Zu Beginn stellen wir einen Momentenschätzer für die Parameter des COGARCH(1,1) Prozesses vor. Es wird gezeigt, dass die so definierten Schätzer konsistent und asymptotisch normalverteilt sind. Ferner wird das empirische Verhalten der Schätzer in zwei Simulationsstudien basierend auf dem compound Poisson und Variance Gamma COGARCH(1,1) Prozess untersucht. Schließlich wird das Modell an reale hochfrequenz Finanzdaten von der New York Stock Exchange angepasst. Als nächstes wird ein neues Modell, der exponentielle COGARCH(p, q) Prozess, zur Modellierung von logarithmierten Aktienpreisen mit stochastischer Volatilität definiert. Es werden Stationaritätseigenschaften sowie die Momentenstruktur untersucht. Für den Fall $p = q$ kann ein unmittelbar auftretender "leverage effect" nachgewiesen werden. Erste Schritte zum Schätzen dieses Modells werden ebenfalls unternommen. Es wird ein Quasi-Maximum Likelihood Schätzer für die Parameter des compound Poisson ECOGARCH(1,1) Prozesses vorgeschlagen. Um lang andauernden Abhängigkeiten in der Volatilität, welche von Zeit zu Zeit in empirischen Daten beobachtet werden, gerecht zu werden muß das ECOGARCH Modell erweitert werden. Zu diesem Zweck wird der fraktionell integrierte ECOGARCH(p, q) Prozess eingeführt. Ähnlich wie zuvor werden Stationaritätseigenschaften und die Momentenstruktur des neuen Modells untersucht. Ferner wird gezeigt, dass das sogenannte "Long Memory" Verhalten, welches per Definition in der logarithmierten Volatilität auftritt auch in der Volatilität selbst vorhanden ist. Im letzten Teil der Arbeit betrachten wir den absolut Betrag des logarithmierten Gewinns als eine Approximation der stochastischen Volatilität zum jeweiligen Zeitpunkt. Dadurch kann der Einfluß von Kovariablen auf den absolut Betrag des logarithmierten Gewinns, und somit die stochastische Volatilität, in "ultra" hochfrequenten Finanzdaten analysiert werden. Genauer gesagt schlagen wir ein neues "mixed effect" Modell für den absolut Betrag des logarithmierten Gewinns vor. Die Kovariablen werden verwendet, um einen fixen Effekt zu modellieren, wohingegen der zufällig Fehler im Modell zerlegt werden kann in einen nichtnegativen zeitstetigen ARMA Prozess und ein unabhängiges Rauschen, resultierend aus der Mikrostruktur des Marktes. Die Parameter des Modells werden geschätzt basierend auf der Zustandsraumdarstellung des Modells und der Verwendung des Kalman Filters. Das empirische Verhalten der Schätzer wird in einer Simulationsstudie untersucht. Abschließend wird das Modell auf IBM Daten angewandt und der Einfluß der Kovariablen "bid-ask spread" und "duration" auf einer täglichen Basis dargestellt.