Technische Universität München

Zentrum Mathematik

Parametric and Nonparametric Estimation of Positive Ornstein-Uhlenbeck Type Processes

Diplomarbeit

von

Maria Graf

Themensteller: Dr. Robert Stelzer
Betreuer: Dr. Robert Stelzer
Abgabetermin: 01. Dezember 2009

Hiermit erkläre ich, dass ich die Diplomarbeit selbstständig angefertigt und nur die angegebenen Quellen verwendet habe.

Garching, den 01. Dezember 2009

Acknowledgments

I would like to take the opportunity to thank all people who have helped and accompanied me during my diploma thesis.

First of all I would like to thank Prof. Dr. Claudia Klüppelberg for creating a comfortable working atmosphere at the Chair of Mathematical Statistics.

I would like to thank my supervisor Dr. Robert Stelzer for all his helpful advice. This thesis would never have existed without his support and encouragement. We had a lot of fruitful and interesting discussions. To him, I offer my most sincere gratitude.

Furthermore, I want to thank all my friends not only for some really influential discussions about this thesis but also for making the past five years of study as enjoyable and varied as they were. In particular, I want to name Oliver and Christian. Thanks a lot to all.

Last but by no mean least, I would like to thank my parents and my brother for their encouragement, support and love during the last five years.

Contents

Chapter 1

Introduction

The Ornstein-Uhlenbeck process, which is named after the physicists Leonard Ornstein $(1900-1988)$ and George Eugene Uhlenbeck $(1880-1941)$, is a stochastic process $(X_t)_{t\geq 0}$ given by the stochastic differential equation

$$dX_t = -aX_t dt + \sigma dL_t$$

where $a > 0$ is the mean reversion parameter, $\sigma > 0$ the variance parameter and $(L_t)_{t\geq 0}$ denotes the background driving Lévy process.

The Ornstein-Uhlenbeck process is the most widely used mean reverting stochastic process in financial mathematics, mostly to model interest rates, commodity prices and currency exchange rates. The name mean reverting comes from the fact that the OU-process is, in contrast to the Wiener process, dependent on the current value of the process. If the current value of the process is less than the mean zero, the drift $-aX_t$ will be positive. If the current process value is greater than zero, the drift is negative.

In particular, positive OU-processes were used for stochastic volatility modeling by Barndorff-Nielson and Shephard, see for instance Barndoff-Nielsen and Shephard (2001a) or Barndoff-Nielsen and Shephard (2001b). This stochastic volatility model has the form

$$\begin{aligned} S_t = S_0 \exp X_t dY_t &= \left(\mu + \beta\sigma_t^2\right), dt + \sigma_t dW_t + \rho dL_t \\ d\sigma_t^2 &= -\lambda\sigma_t^2 dt + dL_t, \end{aligned}$$

where $(S_t)_{t\geq 0}$ is the price of an asset, $(Y_t)_{t\geq 0}$ the corresponding log-return, $\rho \leq 0$ and $\lambda > 0$. Let $(W_t)_{t\geq 0}$ be a standard Brownian motion and $(L_t)_{t\geq 0}$ be the background driving Lévy process which is assumed to be a subordinator without drift.

In Benth et al. (2007) a model for the electricity spot price dynamics is introduced. In this paper they propose to model the spot price dynamics directly by an non-Gaussian Ornstein-Uhlenbeck process. Then electricity forward and futures prices can be calculated based on the proposed spot price dynamics.

This diploma thesis is concerned with the estimation of positive Lévy driven Ornstein-Uhlenbeck type processes like in the model by Barndorff-Nielson and Shephard. In this thesis we assume that the Lévy process is a subordinator without drift component. This work discusses both parametric and nonparametric inference. Although it is not possible to observe e.g. the volatility and so the Ornstein-Uhlenbeck process in the Barndorff-Nielson and Shephard model cannot be estimated, nevertheless it is a step towards estimating this models.

In the first part of the thesis we estimate the parameters of an Ornstein-Uhlenbeck process and its background driving Lévy process. For the estimation of the mean reversion parameter the highly efficient Davis-McCormick estimator is used. In the second one, by using Lévy Khintchine representation, we rewrite the characteristic function of Ornstein-Uhlenbeck type processes with a so-called canonical function which we estimate with the Support Reduction Algorithm. Note that the Support Reduction Algorithm can also be used for a parametric estimation. For this the parameters of a specified distribution are estimated by this algorithm, see Jongbloed and Van Der Meulen (2006).

The fundamental difference between this two estimation methods is not only the way of estimation. One time we estimate the background driving Lévy process, the other time we estimate the stationary distribution of the Ornstein-Uhlenbeck process.

The remainder of this thesis is organized as follows. In Chapter 2 we discuss some properties of Lévy processes. Furthermore, we show the relationship between self decomposability and the representation of the characteristic function. We outline some definitions and theorems as well, which we need to show the consistency of the estimates. In Chapter 3 the parametric estimation method is presented. First we define the Davis-McCormick estimator for the mean reversion coefficient a. Next the consistency of the estimator of a is shown by using point processes like in Davis and McCormick (1989). Then we estimate σ and the distribution of the background driving Lévy process. In Chapter 4 we give some examples for the parametric inference. First a compound Poisson process driven Ornstein-Uhlenbeck process, second a gamma driven Ornstein-Uhlenbeck process. In Chapter 5 we present the nonparametric inference. Therefore, we write the characteristic function of the Ornstein-Uhlenbeck process in terms of a canonical function. Then the cumulant M-estimator can be defined as the projection of a preliminary estimate onto the class of cumulant functions of self-decomposable distributions, relative to a weighted L_2-distance. Then we show the consistency of this estimator. A possible preliminary estimator for the cumulant function is given. In Chapter 6 we introduce the Support Reduction Algorithm and show the efficiency of this algorithm for different examples. We also show a way to determine the density of the stationary distribution of the Ornstein-Uhlenbeck process. The show the nonparametric estimation for two different types of data. One time we estimate the canonical and density function using data from the Ornstein-Uhlenbeck process, the other we estimate from i.i.d. data. The estimation from i.i.d. data shows, that the nonparametric estimation method is not limited to Ornstein Uhlenbeck processes but can also used for positive self-decomposable distributions

Chapter 2

Preliminaries

In this chapter we will briefly recall some important facts about Ornstein-Uhlenbeck processes, Lévy processes and some fundamental information about point processes as well. We will omit the proofs completely and refer to the well-known standard literature like Sato (1999), Protter (2004) or Resnick (2007).

In this thesis we assume that $(\Omega, \mathcal{F}, P, (\mathcal{F}_t)_{t\geq 0})$ is a filtered probability space.

2.1 Lévy-driven Ornstein-Uhlenbeck Type Process

Definition 2.1. *(Lévy driven OU process) An Ornstein-Uhlenbeck (OU) process driven by a Lévy process $(L_t)_{t\geq 0}$ with parameters $a \in \mathbb{R}$ and $\sigma > 0$ is defined to be the solution to the stochastic differential equation*

$$dX_t = -aX_t dt + \sigma dL_t. \tag{2.1}$$

Continuous time autoregressive process $CAR(1)$ is another expression for the OU-process.

Provided that the process is strictly stationary and exists, the solution of the stochastic differential equation is given by

$$X_t = e^{-a(t-s)}X_s + \sigma \int_s^t e^{-a(t-u)} dL_u, \tag{2.2}$$

since with partial integration

$$\begin{aligned} e^{at}X_t &= e^{as}X_s + \int_s^t ae^{au}X_u du + \int_0^t e^{au} dX_u \\ &= e^{as}X_s + \int_s^t ae^{au}X_u du + \int_s^t e^{au}(-aX_u du + \sigma dL_u) \\ &= e^{as}X_s + \int_s^t ae^{au}X_u du + \int_s^t -e^{au}aX_u du + \sigma \int_0^t e^{au} L_u du \\ &= e^{as}X_s + \sigma \int_s^t e^{au} dL_u. \end{aligned}$$

2.2 Lévy process

Let us define some definitions and results for Lévy processes.

Definition 2.2. *A random variable X is said to be self-decomposable if for every $c \in (0,1)$ there exists a random variable X_c independent of X, such that $X \stackrel{d}{=} cX + X_c$.*

Definition 2.3. *(Applebaum, 2004) Let X be a random variable taking values in $\mathbb{R}$ with law μ_X. Then X is said to be infinitely divisible if for all $n \in \mathbb{N}$ there exist i.i.d. random variables $X_1^{(n)}, X_2^{(n)}, \ldots, X_n^{(n)}$ such that*

$$X \stackrel{d}{=} X_1^{(n)} + X_2^{(n)} + \ldots + X_n^{(n)}.$$

Definition 2.4. *(Sato, 1999, cf. Definition 24.16.) A measure ρ on $\mathbb{R}$ is degenerate if there are $a \in \mathbb{R}$ and a proper linear subspace V of $\mathbb{R}$ (that is a linear subspace with $\dim V \leq d-1$) such that $S_\rho \subset a + V$, where S_ρ is the support of ρ. Otherwise, ρ is called non-degenerate.*
Let X_t be a Lévy process on $\mathbb{R}$. X_t is said to be degenerate if P_{X_t} is degenerate for every t. Otherwise X_t is non-degenerate.

Remark 2.5. In particular one can show that all degenerate random variables in $\mathbb{R}$ are self-decomposable.

Definition 2.6. *The characteristic function $\psi_\mu(z)$ of a probability measure μ on $\mathbb{R}$ is defined by*

$$\psi_\mu(z) = \int_{\mathbb{R}} e^{izx} \mu(dx).$$

The characteristic function of the distribution P_X of a random variable X on $\mathbb{R}$ is denoted by $\psi_X(z)$, i.e.

$$\psi_X(z) = E\left(e^{izX}\right) = \int_{\mathbb{R}} e^{izx} P_X(dx).$$

The class of self-decomposable distributions is a subclass of the infinitely divisible distributions. Since every Lévy process is infinitely divisible, the Lévy Khintchine representation can be used to obtain the characteristic function.

Theorem 2.7. ***Lévy Khintchine representation*** *(Sato, 1999, Theorem 8.1.)*
(i) If μ is an infinitely divisible distribution on $\mathbb{R}$, then

$$\psi_\mu(z) = \exp\left(-\frac{1}{2}Az^2 + i\gamma z + \int_{\mathbb{R}} (e^{izx} - 1 - izx1_D(x))\rho(dx)\right), \tag{2.3}$$

$z \in \mathbb{R}$, where $A \geq 0$ and $D = \{x : |x| \leq 1\}$, ρ is a measure on $\mathbb{R}$ satisfying

$$\rho(\{0\}) = 0 \tag{2.4}$$

$$\int_{\mathbb{R}} (|x|^2 \wedge 1)\rho(dx) < \infty \tag{2.5}$$

and $\gamma \in \mathbb{R}$.
(ii) The representation of $\psi(z)$ in (i) by A, ρ, and γ is unique.
(iii) Conversely, if $A \geq 0$, ρ is a measure satisfying (2.4) and (2.5) and $\gamma \in \mathbb{R}$, then there exists an infinitely divisible distribution μ whose characteristic function is given by (2.3).

Definition 2.8. *(Sato, 1999, Definition 8.2.) We call (A, ρ, γ) in Theorem 2.7 the generating triplet of μ. A and ρ are called, respectively, the Gaussian covariance matrix and the Lévy measure of μ. When $A = 0$, μ is called purely non-Gaussian.*

Remark 2.9. (Sato, 1999, cf. Remark 8.4.) The integrand of the integral in the right hand side of (2.3) has to be integrable with respect to ρ. Let $c(x)$ be a bounded measurable function from $\mathbb{R}$ to $\mathbb{R}$ satisfying

$$\begin{aligned} c(x) &= 1 + o(|x|) \text{ as } |x| \to 0 \\ c(x) &= O(1/|x|) \text{ as } |x| \to \infty. \end{aligned}$$

Then (2.3) can be rewritten as

$$\psi(z) = \exp\left(-\frac{1}{2}Az^2 + i\gamma_c z + \int_{\mathbb{R}} \left(e^{izx} - 1 - izc(x)\right)\rho(dx)\right) \tag{2.6}$$

with γ_c defined as

$$\gamma_c = \gamma + \int_{\mathbb{R}} x\left(c(x) - 1_D(x)\right)\rho(dx). \tag{2.7}$$

The notation of the triplet in (2.6) is $(A, \rho, \gamma_c)_c$. It is also called generating triplet and (2.6) is also called the Lévy-Khintchine representation. If the triplet is used without subscript c, then the representation from (2.3) is taken.
More generally, if $c(x)$ is a measurable function and if, for every z, $e^{izx} - 1 - izxc(x)$ is integrable with respect to a given Lévy measure ρ, then we have (2.6) with (2.7). The triplet is again called a generating triplet, written as $(A, \rho, \gamma_c)_c$. Hence, if ρ satisfies the additional condition

$$\int_{|x|\le 1} |x|\rho(dx) < \infty, \tag{2.8}$$

we can use the zero function as truncation function c and get

$$\psi(z) = \exp\left(-\frac{1}{2}Az^2 + i\gamma_0 z + \int_{\mathbb{R}} \left(e^{izx} - 1\right)\rho(dx)\right) \tag{2.9}$$

with $\gamma_0 \in \mathbb{R}$. The triplet is denoted by $(A, \rho, \gamma_0)_0$. The constant γ_0 is called drift of μ.

Theorem 2.10. ***Subordinator*** *(Sato, 1999, Theorem 24.11.) Let L_t be a Lévy process on $\mathbb{R}$. Then the following four conditions are equivalent to each other:*

- *L_t is a subordinator*
- *$S(L_t) \subset [0, \infty)$ for every $t > 0$*
- *$S(L_t) \subset [0, \infty)$ for some $t > 0$*
- *$A = 0$, $S_\rho \subset [0, \infty)$, $\int_0^1 x\rho(dx) < \infty$ and $\gamma_0 \ge 0$,*

where $S(L_t)$ is the support of L_t. If L_t is a subordinator, then

$$E(e^{-uL_t}) = \exp\left(t\left(\int_0^\infty (e^{-ux} - 1)\rho(dx) - \gamma_0 z\right)\right),$$

for $u \ge 0$.

For the characteristic function of a Lévy process it follows with Theorem 2.7

$$\psi_{L_1}(z) = \exp\left(-\frac{1}{2}Az^2 + i\gamma z + \int_{\mathbb{R}} (e^{izx} - 1 - izx1_D(x))\rho(dx)\right).$$

Then, since we have assumed an increasing Lévy process, which is another expression for a subordinator, we get by applying Theorem 2.10

$$\begin{aligned} A &= 0 \\ \int_0^1 x\rho(dx) &< \infty, \end{aligned}$$

where the support of ρ is positive. Then it follows with Remark 2.9 and due to the fact that L is a Lévy process without drift component

$$\psi_{L_1}(z) = \exp\left(\int_0^\infty (e^{izx} - 1)\rho(dx)\right). \tag{2.10}$$

Summing up, the Lévy measure ρ does fulfill the following conditions

$$\rho(\{0\}) = 0 \tag{2.11}$$

$$\int_0^\infty (x \wedge 1)\rho(dx) < \infty. \tag{2.12}$$

Lemma 2.11. *(Sato, 1999, Lemma 17.1.)*
Let L be a Lévy process on $\mathbb{R}$ generated by (G, ρ, β). Let $a \in \mathbb{R}$. Let X be a temporally homogeneous Markov process with

$$E\left(e^{izX_z}|X_0 = x\right) = \int_{\mathbb{R}} e^{izy} P_t(x, dy) = \exp\left(ie^{-at}xz + \int_0^t g(e^{-as}z)ds\right) \tag{2.13}$$

for $z \in \mathbb{R}$, where $g(z) = \log \psi_{L_1}(z)$ is the cumulant function of L_1. For each t and x, the transition kernel $P_t(x, \cdot)$ is infinitely divisible with generating triplet $(A_t, \nu_t, \gamma_{t,x})$ given by

$$\begin{aligned} A_t &= G\int_0^t e^{-2as}ds \\ \nu_t(B) &= \int_{\mathbb{R}}\int_0^t 1_B(e^{-as}y)ds\ \rho(dy),\ B \in \mathcal{B}(\mathbb{R}), \\ \gamma_{t,x} &= e^{-at}x + \beta\int_0^t e^{-as}ds + \int_{\mathbb{R}}\int_0^t e^{-as}y(1_D(e^{-as}y) - 1_D(y))ds\ \rho(dy), \end{aligned}$$

where $D = \{x : |x| < 1\}$. X is called the process of Ornstein-Uhlenbeck type generated by (G, ρ, β, a).

A probability measure μ in $\mathbb{R}$ is the limit distribution of a temporally homogeneous Markov process on $\mathbb{R}$ with a transition function $P_t(x, B)$ if

$$P_t(x, \cdot) \to \mu$$

as $t \to \infty$ for any $x \in \mathbb{R}$.

Theorem 2.12. *(Sato, 1999, Theorem 17.5.)*
Fix $a > 0$.
(i) If ρ satisfies

$$\int_{|x|>2} \log(|x|)\rho(dx) < \infty, \tag{2.14}$$

the process of Ornstein-Uhlenbeck type on $\mathbb{R}$ generated by (G, ρ, β, a) has a limit distribution μ with

$$\psi_\mu(z) = \exp\left(\int_0^\infty g(e^{-as}z)ds\right). \tag{2.15}$$

The distribution μ is self-decomposable and the generating triplet (A, ν, γ) of μ is given by

$$\begin{aligned} A &= \frac{1}{2a}G, \\ \nu(B) &= \frac{1}{a}\int_{\mathbb{R}}\int_0^\infty 1_B(e^{-s}y)ds\ \rho(dy), B \in \mathcal{B}(\mathbb{R}), \\ \gamma &= \frac{1}{a}\beta + \frac{1}{a}\int_{|y|>1} \frac{y}{|y|}\rho(dy). \end{aligned}$$

(ii) For any self-decomposable distribution μ on $\mathbb{R}$, there exists a unique triplet (G, ρ, β) satisfying (2.14) such that μ is the limit distribution of the process of the Ornstein-Uhlenbeck type generated by (G, ρ, β, a).

A probability measure μ on $\mathbb{R}$ is an stationary distribution of the temporally homogeneous Markov process on $\mathbb{R}$ with transition function $P_t(x, B)$ if

$$\int_{\mathbb{R}} \mu(dx)P_t(x, B) = \mu(B)$$

for $t > 0$ and $B \in \mathcal{B}(\mathbb{R})$.

Theorem 2.13. *(Sato, 1999, Theorem 17.9.) A process of Ornstein-Uhlenbeck type process satisfying (2.14) has a unique stationary distribution which is given by 2.15.*

As shown before $(0, \rho, 0)$ is the generating triplet of the subordinator L and the Lévy measure ρ satisfies conditions $(2.11) - (2.12)$. Thus, if condition (2.14) holds, the limit distribution of the OU-process with generating triplet $(0, \rho, 0, a)$ has characteristic function

$$\psi_{X_1}(z) = \exp\left(\int_0^\infty g(e^{-as}z)ds\right), \tag{2.16}$$

where g is the cumulant function with $g(z) = \log \psi_{L_1}(z) = \int_0^\infty (e^{izx} - 1)\rho(dx)$.

Then the characteristic function can be expressed by

$$\psi_{X_1}(z) = \exp\left(\int_0^\infty \left(e^{izx} - 1\right)\nu(dx)\right). \tag{2.17}$$

Theorem 2.14. *(Sato, 1999, Theorem 15.11.) A probability measure μ on $\mathbb{R}$ is self-decomposable if and only if*

$$\psi(t) = \exp\left(-\frac{1}{2}At^2 + i\gamma t + \int_{-\infty}^{\infty}(e^{itx} - 1 - itx1_D(x))\frac{k(x)}{x}dx\right), \tag{2.18}$$

where $A \geq 0$, $D = \{x : |x| \leq 1\}$ and $k(x) \geq 0$, $\int_{-\infty}^{\infty}(|x|^2 \wedge 1)\frac{k(x)}{x}dx < \infty$ and $k(x)$ is increasing on $(-\infty, 0)$ and decreasing on $(0, \infty)$.

Since X is self-decomposable, the measure ν has a special representation. It has a density with respect to Lebesgue measure and with Theorem 2.14 we have

$$\nu(dx) = \frac{k(x)}{x}dx,$$

where k is a decreasing function on $(0, \infty)$, known as the canonical function which is assumed to be right-continuous. Thus, each non-degenerate, positive, self-decomposable random variable is characterized by a decreasing function k on $(0, \infty)$.

The Ornstein-Uhlenbeck process driven by an increasing Lévy process without drift component is characterized by the following characteristic function

$$\psi_{X_1}(z) = \exp\left(\int_0^{\infty}\frac{(e^{izx} - 1)}{x}k(x)dx\right). \tag{2.19}$$

2.3 Regular Variation

Definition 2.15. *(Resnick, 1987, cf. p.13) A measurable function $F : \mathbb{R}_+ \to \mathbb{R}_+$ is regularly varying at zero with index α (written $F \in RV_\alpha$) if for $x > 0$*

$$\lim_{t \searrow 0}\frac{F(tx)}{F(t)} = x^\alpha. \tag{2.20}$$

α is called the exponent of variation.

Note that regular variation at zero of F is equivalent to regular variation at ∞ (with exponent $-\alpha$) of the function $F(\frac{1}{x})$, since

$$\lim_{t\to\infty}\frac{F\left(\frac{1}{tx}\right)}{F\left(\frac{1}{t}\right)} = \lim_{t\searrow 0}\frac{F\left(\frac{t}{x}\right)}{F\left(t\right)} = \left(\frac{1}{x}\right)^\alpha = x^{-\alpha} \tag{2.21}$$

for all $x > 0$.

Lemma 2.16. *(Davis and McCormick, 1989, Lemma 2.1.) Let Z and Y be two positive, independent random variables with F and G denoting the corresponding distribution functions. Assume F is regularly varying at zero with exponent α and $EY^\beta < \infty$ for some $\beta > \alpha$. Then*

$$\lim_{x\searrow 0}\frac{P\left(ZY^{-1} \leq x\right)}{P\left(Z \leq x\right)} = EY^\alpha.$$

Proof. The left hand side can be rewritten as

$$\lim_{x\searrow 0}\frac{P\left(ZY^{-1}\leq x\right)}{P\left(Z\leq x\right)}=\lim_{x\to\infty}\frac{P\left(ZY^{-1}\leq \frac{1}{x}\right)}{P\left(Z\leq \frac{1}{x}\right)}=(*)\,.$$

Note that the distribution of Z^{-1} is regular varying at infinity with exponent $-\alpha$. It follows by dominated convergence

$$(*)=\lim_{x\to\infty}\int_0^\infty\frac{P\left(Z^{-1}\geq x/y\right)}{P\left(Z^{-1}\geq x\right)}dG(y)=\int_0^\infty\left(\frac{1}{y}\right)^{-\alpha}dG(y)=EY^\alpha,$$

cf. Proposition 3 (Breiman, 1965). □

2.4 Point Process

Let E be locally compact Hausdorff space with a countable basis, i.e. every $x\in E$ has a compact neighborhood, and there exists a sequence of open sets $\left(G_n\right)_{n\geq 1}$ such that any open G can be written as $G=\cup_{\alpha\in I}G_\alpha$ for a countable index set I. Furthermore, let $\mathcal{E}$ be the Borel σ-algebra over E, cf. Resnick (1987).

For $x\in E$ define the measure ϵ_x on $\mathcal{E}$ by

$$\epsilon_x\left(A\right)=\begin{cases}1 & x\in A\\ 0 & x\notin A\end{cases}\qquad\text{for } A\in\mathcal{E}.$$

Definition 2.17. *(Resnick, 1987, p. 123) A point measure on E is a measure ξ of the following form: Let $(x_i)_{i\geq 1}$ be a countable collection of (not necessarily distinct) points of E. Then*

$$\xi:=\sum_{k=1}^{\infty}\epsilon_{x_i},\tag{2.22}$$

and if $K\in\mathcal{E}$ is compact, then $\xi(K)<\infty$ (i.e. ξ is Radon, meaning the measure of compact sets is always finite).

Let $M_+\left(E\right)$ be the class of $\mathbb{N}_0$-valued Radon measures on E and $C_c^+\left(E\right)$ the collection of continuous functions $E\to[0,\infty)$ with compact support. $\mathcal{M}_+\left(E\right)$ is the corresponding σ-algebra which is the smallest σ-algebra of subsets of $M_+\left(E\right)$ making the maps $\mu\to\mu(f)=\int_E fd\mu$ from $M_+\left(E\right)\to\mathbb{R}$ measurable for all $f\in C_c^+\left(E\right)$.

Definition 2.18. *(Bauer, 1992) The vague topology is the coarsest topology such that the functions*

$$M_+(E)\to\mathbb{R},\quad \mu\longmapsto\int_X fd\mu$$

are continuous for every function $f\in C_c^+(E)$.

$M_p(E)$ is the set of all Radon point measures of the form (2.22) and $\mathcal{M}_p\left(E\right)$ the σ-algebra generated by the vague topology.

Definition 2.19. *(Resnick, 1987, p. 124) A point process on E is a measurable map*

$$\xi : (\Omega, \mathcal{A}, P) \to (M_p(E), \mathcal{M}_p(E)),$$

i.e. a point process is a random element of $M_p(E)$. The probability law, denoted by P_ξ of the point process ξ, is the measure $P \circ \xi^{-1} = P(\xi \in \cdot)$ on $\mathcal{M}_p(E)$.

Definition 2.20. *(Resnick, 2007, Definition 5.1.) Let ξ be a point process. ξ is a Poisson process on $(E, \mathcal{E})$ with mean measure μ or synonymously a Poisson random measure with mean μ or $PRM(\mu)$ for short if*

1. *For $A \in \mathcal{E}$*
$$P(\xi(A) = k) = \begin{cases} \frac{\exp\{-\mu(A)\}(\mu(A))^k}{k!} & \text{if } \mu(A) < \infty \\ 0 & \text{if } \mu(A) = \infty \end{cases}$$

2. *If $A_1, \ldots, A_k$ are disjoint subsets of E in $\mathcal{E}$, then $\xi(A_1), \ldots, \xi(A_k)$ are independent random variables.*

Remark 2.21. (Resnick, 2007, p. 49) If $\mu_n \in M_+(E)$ for $n \geq 0$, then μ_n converges vaguely, i.e. converges in the vague topology, to μ_0, written $\mu_n \xrightarrow{v} \mu_0$, if for all $f \in C_c^+(E)$, we have

$$\mu_n(f) := \int_E f(x)\mu_n(dx) \longrightarrow \mu_0(f) := \int_E f(x)\mu_0(dx) \tag{2.23}$$

as $n \to \infty$.

Proposition 2.22. *(Resnick, 1987, Proposition 3.17.) The vague topology on M_+ is metrizable as a complete, separable metric space.*

Remark 2.23. (Resnick, 2007, p. 51) There exist some sequence of functions $f_i \in C_c^+(E)$ such that for $\xi_1, \xi_2 \in M_+(E)$ the vague metric is given by

$$\rho(\xi_1, \xi_2) = \sum_{i=1}^{\infty} \frac{|\xi_1(f_i) - \xi_2(f_i)| \wedge 1}{2^i},$$

where $\xi(f) = \int f d\xi$.

For the following theorem we need a condition D* which we now define.

Condition D* (Davis and Resnick, 1988, p. 47) For each $N \geq 0$, let $(W_{N,i})_{i\geq 1}$ be a stationary sequence of random elements of E. In order to define the mixing condition, let $T > 0$ be fixed and let $\mathcal{C}$ be a finite collection of functions

$$\mathcal{C} = \{h_0, h_1, \ldots, h_m\},$$

where $h_0 \equiv 0$, $h_i \in C_c^+(E)$, $i = 1, \ldots, m$. The array $(W_{N,j})_{N\geq 1, j\geq 1}$ is said to satisfy condition D* if for any two disjoint intervals of integers I_1 and I_2, which are contained in $1, 2, \ldots, \lfloor nT \rfloor$ and separated by l,i.e. $\max\{|i_1 - i_2| : i_1 \in I_1, i_2 \in I_2\} \geq l$, we have

$$\left| E \prod_{j=1}^{2} \prod_{i \in I_j} g_i(W_{N,i}) - \prod_{j=1}^{2} E \prod_{i \in I_j} g_i(W_{N,i}) \right| \leq \alpha_{N,l},$$

where $1 - g_i \in \mathcal{C}$ and $\alpha_{N,l(N)} \to 0$ as $N \to \infty$ for some subsequence $l(N) \to \infty$ with $l(N) = o(N)$. The function $\alpha_{N,l(N)}$ may depend on both $\mathcal{C}$ and T.

Theorem 2.24. *(Davis and Resnick, 1988, Theorem 2.1.)*
Suppose for each $N \geq 1$, $(W_{N,i})_{i\geq 1}$ is a stationary sequence of random elements of E and that the array $\{W_{N,i}, i \geq 1, N \geq 1\}$ satisfies condition D^*. Further assume that there exists a Radon measure ν on E such that

$$NP\left(W_{N,1} \in \cdot\right) \xrightarrow{v} \nu,$$

where $\xrightarrow{v}$ denotes vague convergence and for any $g \in C_c^+(E), g \leq 1$,

$$\lim_{k\to\infty} \limsup_{N\to\infty} N \sum_{i=2}^{\lfloor N/k \rfloor} E\left(g\left(W_{N,1}\right) g\left(W_{N,i}\right)\right) = 0.$$

Then in $M_p([0,\infty) \times E)$,

$$\sum_{k=1}^{\infty} \epsilon_{\left(kN^{-1}, W_{N,k}\right)} \Rightarrow \sum_{k=1}^{\infty} \epsilon_{(t_k, j_k)},$$

where the limit is $PRM\,(dt \times d\nu)$ and $\Rightarrow$ denotes convergence in distribution.

Theorem 2.25. *(Billingsley, 1968, Theorem 4.2.) Let $M_+(E)$ be equipped with the vague metric ρ. Assume that, for each q, $\xi_{N,q} \Rightarrow \xi_q$ as $N \to \infty$ and that $\xi_q \Rightarrow \xi$ as $q \to \infty$. Suppose further that*

$$\lim_{q\to\infty} \limsup_{N\to\infty} P\left(\rho\left(\xi_{N,q}, \xi_N\right) \geq \epsilon\right) = 0 \tag{2.24}$$

for each positive ϵ. Then $\xi_N \Rightarrow \xi$ as $N \to \infty$.

Proposition 2.26. *(Resnick, 1987, Proposition 3.18.) Suppose that E, E' are two spaces which are locally compact with countable bases.*
Suppose $T : E \to E'$ is continuous and satisfies

$$T^{-1}(K') \text{ is compact in } E \text{ for every compact } K' \text{ in } E'. \tag{2.25}$$

Then $\widehat{T} : M_+(E) \to M_+(E')$ defined by

$$\widehat{T}\mu = \mu \circ T^{-1}$$

is continuous.
Note that $\widehat{T}$ restricted to $M_p(E)$ is of the form

$$\widehat{T}\left(\sum \epsilon_{x_i}\right) = \sum \epsilon_{Tx_i}$$

so that a continuous function on the points which also satisfies (2.25) induces a continuous function on the point measures.

Theorem 2.27. ***Continuous Mapping Theorem*** *(Resnick, 1987, p. 152)*
Let (S_i, d_i), $i = 1, 2$ be two metric spaces and suppose $X_n, n \geq 0$, are random elements of $(S_1, \mathcal{S}_1)$ and $X_n \Rightarrow X$ as $n \to \infty$. If $h : S_1 \to S_2$ satisfies

$$P\left(X \in D_h\right) = 0,$$

where $D_h = \{s_1 \in S_1 :\ h \text{ is discontinuous at } s_1\}$, then

$$h(X_n) \Rightarrow h(X)$$

in S_2.

2.5 Probability Theory on Metric Spaces

Let $(S, \mathcal{S})$ be a metric space.

Definition 2.28. *(Billingsley, 1968) Let Π be a family of probability measures on $(S, \mathcal{S})$. Π is said to be relatively compact if every sequence in Π contains a weakly convergent subsequence, that is if, for every $\mu_n \in \Pi$, there exists a subsequence (n_k) and a proper probability measure μ_0 such that $\mu_{n_k} \Rightarrow \mu_0$.*

Definition 2.29. *(Billingsley, 1968) A family Π of probability measures on $(S, \mathcal{S})$ is said to be tight if for all $\epsilon > 0$, there exists a compact set K such that*

$$\mu(K) > 1 - \epsilon$$

for all $\mu \in \Pi$.

Definition 2.30. *(Klenke, 2008, Definition 15.20.) Let (S, d) be a metric space. A family $(F_i)_{i \in I}$ of maps $S \to \mathbb{R}$ is called uniformly equicontinuous if, for every $\epsilon > 0$, there exists a $\delta > 0$ such that $|F_i(t) - F_i(s)| < \epsilon$ for all $i \in I$ and all $t, s \in S$ with $d(s, t) < \delta$.*

Theorem 2.31. *(Klenke, 2008, Theorem 15.21.) Let $\mathcal{M}_1(\mathbb{R})$ be the set of probability measures on $\mathbb{R}$. If $\mathcal{F} \subset \mathcal{M}_1(\mathbb{R})$ is a tight family, then $\{\psi_F : F \in \mathcal{F}\}$ is uniformly equicontinuous. In particular, every characteristic function is uniformly continuous.*

Lemma 2.32. *(Klenke, 2008, Lemma 15.22.)*
Let (S, d) be a metric space and $F_0, F_1, F_2, \ldots$ be maps $S \to \mathbb{R}$ with $F_N \to F_0$ pointwise as $N \to \infty$. If $(F_N)_{N \in \mathbb{N}}$ is uniformly equicontinuous, then F_0 is uniformly continuous and $(F_N)_{N \in \mathbb{N}}$ converges to F_0 uniformly on compact sets, i.e. for every compact set $K \subset S$, we have

$$\sup_{s \in K} |F_N(s) - F_i(s)| \to 0$$

as $N \to \infty$.

Theorem 2.33. ***Prokhorov's Theorem*** *(Billingsley, 1968) Let $(S, \mathcal{S})$ be a separable and complete metric space. The family Π of probability measures is relatively compact if and only if Π is tight.*

Chapter 3

Parametric Inference

In this chapter we present a method for estimating the parameters of an Ornstein-Uhlenbeck process. For this, we use the highly efficient Davis-McCormick estimator for the Ornstein-Uhlenbeck parameter a. We will show the convergence of this estimator. Afterwards we estimate the variance of the OU-process as well as the increments of the background driving Lévy process. This estimation method is shown in Brockwell et al. (2007).

First we need some assumptions on the Ornstein-Uhlenbeck process to present the two different estimation methods.

Assumption 1. We assume that the parameter $a > 0$ and that the background driving Lévy process (BDLP) L is a subordinator, i.e. an increasing Lévy process, without drift component. The Lévy measure ρ from the Lévy process L satisfies

$$\int_2^\infty \log x \rho(dx) < \infty.$$

Another assumption is that $(L_t)_{t\geq 0}$ has a.s. right continuous sample paths and existing left hand limits.

3.1 Parameter Estimation based on the Discretely Sampled Process

Setting $t = nh$ and $s = (n-1)h$ in (2.2) for any $h > 0$, the sampled process $(X_n^{(h)})_{n\geq 0}$ is the discrete time $AR(1)$ process satisfying

$$X_n^{(h)} = \phi X_{n-1}^{(h)} + Z_n, \tag{3.1}$$

where $0 < \phi < 1$ with

$$\phi = e^{-ah}, \tag{3.2}$$

and positive i.i.d. distributed Z_n given by

$$Z_n = \sigma \int_{(n-1)h}^{nh} e^{-a(nh-u)} dL_u. \tag{3.3}$$

In addition to Assumption 1 we need some more assumptions for this estimation method.

Assumption 2. We assume that the distribution function F of Z_n is regularly varying at zero with exponent α and $F(0) = 0$. Moreover, we assume that the moment condition

$$\int_{\mathbb{R}} x^\beta F(dx) < \infty \tag{3.4}$$

holds for some $\beta > \alpha$.

Since $1-\phi z \neq 0$ for all $|z| \leq 1$ and $\int_2^\infty \log x F(dx) < \infty$, the assumptions of Theorem 1 in Brockwell and Lindner (2009) are satisfied and the strictly stationary solution of (3.1) is given by

$$X_n = \sum_{j=0}^{\infty} \phi^j Z_{n-j}.$$

3.2 Estimator for the Mean Reversion Parameter

Let the process $(X_t)_{0\leq t\leq T}$ be observed at times $0, h, 2h, \ldots, Nh$, where $N = \lfloor T/h \rfloor$. Then, since $0 < \phi < 1$ and Z_n are nonnegative, an intuitive estimator for ϕ is the Davis-McCormick estimator, given by

$$\widehat{\phi}_N^{(h)} = \min_{1\leq n\leq N} \frac{X_n}{X_{n-1}}. \tag{3.5}$$

To simplify notation we write $\widehat{\phi}_N$ instead of $\widehat{\phi}_N^{(h)}$, but keep in mind that the Davis-McCormick estimator depends on N and h.

The convergence of the estimator for ϕ can be shown with some results from point processes as done by Davis and McCormick (1989). In the following we present an extended version of that proof.

We show that ${k_N}^{-1}(\widehat{\phi}_N^{(h)} - \phi)c_\alpha$ converges in distribution with

$$k_N = F^{\leftarrow}(N^{-1}) = \inf\left\{x : F(x) \geq N^{-1}\right\} \tag{3.6}$$

and

$$c_\alpha = (EX_1^\alpha)^{1/\alpha} \tag{3.7}$$

by rewriting $P\left({k_N}^{-1}(\widehat{\phi}_N^{(h)} - \phi)c_\alpha \leq x\right)$ in terms of a point process and by showing the convergence in distribution of these point processes. Hence, it follows by probability theory

$$\widehat{\phi}_N^{(h)} \to \phi, \quad \text{a.s.}$$

as $N \to \infty$.

Lemma 3.1. *(Davis and McCormick, 1989, Theorem 2.2.) Let $(X_n)_{n\geq 0}$ be a stationary $AR(1)$ process where F satisfies (2.20) and (3.4). Let ξ_N and ξ be point processes on the space $E = [0,\infty) \times (0,\infty]$ defined by*

$$\xi_N = \sum_{n=1}^{N} \epsilon_{(k_N{}^{-1} Z_n, X_{n-1})}$$

and

$$\xi = \sum_{n=1}^{\infty} \epsilon_{(j_k, Y_k)},$$

where $\sum_{k=1}^{\infty} \epsilon_{j_k}$ is $PRM\,(\alpha x^{\alpha-1}dx)$ and $(Y_k)_{k\geq 0}$ is an i.i.d. sequence of random variables which are independent of $\sum_{k=1}^{\infty} \epsilon_{j_k}$ with $Y_1 \overset{d}{=} X_1$. So, ξ is $PRM\,(\alpha x^{\alpha-1}dx \times G(dy))$, where $G(y) = P(X_1 \leq y)$. Then in $M_p(E)$,

$$\xi_N \Rightarrow \xi.$$

Proof. Since F is regular varying at zero with exponent α and $k_N = \inf\left\{x : F(x) \geq \frac{1}{N}\right\}$, we have

$$F\left(k_N\right) = F\left(F^{\leftarrow}\left(\frac{1}{N}\right)\right) \sim \frac{1}{N}$$

as $N \to \infty$. So

$$\lim_{N\to\infty} N \cdot F(k_N x) = \lim_{N\to\infty} \frac{F(k_N x)}{\frac{1}{N}} == \lim_{N\to\infty} \frac{F(k_N x)}{F(k_N)} = x^{\alpha}. \tag{3.8}$$

for $x > 0$, since $k_N \to 0$ for $N \to \infty$. For $q > 1$ let $X_{n,q}$ be the moving average process

$$X_{n,q} = \sum_{j=0}^{q} \phi^j Z_{n-j},$$

and set

$$\xi_{N,q} = \sum_{n=1}^{N} \epsilon_{(\alpha_N^{-1} Z_n, X_{n-1,q})}.$$

First we show that $\xi_{N,q} \Rightarrow \xi_q$. This can be shown using Theorem 2.24. For this, we have to prove two conditions. By defining $W_{N,i} := (k_N^{-1} Z_i, X_{i-1,q})$, we have to show

$$NP(W_{N,1} \in \cdot) \xrightarrow{v} \nu,$$

where ν is a Radon measure. For any $(c_1, d_1] \times (c_2, d_2] \in E$ we have, since Z_n and $X_{n-1,q}$ are independent and (3.8),

$$\begin{aligned}
&NP(W_{N,1} \in (c_1, d_1] \times (c_2, d_2]) \\
&\quad = NP(k_N^{-1} Z_1 \in (c_1, d_1]\,, X_{0,q} \in (c_2, d_2]) \\
&\quad = NP(k_N^{-1} Z_1 \in (c_1, d_1]) P(X_{0,q} \in (c_2, d_2]) \\
&\quad = N\left(P\left(a_N^{-1} Z_1 \leq d_1\right) - P\left(a_N^{-1} Z_1 \leq c_1\right)\right) P\left(X_{0,q} \in (c_2, d_2]\right) \\
&\quad \longrightarrow \left(d_1^{\alpha} - c_1^{\alpha}\right) P\left(Y_{0,q} \in (c_2, d_2]\right),
\end{aligned}$$

as $N \to \infty$, where $Y_{0,q} := \sum_{j=0}^{q} \phi^j Z_j \left(\stackrel{d}{=} X_{0,q} \right)$. Define $Y_{k,q} := \sum_{j=0}^{q} \phi^j Z_{n-j}$.

Let g be in $C_c^+(E)$ with $g \leq 1$. Since the stationary sequence $\{(Z_n, X_{n-1,q}), n = 0, \pm 1, \ldots\}$ is $(q+1)$ dependent, condition D^* is automatically satisfied, cf. Davis and Resnick (1988, p. 51). For all $(x,y) \in [0,\infty) \times (0,\infty]$ it follows

$$
\begin{aligned}
&\limsup_{N\to\infty} N \sum_{i=2}^{\lfloor N/k \rfloor} E\left(g(W_{N,1})g(W_{N,i})\right) \\
&\leq \limsup_{N\to\infty} N \sum_{i=2}^{\lfloor N/k \rfloor} E\left(\epsilon_{(a_N^{-1}Z_1, X_{0,q})}\left((0,x]\times(0,y]\right) \epsilon_{(a_N^{-1}Z_i, X_{i-1,q})}\left((0,x]\times(0,y]\right)\right) \\
&\leq \limsup_{N\to\infty} N \sum_{j=1}^{\lfloor N/k \rfloor} P\left(k_N^{-1} Z_1 \leq x, X_{0,q} \leq y, k_N^{-1} Z_{j+1} \leq x, X_{j,p} \leq y\right) \\
&\leq \limsup_{N\to\infty} N(N/k)\left(P\left(Z_1 \leq k_N x\right)\right)^2 \\
&= x^{2\alpha}/k \\
&\quad \longrightarrow 0, \qquad (3.9)
\end{aligned}
$$

as $k \to \infty$.
Therefore, from Theorem 2.24,

$$\xi_{N,q} \Rightarrow \xi_q, \qquad (3.10)$$

where $\xi_q = \sum_{k=1}^{\infty} \epsilon_{\left(j_k, Y_{k,q}\right)}$ and $(Y_{k,q})_{k\geq 0, q \geq 0}$ is i.i.d. with $Y_{k,q} \stackrel{d}{=} Y_{1,q}$.

Since ξ_q is PRM with intensity measure $\alpha x^{\alpha-1}dx \times P\left(Y_{1,q} \in dy\right)$ which converges vaguely to the intensity measure of ξ, it follows that

$$\xi_q \Rightarrow \xi, \quad \text{as } q \to \infty.$$

To finish the proof of this lemma, it is left to show the convergence $\xi_N \Rightarrow \xi$. For this we use Theorem 2.25 and the definition of the vague metric. Since

$$\rho(\xi_{N,q}, \xi_N) = \sum_{i=1}^{\infty} \frac{|\xi_{N,q}(f_i) - \xi_N(f_i)| \wedge 1}{2^i},$$

we have to show that

$$P\left(\sum_{i=1}^{\infty} \frac{|\xi_{N,q}(f_i) - \xi_N(f_i)| \wedge 1}{2^i} > \epsilon\right)$$

converges to 0 as $N \to \infty$ and $q \to \infty$. Since $\sum_{i=1}^{\infty} 1/2^i < \infty$ there exists an $i_0 \in \mathbb{N}$ such that

$$\sum_{i=i_0+1}^{\infty} \frac{1}{2^i} < \frac{\epsilon}{2},$$

and we have for the probability

$$\begin{aligned}
P\left(\sum_{i=1}^{\infty}\frac{|\xi_{N,q}(f_i)-\xi_N(f_i)|\wedge 1}{2^i}>\epsilon\right) &\le P\left(\sum_{i=1}^{i_0}\frac{|\xi_{N,q}(f_i)-\xi_N(f_i)|\wedge 1}{2^i}>\frac{\epsilon}{2}\right)\\
&\le \sum_{i=1}^{i_0}P\left(\frac{|\xi_{N,q}(f_i)-\xi_N(f_i)|\wedge 1}{2^i}>\frac{\epsilon}{2}\right)\\
&\le \sum_{i=1}^{i_0}P\left(\frac{|\xi_{N,q}(f_i)-\xi_N(f_i)|}{2^i}>\frac{\epsilon}{2}\right)\\
&= \sum_{i=1}^{i_0}P\left(|\xi_{N,q}(f_i)-\xi_N(f_i)|>\epsilon 2^{i-1}\right).
\end{aligned}$$

If

$$P\left(|\xi_{N,q}(f_i)-\xi_N(f_i)|>\eta\right)\to 0, \tag{3.11}$$

as first $N\to\infty$ and then $q\to\infty$ for every $f_i\in C_c^+(E)$ and for every $\eta>0$, the condition from Theorem 2.25 is satisfied. So, we have to show that for all $\eta>0$ and $f\in C_c^+(E)$, $f\le 1$,

$$\lim_{q\to\infty}\limsup_{N\to\infty}P\left(\left|\sum_{k=1}^{N}f(k_N^{-1}Z_k,X_{k-1})-\sum_{k=1}^{N}f(k_N^{-1}Z_k,X_{k-1,q})\right|\ge\eta\right)=0. \tag{3.12}$$

Suppose that f is supported on the compact set $[0,c]\times[0,\infty]$. Since f is uniformly continuous, there exists for every given $\epsilon>0$ a $\delta>0$ such that

$$|f(x,y)-f(x,z)|<\epsilon,$$

whenever $|y-z|<\delta$.

Then divide the event in (3.12) in two disjoint sets, the set

$$V=\bigcap_{k=1}^{N}\left(\{k_N^{-1}Z_k>c\}\cup\{|X_{k-1}-X_{k-1,q}|<\delta\}\right),$$

and its complement

$$V^c=\bigcup_{k=1}^{N}\left(\{k_N^{-1}Z_k\le c\}\cap\{|X_{k-1}-X_{k-1,q}|\ge\delta\}\right).$$

Consider

$$\left|\sum_{k=1}^{N}f(k_N^{-1}Z_k,X_{k-1})-\sum_{k=1}^{N}f(k_N^{-1}Z_k,X_{k-1,q})\right|<\sum_{k=1}^{N}\epsilon_{k_N^{-1}Z_k}\left([0,c]\right)\cdot\epsilon,$$

if $|X_{k-1} - X_{k-1,q}| < \delta$. Then the probability in (3.12) is bounded by

$$\begin{aligned}
& P\left(\left|\sum_{k=1}^{N} f(k_N^{-1} Z_k, X_{k-1}) - \sum_{k=1}^{N} f(k_N^{-1} Z_k, X_{k-1,q})\right| \geq \eta, V^c\right) \\
+ \quad & P\left(\left|\sum_{k=1}^{N} f(k_N^{-1} Z_k, X_{k-1}) - \sum_{k=1}^{N} f(k_N^{-1} Z_k, X_{k-1,q})\right| \geq \eta, V\right) \\
\leq \quad & P\left(\bigcup_{k=1}^{N} \left\{k_N^{-1} Z_k \leq c, |X_{k-1} - X_{k-1,q}| \geq \delta\right\}\right) + P\left(\sum_{k=1}^{N} \epsilon_{k_N^{-1} Z_k}([0,c]) > \eta/\epsilon\right) \\
\leq \quad & NP\left(k_N^{-1} Z_1 \leq c\right) P\left(|X_0 - X_{0,q}| \geq \delta\right) + P\left(\sum_{k=1}^{\infty} \epsilon_{k_N^{-1} Z_k}([0,c]) > \eta/\epsilon\right) \\
\longrightarrow \quad & c^{\alpha} P\left(|X_0 - X_{0,q}| \geq \delta\right) + P\left(\xi\left([0,c] \times [0,\infty]\right) > \eta/\epsilon\right),
\end{aligned}$$

as $N \to \infty$. By choosing $\epsilon > 0$ small and then q large, this bound can be made arbitrarily small.
So, with Theorem 2.25

$$\xi_N \Rightarrow \xi.$$

□

Theorem 3.2. *(Davis and McCormick, 1989, Theorem 2.3.) Let $(X_n)_{n\geq 0}$ be the stationary AR(1) process (3.1) where F is regularly varying at zero with exponent α and satisfies the moment condition $\int x^{\beta} F(dx) < \infty$ for some $\beta > \alpha$. Define the point processes η_N and η on $[0,\infty)$ by*

$$\eta_N = \sum_{n=1}^{n} \epsilon_{a_n^{-1}\left(\frac{X_n}{X_{n-1}} - \phi\right)}$$

and

$$\eta = \sum_{k=1}^{\infty} \epsilon_{j_k / Y_k}$$

with $\sum_{k=1}^{\infty} \epsilon_{j_k}$ and $(Y_k)_{k\geq 0}$ as defined in Lemma 3.1. In particular, η is $PRM\left(EX_1^{\alpha}(\alpha x^{\alpha-1} dx)\right)$. Then in $M_p\left([0,\infty)\right)$, the class of nonnegative integer valued Radon measures on E,

$$\eta_N \Rightarrow \eta.$$

Proof. Let $M > 0$ be a continuity point of the distribution of X_1 and consider the map $T : [0, M] \times (0, \infty] \to [0, \infty)$ given by

$$T(x, y) = \frac{x}{y}.$$

Since T is continuous and $T^{-1}(K)$ is compact for K compact in $[0, M] \times (0, \infty]$, the assumptions of Theorem 2.26 are fulfilled and there exists a continuous map $\widehat{T} : M_+([0, M] \times (0, \infty]) \to M_+([0, \infty))$, where

$$\widehat{T}\left(\sum \epsilon_{(x,y)}\right) = \sum \epsilon_{T(x,y)}.$$

Then we have

$$\begin{aligned}\sum_{n=1}^{N}\epsilon_{k_N^{-1}Z_n/X_{n-1}}\mathbf{1}_{\left(k_N^{-1}Z_n\le M\right)} &= \sum_{n=1}^{N}\epsilon_{T(k_N^{-1}Z_n,X_{n-1})}\mathbf{1}_{\left(k_N^{-1}Z_n\le M\right)}\\ &= \widehat{T}\left(\sum_{n=1}^{N}\epsilon_{(k_N{}^{-1}Z_n,X_{n-1})}\right)\mathbf{1}_{\left(k_N^{-1}Z_n\le M\right)}\\ &= \widehat{T}\left(\xi_N\right).\end{aligned}$$

Since $\widehat{T}$ is continuous and $\xi_n \Rightarrow \xi$ by Lemma 3.1, it follows from the Continuous Mapping Theorem

$$\widehat{T}\left(\xi_N\right) \Rightarrow \widehat{T}\left(\xi\right).$$

Hence,

$$\widehat{T}\left(\xi\right) = \widehat{T}\left(\sum_{k=1}^{\infty}\epsilon_{(j_k,Y_k)}\right)\mathbf{1}_{(j_k\le M)} = \sum_{k=1}^{\infty}\epsilon_{T(j_k,Y_k)}\mathbf{1}_{(j_k\le M)} = \sum_{k=1}^{\infty}\epsilon_{(j_k/Y_k)}\mathbf{1}_{(j_k\le M)}.$$

So, we showed that in $M_p\left([0,\infty)\right)$

$$\sum_{n=1}^{N}\epsilon_{k_N^{-1}Z_n/X_{n-1}}\mathbf{1}_{\left(k_N^{-1}Z_n\le M\right)} \Rightarrow \sum_{k=1}^{\infty}\epsilon_{(j_k/Y_k)}\mathbf{1}_{(j_k\le M)}.$$

As $M\to\infty$,

$$\sum_{k=1}^{\infty}\epsilon_{(j_k,Y_k)}\mathbf{1}_{(j_k\le M)} \to \sum_{k=1}^{\infty}\epsilon_{(j_k,Y_k)} \qquad \text{a.s..}$$

Similarly to the proof of Lemma 3.1 we show by using Theorem 2.25

$$\eta_N \Rightarrow \eta.$$

Then, instead of showing (3.11) we apply the Markov inequality and prove that for all $f \in C_c^+\left([0,\infty)\right)$, $f\le 1$,

$$\lim_{M\to\infty}\limsup_{N\to\infty} E\left(\left|\int f d\eta_N - \sum_{n=1}^{N} f\left(k_N^{-1}Z_n/X_{n-1}\right)\mathbf{1}_{\left(k_N^{-1}Z_n\le M\right)}\right|\right) = 0.$$

Consider

$$
\begin{aligned}
& E\left(\left|\int f d\eta_N - \sum_{n=1}^{N} f\left(k_N^{-1} Z_n / X_{n-1}\right) \mathbf{1}_{\left(k_N^{-1} Z_n \leq M\right)}\right|\right) \\
& = E\left(\left|\sum_{n=1}^{N} f\left(k_N^{-1} Z_n / X_{n-1}\right) - \sum_{n=1}^{N} f\left(k_N^{-1} Z_n / X_{n-1}\right) \mathbf{1}_{\left(k_N^{-1} Z_n \leq M\right)}\right|\right) \\
& = E\left(\left|\sum_{n=1}^{N} f\left(k_N^{-1} Z_n / X_{n-1}\right) \mathbf{1}_{\left(k_N^{-1} Z_n > M\right)}\right|\right) \\
& = N\, E\left(f\left(k_N^{-1} Z_1 / X_0\right) \mathbf{1}_{\left(k_N^{-1} Z_1 > M\right)}\right) \\
& \leq N\, E\left(\mathbf{1}_{\left(k_N^{-1} Z_1 / X_0 < c\right)} \mathbf{1}_{\left(k_N^{-1} Z_1 > M\right)}\right) \\
& \leq N\, P\left(k_N^{-1} Z_1 / X_0 < c, k_N^{-1} Z_1 > M\right),
\end{aligned}
$$

where we have assumed that the support of f is contained in $[0, c]$. Then the last term can be bounded by

$$
N\, P\left(k_N^{-1} Z_1 < c X_0, k_N^{-1} Z_1 > M\right) \leq N\, P\left(k_N^{-1} Z_1 < c X_0, c X_0 > M\right).
$$

By defining $Y := c X_0 \mathbf{1}_{(c X_0 > M)}$ we can apply Lemma 2.16

$$
\begin{aligned}
\lim_{N \to \infty} N\, P\left(k_N^{-1} Z_1 < Y\right) &= \lim_{N \to \infty} N\, P\left(Z_1 Y^{-1} < k_N\right) \\
&= \lim_{N \to \infty} \frac{N\, P\left(Z_1 Y^{-1} < k_N\right)}{N\, P\left(Z_1 < k_N\right)} \\
&= E Y^{\alpha} \\
&= c^{\alpha} E X_0^{\alpha} \mathbf{1}_{(c X_0 > M)}.
\end{aligned}
$$

Since $E X^{\beta} < \infty$ holds for some $\beta > \alpha$, we have $E X_0^{\alpha} < \infty$ and the limit converges to zero as $M \to \infty$.
Hence,

$$
\eta_N \Rightarrow \eta.
$$

□

Corollary 3.3. *(Davis and McCormick, 1989, Corollary 2.4.)*
With $\widehat{\phi}_N = \min_{1 \leq n \leq N} \frac{X_n}{X_{n-1}}$, *we have*

$$
\lim_{N \to \infty} P\left(k_N^{-1}\left(\widehat{\phi}_N - \phi\right) c_{\alpha} \leq x\right) = 1 - exp(-x^{\alpha}), \qquad x > 0,
$$

where $c_{\alpha} = (E X_1^{\alpha})^{1/\alpha}$ *and* $\widehat{\phi} \to \phi$ *a.s.. In particular,* ϕ_N *is consistent.*

Proof. Since

$$
\begin{aligned}
\eta_N\left([0, x / c_{\alpha}]\right) &= \sum_{n=1}^{N} \epsilon_{a_n^{-1}(X_n / X_{n-1} - \phi)}\left([0, x / c_{\alpha}]\right) \\
& \begin{cases} = 0 & \text{if } k_N^{-1}(\widehat{\phi}_N - \phi) > x / c_{\alpha} \\ > 0 & \text{else} \end{cases},
\end{aligned}
$$

we have using Theorem 3.2

$$\begin{aligned} P\left(k_N^{-1}(\widehat{\phi}_N - \phi)c_\alpha > x\right) &= P\left(\eta_N\left([0, x/c_\alpha]\right) = 0\right) \\ &\longrightarrow P\left(\eta\left([0, x/c_\alpha]\right) = 0\right) \end{aligned}$$

as $N \to \infty$. Since η is $PRM\left(EX_1^\alpha\left(\alpha x^{\alpha-1}dx\right)\right)$ and μ the appropriate intensity measure, it follows

$$\mu\left([0, x/c_\alpha)\right) = \int_0^{x/c_\alpha} EX_1^\alpha \alpha y^{\alpha-1} dy = x^\alpha,$$

where $c_\alpha = (EX_1^\alpha)^{1/\alpha}$. So, by Definition 2.20

$$P\left(\eta\left([0, x/c_\alpha]\right) = 0\right) = \exp\left\{-x^\alpha\right\}.$$

Since $k_N^{-1} \to \infty$ we must have $\widehat{\phi}_N \xrightarrow{P} \phi$ for $N \to \infty$. But this implies $\widehat{\phi}_N \longrightarrow \phi$ a.s. since $\widehat{\phi}_N \geq \phi$ and $\widehat{\phi}_N$ is non increasing. □

In summary,

$$\lim_{N\to\infty} P\left(k_N^{-1}\left(\widehat{\phi}_N - \phi\right)c_\alpha \leq x\right) = G_\alpha(x),$$

where G_α is the Weibull distribution function,

$$G_\alpha(x) = \begin{cases} 1 - \exp(-x^\alpha) & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases}.$$

The estimator $\widehat{\phi}_N^{(h)}$ can be obtained from the observation $\left\{X_n^{(h)}, n = 0, 1, 2, \ldots, N\right\}$ and from relation (3.2), an estimator for the mean reversion coefficient a is given by

$$\widehat{a}_N^{(h)} = -h^{-1} \log \widehat{\phi}_N^{(h)}. \tag{3.13}$$

By defining $g(\phi) := -h^{-1}\log(\phi)$ and by Chapter 6 in Brockwell and Davis (1991) it follows from

$$k_N^{-1}\left(\widehat{\phi}_N - \phi\right)c_\alpha \Longrightarrow V,$$

where V is Weibull distributed with parameter α that

$$k_N^{-1}\left(g(\widehat{\phi}_N) - g(\phi)\right)c_\alpha \Longrightarrow g'(\phi)V.$$

Thus,

$$k_N^{-1}(-h)e^{-ah}\left(\widehat{a}_N - a\right)c_\alpha \Longrightarrow V.$$

So we have for the estimator $\widehat{a}_N$ again a Weibull distribution function and we get

$$\lim_{N\to\infty} P\left(k_N^{-1}(-h)e^{-ah}(\widehat{a}_N - a)c_\alpha \leq x\right) = G_\alpha(x).$$

With the same argumentation like in Corollary 3.3 it follows that $\widehat{a}_N$ is consistent and $\widehat{a}_N \to a$ a.s..

3.3 Estimator for the Variance of the OU-Process

The variance of the AR(1)-process X_n var(X_n) can be obtained if some additional assumptions on L hold.

Let L be a second-order Lévy process, i.e. $E\left({L_1}^2\right) < \infty$.Then there exist real constants μ and σ with $E\left(L_t\right) = \mu t$ and var$L_t = \sigma^2 t$ for $t \geq 0$. In order to avoid confusion with OU-parameter σ the Lévy process L is assumed to be scaled with var$L_t = t$ for all $t \geq 0$ instead of var$L_t = \sigma$ for all $t \geq 0$. Then L is called a standardized second-order Lévy process.

Lemma 3.4. *Let L_t be a standardized second-order Lévy process. Then*

$$\mathrm{var}(X_n) = \frac{\sigma^2}{2a}. \tag{3.14}$$

Proof. Since $X_n = \sum_{j=0}^{\infty} \phi^j Z_{n-j}$ the variance can be obtained in the following way.

$$\mathrm{var}\left(X_n\right) = \mathrm{var}\left(\sum_{j=0}^{\infty} \phi^j Z_{n-j}\right) = \sum_{j=0}^{\infty} \phi^{2j} \mathrm{var}\left(Z_{n-j}\right). \tag{3.15}$$

The second identity is true since $\sum_{j=0}^{\infty} \phi^j < \infty$, $|EZ_1| < \infty$, and $|EZ_1^2| < \infty$, see for instance Brockwell and Davis (1991). The last condition will be proved below. Since we have assumed that Z is i.i.d. distributed, we have

$$\mathrm{var}\left(X_n\right) = \sum_{j=0}^{\infty} \phi^{2j} \mathrm{var}\left(Z_1\right). \tag{3.16}$$

So, by (3.3), Proposition 4.44 inJacod and Shiryaev (2003), and the dominated convergence theorem,

$$\begin{aligned}
\mathrm{var}(Z_1) &= \mathrm{var}\left(\sigma \int_0^h e^{-a(h-u)} dL_u\right) \\
&= \sigma^2 e^{-2ah} \mathrm{var}\left(\int_0^h e^{au} dL_u\right) \\
&= \sigma^2 e^{-2ah} \mathrm{var}\left(\lim_{N\to\infty} \sum_{j=1}^{N} e^{ah\left(\frac{j}{N} - \frac{j-1}{N}\right)} \left(L_{h\frac{j}{N}} - L_{h\frac{j-1}{N}}\right)\right) \\
&= \sigma^2 e^{-2ah} \mathrm{var}\left(\lim_{N\to\infty} \sum_{j=1}^{N} e^{ah\left(\frac{j}{N} - \frac{j-1}{N}\right)} L_{h\left(\frac{j}{N} - \frac{j-1}{N}\right)}\right)
\end{aligned}$$

$$
\begin{aligned}
&= \sigma^2 e^{-2ah} \lim_{N\to\infty} \operatorname{var}\left(\sum_{j=1}^{N} e^{ah\left(\frac{j}{N}-\frac{j-1}{N}\right)} L_{h\left(\frac{j}{N}-\frac{j-1}{N}\right)}\right) \\
&= \sigma^2 e^{-2ah} \lim_{N\to\infty} \sum_{j=1}^{N} e^{2ah\left(\frac{j}{N}-\frac{j-1}{N}\right)} \operatorname{var}\left(L_{h\left(\frac{j}{N}-\frac{j-1}{N}\right)}\right) \\
&= \sigma^2 e^{-2ah} \lim_{N\to\infty} \sum_{j=1}^{N} e^{2ah\left(\frac{j}{N}-\frac{j-1}{N}\right)} h\left(\frac{j}{N}-\frac{j-1}{N}\right) \\
&= \sigma^2 e^{-2ah} \int_0^h e^{2au} du \\
&= \sigma^2 e^{-2ah} \frac{1}{2a}\left(e^{2ah}-1\right) \\
&= \frac{\sigma^2}{2a}\left(1-e^{2ah}\right). \qquad (3.17)
\end{aligned}
$$

Thus, by setting (3.17) in (3.16) and applying (3.2),

$$
\begin{aligned}
\operatorname{var}(X_n) &= \frac{\sigma^2}{2a}\left(1-e^{2ah}\right)\sum_{j=0}^{\infty}\phi^{2j} = \frac{\sigma^2}{2a}\left(1-e^{2ah}\right)\frac{1}{1-\phi^2} \\
&= \frac{\sigma^2}{2a}\left(1-e^{2ah}\right)\frac{1}{1-e^{2ah}} = \frac{\sigma^2}{2a}.
\end{aligned}
$$

□

By defining $\bar{X}_N^{(h)} = \frac{1}{N+1}\sum_{n=0}^{N} X_n^{(h)}$ an estimator for σ^2 is

$$
\widehat{\sigma}_N^2 = \frac{2\widehat{a}_N^{(h)}}{N}\sum_{n=0}^{N}\left(X_n^{(h)}-\bar{X}_N^{(h)}\right)^2. \qquad (3.18)
$$

Remark 3.5. By Theorem 4.3. in Masuda (2004) it follows that the Ornstein-Uhlenbeck process is ergodic and by Theorem 3.1. in Jongbloed et al. (2005) X is β-mixing. Therefore, it follows that the estimator for σ is convergent.

3.4 Estimation of the Lévy Increments

If the $CAR(1)$ process is observed continuously on $[0,T]$, then the integrated form of (2.2) immediately gives

$$
L_t = \sigma^{-1}\left(X_t - X_0 + a\int_0^t X_s ds\right).
$$

The increments of the driving Lévy process on the interval $((n-1)h, nh]$ can be expressed by

$$
\Delta L_n^{(h)} := L_{nh} - L_{(n-1)h} = \sigma^{-1}\left(X_{nh} - X_{(n-1)h} + a\int_{(n-1)h}^{nh} X_u du\right). \qquad (3.19)
$$

There exists different methods of approximating the integral in (3.19). We introduce two numerical integration methods, an trapezoidal approximation and a method using Simpson's rule.

Trapezoidal Approach

First, replacing the integral by a trapezoidal approach and the Ornstein-Uhlenbeck parameters by their estimators, the estimated Lévy increments can be obtained by

$$\Delta\widehat{L}_n^{(h)} = \widehat{\sigma}_N^{-1}\left(X_n^{(h)} - X_{n-1}^{(h)} + \frac{\widehat{a}_N^{(h)} h \left(X_n^{(h)} + X_{n-1}^{(h)}\right)}{2}\right). \tag{3.20}$$

If one intends to compare different methods as we plan to, it is useful to have estimates for one time unit. Therefore, we add these estimates in blocks of length $1/h$ to obtain estimated increments of L in one time unit. Let $p = 1/h$ be the number of estimated increments in one time unit. Then

$$\begin{aligned}
&\Delta\widehat{L}_k^{(1)} \\
&:= \sum_{j=1}^{p} \Delta\widehat{L}_{(k-1)p+j}^{(h)} \\
&= \sum_{j=1}^{p} \widehat{\sigma}^{-1}\left(X_{(k-1)p+j}^{(h)} - X_{(k-1)p+j-1}^{(h)} + \frac{\widehat{a}_N^{(h)} h \left(X_{(k-1)p+j}^{(h)} + X_{(k-1)p+j-1}^{(h)}\right)}{2}\right) \\
&= \widehat{\sigma}^{-1}\left(X_{kp}^{(h)} + X_{(k-1)p}^{(h)} + \frac{\widehat{a}_N^{(h)} h}{2}\left(X_{(k-1)p}^{(h)} + 2X_{(k-1)p+1}^{(h)} + \ldots + 2X_{kp-1}^{(h)} + X_{kp}^{(h)}\right)\right) \\
&= \widehat{\sigma}^{-1}\left(X_{kp}^{(h)} + X_{(k-1)p}^{(h)} + \widehat{a}_N^{(h)} h\left(\frac{1}{2}X_{(k-1)p}^{(h)} + X_{(k-1)p+1}^{(h)} + \ldots + X_{kp-1}^{(h)} + \frac{1}{2}X_{kp}^{(h)}\right)\right).
\end{aligned} \tag{3.21}$$

Simpson's rule

In the other approach we use the Simpson rule. In contrast to the trapezoidal approximation we look at two time intervals. Then the estimator for the Lévy increment is

$$\Delta\widehat{L}_n^{(2h)} = \widehat{\sigma}^{-1}\left(X_{n+1}^{(h)} - X_{n-1}^{(h)} + \frac{\widehat{a}_N^{(h)} h \left(X_{n+1}^{(h)} + 4X_n^{(h)} + X_{n-1}^{(h)}\right)}{6}\right). \tag{3.22}$$

Then, after adding the estimator in blocks of length $1/2h$, we get

$$
\begin{aligned}
&\Delta\widehat{L}_k^{(2h)} \\
&:= \sum_{j=1}^{q} \Delta\widehat{L}^{(h)}_{(k-1)q+j} \\
&= \sum_{j=1}^{q} \widehat{\sigma}^{-1}\Bigg(X^{(h)}_{(k-1)q+j+1} - X^{(h)}_{(k-1)q+j-1} \\
&\qquad + \frac{\widehat{a}_N^{(h)} h \left(X^{(h)}_{(k-1)q+j-1} + 4X^{(h)}_{(k-1)q+j} + X^{(h)}_{(k-1)q+j+1}\right)}{6}\Bigg) \\
&= \widehat{\sigma}^{-1}\Bigg(X^{(h)}_{kq} - X^{(h)}_{kq-q} \\
&\quad + \frac{\widehat{a}_N^{(h)} h \left(X^{(h)}_{kq-q} + 4X^{(h)}_{kq-q+j+1} + 2X^{(h)}_{kq-q+j+2} + \ldots + 2X^{(h)}_{kq-2} + 4X^{(h)}_{kq-1} + X^{(h)}_{kq}\right)}{6}\Bigg),
\end{aligned}
\tag{3.23}
$$

where $q = 1/2h$.

Remark 3.6. Note that not all distribution functions F of Z_n, cf. (3.3), are regularly varying. Since we assumed that F has to fulfill the regular variation condition, the convergence of $\widehat{\phi}$ is no longer clear. For example, if the background driving Lévy process is a compound Poisson process, the distribution is not regular varying. So, we have to show the consistency of the estimators stand-alone, but without a distribution limit theorem. We will show this in Chapter 4. Otherwise, it is enough to verify the regular variation condition. For this case we will also give an example, cf. Example 3.

Chapter 4

Applications and Examples

In this section we want to introduce two different methods to simulate an Ornstein Uhlenbeck process - the Euler method and a direct simulation. Besides, we want to apply different background driving Lévy processes, so Example 1 and Example 2 will present a compound Poisson process and Example 3 will provide a standardized gamma process.

4.1 Compound Poisson Process

Definition 4.1. *Let $(N_t)_{t\in\mathbb{R}_+}$ be a Poisson process with rate $\lambda > 0$ and $Y_1, Y_2, \ldots$ i.i.d-distributed random variables with distribution function Q which is also independent of $(N_t)_{t\in\mathbb{R}_+}$. Define*

$$S_t = \sum_{i=1}^{N_t} Y_i.$$

Then $(S_t)_{t\in\mathbb{R}_+}$ is said to be a compound Poisson process with rate λ and jump size distribution Q.

4.1.1 Consistency

Since a compound Poisson process does not satisfy the regular variation condition (2.20), one cannot use the results from Chapter 3. Thus, the consistency has to be proven for the compound Poisson process stand-alone.

Let $(L_n)_{0\le n\le N}$ be the BDLP observed at times $0, h, 2h, \ldots, Nh$. The Lévy increments for the time interval h are defined as

$$\Delta L_n = L_{nh} - L_{(n-1)h}$$

for $n = 1, \ldots, N$.

Since for a compound Poisson process $P(\Delta L_n = 0) > 0$ and the increments of a Lévy process are independent, it follows that

$$\begin{aligned}
\lim_{N\to\infty} P\left(\exists n \in \{1,\dots,N\} : \Delta L_n = 0\right) &= \lim_{N\to\infty} \left(1 - P\left(\Delta L_n = 0 \ \forall n \in \{1,\dots,N\}\right)\right) \\
&= \lim_{N\to\infty} \left(1 - \big(\underbrace{P\left(\Delta L_1 = 0\right)}_{<1}\big)^N\right) \\
&= 1
\end{aligned}$$

4.1.2 Simulation and Estimation

For the compound Poisson process we want to introduce two methods to simulate the Ornstein-Uhlenbeck process, the Euler method and a direct approach. Let the OU-process X_t be given by

$$dX_t = -aX_t dt + \sigma dL_t, \tag{4.1}$$

where L_t is a compound Poisson process with intensity λ and jump distribution $\exp(\xi)$ and mean reversion parameter $a > 0$. We want to introduce these two simulation methods and illustrate the estimation procedure first using an Euler method for the simulation of the OU-process and second a direct approach.

Euler Schema

The general idea of the Euler method is a central formula. The stochastic differential equation (4.1) is substituted by the following simple recursive formula

$$X_{t_n} = X_{t_{n-1}} - aX_{t_{n-1}}(t_n - t_{n-1}) + \sigma\left(L_{t_n} - L_{t_{n-1}}\right), \tag{4.2}$$

where L_{t_n} is a Lévy process obtained at times $t_0, t_1, t_2, \dots, t_N$. For ease of notation we write $X_{t_n} = X_n, i = 1, \dots, N$ and $h = t_n - t_{n-1}$. Here we apply the Euler approximation for a compound Poisson process, but the process can be replaced by any other Lévy process. $a > 0$ and $\sigma > 0$ are the parameters of the OU-process.

Example 1. The compound Poisson-driven OU-processes defined by

$$dX_t = -2X_t dt + dL_t, \tag{4.3}$$

was simulated at times $0, 0.001, 0.002, \dots, 5000$. We simulated a compound Poisson process with intensity parameter $\lambda = 2$ and with a jump size which is exponential distributed with mean $\frac{1}{2}$. Figure 4.1 shows the compound Poisson process up to $t = 10$ and Figure 4.2 the corresponding Ornstein-Uhlenbeck process. Figure 4.3 displays the OU-process for a longer time horizon.

Estimation of a From the $N = 5000000$ simulated process values $X_0, \dots, X_N$ we obtained the Davis-McCormick estimator

$$\widehat{\phi}_N^{0.001} = \min_{1\le n\le N} \frac{X_n}{X_{n-1}} = 0.9980, \tag{4.4}$$

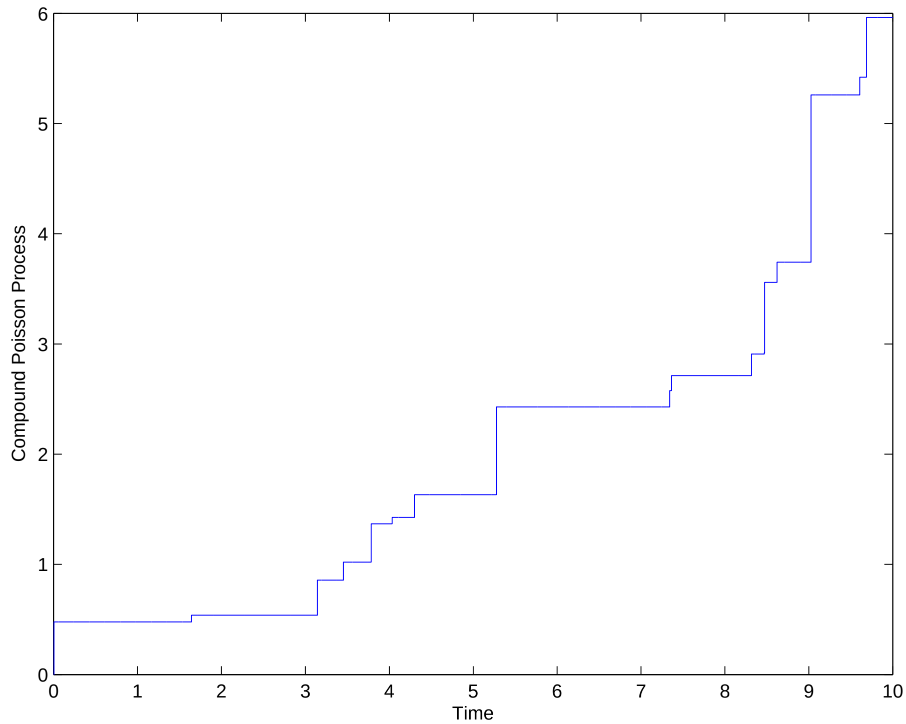

Figure 4.1: Compound Poisson process of intensity 2 with exponential jumps of expectation 1/2

and with

$$\widehat{a}_N^{0.001} = -\frac{1}{h} \log \widehat{\phi} = 2.002 \tag{4.5}$$

we get an estimator for the CAR(1) coefficient a.

In Figure 4.4 we computed the Davis McCormick estimator at different subintervals, $h = 0.01, 0.02, 0.03 \ldots$. There $\widehat{\phi}_{n^*}^h$ or equivalently $\widehat{a}_{n^*}^h$ are obtained as follows. If, for example, $h = 0.01$, we only deal with the sample $X_0, X_{100}, X_{200}, \ldots, X_N$. So we have in total $n^* = 50000$ observation points, denoted by $\overline{X_1}, \overline{X_2}, \ldots, \overline{X_{n^*}}$

$$\overline{\phi}_{n^*}^{(h)} = \min_{1 \le n \le n^*} \frac{\overline{X_n}}{\overline{X_{n-1}}}.$$

Then the OU-estimator is $\overline{a}_{50000}^{(0.01)} = 2.002$. If $n^* = 250$, then 250 values from the times series were used for the estimation and so $h = 20$ or equivalently every 50000th value was used for the estimation for a, $X_0, X_{20000}, X_{40000}, \ldots$. Here $\bar{a}_{250}^{(20)} = 0.5159$.

In general, if n^* in Figure 4.4 is small, then the interval h is large. In this case only a few data points are used for the estimation of a and the estimator $\widehat{a}$ is according to this imprecise. Otherwise, if n^* is large, the grid of the corresponding sample is small. From the structure of the Davis-McCormick estimator we would expect

$$\widehat{\phi} \ge \phi,$$

where ϕ is the true value. Then with (3.2)

$$\widehat{a} \le a.$$

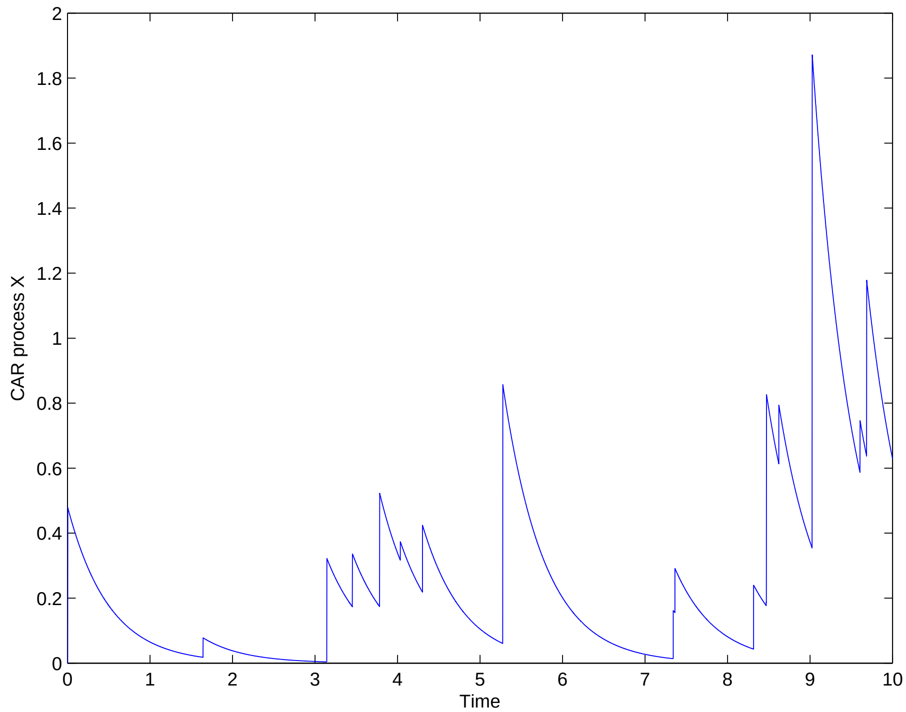

Figure 4.2: Ornstein-Uhlenbeck Process driven by a compound Poisson process (cf. Figure 4.1)

But (4.5) (or Figure 4.4) shows the contrary, since we know the true value of a. This fact is a result from the structure of the Euler Schema. If h is chosen small enough, there exist time periods without a jump in between. Hence, the estimator can be obtained directly. This is the case in our example. The intensity measure is 2, so there are on average two jumps in one unit. Since we have 5000 units with 1000 time points for each time unit, so totally 5000000 observations, there exist many time grids without jumps to compute the estimator exactly. Thus, with the definition of the Davis-McCormick estimator and (4.2) it follows for an interval without any jumps

$$\widehat{\phi} = \frac{X_n}{X_{n-1}} = \frac{X_{n-1} - ahX_{n-1}}{X_{n-1}} = 1 - ah.$$

For the real value ϕ we know

$$\phi = e^{-ah} = \sum_{j=0}^{\infty} \frac{(-ah)^j}{j!} = 1 - ah + \frac{(ah)^2}{2!} - \frac{(ah)^3}{3!} + \dots .$$

Since

$$\sum_{j=2}^{\infty} \frac{(-ah)^j}{j!} = \sum_{j=1}^{\infty} \frac{(ah)^{2j}(1+2j-ah)}{(1+2j)!} \geq 0,$$

it follows

$$1 - ah \leq e^{-ah}.$$

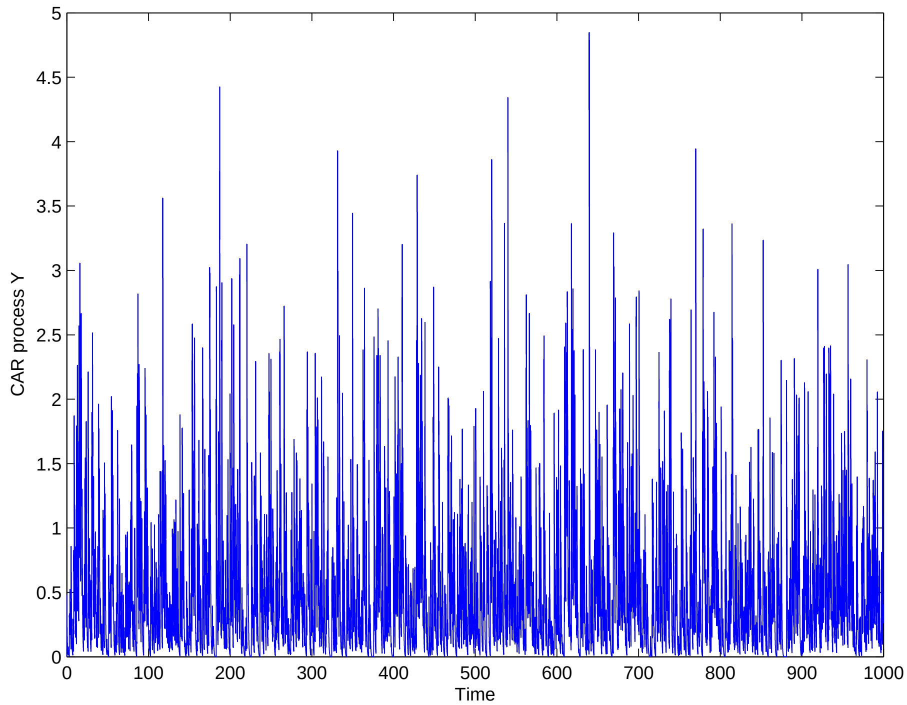

Figure 4.3: Ornstein-Uhlenbeck Process for a longer time horizon

Then

$$\widehat{\phi} \leq \phi$$

and

$$\widehat{a} \geq a.$$

Note that the estimated value $\widehat{a}$ we obtained, cf. (4.5), is exactly the one we would expect from our calculations above

$$\widehat{a} = -\frac{1}{0.001} \log\left(1 - 2 \cdot 0.001\right) = 2.002.$$

In order to get more detailed information of the estimator we concentrate on the samples of the process at six different intervals, $h = 0.01$, $h = 0.1$, $h = 1$, $h = 5$, $h = 10$ and $h = 100$ by selecting every $10th$, $100th$, $1000th$, $5000th$, $10000th$, and $100000th$ value respectively. We generated 100 such realizations of the process and applied the above estimation procedure to generate 100 independent estimates, for each h, of the parameters a and σ. The sample means and standard deviation of these estimators are shown in Table 4.1. Figure 4.5 shows the Weibull probability plot comparing the distributions of the data to the Weibull distribution where we used different grids for the data. The plot induces also a reference line useful for judging whether the data follows a Weibull distribution.

Even though the distribution of a compound Poisson process is not regularly varying, we wanted to check if the distribution of $\widehat{a}$ is similar to a Weibull distribution. Figure 4.5 illustrates, that the distribution of $\widehat{a}$ is no longer as a Weibull distribution recognizable if you look to the tails of the samples with $h \geq 5$.

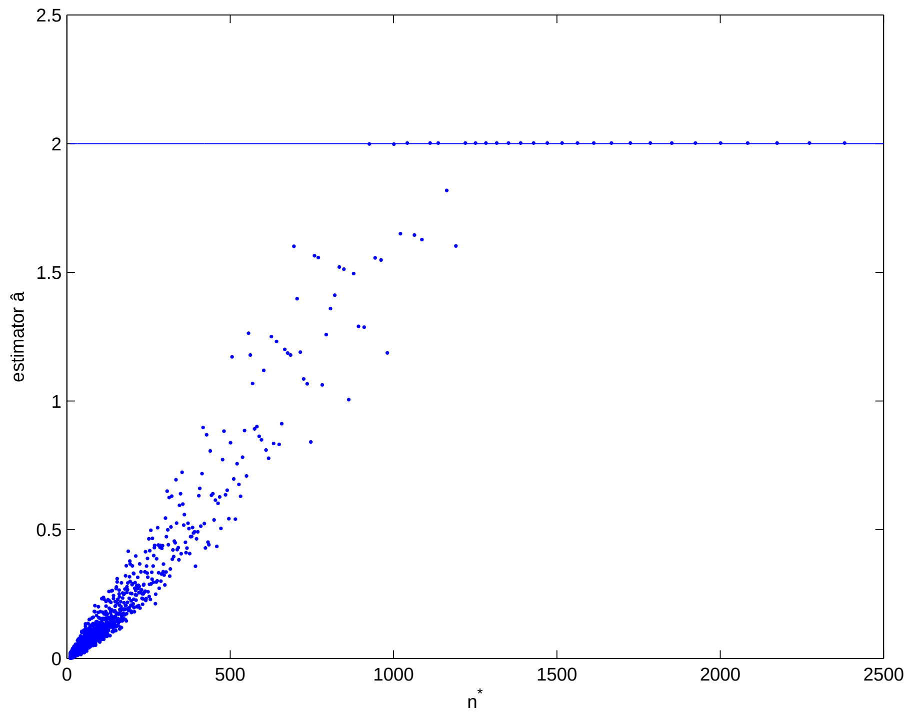

Figure 4.4: Computing the estimator $\widehat{a}$ at different subintervals h, n^* number of observation points used for estimation, $n^* = \frac{5000}{h}$.

Table 4.1: Estimated parameters based on 100 replicates on $[0, 5000]$ of the Compound-Poisson driven OU-process, observed at times nh, $n = 0, \ldots, \lfloor T/h \rfloor$.

h	Parameter	Sample mean of estimator	Sample standard deviation of estimator
0.01	$\widehat{a}$	2.002	0.0
	$\widehat{\sigma}$	1.0033	0.0344
0.1	$\widehat{a}$	2.002	0.0
	$\widehat{\sigma}$	1.0032	0.0343
1	$\widehat{a}$	2.002	0.0
	$\widehat{\sigma}$	0.9907	0.0497
5	$\widehat{a}$	1.1786	0.2432
	$\widehat{\sigma}$	0.7557	0.1110
10	$\widehat{a}$	0.5159	0.1136
	$\widehat{\sigma}$	0.4962	0.0926
100	$\widehat{a}$	0.0293	0.0135
	$\widehat{\sigma}$	0.1076	0.0523

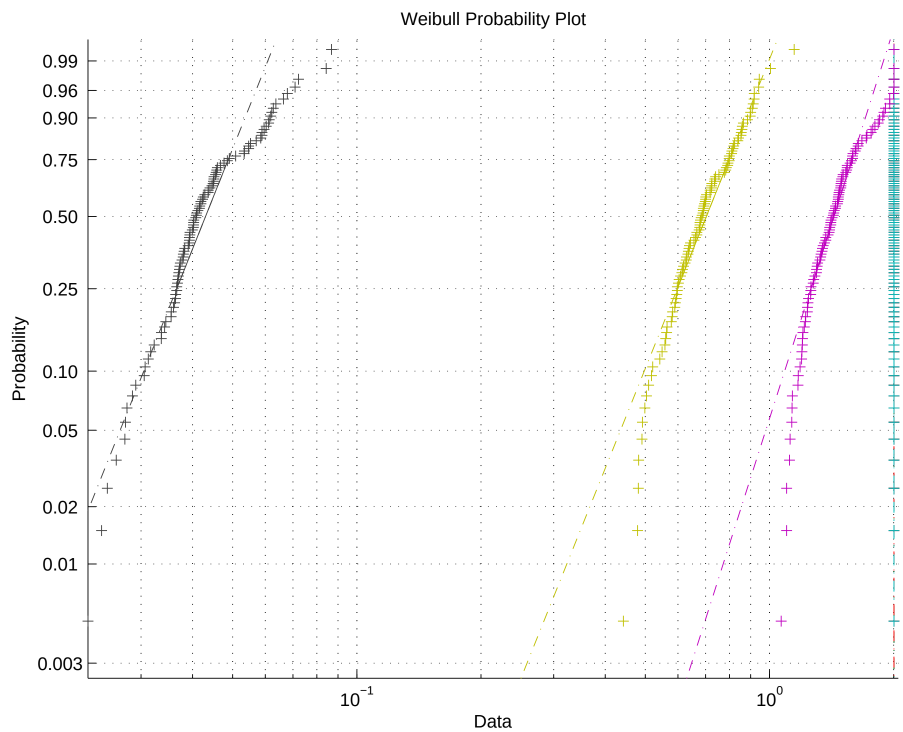

Figure 4.5: Weibull probability plot for three different step sizes, from left to right: $h = 100$, $h = 10$, $h = 5$ and $h = 1$

Exact Simulation

Another possibility to simulate a Compound-Poisson Process driven OU-process is an exact simulation. Using the solution of the stochastic differential equation (2.2)

$$X_t = e^{-a(t-s)}X_s + \sigma \int_s^t e^{-a(t-u)} dL_u,$$

the process is simulated.

If there is no jump before the next time point, we have

$$X_{t_n} = e^{-ah} X_{t_{n-1}}, \tag{4.6}$$

where $h = t_n - t_{n-1}$ for all n, since the time-lags are assumed to be equidistant. If there is a jump within the next time point, more calculations have to be done. First we calculate $X_{jumptime}$, the value of the OU-process just before the jump

$$X_{jumptime} = e^{-a(jumptime - t_{n-1})} X_{t_{n-1}}, \tag{4.7}$$

then

$$X_{jumptime} = X_{jumptime} + Y, \tag{4.8}$$

where Y is the jump size, here in our example exponential distributed with mean $\frac{1}{\xi}$. If there is another jump before the next time point t_n another iteration like (4.7) and (4.8) has to be done. Otherwise the process value for the next observation point is

$$X_{t_n} = e^{-a(t_n - jumptime)} X_{jumptime}. \tag{4.9}$$

In the following we apply the exact method by using the same parameters like in Example 1.

Example 2. We simulate the Ornstein-Uhlenbeck process characterized by

$$dX_t = -2X_t dt + dL_t.$$

at times $0, 0.1, 0.2, \ldots, 5000$. The background driving Lévy process L_t is a compound Poisson process of intensity 2 with exponential jumps of expectation $\frac{1}{2}$. Figure 4.6 and Figure 4.7 show the OU-process simulated with the exact method for a short and long period, respectively.

The Davis-McCormick estimator is

$$\widehat{a}_N^{0.1} = 2. \tag{4.10}$$

So, in comparison to the Euler approximation the Davis-McCormick estimator works for the exact method precisely. The following is the reason for this. The time grid is fine and we have only two jumps on average per time unit. Then there exists two observations X_n and X_{n-1} with no jump in between and the estimator can be obtained directly by using (4.6).

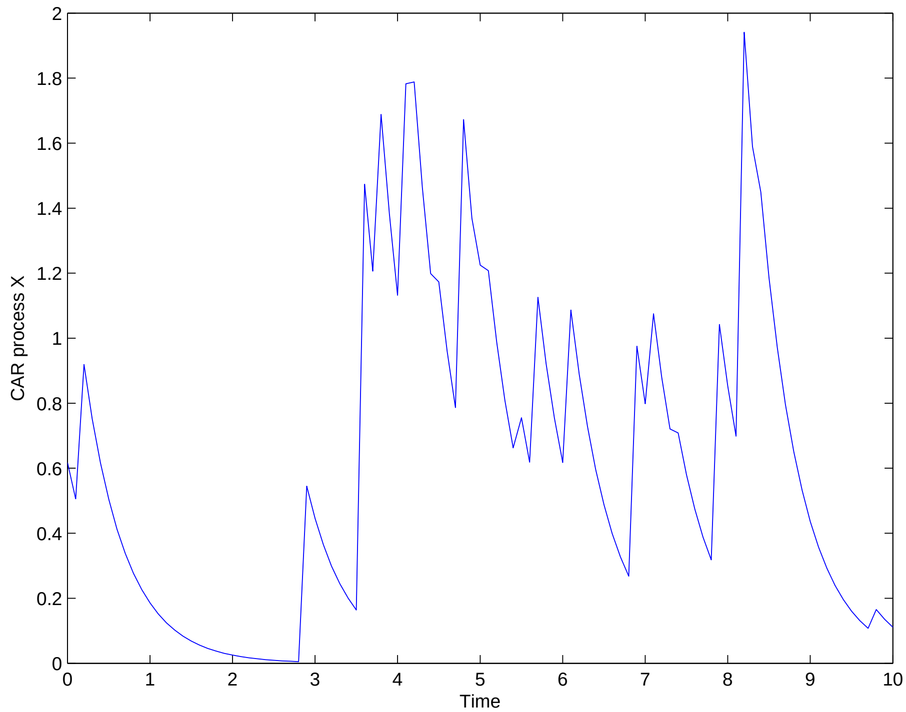

Figure 4.6: Compound Poisson driven Ornstein Uhlenbeck Process with intensity $\lambda = 2$ and exponential jumps of expectation 1/2

In Figure 4.8 the development of the estimator $\widehat{a}$ is pointed out. So already with 1500 data points, i.e. about every 35th process value is used for the calculation of $\widehat{a}$, the estimator is equal to the real value $a = 2$.

We generated 100 replicates of the Ornstein-Uhlenbeck process in order to get the estimators' sample means and standard deviations for a variety of time grids. Table 4.2 illustrates these quantities for $h = 0.1, h = 1, h = 2, h = 5, h = 10$ and $h = 10$.

Figure 4.9 shows the Weibull probability plot comparing the distributions of the data to the Weibull distribution where we used different grids for the data. The plot induces also a reference line useful for judging whether the data follows a Weibull distribution. Even though we do not know the asymptotic distribution of the estimator $\widehat{a}_N$ we wanted to check if the distribution of $\widehat{a}_N$ is similar to a Weibull distribution. Figure 4.9 displays that the distribution of $\widehat{a}_N$ is no longer as a Weibull distribution recognizable if you look to the tails of the samples with $h = 5$ and $h = 10$ compared to the reference line.

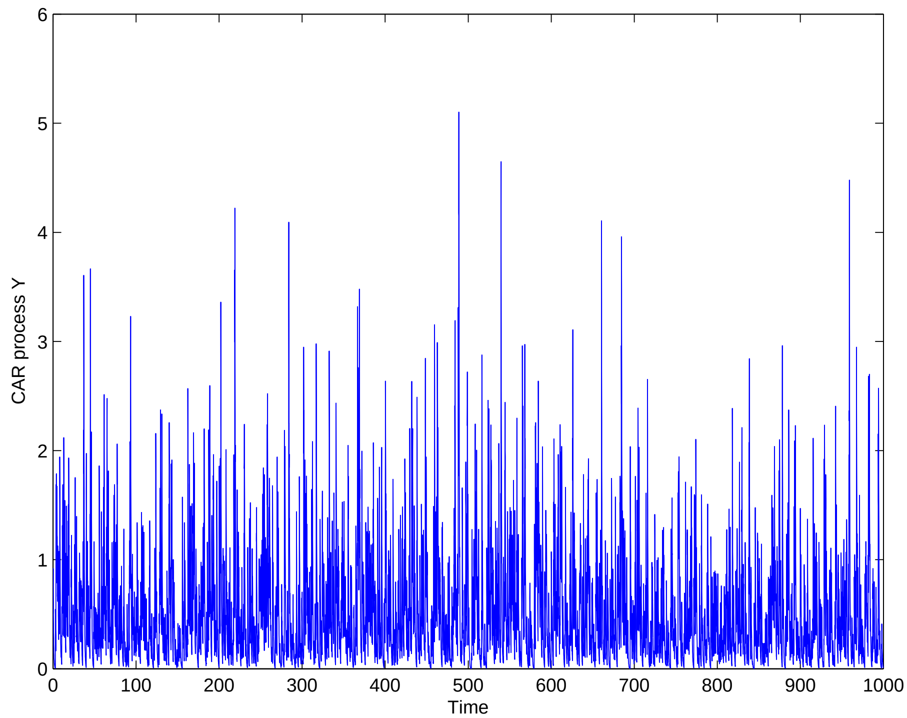

Figure 4.7: Ornstein Uhlenbeck Process for a longer time horizon

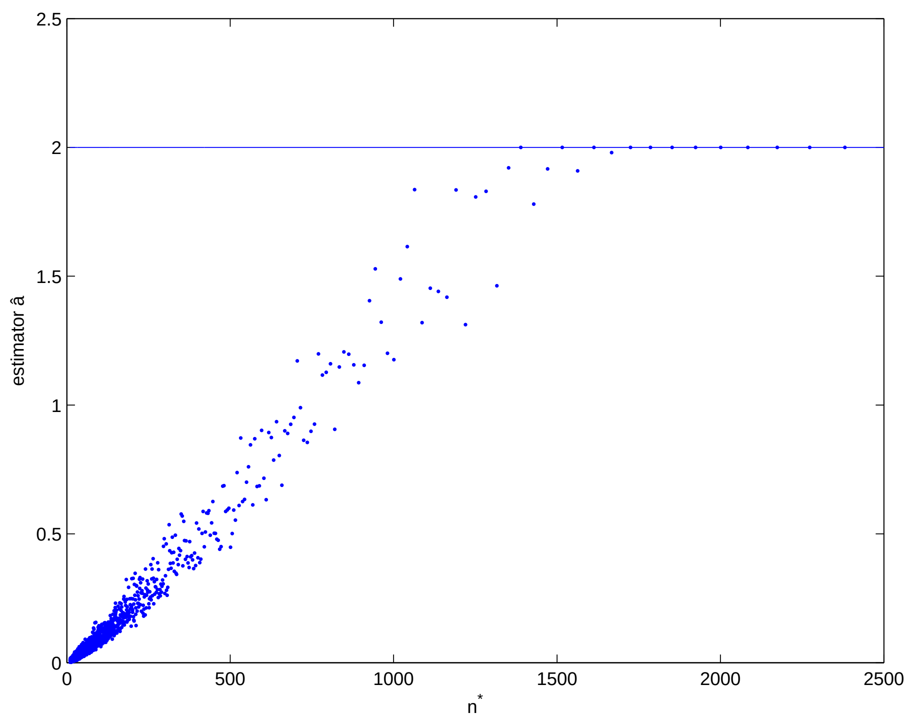

Figure 4.8: Computing the estimator $\widehat{a}$ at different subintervals h, n^* number of observation points used for estimation, $n^* = \frac{5000}{h}$.

Table 4.2: Estimated parameters based on 100 replicates on $[0, 5000]$ of the compound Poisson driven OU-process, observed at times nh, $n = 0, \ldots, \lfloor T/h \rfloor$.

h	Parameter	Sample mean of estimator	Sample standard deviation of estimator
0.1	$\widehat{a}$	2.0	0.0
	$\widehat{\sigma}$	1.0013	0.0145
1	$\widehat{a}$	2.0	0.0
	$\widehat{\sigma}$	1.0039	0.0221
2	$\widehat{a}$	2.0	0.0
	$\widehat{\sigma}$	1.0058	0.0324
5	$\widehat{a}$	1.4485	0.2318
	$\widehat{\sigma}$	0.8590	0.0804
10	$\widehat{a}$	0.6571	0.1217
	$\widehat{\sigma}$	0.5796	0.0637

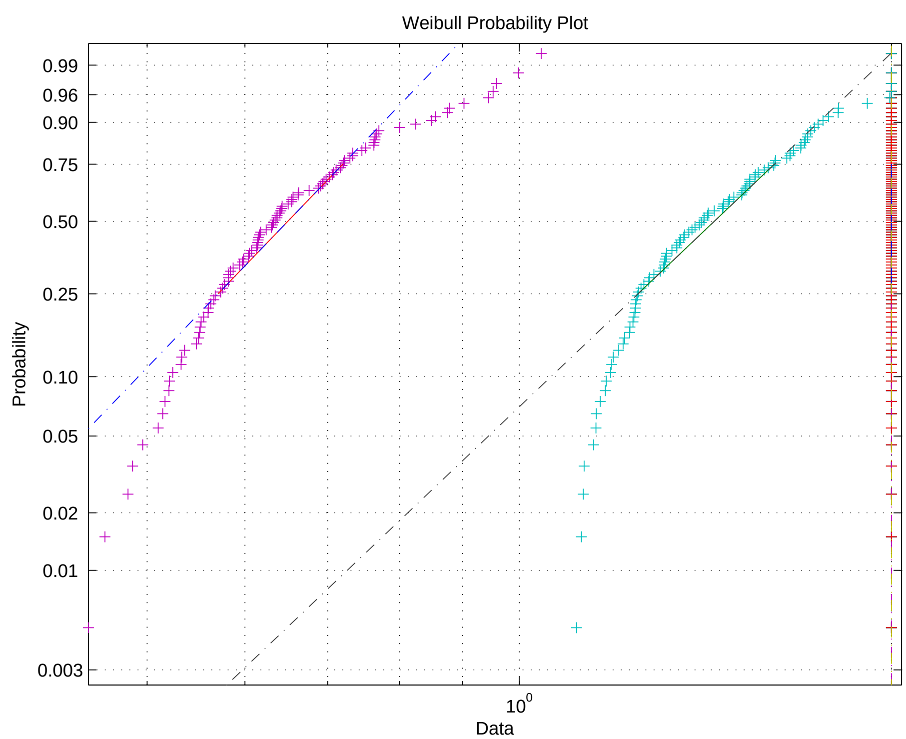

Figure 4.9: Weibull probability plot for three different step sizes, from left to right: $h = 10$, $h = 5$ and $h = 2$

4.2 Gamma Process

In this section we want to illustrate the parameter estimation method for the case in which the background Lévy process is a standardized Gamma process. So the Gamma-driven OU-process X_t is defined by the probability density function

$$dX_t = -aX_t dt + \sigma dG_t,$$

where the background Lévy process G_t is a standardized Gamma process.

First we give some properties of the gamma process. The Gamma distribution with shape parameter α and inverse scale parameter β is defined by

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-x\beta} \mathbf{1}_{[0,\infty)}.$$

The standardized gamma process G_t has the density f_{G_t} given by

$$f_{G_t}(x) = \frac{\gamma^{1/2\gamma t}}{\Gamma(\gamma t)} x^{\gamma t-1} e^{-x\gamma^{1/2}} \mathbf{1}_{[0,\infty)}$$

and the process has mean $\gamma^{1/2}t$ and variance t. Then, since the increments of the Gamma process are gamma distributed, we have increments with

$$G_n - G_{n-1} \sim \text{Gamma}\left(\gamma h, \gamma^{1/2}\right),$$

where $h = t_n - t_{n-1}$.
The Laplace transform of G_t is

$$\varphi_{G_t}(s) := Ee^{-sG_t} = \int_0^\infty e^{-sx} f_{G_t}(x)dx = \frac{\gamma^{1/2}}{(\gamma^{1/2}+s)^{\gamma t}} = \frac{1}{(1+\gamma^{-1/2}s)^{\gamma t}} = e^{-t\Phi(s)} \quad (4.11)$$

for $\Re(s) \geq 0$, where $\Phi(s) = \log(1+\beta s)^\gamma = \gamma \log(1+\beta s)$ with $\beta = \gamma^{-1/2}$ and $\gamma > 0$.

We estimate the discrete-time autoregression coefficient ϕ and the Ornstein-Uhlenbeck process parameters a and σ^2 using (3.13) and (3.18) based on h-spaced observations $(X_n^{(h)})_{0\leq n\leq N}$. Afterwards, the Lévy increments will be estimated in two different ways, one time using a trapezoidal approximation and the other time a Simpson's rule approach. From this estimated parameters we will estimate the parameter γ of the standardized gamma process.

4.2.1 Consistency

To obtain the asymptotic distribution of $\phi_N^{(h)}$ and $a_N^{(h)}$ as N tends to infinity with h fixed we check the regular variation condition of F, the distribution function of Z_n in (3.3), where $Z_n = \sigma \int_{(n-1)h}^{nh} e^{-a(h-t)} dG_t$ first. Then we calculate the coefficients k_N and c_α. We follow here Brockwell et al. (2007, p. 982).

In order to show the regular variation condition we first give an appropriate representation of the Laplace transform of Z_1 by using power series expansion.

By defining $W_h := Z_1/\sigma$ the Laplace transform of W_h is

$$
\begin{aligned}
\varphi_{W_h}(s) &= Ee^{-s\frac{Z_1}{\sigma}} \\
&= Ee^{-\frac{1}{\sigma}\int_0^h se^{-a(h-t)}dG_t} \\
&= \exp\left(-\int_0^h \Phi\left(se^{-at}\right)dt\right) \\
&= \exp\left(-\int_0^h \gamma\log\left(1+\beta se^{-at}\right)dt\right), \qquad (4.12)
\end{aligned}
$$

where Φ is defined as in (4.11). The second equality holds with Cont and Tankov (2004, Theorem 15.1).
Thus,

$$
\begin{aligned}
-\int_0^h \gamma\log\left(1+\beta se^{-at}\right)dt &= -\gamma\int_0^h \log\left(\beta se^{-at}\left(1+\frac{1}{\beta se^{-at}}\right)\right)dt \\
&= -\gamma\int_0^h \left(\log\left(\beta se^{-at}\right)+\log\left(1+\frac{1}{\beta se^{-at}}\right)\right)dt.
\end{aligned}
$$

Next we simplify the terms in the last line as follows

$$
\begin{aligned}
\int_0^h \log\left(\beta se^{-at}\right)dt &= \int_0^h \left(\log(\beta s)+\log\left(e^{-at}\right)\right)dt \\
&= h\log(\beta s)-\frac{ah^2}{2}.
\end{aligned}
$$

Using a Taylor series expansion $\log(1+x)=\sum\limits_{n=0}^{\infty}(-1)^n\frac{x^{n+1}}{n+1}$ for $|x|<1$ and $|x|\neq -1$ and applying monotone convergence we get for the other part

$$
\begin{aligned}
\int_0^h \log\left(1+\frac{1}{\beta se^{-at}}\right)dt &= \int_0^h \sum_{n=0}^{\infty}\frac{(-1)^n}{n+1}\left(\frac{1}{\beta se^{-at}}\right)^{n+1}dt \\
&= \sum_{n=0}^{\infty}\frac{(-1)^n}{n+1}\left(\frac{1}{\beta s}\right)^{n+1}\int_0^h e^{at(n+1)}dt \\
&= \sum_{n=0}^{\infty}\frac{(-1)^{n+1}}{(n+1)^2}\left(\frac{1}{\beta s}\right)^{n+1}\frac{1}{a}\left(1-e^{(n+1)ah}\right).
\end{aligned}
$$

With the previous calculations the exponent in (4.12) can be rewritten as

$$
\begin{aligned}
&-\gamma\int_0^h \log\left(1+\beta se^{-at}\right)dt \\
&\quad = \log(\beta s)^{-\gamma h}+\frac{a\gamma h^2}{2}-\gamma\sum_{n=0}^{\infty}\frac{(-1)^{n+1}}{(n+1)^2}\left(\frac{1}{\beta s}\right)^{n+1}\frac{\left(1-e^{(n+1)ah}\right)}{a} \\
&\quad = \log(\beta s)^{-\gamma h}+\frac{a\gamma h^2}{2}+\frac{\gamma\left(1-e^{ah}\right)}{\beta sa}-\frac{\gamma\left(1-e^{2ah}\right)}{4\beta^2s^2a}+\mathrm{o}\left(\frac{1}{s^2}\right).
\end{aligned}
$$

Applying another Taylor series expansion for the exponential function it follows

$$
\begin{aligned}
\varphi_{W_h}(s) &= \exp\left(\log(\beta s)^{-\gamma h} + \frac{a\gamma h^2}{2} + \frac{\gamma}{\beta s a}\left(1-e^{ah}\right) - \frac{\gamma}{4\beta^2 s^2 a}\left(1-e^{2ah}\right) + \mathrm{o}\left(\frac{1}{s^2}\right)\right) \\
&= (\beta s)^{-\gamma h}\exp\left(\frac{a\gamma h^2}{2}\right)\exp\left(\frac{\gamma}{\beta s a}\left(1-e^{ah}\right) - \frac{\gamma}{4\beta^2 s^2 a}\left(1-e^{2ah}\right) + \mathrm{o}\left(\frac{1}{s^2}\right)\right) \\
&= (\beta s)^{-\gamma h}\exp\left(\frac{a\gamma h^2}{2}\right)\left(1 + \left(\frac{\gamma\left(1-e^{ah}\right)}{\beta s a} + \frac{\gamma\left(1-e^{2ah}\right)}{4\beta^2 s^2 a} + \mathrm{o}\left(\frac{1}{s^2}\right)\right)\right. \\
&\quad \left. + \frac{1}{2}\left(\frac{\gamma(1-e^{ah})}{\beta s a} + \frac{\gamma\left(1-e^{2ah}\right)}{4\beta^2 s^2 a} + \mathrm{o}\left(\frac{1}{s^2}\right)\right)^2 + \mathrm{o}\left(\frac{1}{s^2}\right)\right) \\
&= \frac{\beta^{-\gamma h}}{s^{\gamma h}}\exp\left(\frac{a\gamma h^2}{2}\right) + \frac{C_1}{s^{\gamma h+1}} + \frac{C_2}{s^{\gamma h+2}} + \mathrm{o}\left(\frac{1}{s^{\gamma h+2}}\right),
\end{aligned}
$$

where $C_1, C_2, \ldots$ are constants depending on γ, β, h and a. Since $\varphi_{Z_1}(s) = \varphi_{W_h}(\sigma s)$, it follows that

$$
\begin{aligned}
s^{\gamma h}\varphi_{Z_1}(s) &= s^{\gamma h}\varphi_{W_h}(\sigma s) \\
&= (\sigma\beta)^{-\gamma h}\exp\left(\frac{a\gamma h^2}{2}\right) + \frac{C_1\sigma^{-\gamma h-1}}{s^1} + \frac{C_2\sigma^{-\gamma h-2}}{s^2} + \mathrm{o}\left(\frac{1}{s^2}\right) \\
&\longrightarrow (\sigma\beta)^{-\gamma h}\exp\left(\frac{\gamma a h^2}{2}\right)
\end{aligned}
$$

as $s \to \infty$. By Theorem 30.2 of (Doetsch, 1976) the density of Z_1, f_{Z_1}, has the following expansion in a neighborhood of 0, namely

$$
f_{Z_1}(x) = (\sigma\beta)^{-\gamma h}\exp\left(\frac{a\gamma h^2}{2}\right)\frac{x^{\gamma h-1}}{\Gamma(\gamma h)} + \frac{C_1\sigma^{-\gamma h-1}x^{\gamma h}}{\Gamma(\gamma h+1)} + \frac{C_2\sigma^{-\gamma h-2}x^{\gamma h+1}}{\Gamma(\gamma h+2)} + \ldots.
$$

Then

$$
\begin{aligned}
\frac{f_{Z_1}(x)}{x^{\gamma h-1}} &= (\sigma\beta)^{-\gamma h}\exp\left(\frac{a\gamma h^2}{2}\right)\frac{1}{\Gamma(\gamma h)} + \frac{C_1\sigma^{-\gamma h-1}x}{\Gamma(\gamma h+1)} + \frac{C_2\sigma^{-\gamma h-2}x^2}{\Gamma(\gamma h+2)} + \ldots \\
&\longrightarrow (\sigma\beta)^{-\gamma h}\exp\left(\frac{a\gamma h^2}{2}\right)\frac{1}{\Gamma(\gamma h)},
\end{aligned}
$$

as $x \to 0$. Since

$$
\frac{F_{Z_1}(x)}{x^{\gamma h}} \longrightarrow \frac{(\sigma\beta)^{-\gamma h}\exp\left(\frac{\gamma a h^2}{2}\right)}{\Gamma(\gamma h+1)}
$$

as $x \to 0$ and

$$
F_{Z_1}(x) \sim \frac{x^{\gamma h}(\sigma\beta)^{-\gamma h}\exp\left(\frac{\gamma a h^2}{2}\right)}{\Gamma(\gamma h+1)} \tag{4.13}
$$

as $x \to 0$, it follows that

$$\frac{F_{Z_1}(\lambda x)}{F_{Z_1}(x)} \sim \lambda^{\gamma h}, \tag{4.14}$$

which means F is regularly varying at 0 with exponent γh.

In the following we give a way to calculate k_n^{-1} and c_α. Recall that $k_N = F^{\leftarrow}(N^{-1})$ then

$$\frac{1}{N} = \int_0^{k_N} F_{Z_1}(du). \tag{4.15}$$

Together with (4.13) we have

$$\frac{1}{N} \sim \frac{k_N^{\gamma h} \left(\sigma\beta\right)^{-\gamma h} \exp\left(\frac{\gamma a h^2}{2}\right)}{\Gamma\left(\gamma h + 1\right)}, \tag{4.16}$$

and hence

$$k_N^{-1} \sim \frac{\exp\left(\frac{ah}{2}\right) N^{\frac{1}{\gamma h}}}{\sigma \ \beta \ \Gamma\left(\gamma h + 1\right)^{\frac{1}{\gamma h}}}, \tag{4.17}$$

as $N \to \infty$.

For computing $c_{\gamma h}$ we need $E\left(\left(X_n^{(h)}\right)^{\gamma h}\right)$, where $X_n^{(h)} = \sum_{j=0}^{\infty} \phi^j Z_{n-j}$. Then the Laplace transform is

$$\varphi_{X_n^{(h)}}(s) = E\left(\exp\left(-sX_n^{(h)}\right)\right) = \prod_{j=0}^{\infty} E\left(\exp\left(-s\phi^j Z_{n-j}\right)\right) = \prod_{j=0}^{\infty} \varphi_{Z_1}(s\phi^j). \tag{4.18}$$

Thus, it follows

$$\begin{aligned}
\log\left(\varphi_{X_n^{(h)}}(s)\right) &= \sum_{j=0}^{\infty} \log\left(\varphi_{Z_1}(s\phi^j)\right) \\
&= \sum_{j=0}^{\infty} \log\left(\varphi_{W_h}(s\sigma\phi^j)\right) \\
&= \sum_{j=0}^{\infty} \log\left(\exp\left(-\gamma \int_0^h \left(\log\left(1 + \beta s\sigma\phi^j e^{-at}\right)\right) dt\right)\right) \\
&= -\gamma \sum_{j=0}^{\infty} \int_0^h \left(\log\left(1 + \beta s\sigma\phi^j e^{-at}\right)\right) dt.
\end{aligned}$$

Then, by an appropriate substitution we have

$$\begin{aligned}
\varphi_{X_n^{(h)}}(s) &= \exp\left(-\gamma\sum_{j=0}^{\infty}\int_0^h \left(\log\left(1+\beta s\sigma\phi^j e^{-at}\right)\right)dt\right)\\
&= \exp\left(-\frac{\gamma}{a}\sum_{j=0}^{\infty}\int_{1+\beta s\sigma\phi^j}^{1+\beta s\sigma\phi^j e^{-ah}}\frac{\log u}{1-u}du\right)\\
&= \exp\left(\frac{\gamma}{a}\sum_{j=0}^{\infty}\operatorname{dilog}\left(1+\beta s\sigma\phi^j\right)-\operatorname{dilog}\left(1+\beta s\sigma\phi^j e^{-ah}\right)\right),
\end{aligned}$$

where dilog denotes the dilogarithm function,

$$\operatorname{dilog}(x) = \int_1^x \frac{\log(u)}{1-u}du.$$

Recall that $\phi = e^{-ah}$, then

$$\begin{aligned}
\varphi_{X_n^{(h)}}(s) &= \exp\left(\frac{\gamma}{a}\sum_{j=0}^{\infty}\left(\operatorname{dilog}\left(1+\beta s\sigma\phi^j\right)-\operatorname{dilog}\left(1+\beta s\sigma\phi^{j+1}\right)\right)\right)\\
&= \exp\left(\frac{\gamma}{a}\operatorname{dilog}(1+\beta s\sigma)\right).
\end{aligned}$$

Applying Theorem 2.1. of Brockwell and Brown (1978) it follows for $\gamma h < 1$

$$\begin{aligned}
E\left(X_n^{(h)}\right)^{\gamma h} &= \frac{1}{\Gamma(1-\gamma h)}\int_0^{\infty} s^{-\gamma h}\left|D\varphi_{X_n^{(h)}}(s)\right| ds\\
&= \frac{\gamma}{a\Gamma(1-\gamma h)}\int_0^{\infty} s^{-\gamma h-1}\exp\left(\frac{\gamma}{a}\operatorname{dilog}(1+\beta s\sigma)\right)\operatorname{dilog}(1+\beta s\sigma)ds,
\end{aligned} \tag{4.19}$$

where $D\varphi$ denotes the derivative of φ. In this way $c_{\gamma h}$ can be numerically computed from formula (4.19) for h fixed. Theorem 2.1. covers also the case $\gamma h \geq 1$, but we deal primarily with small h. With k_N^{-1} and c_α it is possible to obtain confidence intervals of the estimators.

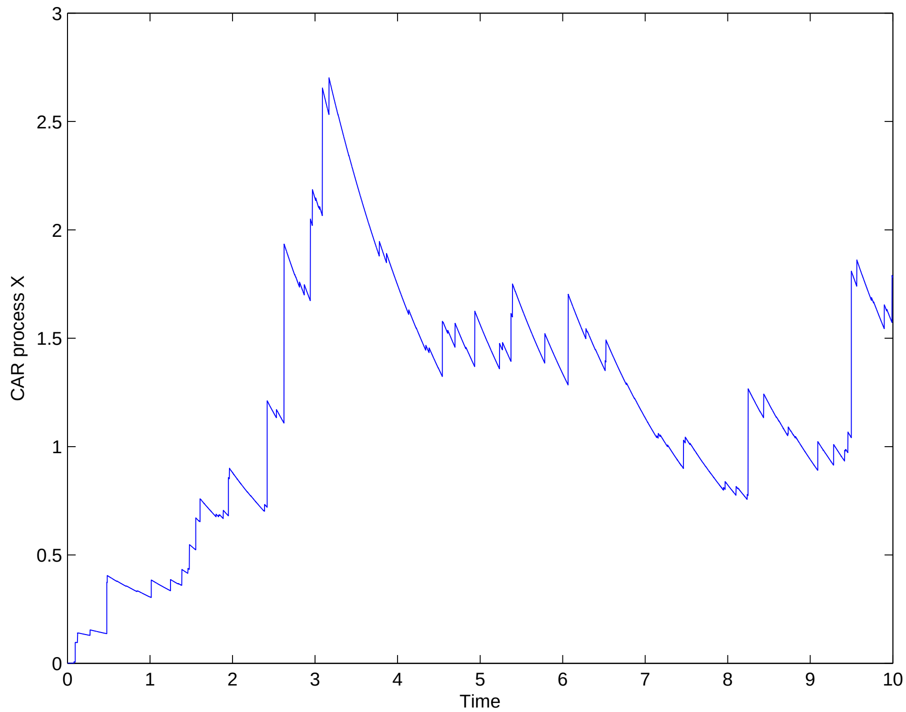

Figure 4.10: Standardized Gamma-driven OU process with $\gamma = 2$ for a short period

4.2.2 Simulation and Estimation

For the following example we use again the Euler scheme, so the following formula analogous to (4.2) is used for the simulation

$$X_n = X_{n-1} - aX_{n-1}h + \sigma \left(G_n - G_{n-1}\right). \tag{4.20}$$

Example 3. Now we simulate a Gamma driven OU-process, defined by

$$dX_t = -0.6X_t dt + dL_t, \tag{4.21}$$

where L_t is the standardized Gamma Process with parameter $\gamma = 2$. We obtained the $CAR(1)$ process using the Euler-Maruyama method at times $0, 0.001, 0.002, \ldots, 5000$. Figure 4.10 and Figure 4.11 shows the Gamma driven OU process for a short distance and for a long time horizon. Figure 4.12 shows the estimator depending on the different subintervals h or alternatively depending on n^*, the numbers of data points used for the estimation. For the estimation of the OU-parameter a and σ we sampled the process at subintervals $h = 0.01, h = 0.1, h = 1, h = 2, h = 5, h = 10$ and $h = 100$. To get for each h independent estimates for a and σ we simulated 100 replicates of the process. The sample means and the standard deviations for the different subintervals are shown in Table 4.3, where a remarkable accuracy is illustrated. Like in Example 1 the estimated values for a are greater than the real value if you choose a small h. Once again, the Euler method is the reason for this.

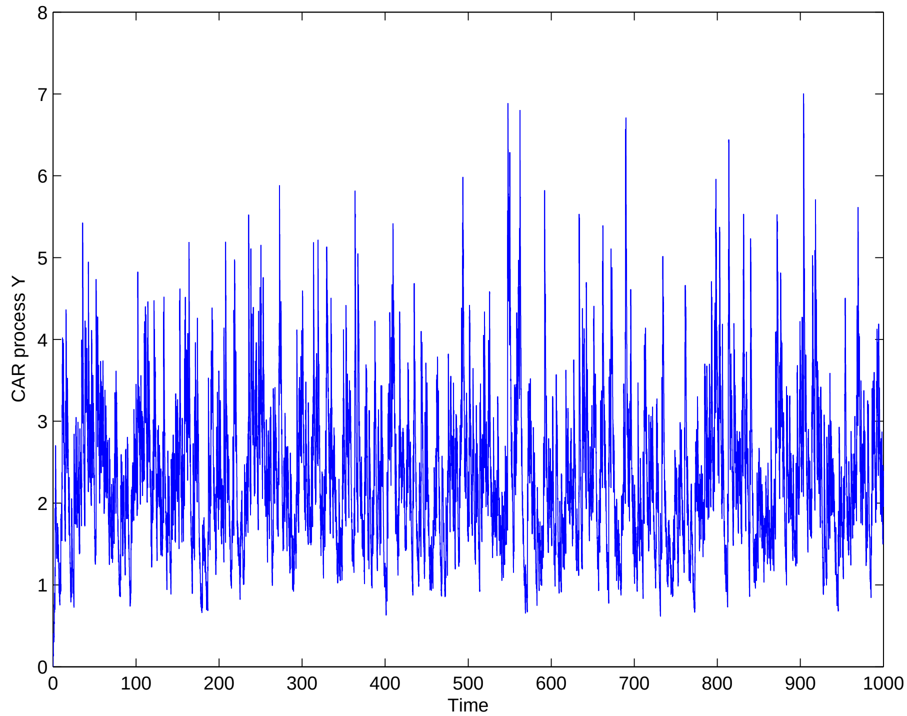

Figure 4.11: Standardized Gamma-driven OU process with $\gamma = 2$ for a longer time horizon

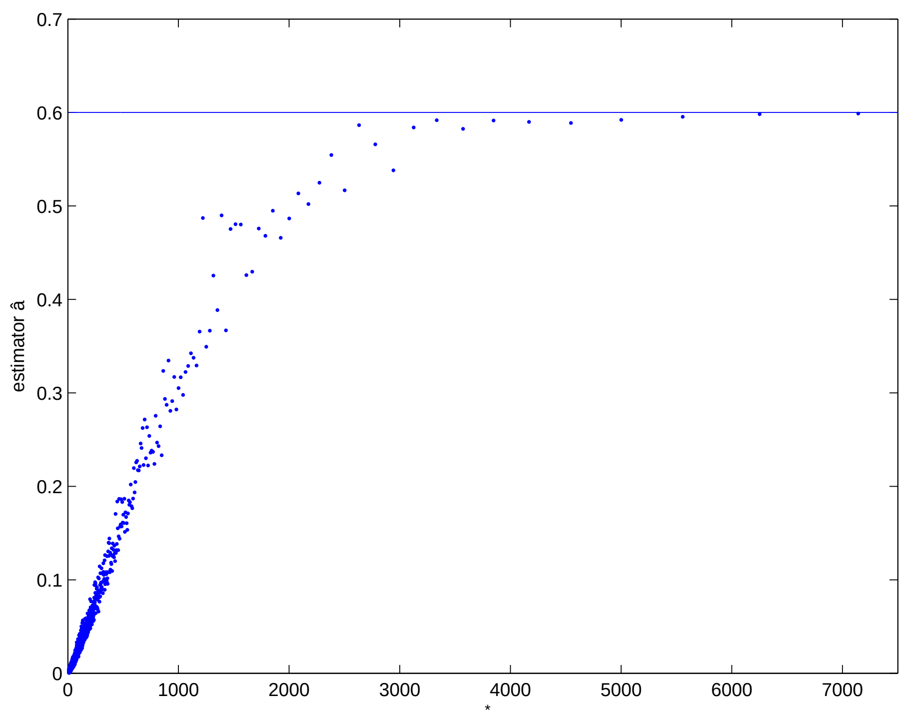

Figure 4.12: Estimator for a, depending on different time grids

Table 4.3: Estimated parameters based on 100 replicates on $[0, 5000]$ of the gamma-driven OU-process, observed at times nh, $n = 0, \ldots, \lfloor T/h \rfloor$.

h	Parameter	Sample mean of estimator	Sample standard deviation of estimator
0.01	$\widehat{a}$	0.6002	0.0
	$\widehat{\sigma}$	0.9990	0.0166
0.1	$\widehat{a}$	0.6002	0.0
	$\widehat{\sigma}$	0.9990	0.0167
1	$\widehat{a}$	0.5927	0.0036
	$\widehat{\sigma}$	0.9930	0.0244
2	$\widehat{a}$	0.5380	0.0188
	$\widehat{\sigma}$	0.9453	0.0170
5	$\widehat{a}$	0.3235	0.0343
	$\widehat{\sigma}$	0.7320	0.0471
10	$\widehat{a}$	0.1629	0.0192
	$\widehat{\sigma}$	0.5196	0.0396
100	$\widehat{a}$	0.0117	0.0024
	$\widehat{\sigma}$	0.1450	0.0304

Thereafter, analogous to Example 1, we want to analyze the distribution of the estimator $\widehat{a}$. For this purpose, we utilized the Weibull probability plot (Figure 4.13) by comparing the independent OU-estimator for the different time grids h with the Weibull-distributed data. Like in the previous example the distribution is for $h \geq 2$ not similar to a Weibull distribution.

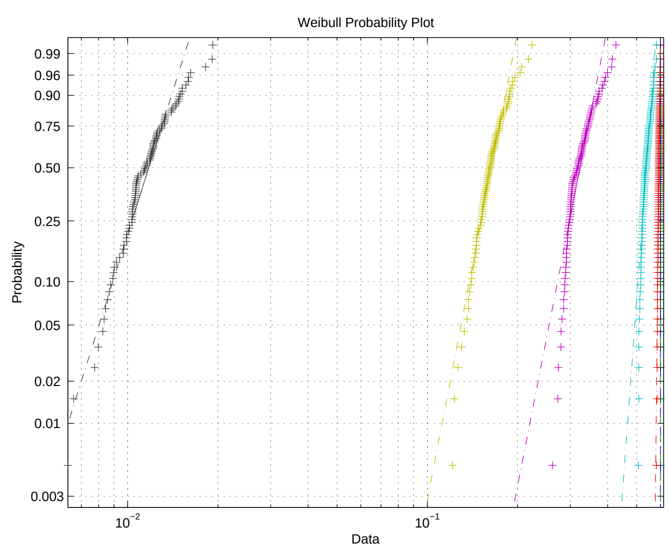

Figure 4.13: Weibull probability plot for different step sizes, from left to right: $h = 100$, $h = 10$, $h = 5$, $h = 2$ and $h = 1$

Now we estimate the parameters of the driving Lévy process. Therefore we first have to estimate the increments. As in Chapter 3 introduced, we apply both the trapezoidal and the Simpson approach.

Trapezoidal approach

For each h and each replicate we used the estimated OU-process parameters in (3.19) to compute the estimated Lévy increments $\Delta\widehat{L}_n^{(h)}$, $n = 1, \ldots, 5000/h$ in (3.20). Then we add these estimates in blocks of length $1/h$. So we get 5000 independent estimated increments of L, $\Delta\widehat{L}_k^1$, $k = 1, \ldots, 5000$, in one time unit, cf. (3.21). In Figure 4.14 the histogram for one realization with $h = 0.01$ is shown, together with the true probability density of G_1, Gamma$(\gamma, \gamma^{\frac{1}{2}})$.

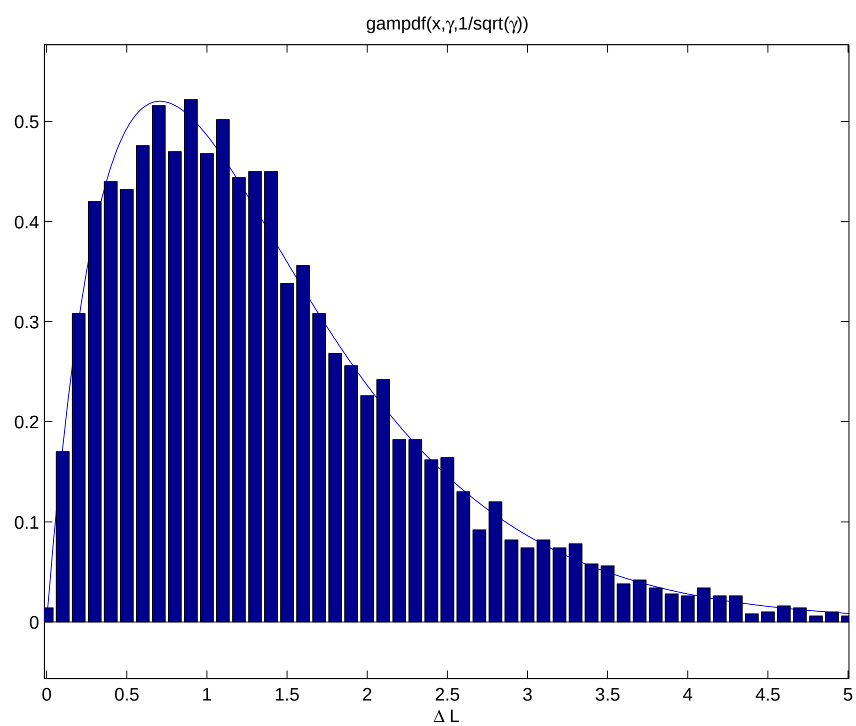

Figure 4.14: Probability density of the increments of the standardized Lévy process with $\gamma = 2$ and the histogram of the estimated increments from one realization of the OU-process, obtained by computing $\Delta L_n^{(0.01)}, n = 1, \ldots, 5,000,000$ with the trapezoidal approximation (3.20) and adding values in blocks of 100 to give estimated increments per unit time.

Table 4.4: Estimated parameters of the standardized driving Lévy process based on 100 replicates on $[0, 5000]$ of the gamma-driven OU-process.

h	Parameter	Sample mean of estimator	Sample standard deviation of estimator
0.01	γ	2.0039	0.0314
0.1	γ	2.0043	0.0340
1	γ	1.9967	0.0539

Table 4.4 shows the estimators for the parameter of Gamma distribution function. Even if we do not know that the BDLP is a gamma process, the histogram in Figure 4.14 strongly suggest that ΔL is gamma distributed. Figure 4.15 illustrates the histogram of the increments $\Delta\widehat{L}_k$ of all replicates for $h = 0.01$. Again without knowing the distribution at all, the histogram is an indicator for the gamma distribution. The histogram is a really smooth approximation of the true density function.

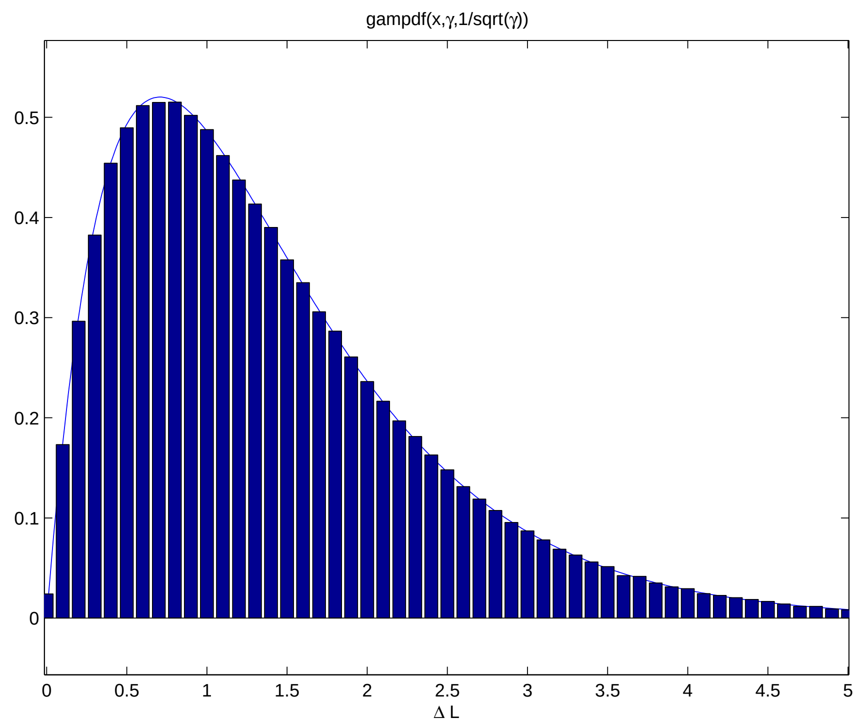

Figure 4.15: Probability density of the increments of the standardized Lévy process with $\gamma = 2$ and the histogram of the estimated increments for all 100 realization of the OU-process, obtained by computing $\Delta L_n^{(0.01)}, n = 1, \ldots, 5000000$ with the trapezoidal approximation (3.20) and adding values in blocks of 100 to give estimated increments per unit time for each realization.

Simpson's rule

Second, the Simpson's rule approach is shown. There, the estimator for ΔL is defined for an interval of length $2h$ by (3.22). Thereafter the estimated values $\Delta\widehat{L}_n^{(2h)}$, $n = 1,\ldots,5000/(2h)$ were summed up in blocks of length 50 to calculate 5000 independent increments of the gamma process in one time unit. Figure 4.16 illustrates the histogram for one of these iterates of G_1 combined with the true density of the increment function.

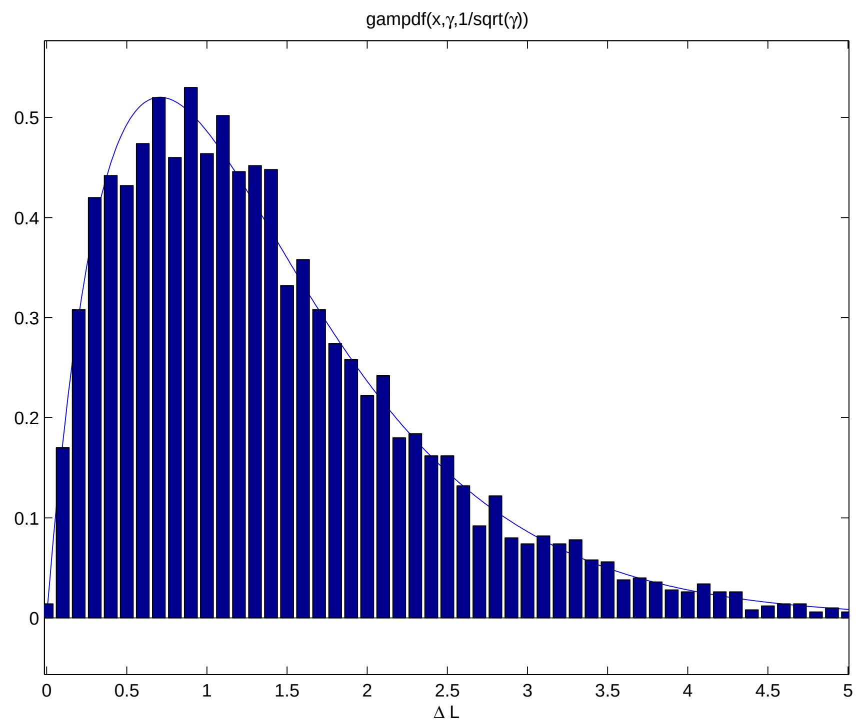

Figure 4.16: Probability density of the increments of the standardized Lévy process with $\gamma = 2$ and the histogram of the estimated increments from one realization of the OU-process, obtained by computing $\Delta L_n^{(0.01)}, n = 1,\ldots,5000000$ with the approximation using Simpson's rule (3.22) and adding values in blocks of 100 to give estimated increments per unit time.

Remark 4.2. Even if we have no results for the convergence of the estimated Lévy increments, the simulations strongly suggest a convergence to the real background driving Lévy process.

Chapter 5

Nonparametric Inference

In this section we introduce another estimation method (suggested by Jongbloed et al. (2005)) based on the characteristic function of the Ornstein-Uhlenbeck process. As shown in Chapter 2 the characteristic function of the OU-process without drift component can be written in terms of a so-called canonical function

$$\psi(z) = \exp\left(\int_0^\infty (e^{izx} - 1)\frac{k(x)}{x}dx\right). \tag{5.1}$$

The cumulant M-estimator we introduce in this chapter is defined as the projection of a preliminary estimate onto the class of cumulant functions of self-decomposable distributions, relative to a weighted L_2-distance. To define this estimator in Section 5.2 we first need to define some sets. At the end of the chapter, in Section 5.3 we show the consistency of the estimator.

5.1 Introduction

Definition of K, Ψ and G

We need the following three sets for the estimation of the canonical function. The set of canonical functions $K \subseteq \mathcal{L}^1(\pi)$ defined by

$$K := \left\{k \in \mathcal{L}^1(\pi) : k(x) \geq 0, k \text{ is decreasing and right-continuous}\right\}$$

is a convex cone which contains precisely the canonical functions of all non-degenerate self-decomposable distributions on $\mathbb{R}_+$ and the degenerate distribution at 0. Here the measure π is a Borel measure on $(0,\infty)$ defined by

$$\pi(dx) = \frac{1 \wedge x}{x}\, dx$$

and $\mathcal{L}^1(\pi)$ is the space of π-integrable functions on $(0,\infty)$. The semi-norm $\|\cdot\|_\pi$ on $\mathcal{L}^1(\pi)$ is defined by $\|k\|_\pi = \int |k|\, d\pi$ Let us define the set of appropriate characteristic functions by

$$\Psi = \left\{\psi : \mathbb{R} \to \mathbb{C}|\ \psi(z;k) = \exp\left(\int_0^\infty (e^{izx} - 1)\frac{k(x)}{x}dx\right) \text{ for some } k \in K\right\}.$$

Equivalently to Ψ there exists the set of cumulant functions

$$G := T(\Psi) = \left\{ g : \mathbb{R} \to \mathbb{C} | g(z) = \left(\int_0^\infty (e^{izx} - 1) \frac{k(x)}{x} dx \right) \text{ for some } k \in K \right\}.$$

Definition of Q, L and T

In the follow we introduce mappings which create relations between the different sets.

The mapping

$$Q : \; K \longrightarrow \Psi$$

assigns to each canonical function $k \in K$ its corresponding characteristic function in Ψ. The mapping Q is onto and one-to-one.

A result from complex analysis shows us the connection between Ψ and G.

Proposition 5.1. *(Chung, 2001, Theorem 7.6.2.) Let ψ be continuous with $\psi(0) = 1$ and $\psi(z) \neq 0$ for all $z \in [-Z, Z]$. Then there exists a unique function $g : \; [-Z, Z] \to \mathbb{C}$, such that $g(0) = 0$ and $e^{g(z)} = \psi(z)$ for all $z \in [-Z, Z]$. Under proper assumptions $[-Z, Z]$ can be replaced by $[-\infty, \infty]$. g is called the distinguished logarithm of ψ. If ψ is a characteristic function, g is called the cumulant function.*

Since ψ has as a characteristic function of an infinitely divisible distribution no zeros, see for instance (Sato, 1999, Lemma 7.5.), we have

$$\psi(z) = e^{g(z)} \text{ and } g(0) = 0.$$

Then the mapping $T : \Psi \to G$ is defined as

$$[T(\psi)](z) = g(z),$$

where $\psi \in \Psi$ and $z \in \mathbb{R}$. By the uniqueness of the distinguished logarithm it follows that T is one-to-one and onto.

Another mapping between canonical and cumulant functions is

$$L : K \longrightarrow G$$

with $L = T \circ Q$ and

$$[L(k)](z) = \int_0^\infty (e^{izx} - 1) \frac{k(x)}{x} \, dx$$

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

Figure 5.1 illustrates the relations between the different sets.

For notation: If we use in the following the parameter k, ψ, g, we always mean

$$\begin{aligned} k &\in K \\ \psi &\in \Psi \\ g &\in G. \end{aligned}$$

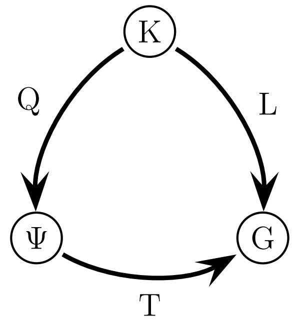

Figure 5.1: Sets K, Ψ and G and mappings Q, L and T

5.2 Definition of the Cumulant M-Estimator

The unique stationary probability distribution of X is denoted by μ_0. Any reference to the true underlying distribution is denoted by a subscript 0. For example, F_0 denotes the true underlying distribution function of X_1 and k_0 the true underlying canonical function. For the estimation of the canonical function based on discrete-time observations from X we need a preliminary estimator $\tilde{\psi}_N$ for ψ_0 first. A natural preliminary estimator is the empirical characteristic function which is introduced in Section 6.

Besides Assumption 1 we need another assumption.

Assumption 3. For the following the estimator $\tilde{\psi}_N$ has to satisfy either

$$\tilde{\psi}_N \text{ is a characteristic function for all } N \text{ and } \tilde{\psi}_N(z) \to \psi_0(z) \ a.s. \text{ for } N \to \infty, \tag{5.2}$$

or

$$\tilde{\psi}_N \text{ is a characteristic function for all } N \text{ and } \tilde{\psi}_N(z) \xrightarrow{P} \psi_0(z) \text{ for } N \to \infty. \tag{5.3}$$

The general idea of the cumulant M-estimator is the minimization of the distance between $\tilde{\psi}_N$ and $Q(k)$. w is the positive weight function which is (Lebesgue) integrable and compactly supported. Let S_w be the support of this weight function, here in this work, in particular in the applications, it is assumed to be the interval $(-z^*, z^*)$.

The M-estimator might be defined by

$$\widehat{k}_N = \operatorname*{argmin}_{k \in K} \int \left| [Q(k)](z) - \tilde{\psi}_N(z) \right|^2 w(z) dz. \tag{5.4}$$

Since the objective function $Q(k)$ is not convex, we use instead the function L which is obviously linear. So the cumulant M-estimator is defined by

$$\widehat{k}_N = \operatorname*{argmin}_{k \in K} \int \left| [L(k)](z) - \tilde{g}_N(z) \right|^2 w(z) dz, \tag{5.5}$$

where $\tilde{g}_N(z) = \log \tilde{\psi}_N(z)$ and log is the distinguished logarithm.

Detailed derivation of the cumulant M-estimator

In the following we want to describe the estimator more precisely. The space of square integrable functions with respect to $w(z)dz$ is defined by

$$L^2(w) := \left\{ f : \mathbb{R} \longrightarrow \mathbb{C} | f \text{ is Borel measurable and } \int |f(z)|^2 w(z)dz < \infty \right\}.$$

The inner-product $< \cdot, \cdot >_w$ on $L^2(w)$ is defined by

$$< f, g >_w = \Re \int f(z)\overline{g(z)}w(z)dz, \tag{5.6}$$

where the bar over g denotes the complex conjugation and $\Re$ is the operation of taking the real part of an element of $\mathbb{C}$. Define the norm by $\|g\|_w = \sqrt{< g, g >_w}$ for $g \in L^2(w)$. The space $(L^2(w), < \cdot, \cdot >_w)$ is a Hilbert space.

Define the estimator for the real cumulant function $g_0 = T(\psi_0)$ as a minimizer of

$$\Gamma_N(g) := \|g - \tilde{g}_N\|_w^2 = \int |g(z) - \tilde{g}_N(z)|^2 w(z)dz.$$

The mapping Γ_N is defined by

$$\Gamma_N : \ G^* \to \mathbb{C}, \tag{5.7}$$

where $G^* \subseteq G$ which is an appropriate subspace $L^2(w)$ that we determine in the following.

If G^* is a nonempty, closed, and convex subset of the Hilbert space $L^2(w)$, then there exists a unique element $g^* \in G^*$ which minimizes $\Gamma_N(g), g \in G^*$. Since Γ_N is a squared norm in a Hilbert space, we only need to specify an appropriate subset of G, which is nonempty, closed and convex.

Lemma 5.2. *(Jongbloed et al., 2005)*
$L : K \longrightarrow G$ is continuous, onto and one-to-one.

Proof. First we show the continuity of L. Let $(k_N)_{N\geq 1}$ be a sequence in K converging to $k^* \in K$ with $\|k_N - k^*\|_\pi \to 0$ as $N \to \infty$. Consider

$$\begin{aligned} |[L(k_N)](z) - [L(k^*)](z)| &= \left| \int_0^\infty \frac{e^{izx} - 1}{x} (k_N(x) - k^*(x))\, dx \right| \\ &\leq \int_0^\infty \left|e^{izx} - 1\right| \frac{|k_N(x) - k^*(x)|}{|x|} dx \\ &\leq (*)\,. \end{aligned}$$

Using the inequality $|e^{iz} - 1| \leq \min\{|z|, 2\}$ we get

$$(*) \ \leq \ |z| \int_0^1 |k_N(x) - k^*(x)|\, dx + \int_1^\infty \frac{2\,|k_N(x) - k^*(x)|}{|x|} dx.$$

Recall the definition of the measure π

$$\pi(dx) = \frac{1 \wedge x}{x} dx,$$

and the definition of $\|\cdot\|_\pi$, then

$$\begin{aligned} |[L(k_N)](z) - [L(k^*)](z)| &\leq \max\{|z|, 2\} \int_0^\infty (k_N(x) - k^*(x))\,\pi(dx) \\ &= \max\{|z|, 2\} \|k_N - k^*\|_\pi . \end{aligned}$$

Then $L(k_N) - L(k^*)$ converges uniformly on S_w which implies $\|L(k_N) - L(k^*)\|_w \to 0$ as $N \to \infty$. The definition of G already shows the surjectivity of L. If g_1 and $g_2 \in G$ and $\|g_1 - g_2\|_w = 0$, then $g_1 = g_2$ on S_w. Then also $\psi_1 := g_1 = g_2 := \psi_2$ on S_w. Section 13 in Loève (1977) implies that $\psi_1 = \psi_2$ on $\mathbb{R}$. Since $Q: \ K \to \Psi$ is one-to-one we have $k_1 = k_2$. □

Since G is not closed we introduce another set G' by

$$G' = \left\{ g : \mathbb{R} \longrightarrow \mathbb{C} : g(z) = \beta_0 iz + \int_0^\infty \frac{e^{izx} - 1}{x} k(x)dx, \beta_0 \geq 0, k \in K \right\},$$

which is closed under weak convergence. To show this, let S be a compact set containing the origin. If there exists a sequence $(g_N)_{N\geq 0} \in G'$ with

$$\sup_{z\in S} |g_N(z) - g(z)| \longrightarrow 0$$

as $N \to \infty$ for some g, then it follows

$$\sup_{z\in S} |\psi_N(z) - \psi(z)| \longrightarrow 0$$

as $N \to \infty$. Denote by X_N the random variable belonging to the characteristic function ψ_N. Applying Lévy's continuity theorem the random variables X_N are uniformly continuous and so uniformly tight.

As the random variables X_N are uniformly tight, it follows with Prohorov's theorem that $(X_N)_{N\geq 0}$ is relatively compact, i.e. there exists a subsequence N_l such that

$$X_{N_l} \Rightarrow X^*.$$

Recall that X_N is a positive self-decomposable random variable. Since the class of positive self-decomposable random variables is closed under weak convergence X^* is self-decomposable as well. Denote the cumulant function of X^* by g^*. Then it follows

$$g^* \in G'$$

and

$$\sup_{z\in S} |g_{N_l}(z) - g^*(z)| \longrightarrow 0$$

as $N \to \infty$. With the continuity of g and g^* on S we have

$$g = g^*$$

on S and

$$g = g^* \in G'$$

Thus, G' is closed in $L^2(w)$.

Example 4. Let S be a compact set containing the origin and let $(k_N)_{N\geq 1} \in K$ be a sequence with $k_N(x) = N \cdot 1_{[0,\frac{1}{N})}(x)$. Then, for each $z \in \mathbb{R}$,

$$g_N(z) = [L(k_N)](z) = N \int_0^{\frac{1}{N}} \frac{e^{izx}-1}{x} dx = \lim_{M\to\infty} \sum_{m=1}^{M} N \frac{\left(iz\frac{1}{N}\right)^m}{m \cdot m!} \longrightarrow iz$$

as $N \to \infty$. Let $g(z) = iz$. Then, as g_N and g are uniformly continuous on the compact set S, we have

$$\sup_{z\in S} |g_N(z) - g(z)| \longrightarrow 0$$

as $N \to \infty$. Recall that $G = \left\{g : \mathbb{R} \to \mathbb{C} | g(z) = \int_0^\infty (e^{izx}-1)\frac{k(x)}{x}dx \text{ for some } k \in K\right\}$ and therefore $g \notin G$ and $\psi(z) = e^{iz}$. This is an example which preclude closedness of G. The example also shows that the set G is not closed in $L^2(w)$.

Since G is not closed under weak convergence, we have to obtain an appropriate closed subset of G. This can be done with envelope functions. Let $R > 0$. Then there exists canonical functions $k_R \in K$ such that $\|k_R\|_\pi \leq R$. A possible choice is $k_R(x) = \frac{R}{4\sqrt{x}}$. The sequence $(k_R)_{R>0}$ defines a set of envelope functions. Then the set of canonical function is defined by

$$K_R := \{k \in K | k(x) \leq k_R(x) \text{ for } x \in (0,\infty)\}$$

and the corresponding set of cumulant functions by

$$G_R := L(K_R),$$

i.e. G_R is the image of K_R under L.

Lemma 5.3. *(Jongbloed et al., 2005, Lemma 4.3.) Let $R > 0$. Then*

1. *K_R is a compact, convex subset of $\mathcal{L}^1(\pi)$.*
2. *G_R is a compact, convex subset of $L^2(w)$.*

Proof. 1. The convexity of K_R is obvious. Let us denote by $(k_N)_{N\geq 1}$ a sequence in K_R. The sequence k_N is bounded on all strictly positive rational numbers, so we can use Cantor's diagonal argument to take a subsequence N_j of N such that

$$\lim_{j\to\infty} k_{N_j}(x) = k^*(x) \quad \text{pointwise}$$

for all $x > 0$ and $x \in \mathbb{Q}$. Then define

$$\tilde{k}(x) = \sup\{k^*(q), x < q, q \in \mathbb{Q}\}$$

for all $x \in (0,\infty)$. k^* is by definition a decreasing and right continuous function and we have $\tilde{k}(x) \leq k_R(x)$ for all $x \in (0,\infty)$. Then it follows $\tilde{k} \in K_R$. Besides,

$$\lim_{j\to\infty} k_{N_j}(x) = \tilde{k}(x) \quad \text{pointwise}$$

at all continuity points of $\tilde{k}$. Given that the discontinuity points of $\tilde{k}$ are at most countable, we have

$$\lim_{j\to\infty} k_{N_j}(x) = \tilde{k}(x) \ \pi - \text{ a.s}$$

on $x \in (0,\infty)$. Since $k_{N_j} \leq k_R$ on $(0,\infty)$ and $k_R \in \mathcal{L}^1(\pi)$, it follows with dominated convergence

$$\left\| k_{N_j} - \tilde{k} \right\|_\pi \to 0$$

as $N_j \to \infty$. So, K_R is sequentially compact, which is equivalent to compact.

2. G_R is compact, since it is the image of the compact set K_R under the continuous mapping L. G_R is convex since K_R is convex, since L is linear.

□

Corollary 5.4. *The inverse operator of L, $L^{-1} : G_R \to K_R$ is continuous.*

Proof. This is a standard result from topology, see for instance Werner (2007, Corollary IV.3.4.). □

One last step is missing for defining the objective function in terms of canonical functions. Until now we only know that Γ_N has unique minimizer over G_R, since G_R is compact and convex. However, since the mapping $L : \ K_R \longrightarrow G_R$ is onto and one-to-one and so to the each $g \in G_R$ there exists a unique $k \in K_R$, then there exists also a unique minimizer of $\Gamma_N \circ L (= \Gamma_N L)$ over K_R. The following theorem shows this more detailed.

Theorem 5.5. *(Jongbloed et al., 2005)*
Let $\widehat{g}_N = \operatorname{argmin}_{g\in G_R} \Gamma_N(g)$. Then $\widehat{k}_N = \operatorname{argmin}_{k\in K_R} [\Gamma_N L](k)$ exists. Moreover, $\widehat{k}_N = L^{-1}(\widehat{g}_N)$ and $\widehat{k}_N$ is unique.

Proof. Recall that $L : \ K_R \longrightarrow G_R$ is onto and one-to-one, so there exists to each $g \in G_R$ a unique $k \in K_R$, such that $L(k) = g$. It follows that

$$\gamma = \min_{g\in G_R} \Gamma_N(g) = \min_{k\in K_R} [\Gamma_N L](k). \tag{5.8}$$

Define $\widehat{k}_N = L^{-1}(\widehat{g}_N)$ and choose an arbitrary $k \in K_R$, but $k \neq \widehat{k}_N$. Then $\widehat{k}_N \in K_R$ and

$$[\Gamma_N L](\widehat{k}_N) = \Gamma_N(\widehat{g}_N) = \gamma < [\Gamma_N L](k).$$

So, $\widehat{k}_N$ is the unique minimizer of $\Gamma_N L$ over K_R. □

5.3 Consistency

We discuss in the following section the consistency of the cumulant M-estimator by strengthening the point wise convergence in (5.2) to uniform convergence.

Theorem 5.6. *(Jongbloed et al., 2005, Lemma 5.2, Theorem 5.3) Let $(\Omega, \mathcal{A}, P)$ be a probability space. Assume that for the sequence of preliminary estimators $\tilde{\psi}_N$ holds that $\tilde{\psi}_N$ is a characteristic function for all N and $\tilde{\psi}_N(z) \stackrel{a.s.}{\to} \psi_0(z)$ for $N \to \infty$. If $k_0 \in K_R$ for some $R > 0$, then the cumulant M-estimator is consistent. That is*

$$\begin{aligned} \|\widehat{g}_N - g_0\|_w &\longrightarrow 0 \ a.s. \ N \to \infty, \\ \left\|\widehat{k}_N - k_0\right\|_\pi &\longrightarrow 0 \ a.s. \ N \to \infty. \end{aligned}$$

The same results hold in probability if we only assume that $\tilde{\psi}_N$ is a characteristic function for all N and $\tilde{\psi}_N(z) \stackrel{P}{\to} \psi_0(z)$ for $N \to \infty$.

Proof. For the proof of the consistency we present an extended version of the proof in Jongbloed et al. (2005). For the convergence in probability which we omit we refer as well to Jongbloed et al. (2005).

Let $A \subseteq \Omega$ be the set with $P(A) = 1$ on which the convergence occurs. Let $\tilde{F}_N(\cdot, \omega)$ and F_0 be the corresponding distribution functions to $\tilde{\psi}_N(\cdot, \omega)$and ψ_0. Let S_w denote the support of the weight function.

If for every subsequence N there exists a further subsequence N_l and a set $A \in \Omega$ with $P(A) = 1$ such that

$$\tilde{F}_N(\cdot, \omega) \to F_0$$

for all $\omega \in A$ along each subsequence, then the family $\mathcal{F} := \left(\tilde{F}_N\right)_{N \geq 0}$ is tight.

By assumption there exists for each sequence N a characteristic function $\tilde{\psi}_N$ such that $\forall z \in \mathbb{R}$

$$\tilde{\psi}_N(z, \omega) \to \psi_0(z) \text{ a.s..}$$

For every $\delta > 0$

$$\begin{aligned} \int_{|x|>2/\delta} \tilde{F}_N(dx, \omega) &= 1 - \int_{-2/\delta}^{2/\delta} \tilde{F}_N(dx, \omega) \\ &\leq 2 - \frac{1}{\delta} \left| \int_{-\delta}^{\delta} \tilde{\psi}_N(z, \omega) dz \right| \\ &= \frac{1}{\delta} \int_{-\delta}^{\delta} dz - \frac{1}{\delta} \left| \int_{-\delta}^{\delta} \tilde{\psi}_N(z, \omega) dz \right| \\ &\leq \frac{1}{\delta} \int_{-\delta}^{\delta} \left| 1 - \tilde{\psi}_N(z, \omega) \right| dz \\ &:= b_N(\delta, \omega), \end{aligned} \tag{5.9}$$

where we used for the first inequality the Lemma in (Chung, 2001, p.170). Then define $b(\delta, \omega) := \frac{1}{\delta}\int_{-\delta}^{\delta} |1 - \psi_0(z)|\, dz$ and with Fubini's theorem it follows

$$\begin{aligned} E\left(\sup_{M\geq N} |b_N(\delta,\cdot) - b(\delta)|\right) &\leq \frac{1}{\delta}\int_{-\delta}^{\delta} E\left(\sup_{M\geq N}\left|\left|1-\tilde{\psi}_M(z,\cdot)\right| - |1-\psi_0(z)|\right|\right) dz \\ &\leq \frac{1}{\delta}\int_{-\delta}^{\delta} E\left(\sup_{M\geq N}\left|\tilde{\psi}_M(z,\cdot) - \psi_0(z)\right|\right) dz \\ &\to 0 \end{aligned} \tag{5.10}$$

as N tends to ∞. This implies that

$$|b_N(\delta,\cdot) - b(\delta)| \to 0 \ a.s. \text{ as } N \to \infty$$

for all $\delta > 0$. Together with (5.9) we have

$$\limsup_{N\to\infty} \int_{|x|>2/\delta} \tilde{F}_N(dx,\omega) \leq b(\delta,\omega)$$

for all $\delta \in \mathbb{Q}$ and $\omega \in A_1$ for some set $A_1 \in \mathcal{F}$ with $P(A_1) = 1$. Since

$$b(\delta) = \frac{1}{\delta}\int_{-\delta}^{\delta} |1-\psi_0(z)|\, dz \leq 2 \max_{z\in(-\delta,\delta)} |1-\psi_0(z)| \to 0$$

as $\delta \to 0$, the sequence $\mathcal{F}$ is tight for $\omega \in A_1$.

By assumption $A_2 \in \mathcal{F}$ is a set of probability one such that $\tilde{\psi}_N(z,\omega) \to \psi_0(z)$, for all $z \in \mathbb{Q}$ and for all $\omega \in A_2$.
If G is a limit of $\tilde{F}_N(\cdot,\omega)$, then

$$\int e^{izx} dG(x) = \lim_{j\to\infty} \int e^{izx} \tilde{F}_{N_j}(dx,\omega) = \int e^{izx} dF(x)$$

for all $z \in \mathbb{Q}$ and for all $\omega \in A_2$.

Hence $F = G$ and $\mathcal{F}$ has only one limit. So, $\tilde{F}_N(\cdot,\omega) \Rightarrow F$ for all $\omega \in A := A_1 \cap A_2$.

By Theorem 2.31 it follows that $\left\{\tilde{\psi}_N : \tilde{F}_N \in \mathcal{F}\right\}$ is uniformly equicontinuous. Since by assumption $\tilde{\psi}_N \to \psi_0$ pointwise and $\left\{\tilde{\psi}_N : \tilde{F}_N \in \mathcal{F}\right\}$ is uniformly equicontinuous, it follows with Lemma 2.32 that $\left\{\tilde{\psi}_N : \tilde{F}_N \in \mathcal{F}\right\}$ converges to ψ_0 uniformly on the compact sets S_w, i.e.

$$\sup_{z\in K}\left|\tilde{\psi}_N(z,\omega) - \psi_0(z)\right| \to 0$$

as $N \to \infty$.

Now fix $\omega \in A$. Since the characteristic function ψ_0 has no zeros there exists an $\epsilon > 0$ such that

$$\inf_{z\in S_w} |\psi_0(z)| > 2\epsilon.$$

For this ϵ there exists also an $N^* \in N^*(\epsilon,\omega) \in \mathbb{N}$ such that

$$\sup_{z\in S_w}\left|\tilde{\psi}_N(z,\omega) - \psi_0(z)\right| < \epsilon$$

for all $N \geq N^*$. Thus, for all $N \geq N^*$ and for all $z \in S_w$ we have

$$\left|\tilde{\psi}_N(z,\omega)\right| \geq |\psi_0(z)| - \left|\tilde{\psi}_N(z,\omega) - \psi_0(z)\right| \geq \epsilon > 0.$$

We define $\tilde{g}_N(\omega) = \log\left(\tilde{\psi}_N(\omega)\right)$ on S_w for $N \geq N^*$. By Theorem 5.1 the uniform convergence of $\tilde{\psi}_N(\omega)$ to ψ_0 on S_w carries over to uniform convergence of $\tilde{g}_N(\omega)$ to g_0 on S_w. With dominated convergence it follows

$$\lim_{N\to\infty} \|\tilde{g}_N(\omega) - g_0\|_w = 0.$$

Since $\widehat{g}_N(\cdot,\omega)$ minimizes Γ_N over G_R, we have

$$\|\widehat{g}_N(\cdot,\omega) - g_0\|_w \leq \|\widehat{g}_N(\cdot,\omega) - \tilde{g}_N(\cdot,\omega)\|_w + \|g_0 - \tilde{g}_N(\cdot,\omega)\|_w \leq 2\,\|g_0 - \tilde{g}_N(\cdot,\omega)\|_w \to 0$$

as $N \to \infty$. By Corollary 5.4 we have

$$\left\|\widehat{k}_N(\cdot,\omega) - k_0\right\|_\pi = \left\|L^{-1}(\widehat{g}_N(\cdot,\omega)) - L^{-1}(g_0)\right\|_\pi \to 0$$

as $N \to \infty$. Thus,

$$\lim_{N\to\infty} \left\|\widehat{k}_N(\cdot,\omega) - k_0\right\|_\pi = 0$$

for all $\omega \in A$ with $P(A) = 1$. □

As already mentioned the empirical characteristic function is a possible preliminary estimator.

Definition 5.7. *The empirical characteristic function is defined by*

$$\tilde{\psi_N}(z) = \int e^{izx} d\tilde{F}_N(x) = \frac{1}{N}\sum_{k=0}^{N-1} e^{izX_k}, \tag{5.11}$$

where $(X_1, \ldots X_N)$ is the sample and N the number of observations. The corresponding empirical cumulant function is defined by

$$\tilde{g_N}(z) = \log\left(\int e^{izx} d\tilde{F}_N(x)\right) = \log\left(\frac{1}{N}\sum_{k=0}^{N-1} e^{izX_k}\right), \tag{5.12}$$

where log *is the distinguished logarithm.*

With an application of Birkhoff's Ergodic Theorem, cf. Krengel (1985), it can be shown that for $z \in \mathbb{R}$

$$\tilde{\psi_N}(z) \xrightarrow{a.s.} \int e^{izx} dF_0(x) = \psi_0(z),$$

as $N \to \infty$. Thus, the consistency of k_N follows directly from Theorem 5.6.I

Chapter 6

Computation of the cumulant M-estimator, Applications and Examples

In this chapter we will present the Support Reduction Algorithm, introduced by Groeneboom et al. (2008), which gives a method to estimate the canonical function. In the first part of this chapter we show the theory of this algorithm and how to calculate the density of the stationary distribution of the Ornstein-Uhlenbeck process whereas in the second part we will give examples for the algorithm's performance.

6.1 Support Reduction Algorithm

6.1.1 Preliminaries

In order to compute the cumulant M-estimator we have to approximate the convex cone

$$K := \left\{k \in \mathcal{L}^1(\pi) : k(x) \geq 0, k \text{ is decreasing and right-continuous}\right\}$$

by a finite-dimensional subset. Therefore, we define a fixed set of positive numbers by

$$\Theta = \{\theta_1, \ldots, \theta_M\}$$

with $0 < \theta_1 < \ldots < \theta_M$, $M \geq 1$. A possible choice is taking an equidistant grid with grid points $\theta_j = jh$, $1 < j < M$, where h is the mesh width. Then define the basis functions by

$$u_\theta(x) = \mathbf{1}_{(0,\theta)}(x),$$

where $x \geq 0$. $u_\theta(x)$ is decreasing on $[0, \infty)$ and positive, so the requirements for the canonical functions are fulfilled. Taking the indicator functions as basis functions is only one possible choice, e.g. exponential function $u_\theta(x) = e^{j\theta x}$ is a possible alternative. We decided to use indicator functions since then integral of the inner-product $< \cdot, \cdot >_w$ can be simplified very well.
Consider, since $k(x) = u_\theta(x)$

$$v_\theta(z) = [Lu_\theta](z) = \int_0^\infty (e^{izx} - 1)\frac{k(x)}{x}dx = \int_0^\theta \frac{e^{izx} - 1}{x}dx.$$

Let $\mathcal{U}_\Theta = \{u_\theta, \theta \in \Theta\}$ be the set of the basis functions and K_Θ is defined as the convex cone generated by $\mathcal{U}_\Theta$

$$K_\Theta = \left\{ k \in K | k = \sum_{j=1}^{M} \alpha_j u_{\theta_j}, \alpha_j \in [0, \infty), 1 \leq i \leq N \right\}.$$

Recall the definition of the cumulant M-estimator (5.5),

$$\widehat{k}_N = \operatorname*{argmin}_{k \in K} \int |[L(k)](z) - \tilde{g}_N(z)|^2 w(z) dz,$$

where $w(z)$ is a weight function, and the definition of the norm $\|\cdot\|_w$,

$$\|g\|_w = \sqrt{< g, g >_w},$$

where $< g, g >_w = \Re \int g(z)\overline{g(z)} w(z) dz$ for $g \in \mathcal{L}^1(\pi)$. In the following we assume that $w(z)$ is an indicator function, i.e. $w(z) = \mathbf{1}_{(-z^*, z^*)}$ with z^* great. But another choice is also possible.

Since we approximate the convex cone K by the finite-dimensional cone K_Θ, the M-estimator for the canonical function k in K_Θ is defined as

$$\breve{k}_N = \operatorname*{argmin}_{k \in K_\Theta} [\Gamma_N L](k) = \operatorname*{argmin}_{k \in K_\Theta} \|L(k) - \tilde{g}_N\|_w^2 = \operatorname*{argmin}_{\alpha_1 \geq 0, \ldots, \alpha_M \geq 0} \left\| \sum_{j=1}^{M} \alpha_j v_{\theta_j} - \tilde{g}_N \right\|_w^2. \quad (6.1)$$

Furthermore, the distinguished logarithm in (6.1) can be defined only for those X_k from the sample space for which $\tilde{\psi}_N$ as a function of z does not hit zero on $(-\infty, \infty)$. For those X_k for which this is not satisfied, $[\Gamma_N L](k)$ has to be assigned an arbitrary value. It is shown in Gugushvili (2009) that the probability of the event that $\tilde{\psi}$ hits zero for z in $(-\infty, \infty)$ vanishes under appropriate conditions as $N \to \infty$.

Then

$$\begin{aligned}
< v_{\theta_j}, v_{\theta_k} >_w &= \Re \int_{-z^*}^{z^*} v_{\theta_j}(z) \overline{v_{\theta_k}(z)} dz \\
&= \Re \int_{-z^*}^{z^*} \int_0^{\theta_j} \frac{e^{izu} - 1}{u} du \overline{\int_0^{\theta_k} \frac{e^{izs} - 1}{s} ds} \, dz \\
&= \Re \int_{-z^*}^{z^*} \int_0^{\theta_j} \frac{e^{izu} - 1}{u} du \int_0^{\theta_k} \frac{e^{-izs} - 1}{s} ds \, dz \\
&= \Re \int_{-z^*}^{z^*} \int_0^{\theta_j} \int_0^{\theta_k} \frac{e^{izu} - 1}{u} \cdot \frac{e^{-izs} - 1}{s} ds \, du \, dz \\
&= \Re \int_{-z^*}^{z^*} \int_0^{\theta_j} \int_0^{\theta_k} \frac{e^{iz(u-s)} - e^{izu} - e^{-izs} + 1}{us} ds \, du \, dz. \quad (6.2)
\end{aligned}$$

Consider as well

$$
\begin{aligned}
< \tilde{g}_N, z_{\theta_j} >_w &= \Re \int_{-z^*}^{z^*} \tilde{g}_N(z)\overline{v_{\theta_j}(z)}dz \\
&= \Re \int_{-z^*}^{z^*} \log \tilde{\psi}_N(z)\overline{v_{\theta_j}(z)}dz \\
&= \Re \int_{-z^*}^{z^*} \log \left(\frac{1}{N} \sum_{k=0}^{N-1} e^{izX_k} \right) \overline{\int_0^{\theta_j} \frac{e^{izu}-1}{u} du} \, dz \\
&= \Re \int_{-z^*}^{z^*} \log \left(\frac{1}{N} \sum_{k=0}^{N-1} e^{izX_k} \right) \int_0^{\theta_j} \frac{e^{-izu}-1}{u} du \, dz, \qquad (6.3)
\end{aligned}
$$

where log is the distinguished logarithm defined as in Tucker (1967, p.92).

In the following we want to show how the distinguished logarithm $\log(\tilde{g}_N(z))$ can be calculated.

Remark 6.1. Numerical calculation of the distinguished logarithm
With Theorem 1 in (Tucker, 1967, Section 4.3.) there exists for each characteristic function of an infinitely divisible distribution a continuous real valued function $\phi(z)$ such that $\phi(0) = 0$ and

$$\psi(z) = |\psi(z)| \, e^{i\phi(z)} \qquad (6.4)$$

for all $z \in (-\infty, \infty)$, since it fulfills $\psi(z) \neq 0 \ \forall z \in (-\infty, \infty)$ and $\psi(0) = 1$. Equation (6.4) can be written as

$$\log \psi(z) = \log |\psi(z)| + i\phi(z).$$

The function ϕ is unique. Since by Theorem 2.31 characteristic functions are uniformly continuous, ψ can be defined as a continuous function over $[0, M]$, where $z^* < M$ and z^* defined as before. Then we have to choose a time grid for the interval $[0, M]$. The continuity condition gives a hint to find an appropriate calibration for the time grid. By continuity, there exists a $\delta > 0$ such that if

$$0 < z_j - z_{j-1} < \delta,$$

then $|\psi(z_j) - \psi(z_{j-1})| < \epsilon$, where $\epsilon > 0$. Then the interval can be divided by

$$0 = z_0 < z_1 < \cdots < z_{m-1} < z_m = M,$$

where

$$0 < \max z_j - z_{j-1}, 1 \leq j \leq m < \delta.$$

We assume in the following examples that the time grid is equidistant.

Then define $\phi(z)$ in $z \in [z_0, z_1]$ with

$$\phi(z) = \arg(\psi(z)),$$

where $\arg(z) = \arctan\left(\frac{\Im(z)}{\Re(z)}\right)$ and $\Im(z)$ denotes the imaginary part of z. If $\phi(z)$ is defined over $[z_0, z_j]$, ϕ is defined on $z \in [z_{j-1}, z_j]$ by

$$\phi(z) = \phi(z_j) + \arg\left(\frac{\psi(z)}{\psi(z_j)}\right).$$

Since the function arctan is symmetric about the origin $(0, 0i)$, it follows

$$\phi(-z) = -\phi(z)$$

for all $z \in [0, M]$ and ϕ is defined on $[-M, M]$. Together with

$$\log|\psi(-z)| = \log\sqrt{\psi(-z)\overline{\psi(-z)}} = \log\sqrt{\overline{\psi(z)}\psi(z)} = \log|\psi(z)|,$$

since $\psi(-z) = \overline{\psi(z)}$, it follows

$$\begin{aligned}
\log\psi(-z) &= \log|\psi(-z)| + i\phi(-z) \\
&= \log|\psi(z)| - i\phi(z) \\
&= \overline{\log\psi(z)}.
\end{aligned}$$

Then we have

$$\tilde{g}_N(-z) = \overline{\tilde{g}_N(z)}.$$

Thus, (6.3) can be written as

$$\begin{aligned}
< \tilde{g}_N, v_{\theta_j} >_w &= \Re\int_{-z^*}^{z^*} \tilde{g}_N(z)\overline{v_{\theta_j}(z)}dz \\
&= \Re\int_0^{z^*} \tilde{g}_N(z)\overline{v_{\theta_j}(z)}dz + \int_{-z^*}^{0} \tilde{g}_N(z)\overline{v_{\theta_j}(z)}dz \\
&= \Re\int_0^{z^*} \tilde{g}_N(z)\overline{v_{\theta_j}(z)}dz + \int_0^{z^*} \tilde{g}_N(-z)\overline{v_{\theta_j}(-z)}dz \\
&= \Re\int_0^{z^*} \tilde{g}_N(z)\overline{v_{\theta_j}(z)}dz + \int_0^{z^*} \overline{\tilde{g}_N(z)}v_{\theta_j}(z)dz \\
&= \Re\int_0^{z^*} \tilde{g}_N(z)\overline{v_{\theta_j}(z)} + \overline{\tilde{g}_N(z)}v_{\theta_j}(z) \\
&= 2\int_0^{z^*} \Re(\tilde{g}_N(z))\Re\big(v_{\theta_j}(z)\big) + \Im(\tilde{g}_N(z))\Im\big(v_{\theta_j}(z)\big)dz \\
&= 2\int_0^{z^*} \log\left|\tilde{\psi}_N(z)\right|\Re\big(v_{\theta_j}(z)\big) + \phi(z)\Im\big(v_{\theta_j}(z)\big)dz,
\end{aligned}$$

where ϕ is defined as in the Remark before.

6.1.2 Optimality Conditions

Let $D_{\Gamma_N L}$ be the directional derivative of $\Gamma_N L$ at $k_1 \in K$ in the direction of $k_2 \in K$ defined by

$$D_{\Gamma_N L}(k_2; k_1) := \lim_{\epsilon\to 0} \epsilon^{-1}\left(\left(\Gamma_N L\right)(k_1 + \epsilon k_2) - \left(\Gamma_N L\right)(k_1)\right).$$

$\Gamma_N L$ is the convex functional on K, where Γ_N, as an L^2-distance on a Hilbert space, is a strictly convex functional on G, and L is linear. Then, with optimization theory (see for instance Lemma 1 in Groeneboom et al. (2008))

$$\widehat{k}_N \text{ minimizes } \Gamma_N L \text{ over } K_\Theta \Longleftrightarrow D_{\Gamma_N L}(u_{\theta_j}; \widehat{k}_N)\begin{cases} \geq 0 & \forall j \in \{1, \ldots, M\}, \\ = 0 & \forall j \in J, \end{cases} \tag{6.5}$$

where

$$\widehat{k}_N = \sum_{j\in J} \alpha_j u_{\theta_j}$$

with $J = \{j \in \{1,\ldots,M\} : \alpha_j > 0\}$.

6.1.3 Algorithm

In the following we show the procedure of the Support Reduction Algorithm. Let the current iterate be given by

$$k^J = \sum_{j\in J} \alpha_j u_{\theta_j}.$$

Step 1: Determine the direction of the descent of $\Gamma_N L$

Define the set of possible directions of descent by

$$\Theta_< := \left\{\theta \in \Theta : \; D_{\Gamma_N L}(u_\theta, k^J) < 0\right\}.$$

If there exists a direction of descent, $\Theta_<$ is non empty. Let θ_{j^*} be the index of the next descent direction. A particular choice is the direction of the steepest descent. So,

$$\theta_{j^*} = \underset{\theta\in\Theta}{\operatorname{argmin}}\, D_{\Gamma_N L}(u_\theta, k^J).$$

But we use an alternative choice. Given the current iterate we want to find a function u_θ which provides a direction of descent for $\Gamma_N L$. By linearity of L we have

$$\begin{aligned}
[\Gamma_N L]\,(k+\epsilon u_\theta) - [\Gamma_N L]\,(k) &= \|L(k+\epsilon u_\theta) - \widetilde{g}_N\|_w^2 - \|L(k) - \widetilde{g}_N\|_w^2 \\
&= \|L(k+\epsilon u_\theta)\|_w^2 - 2 < L(k+\epsilon u_\theta), \widetilde{g}_N >_w + \|\widetilde{g}_N\|_w^2 \\
&\quad - \|Lk\|_w^2 + 2 < Lk, \widetilde{g}_N >_w - \|\widetilde{g}_N\|_w^2 \\
&= \|Lk\|_w^2 + 2\epsilon < Lk, Lu_\theta >_w + \epsilon^2 \|Lu_\theta\|_w^2 \\
&\quad -2 < L(k+\epsilon u_\theta), \widetilde{g}_N >_w - \|Lk\|_w^2 + 2 < Lk, \widetilde{g}_N >_w \\
&= 2\epsilon < Lk, Lu_\theta >_w + \epsilon^2 \|Lu_\theta\|_w^2 \\
&\quad -2 < Lk, \widetilde{g}_N >_w - 2\epsilon < Lu_\theta, \widetilde{g}_N >_w + 2 < Lk, \widetilde{g}_N >_w \\
&= 2\epsilon < Lk - \widetilde{g}_N, Lu_\theta >_w + \epsilon^2 \|Lu_\theta\|_w^2 \\
&= \epsilon c_1(\theta,k) + \frac{1}{2}\epsilon^2 c_2(\theta),
\end{aligned}$$

where

$$c_2(\theta) := 2\,\|Lu_\theta\|_w^2 = 2\,\|v_\theta\|_w^2 > 0 \tag{6.6}$$

and

$$c_1(\theta,k) := 2 < Lk - \widetilde{g}_N, Lu_\theta >_w = 2 < \sum_{j\in J} \alpha_j v_{\theta_j} - \widetilde{g}_N, v_\theta >_w . \tag{6.7}$$

We have $c_2(\theta) > 0$ for all $\theta \in \Theta$ and $\epsilon > 0$. So, to get a descent direction the parameter θ_{j^*} has to fulfill $c_1(\theta,k) < 0$. If $c_1(\theta,k) < 0$, then

$$\underset{\epsilon>0}{\operatorname{argmin}}\, \epsilon c_1(\theta,k) + \frac{1}{2}\epsilon^2 c_2(\theta) = -\frac{c_1(\theta,k)}{c_2(\theta)} =: \widehat{\epsilon}_\theta. \tag{6.8}$$

Minimizing $[\Gamma_N L](k+\widehat{\epsilon}_\theta u_\theta) - [\Gamma_N L](k)$ over all points $\theta \in \Theta$ with $c_1(\theta,k)<0$ gives

$$\begin{aligned}
\theta^* &= \underset{\theta\in\Theta, c_1(\theta,k)<0}{\operatorname{argmin}} [\Gamma_N L](k+\widehat{\epsilon}_\theta u_\theta) - [\Gamma_N L](k) \\
&= \underset{\theta\in\Theta, c_1(\theta,k)<0}{\operatorname{argmin}} \widehat{\epsilon}_\theta c_1(\theta,k) + \frac{1}{2}\,\widehat{\epsilon}_\theta^{\,2} c_2(\theta) \\
&= \underset{\theta\in\Theta, c_1(\theta,k)<0}{\operatorname{argmin}} -\frac{c_1(\theta,k)}{c_2(\theta)} c_1(\theta,k) + \frac{1}{2}\left(-\frac{c_1(\theta,k)}{c_2(\theta)}\right)^2 c_2(\theta) \\
&= \underset{\theta\in\Theta, c_1(\theta,k)<0}{\operatorname{argmin}} -\frac{c_1(\theta,k)^2}{2c_2(\theta)} \\
&= \underset{\theta\in\Theta}{\operatorname{argmin}} \frac{c_1(\theta,k)}{\sqrt{c_2(\theta)}}.
\end{aligned}$$

The last equality holds because of the monotonicity of the root function.
For implementation you can use a modified expression for $c_1(\theta,k)$ and $c_2(\theta)$ respectively.

$$\begin{aligned}
c_1(\theta,k) &= 2 < \sum_{j\in J} \alpha_j v_{\theta_j} - \widetilde{g}_N, v_\theta >_w \\
&= 2 < \sum_{j\in J} \alpha_j v_{\theta_j}, v_\theta >_w - 2 < \widetilde{g}_N, v_\theta >_w \\
&= 2\sum_{j\in J} \alpha_j < v_{\theta_j}, v_\theta >_w - 2 < \widetilde{g}_N, v_\theta >_w
\end{aligned}$$

and

$$c_2(\theta) = 2\,\|v_\theta\|_w^2 = 2 < v_\theta, v_\theta >_w . \tag{6.9}$$

Then

$$\theta^* = \underset{\theta\in\Theta}{\operatorname{argmin}} \frac{\sqrt{2}\left(\sum_{j\in J}\alpha_j < v_{\theta_j}, v_\theta >_w - < \widetilde{g}_N, v_\theta >_w\right)}{\sqrt{< v_\theta, v_\theta >_w}}.$$

Step 2: Computing the weights and support reduction step

Computation of the weights
Let j^* be the index which belongs to θ^*. Then the new iterate is given by

$$k^{J^*} = \sum_{j\in J^*} \beta_j u_{\theta_j}, \tag{6.10}$$

where $J^* = J \cup \{j^*\}$ and $\{\beta_j, j \in J^*\}$ are the weights. Then these unknown weights have to be determined by minimizing $\Gamma_N L(k^{J^*})$ with respect to $\{\beta_j, j \in J^*\}$ without positivity constraints. As shown below this is a quadratic unconstrained optimization problem or in other words a standard least-square problem.

Recall that

$$\min_{\beta_j\in\mathbb{R}, j\in J^*} \Gamma_N L\left(k^{J^*}\right) = \min_{\beta_j\in\mathbb{R}, j\in J^*} \left\| \sum_{j\in J^*} \beta_j v_{\theta_j} - \tilde{g}_N \right\|_w^2$$

then the partial derivatives are as follows

$$\begin{aligned}\frac{\partial \Gamma_N L\left(k^{J^*}\right)}{\partial \beta_i} &= 2 < \sum_{j\in J^*} \beta_j v_{\theta_j} - \tilde{g}_N, v_{\theta_i} >_w \\ &= 2\sum_{j\in J^*} \beta_j < v_{\theta_j}, v_{\theta_i} >_w - 2 < \tilde{g}_N, v_{\theta_i} >_w,\end{aligned}$$

with $i \in J^*$. Setting the partial derivative equal to zero we get the following linear equation system $A\beta = b$ with

$$A_{j,i} =< v_{\theta_j}, v_{\theta_i} >_w, \quad j, i \in J^* \tag{6.11}$$

$$b_j =< v_{\theta_j}, \tilde{g}_N >_w, \quad j \in J^* \tag{6.12}$$

and β the vector with the weights. The matrix A is symmetric. So, α is the unique solution of the system $A\beta = b$.

Support Reduction Step

If $\min\{\beta_j, j \in J^*\} \geq 0$, then $k^{J^*} \in K_\Theta$ and k^{J^*} satisfies the equality part of (6.5). The second conclusion of (6.5) holds since

$$\begin{aligned}D_{\Gamma_N L}(u_{\theta_i}; k^{J^*}) &= \lim_{\epsilon\to 0} \epsilon^{-1}\left(\left(\Gamma_N L\right)\left(k^{J^*} + \epsilon u_{\theta_i}\right) - \left(\Gamma_N L\right)\left(k^{J^*}\right)\right) \\ &= c_1\left(\theta_i, k^{J^*}\right) \\ &= 2 < \sum_{j\in J^*} \beta_j v_{\theta_j} - \widehat{g}_N, v_{\theta_i} >_w \\ &= 0. \end{aligned} \tag{6.13}$$

The last equation is valid due to the definition of system $A\beta = b$. Then the inequality part of (6.5) has to be proved and if a descent direction exists returned to step 1.
Otherwise, if $\min\{\beta_j, j \in J^*\} < 0$, we apply a support-reduction step. Then we can make a move from k^J towards $k^{J^*} \notin K_\Theta$ and stay within the cone K_Θ initially. As a next iterate, we take $k := k^J + \widehat{c}(k^{J^*} - k^J)$, where

$$\begin{aligned}\widehat{c} &= \max\left\{c \in [0,1] : k^J + c\left(k^{J^*} - k^J\right) \in K_\Theta\right\} \\ &= \max\left\{c \in [0,1] : \sum_{j\in J}\left(\alpha_j + c\left(\beta_j - \alpha_j\right)\right) u_{\theta_j} + c\beta_{j^*} u_{\theta_{j^*}} \in K_\Theta\right\} \\ &= \max\left\{c \in [0,1] : \sum_{j\in J}\left(c\beta_j + (1-c)\,\alpha_j\right) u_{\theta_j} + c\underbrace{\beta_{j^*}}_{>0} u_{\theta_{j^*}} \in K_\Theta\right\} \\ &= \max\left\{c \in [0,1] : c\beta_j + (1-c)\,\alpha_j \geq 0 \text{ for all } \beta_j, j \in J \text{ with } \beta_j < 0\right\}\end{aligned}$$

For the last equation consider, if $\beta_j \geq 0$, then

$$\underbrace{c\beta_j}_{\geq 0} + \underbrace{(1-c)}_{\geq 0}\underbrace{\alpha_j}_{\geq 0} \geq 0,$$

and the direction u_{θ_j} is not removed. Moreover,

$$\begin{aligned} c\beta_j + (1-c)\,\alpha_j &\geq 0 \\ \Longleftrightarrow \quad c &\leq \frac{\alpha_j}{\alpha_j - \beta_j}. \end{aligned} \tag{6.14}$$

Then

$$\widehat{c} = \min\left\{\frac{\alpha_j}{\alpha_j - \beta_j}, j \in J, \text{ for which } \beta_j < 0\right\}. \tag{6.15}$$

Let j^{**} be the index which belongs to the minimum in (6.15). The coefficient for j^{**} is $\widehat{c}\beta_{j^{**}} + (1-\widehat{c})\,\alpha_{j^{**}} = 0$, the coefficient for $\beta_j \geq 0$ is in anyway greater or equal than zero and finally the coefficient for $\beta_j < 0$ is non-negative since with 6.14

$$\begin{aligned} \widehat{c}\beta_j + (1-c)\,\alpha_j &= \widehat{c}(\beta_j - \alpha_j) + \alpha_j \\ &\geq \widehat{c}(\beta_j - \alpha_j) + \widehat{c}(\beta_j - \alpha_j) \\ &= 0. \end{aligned}$$

Thus, it holds $k \in K_\Theta$. The support point $\theta_{j^{**}}$ is removed whereas the other support points are kept in the current support set. This set of support points is denoted by $J^{**} = J^* \setminus \{j^{**}\}$. Thereafter we compute the optimal weights without constraint qualification for the new iterate $k^{J^{**}} = \sum_{j \in J^{**}} \gamma_j u_{\theta_j}$. If all weights γ_j are non-negative, the equality part of (6.5) is satisfied, cf. (6.13) and we can check the inequality part of (6.5). If $D_{\Gamma_N L}(u_{theta_j}; k^{J^{**}}) \geq 0$ for all $j \in \{1, \ldots, N\}$, the iterate $k^{J^{**}}$ minimizes $\Gamma_N L$ over K_Θ. Otherwise return to step 1. If $\min\{\gamma_j, j \in J^{**}\} < 0$, another support reduction step has to be applied. Finally, the iterate k will satisfy both the equality and the inequality part in (6.5).

To start the algorithm, we choose a starting value $\theta^{(0)} \in \Theta$. Then we determine the function $du_{\theta^{(0)}}$ by minimizing $\Gamma_N L$ as a function of $d > 0$.

$$\alpha_1 = \underset{\alpha \geq 0}{\operatorname{argmin}} \|\alpha v_{\theta^{(0)}} - \tilde{g}_N\|_w^2. \tag{6.16}$$

We obtain the minimum by setting the derivative with respect to α equal to zero.

$$\begin{aligned} \frac{\partial \|\alpha v_{\theta^{(0)}} - \tilde{g}_N\|_w^2}{\partial \alpha} &= 2 < \alpha v_{\theta^{(0)}} - \tilde{g}_N, v_{\theta^{(0)}} >_w \overset{!}{=} 0 \\ \alpha &= \frac{< v_{\theta^{(0)}}, \tilde{g}_N >_w}{< v_{\theta^{(0)}}, v_{\theta^{(0)}} >_w} \end{aligned}$$

Once the algorithm has been initialized it starts iteratively adding and removing support points, while in between computing optimal weights.

Remark 6.2. To show that the algorithm is convergent we refer to Groeneboom et al. (2008). There, in Theorem 1, the convergence under some assumptions is shown. In Jongbloed et al. (2005) a proof of that is given.

6.2 Determination of the Probability Density Function

In this chapter we give a way to calculate the stationary distribution function of the Ornstein-Uhlenbeck process.

Due to the uniqueness of the characteristic function there exists to each characteristic function ψ a corresponding distribution function. By an appropriate inversion theorem, see for instance Tucker (1967) or Chung (2001), the probability density function f can be expressed by the characteristic function

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-iux} \psi(u) du. \tag{6.17}$$

Since the characteristic function of the Ornstein-Uhlenbeck process X_1 is defined by

$$\psi(u) = \exp\left(\int_0^{\infty} (e^{iux} - 1) \frac{k(x)}{x} dx \right),$$

it follows for the estimated characteristic function

$$\begin{aligned} \widehat{\psi}(u) &= \exp\left(\int_0^{\infty} (e^{iux} - 1) \frac{\widehat{k}(x)}{x} dx \right) \\ &= \exp\left(\int_0^{\infty} (e^{iux} - 1) \frac{\sum_{j=1}^{M} \alpha_j u_{\theta_j}(x)}{x} dx \right) \\ &= \exp\left(\sum_{j=1}^{M} \alpha_j v_{\theta_j}(u) \right), \end{aligned}$$

where we use the same notation like in the Support Reduction Algorithm.

Thus, it follows for the estimated density of X_1 based on the sample $X_1, \ldots, X_N$

$$\widehat{f}(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-iux} e^{\sum_{j=1}^{M} \alpha_j v_{\theta_j}(u)} du. \tag{6.18}$$

Since equation (6.18) is quite difficult to implement, we use an approximation for the density function. Schorr (1975) proposed to use a Fourier approximation. Consider

$$f(x) = \sum_{k=-\infty}^{\infty} a_k e^{ik\omega x} \tag{6.19}$$

for $x \in (T_-, T_+)$, where

$$\begin{aligned} a_k &= \frac{\omega}{2\pi} \int_{-\infty}^{\infty} f(u) e^{-ik\omega u} du = \frac{\omega \psi(-k\omega)}{2\pi}, \\ \omega &= \frac{2\pi}{T}, \\ T &= T_+ - T_-, \end{aligned}$$

and $T_- < T_+$ are such that the probability density function $f(x)$ is 'small' for $x < T_-$ and $x > T_+$. In Schorr (1975) a method to determine T_- and T_+ is given more precisely.

Then it follows

$$\widehat{f}(x) = \sum_{k=-\infty}^{\infty} \frac{1}{T} \exp\left(\sum_{j=1}^{M} \alpha_j \int_0^{jh\omega k} \frac{e^{-iz} - 1}{z} dz + ik\omega x \right). \tag{6.20}$$

6.3 Examples

In this section we show the efficiency of the Support Reduction Algorithm for time for a gamma Ornstein-Uhlenbeck process, the other time for an i.i.d. simulated Inverse Gaussian distributed process. First we give some calibration details of the Support Reduction Algorithm.

Remark 6.3. For the following examples we used 60 basis functions with a time grid of 0.05. We set $z^* = 10$. The limit of tolerance in equations (6.5) and in the positivity condition of the coefficient vector α was set to 10^{-8}. We need this tolerance limit to get a termination condition for the algorithm. The parameter M we needed for the calculation of the distinguished logarithm was chosen to be 10.

The fundamental advantage of the Support Reduction Algorithm is the run time. Even though the calculation of matrix A, cf. (6.11), needs a long time, the algorithm works very efficiently. If once initialized, the determination of the vector α is fast. The elapsed time of the Support Reduction Algorithm was in the examples about 0.01 seconds.

6.3.1 Estimation from the Ornstein-Uhlenbeck process

Example 5. We simulated a compound Poisson driven Ornstein-Uhlenbeck process defined by

$$dX_t = -2X_t dt + dL_t,$$

where L_t is a compound Poisson process with intensity measure 4 and with exponential jumps of expectation $\frac{1}{3}$. We simulated the OU-process at times $0, 0.1, 0.2, \dots, 10000$. Then we estimated with the Support Reduction Algorithm the canonical function $\widehat{k}$. In Figure 6.1 the true and the estimated canonical function of the gamma-OU process is illustrated. Figure 6.2 shows the true and the estimated density function of the stationary distribution of the Ornstein-Uhlenbeck process. For computing the probability density function we used the method from Schorr (1975), where we set $T_- = 0$ and $T_+ = 6.75$ and the sum in (6.20) is approximated by a sum with limits ± 100.

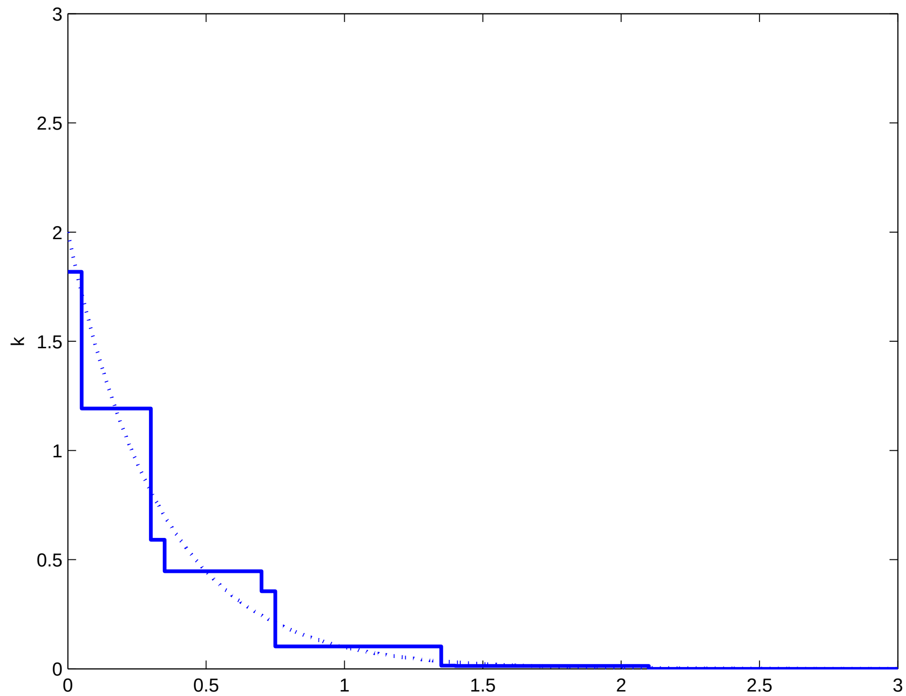

Figure 6.1: Gamma(2,3) distribution: Estimated (solid) and true(dotted) canonical function

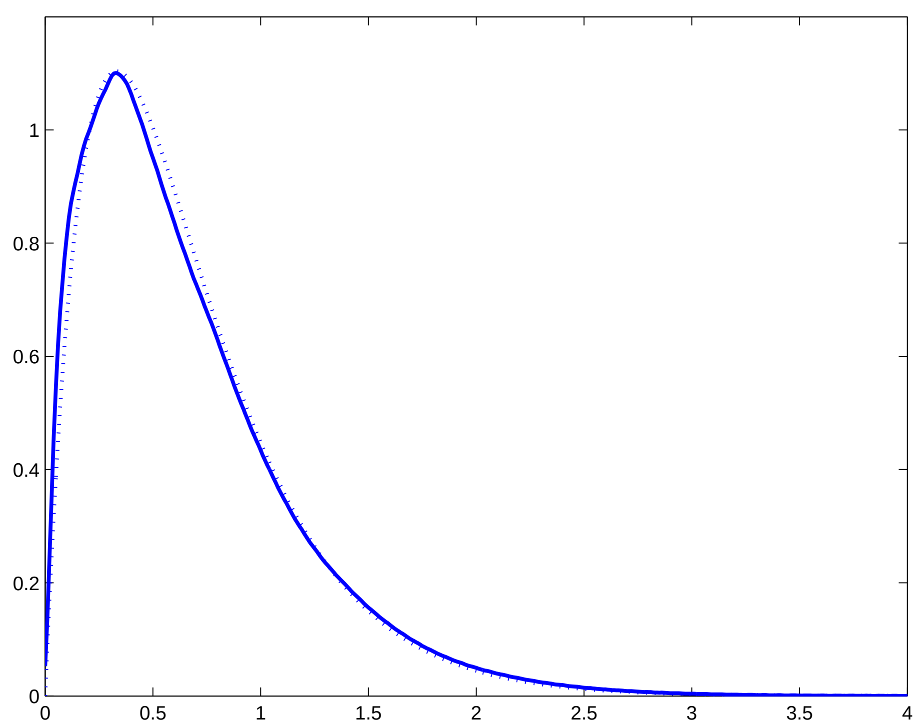

Figure 6.2: Gamma(2,3) distribution: Estimated (solid) and true(dotted) density function

6.3.2 Estimation from i.i.d. Data

Here we simulate, in contrast to the previous example, from i.i.d. data. Let $(X_n)_{1 \leq n \leq N}$ be independent random variables with common distribution function F. We give two examples, one time using the same gamma distribution like in the previous example, the other time using an Inverse Gaussian distribution.

Example 6. In this example we simulated i.i.d data from an Gamma(2,3) process at the times $0, 0.1, \ldots 10000$ like in the previous example. Then we used the Support Reduction Algorithm to compute the estimated canonical fucntion. Figure 6.3 and Figure 6.4 show the plots for the canonical function and the density function. Comparing the plots in Example 5 and in Example 6 one can observe that the i.i.d. data gives an infinitesimally better approximation of the gamma distribution.

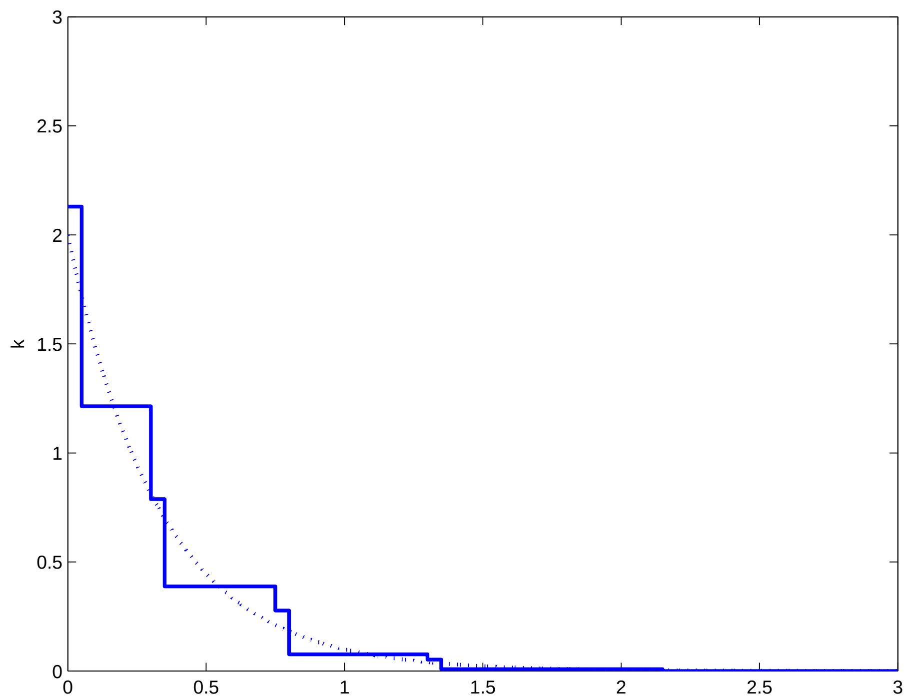

Figure 6.3: Gamma(2,3) distribution: Estimated (solid) and true(dotted) canonical function

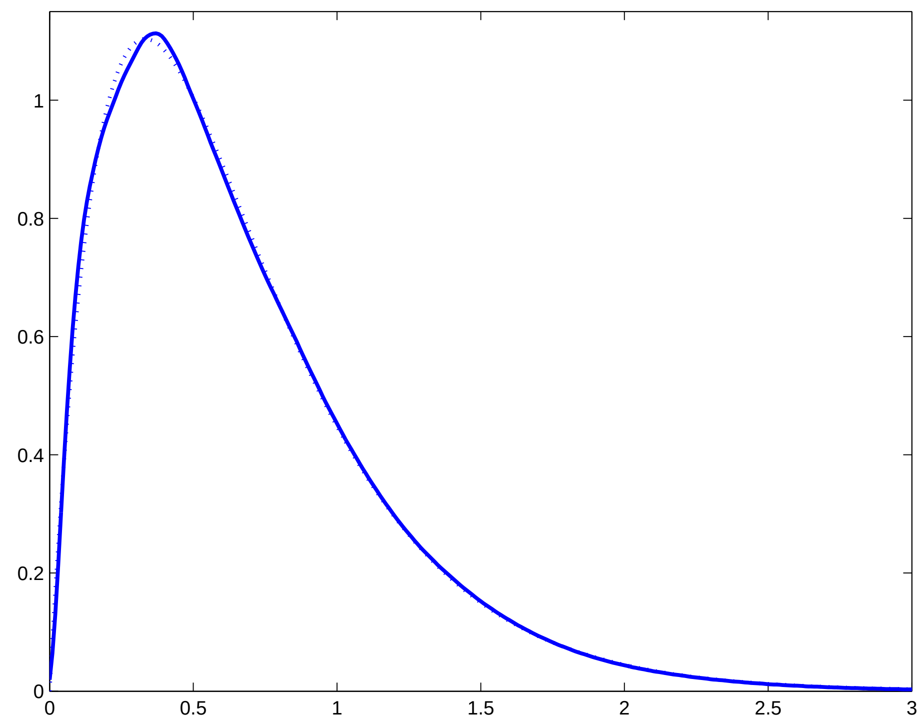

Figure 6.4: Gamma(2,3) distribution: Estimated (solid) and true(dotted) density function

In the following we give the definition of the Inverse Gaussian distribution. The Inverse Gaussian distribution $IG(\delta, \gamma)$ has the following probability density function

$$f(x) = \left(\frac{\delta^2}{2\pi x^3}\right)^{\frac{1}{2}} e^{\delta\gamma - \frac{\left(\delta^2+\gamma^2 x^2\right)}{2x}} \mathbf{1}_{(x>0)},$$

where $\delta > 0$ and $\gamma \geq 0$.

Then the corresponding canonical function can be expressed by

$$k(x) = \left(\frac{\delta^2}{2\pi x}\right)^{\frac{1}{2}} e^{-\frac{\gamma^2 x}{2}} \mathbf{1}_{(x>0)}.$$

In Example 7 we show the nonparametric estimation of an i.i.d. $IG\,(\delta, \gamma)$.

Example 7. We simulated 200000 independent $IG(2,3)$ distributed random variables at times $0, 0.5, 1, \ldots, 100000$. Then we applied the Support Reduction Algorithm to get the canonical function for this data set. Figure 6.5 shows the estimated and the true canonical function for this process. In Figure 6.6 the estimated and true density functions of Inverse Gaussian distribution $IG(2,1)$ are plotted by using equation (6.20). In this example the calibration parameters were set to $T_{-} = 0$, $T_{+} = 4.2766$ and again the sum in equation (6.20) was approximated by a sum with limits ± 100.

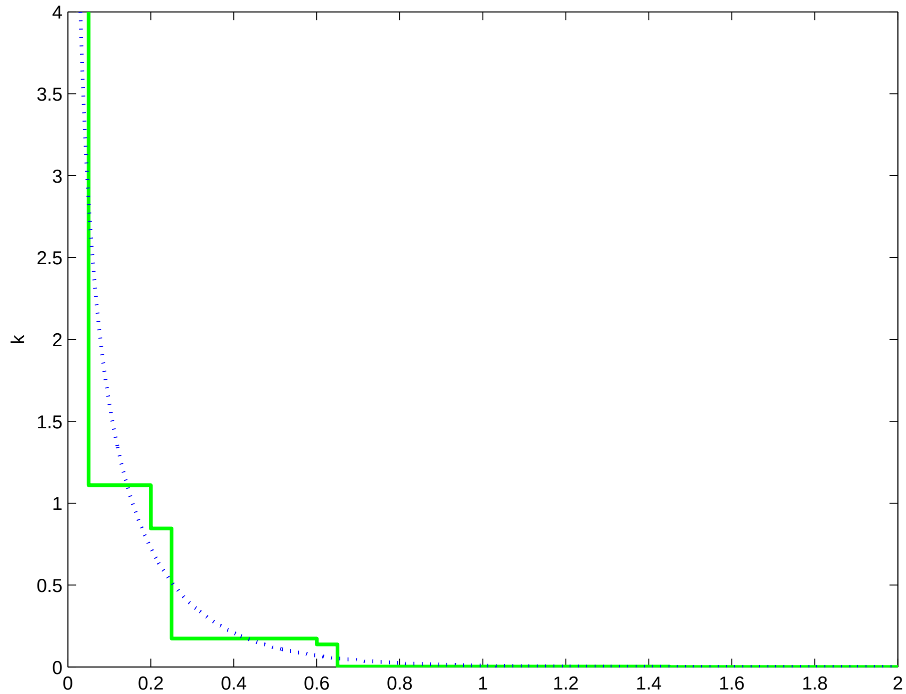

Figure 6.5: IG(2,1) distribution: Estimated (solid) and true(dotted) canonical function

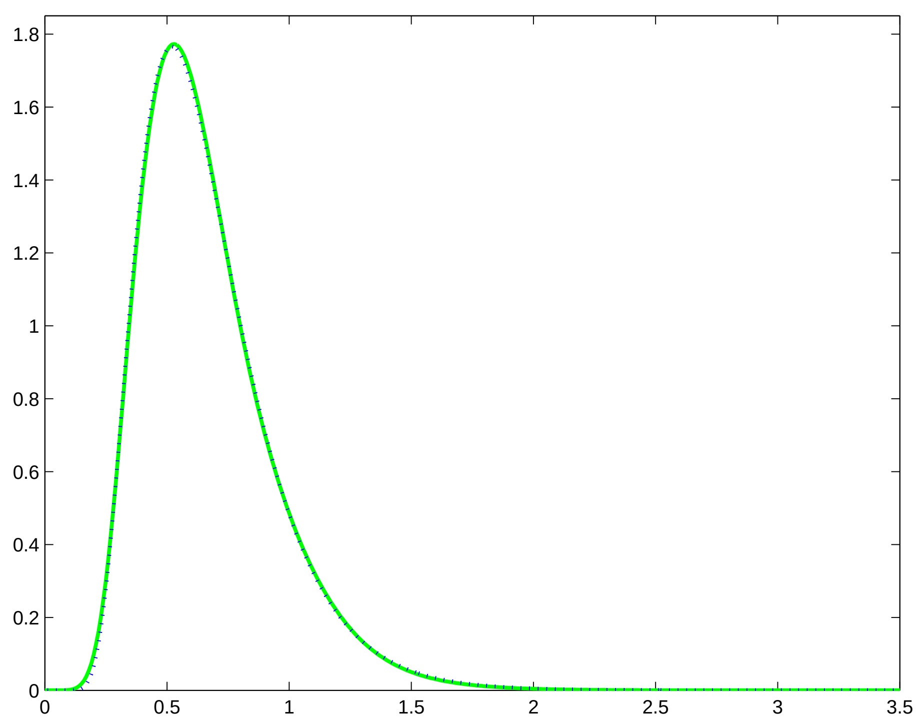

Figure 6.6: IG(2,1) distribution: Estimated (solid) and true(dotted) density function

Appendix A

Support Reduction Algorithm

A.1 Calculation of the basis functions

A.1.1 Calculation of $< v_{\theta_j}, v_{\theta_k} >_w$

```
% Calculation of the Basisfunctions (1): <vj_vk>
%------------------------------------------------------------------
% h = time grid of the basis functions
% M = number of basis functions
% z_star = boundaries in the indicator function 1_(-z_star,z_star)
%------------------------------------------------------------------
h = 0.05;
M = 60;
z_star = 10;

% integrand for computation of <v_(theta_j),v_(theta_k)>
integrandvj_vk = @(s,u,z)
((exp(i*z.*(u-s))-exp(i*z.*u)-exp(-i*z.*s)+ 1)./(u.*s));

% matrix with entries <v_(theta_j),v_(theta_k)>
vj_vk = zeros(M,M);
for j = 1 : M
    for k = j : M
        vj_vk(j,k) =
        real(triplequad
        (integrandvj_vk,10^-8,h*k,10^-8,h*j,-z_star,z_star))
        if (k > j)
            vj_vk(k,j) = vj_vk(j,k);
        end
    end
end

vj_vj = diag(vj_vk);
save ('basis functions','vj_vk','vj_vk','M','h','z_star')
```

A.1.2 Calculation of $< v_{\theta_j}, g >_w$

```
% Given X_0,X_1,...X_N, N observations from the stationary
% OU process
z=0:0.01:10;
```

```
function empfun = empcharfun (X,z);
% Calcualation of the distinguished logarithm
Z = length(z);
% Real and imaginary part of the characteristic function
for j = 1 : Z
    z1(j) = mean(cos(z(j)*X));
    z2(j) = mean(sin(z(j)*X));
end
psi_tilde = complex(z1,z2);
index = 1;
% Calculation of phi (cf. Remark 6.1.)
phi(index) = 0;
phi(index+1) = atan(h2(index+1)/h1(index+1));
for j = index+2:T
    im_part = z2(j)*z1(j-1)-z1(j)*z2(j-1);
    re_part = z1(j)*z1(j-1)+z2(j)*z2(j-1);
    phi(j) = phi(j-1) + atan(im_part/re_part);
end
empfun=complex(log(abs(psi_tilde)),phi);
end
```

```
% Calculation of the Basisfunctions (2): <vj_g>

integrand_vj_g =@(u,z) (exp(i*u*(0:0.01:z))-1) / u;

v_theta=zeros(1,length(t));
for j = 1 : M
     v_theta(:)=quadv(@(u)integrand_vj_g(u,z_star),10^-8,h*j);
     vj_g(j) = 2*sum(real(empfun).*real(v_theta)
     +imag(empfun).*imag(v_theta))*0.01;
end
```

A.2 Support Reduction Algorithm

```
function [J, alpha, m] = supportreduction(vj_vk,vj_vj,vj_g,M)
%-----------------------------------------------------------------
% INPUT:
% vj_vk = Matrix (M x M)
% vj_vj = diag(vj_vk) (1 x M)
% vj_g = (1 x M)
% M = number og basis function
%-----------------------------------------------------------------
% OUTPUT:
% J = logical vector, active directions (1 if basis function is
%     active and 0 if inactive) (1 x M)
% alpha = vector with weights/coefficients, belonging to J (1 x M)
% m = number of iterations
%-----------------------------------------------------------------
% Variables used in the algorithm:
% index vector = just the numbers from 1 to M
% descent = descent directions evaluated at each theta_j
% max_iter = maximum number of iterations
%-----------------------------------------------------------------

index_vector = 1:M;
J = logical(zeros(M,1));
alpha = zeros(M,1);
descent = zeros(M,1);

% Starting value:
% current iterate consists of only one direction theta_1
J(1) = 1;
alpha(1) = vj_g(1) / vj_vk(1,1);

max_iter = 100000;

for m=1:max_iter
         % compute descent direction (theta^*)
     descent = sqrt(2) * (vj_vk*alpha - vj_g) ./ sqrt(vj_vj);

     % find minimum over descent
     [min_desc, min_desc_index] = min(descent);
     theta_star = min_desc_index;

     if  min_desc >= 10^-8;
              % no further descent possible; done
              return;
     end

     % add new basis function theta_star
```

```
    J(theta_star) = 1;

    % compute new weights
    % length of beta = number of active thetas in J
    beta = (vj_vk.*(((+J)*(+J)'))+diag(-1*(+J)+1))\(vj_g.*(+J));
    betaJ = beta(J);
    min_beta = min(betaJ);

    % support reduction, if necessary
    while min_beta < -10^-8
            active_indices = index_vector(J);
            b_negative = active_indices(betaJ<-10^-8);

            % compute c and j_**
            c = alpha(b_negative)
                    ./(alpha(b_negative)-beta(b_negative));
            [min_c, min_c_index] = min(c);
            j_star_star = b_negative(min_c_index);

            % remove index j_**
            J(j_star_star) = 0;

            % compute new weights
            beta = (vj_vk.*(((+J)*(+J)'))
                         + diag(-1*(+J)+1))\(vj_g.*(+J));
            betaJ = beta(J);
            min_beta = min(betaJ);
    end

    if abs(alpha-beta) < 10^-8;
            return
    else
            % set alpha to new weights
            alpha = beta;
    end

end
% maximum number of iterations exceeded
disp('Maximum number of iterations exceeded');

end
```

Bibliography

David Applebaum. *Lévy Processes and Stochastic Calculus.* Cambridge University Press, Cambridge, 2004.

Ole E. Barndoff-Nielsen and Neil Shephard. Non-Gaussian Ornstein-Uhlenbeck based models and some of their uses in financial econometrics. *Journal of the Royal Statistical Society Series B*, 63:167–241, 2001a.

Ole E. Barndoff-Nielsen and Neil Shephard. Modelling by Lévy processes for financial econometrics. In *Lévy processes - Theory and Applications*, pages 283–318. Barndorff-Nielsen, Ole E. and Mikosch, Thomas and Resnick, Sidney I., 2001b.

Heinz Bauer. *Maß- und Integrationstheorie.* Walter de Gruyter, Berlin, 1992.

Fred Espen Benth, Thilo Meyer-Brandis, and Jan Kallsen. A non-Gaussian Ornstein-Uhlenbeck process for electricity spot price modeling and derivatives pricing. *Applied Mathematical Finance*, 14:153–169, 2007.

Patrick Billingsley. *Convergence of Probability Measures.* John Wiley & Sons, New York, 1968.

Leo Breiman. On some limit theorems similar to the arc-sin law. *Theory of Probability and Its Applications*, 10:323–331, 1965.

Peter J. Brockwell and B. Malcom Brown. Expansions for the positive stable laws. *Probability Theory and Related Fields*, 45:213–224, 1978.

Peter J. Brockwell and Richard A. Davis. *Time Series: Theory and Methods.* Springer, New York, 1991.

Peter J. Brockwell and Alexander Lindner. Strictly stationary solutions of autoregressive moving average equations. submitted for publication, 2009.

Peter J. Brockwell, Richard A. Davis, and Yu Yang. Estimation for nonnegative Lévy-driven Ornstein-Uhlenbeck processes. *Journal of Applied Probability*, 44:977–989, 2007.

Kai Lai Chung. *A Course in Probability Theory.* Academic Press, San Diego, 2001.

Rama Cont and Peter Tankov. *Financial Modelling With Jump Processes.* Chapman & Hall, Boca Raton, Fla., 2004.

Richard A. Davis and William P. McCormick. Estimation for first-order autoregressive processes with positive and bounded innovations. *Stochastic Processes and their Applications*, 31:237–250, 1989.

Richard A. Davis and Sidney I. Resnick. Extremes of moving averages of random variables from the domain of attraction of the double exponential dstribution. *Stochastic Processes and their Applications*, 30:41–68, 1988.

Gustav Doetsch. *Einführung in Theorie und Anwendung der Laplace-Transformation*. Birkhäuser Verlag, Basel, 1976.

Piet Groeneboom, Geurt Jongbloed, and Jon A. Weller. The Support Reduction Algorithm for computing nonparamtetric function estimates in mixture models. *Scandinavian Journal of Statistics*, 35:385–399, 2008.

Shota Gugushvili. Nonparametric estimation of the characteristic triplet of a discretely observed Lévy process. *Journal of Nonparametric Statistics*, 21:321 – 343, 2009.

Jean Jacod and Albert N. Shiryaev. *Limit Theorems For Stochastic Processes*. Springer, Berlin, 2003.

Geurt Jongbloed and Frank H. Van Der Meulen. Parametric estimation for subordinators and induced OU processes. *Scandinavian Journal of Statistics*, 33:825–847, 2006.

Geurt. Jongbloed, Frank H. Van Der Meulen, and Aad W. Van Der Vaart. Nonparametric inference for Lévy-driven Ornstein-Uhlenbeck processes. *Bernoulli*, 11:759–791, 2005.

Achim Klenke. *Probability Theory, A Comprehensive Course*. Springer, London, 2008.

Ulrich Krengel. *Ergodic Theorems*. Walter de Gruyter, Berlin, 1985.

Michel Loève. *Probability Theory I*. Springer, New York, 1977.

Hiroki Masuda. On multidimensional Ornstein-Uhlenbeck processes driven by a general Lévy process. *Bernoulli*, 10:97 – 120, 2004.

Philip E. Protter. *Stochastic Integration and Differential Equations*. Springer, Berlin, 2004.

Sidney I. Resnick. *Extreme Values, Regular Variation, and Point Processes*. Springer, New York, 1987.

Sidney I. Resnick. *Heavy- Tail Phenomena*. Springer, New York, 2007.

Ken-Iti Sato. *Lévy Processes and Infinitely Divisible Distributions*. Cambridge University Press, Cambridge, 1999.

Benno Schorr. Numerical inversion of a class of characteristic functions. *BIT*, 15:94–102, 1975.

Howard G. Tucker. *A graduate course in probability*. Academic Press, New York, 1967.

Dirk Werner. *Funktionalanalysis*. Springer, Berlin, 2007.