Nonparametric inference for Lévy driven Ornstein-Uhlenbeck processes

G. Jongbloed, F.H. van der Meulen*, A.W. van der Vaart
Vrije Universiteit Amsterdam
Department of Mathematics
De Boelelaan 1081a, 1081 HV Amsterdam, The Netherlands
{geurt,meulen,aad}@cs.vu.nl

Abstract

We consider nonparametric estimation of the Lévy measure of a hidden Lévy process driving a stationary Ornstein-Uhlenbeck process, which is observed at discrete time points. This Lévy measure can be expressed in the canonical function of the stationary distribution of the Ornstein-Uhlenbeck process, which is known to be self-decomposable. We propose an estimator for this canonical function based on a preliminary estimator of the characteristic function of this stationary distribution. We provide a suppport-reduction algorithm for the numerical computation of the estimator, and show that the estimator is asymptotically consistent under various sampling schemes. We also define a simple consistent estimator of the intensity parameter of the process. As sidesteps, a non-parametric procedure for estimating a self-decomposable density function is constructed, and it is shown that the Ornstein-Uhlenbeck process is β-mixing. Some general results on uniform convergence of random characteristic functions are included.

Key words and Phrases: Lévy process, self-decomposability, support-reduction algorithm, uniform convergence of characteristic functions.

1 Introduction

For a given positive number λ and a given increasing Lévy process Z without drift component consider the stochastic differential equation

$$dX(t) = -\lambda X(t)dt + dZ(\lambda t), \quad t \geq 0\,. \tag{1.1}$$

A solution X to this equation is called a Lévy driven *Ornstein-Uhlenbeck (OU) process*, and the process Z is referred to as the *background driving Lévy process (BDLP).* The auto-correlation of X at lag h can be expressed in the "intensity parameter" λ as $e^{-\lambda|h|}$.

By the Lévy-Khintchine representation (Sato (1999), theorem 8.1), the distribution of Z is characterized by its Lévy measure ρ. If $\int_{(2,\infty)} \log x\, \rho(dx) < \infty$, then a unique stationary solution to (1.1) exists (Sato (1999), theorem 17.5 and corollary 17.9). Moreover, the stationary distribution π of $X(1)$ is *self-decomposable* with characteristic function (ch.f.)

$$\psi(t) := \int e^{itx}\pi(dx) = \exp\left(\int_0^\infty (e^{itx} - 1)\frac{k(x)}{x}dx\right)\,, \tag{1.2}$$

where $k(x) = \rho(x, \infty)$. This shows that π is characterized by the decreasing function k, which is called the *canonical function.* Conversely, if we presuppose that π satisfies (1.2), then there exists an increasing Lévy process Z, unique in law, such that (1.1) holds for all $\lambda > 0$. Due to the special scaling in (1.1), π does not depend on λ.

Assume that we have discrete-time observations $X_0, X_\Delta, \ldots, X_{(n-1)\Delta}$ $(\Delta > 0)$ from $(X_t,\, t \geq 0)$, as defined by (1.1), where the sampling interval Δ may depend on n. Based on these observations we aim to estimate the parameters of the model. From the previous remarks this comes down to *(i)* estimating the intensity parameter λ and *(ii)* estimating the canonical function k. In this paper we deal with both estimation problems. Our approach for *(ii)* is nonparametric, although parametric submodels can be handled with our method as well (see Jongbloed and Van der Meulen (2004)).

One motivation for studying this problem comes from stochastic volatility models in financial mathematics. In Barndorff-Nielsen and Shephard (2001), stock price is modeled as a geometric Brownian motion. The diffusion coefficient of this motion, referred to as the volatility, is assumed to be a Lévy driven OU-process. Based on stock prices, the objective is to estimate the Lévy measure of the BDLP and λ. Although related, this estimation problem is intrinsically harder than the one we consider, since volatility is unobservable in practice. Despite this, the present work may be extended to handle these models by addition of a deconvolution step, and hence may provide a first step towards estimating these models nonparametrically. Another motivation comes from storage theory, where equation (1.1) is often referred to as the "storage equation" (see for example Çinlar and Pinsky (1971)).

Rubin and Tucker (1959) have considered nonparametric estimation for general Lévy processes, based on both continuous and discrete time observations, and Basawa and Brockwell (1982) has considered estimation for the subclass of continuously observed increasing Lévy processes. In this paper we consider indirect estimation through the observation of the Ornstein-Uhlenbeck process X at discrete time instants. Thus we deal with an inverse problem, and correspondingly our estimation techniques are quite different from the ones in these papers. Another paper on estimation for OU-processes is Papaspiliopoulos et al. (2004), where Bayesian estimation for parametric models is considered. Other papers on empirical characteristic function procedures include Knight and Satchell (1997), Feuerverger and McDunnough (1981) and, in a more general framework, Luong and Thompson (1987).

The organization of the paper is as follows.

In section 2 we discuss self-decomposability via the Lévy-Khintchine representation theorem. We show that a self-decomposable distribution is characterized by the logarithm of its ch.f., which is called the *cumulant function*. Furthermore, we state the close relationship between self-decomposability and Lévy driven OU-processes. Additional details on this can be found in section 17 in Sato (1999). We show that the process $(X_t,\, t \geq 0)$ is a Feller process (proposition 2.2) and hence satisfies the strong Markov-property. We also give some examples of self-decomposable distributions and related OU-processes. In section 3 we prove that the OU-process is β-mixing. In the proof, we use theory as developed in Meyn and Tweedie (1993) and Meyn and Tweedie (1993a) and a result from Shiga (1990).

Section 4 explains our method to estimate the canonical function. The method uses a given preliminary, consistent estimator $\tilde{\psi}_n$ for the characteristic function ψ_0 of $X(1)$, a typical example being the empirical characteristic function of the observations. Any characteristic function ψ without zeros possesses a unique (distinguished) logarithm, its associated cumulant function, which we denote by $T\psi$. Our estimator of the cumulant function $T\psi_0$ is now defined as the projection of the preliminary estimate $T\tilde{\psi}_n$ onto the class of cumulant functions of self-decomposable distributions, relative to a weighted L_2-distance. The estimates of ψ_0 and its associated canonical function are defined by inverting the respective maps. Under a "compactness condition" on the set of canonical functions, this *cumulant M-estimator* exists uniquely (theorem 4.5). In section 5 we prove two uniform convergence results on random characteristic functions, which may be of independent interest. We then use these results to provide conditions under which the cumulant M-estimator is consistent (theorem 5.4). The estimator can numerically be approximated by a support-reduction algorithm, as discussed in Groeneboom et al. (2003). In section 6 we explain how this algorithm fits within our setup.

Section 7 contains applications and examples of estimators under different observation schemes and present some simulation results. As a side-step we consider the estimation of a self-decomposable distribution based on independent and identically distributed (i.i.d.) data. This problem is difficult to handle by standard estimation techniques, as there exists no general closed form expression for the density of a self-decomposable distribution. The approach is to first estimate the canonical function by our cumulant-M-estimator and then apply Fourier-inversion.

For the intensity parameter λ, a simple explicit estimator is defined in section 8. This estimator is shown to be asymptotically consistent, although biased upward.

The appendix contains proofs of more technical lemmas.

2 Preliminaries

In this section we discuss self-decomposable distributions on $\mathbb{R}_+$ and Lévy driven OU-processes. Furthermore, we introduce notation that will be used throughout the rest of the paper.

2.1 Self-decomposable distributions on $\mathbb{R}_+$

A random variable X, with distribution function F, is said to be *self-decomposable (SD)* if for every $c \in (0,1)$ there exists a random variable X_c, independent of X, such that $X \stackrel{d}{=} cX + X_c$. In particular, all degenerate random variables are SD. Since the concept of self-decomposability only involves the distribution of a random variable (r.v.), we define a probability measure, or a characteristic function to be SD if its corresponding random variable is SD.

The class of self-decomposable distributions is a subclass of the class of infinitely divisible distributions. For the latter type of distributions, there is a powerful characterization in terms of characteristic functions: the Lévy Khintchine representation. A random variable Y with values in $\mathbb{R}_+$ $(= [0,\infty))$ is infinitely divisible if and only if its characteristic function has the form

$$\psi(t) = \mathbb{E}e^{itY} = \exp\left(i\gamma_0 t + \int_0^\infty (e^{itx} - 1)\nu(dx)\right), \quad \forall t \in \mathbb{R}, \tag{2.1}$$

where $\gamma_0 \geq 0$. The measure ν is called the *Lévy measure* of Y and satisfies the integrability condition $\int_0^\infty (x \wedge 1)\, \nu(dx) < \infty$. The parameter γ_0 is called the *drift*.

In case Y is self-decomposable, the measure ν takes a special form. It has a density with respect to Lebesgue measure (Sato (1999), corollary 15.11) and

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

where k is a decreasing function on $(0,\infty)$, known as the *canonical function*. We take this function to be right continuous. The integrability condition on ν reads as $\int_0^1 k(x)dx +$

$\int_1^\infty x^{-1}k(x)dx < \infty$. By proposition V.2.3. in Van Harn and Steutel (2004), the class of SD-distributions on $\mathbb{R}_+$ is closed under weak convergence. By theorem 27.13 in Sato (1999), the distribution of Y is either absolutely continuous with respect to Lebesgue measure or degenerate.

Thus each non-degenerate positive, self-decomposable random variable is characterized by a couple (γ_0, k) of a nonnegative number γ_0 and decreasing function k. In the next section we shall see that the variable $X(1)$ of the process X solving (1.1) is self-decomposable. Due to our assumption that the BDLP Z in (1.1) possesses no drift the parameter γ_0 corresponding to $X(1)$ is zero.

Next, we introduce some notation. Define a measure μ on the Lebesgue measurable sets in $(0,\infty)$ by

$$\mu(dx) = \frac{1 \wedge x}{x} dx\,, \quad x \in (0,\infty)\,.$$

Let $\mathcal{L}^1(\mu)$ be the space of μ-integrable functions on $(0,\infty)$. Define a semi-norm $\|\cdot\|_\mu$ on $\mathcal{L}^1(\mu)$ by $\|k\|_\mu = \int |k| d\mu$. Note that the definition of the measure μ precisely suits the integrability condition on k, which can now be formulated by $\|k\|_\mu < \infty$.

Define a set of functions by

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

Theset $K \subseteq \mathcal{L}^1(\mu)$ 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.

Let Ψ be the corresponding set of characteristic functions

$$\Psi := \left\{ \psi : \mathbb{R} \to \mathbb{C} \mid \psi(t;k) = \exp\left(\int_0^\infty (e^{itx} - 1) \frac{k(x)}{x} dx \right) \text{ for some } k \in K \right\}. \tag{2.2}$$

By the definition of Ψ the mapping $Q : K \mapsto \Psi$, assigning to each function $k \in K$ its corresponding ch.f. in Ψ, is onto. As a consequence of the Lévy-Khintchine theorem Q is also one-to-one.

The following result from complex analysis can be found for example in Chung (2001), section 7.6. Suppose $\varphi : \mathbb{R} \to \mathbb{C}$ is continuous, $\varphi(0) = 1$ and $\varphi(x) \neq 0$ for all $x \in [-T, T]$.

Then there exists a unique continuous function $f : [-T, T] \to \mathbb{C}$ such that $f(0) = 0$ and $\exp(f(x)) = \varphi(x)$. The corresponding statement when $[-T, T]$ is replaced by $(-\infty, \infty)$ is also true. The function f is referred to as the *distinguished logarithm.* In case φ is a ch.f. the function f is called a *cumulant function.*

Since an infinitely divisible ch.f. has no real zeros (see e.g. Sato (1999), lemma 7.5), we can attach to each $\psi \in \Psi$ a unique continuous function g such that $e^{g(t)} = \psi(t)$ and $g(0) = 0$. Since we will switch between sets of characteristic functions and cumulant functions throughout, we define a mapping T from Ψ onto its range by

$$[T(\psi)](t) = g(t), \quad \psi \in \Psi,\, t \in \mathbb{R}\,,$$

By uniqueness of the distinguished logarithm and the Lévy-Khintchine representation it follows that

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

By now, we have defined three sets, each parametrizing the class of SD distributions: *(i)* K: the set of canonical functions, *(ii)* Ψ: the set of ch.f., *(iii)* G: the set of cumulant functions. Typical members of each will be denoted by k, ψ and g respectively.

In order to switch easily between canonical functions and cumulants, we define the mapping $L : K \to G$ by $L = T \circ Q$. That is, for $k \in K$,

$$[L(k)](t) = \int_0^\infty (e^{itx} - 1)\frac{k(x)}{x}dx, \quad t \in \mathbb{R}\,.$$

The following diagram helps to clarify the relations between the operators defined so far:

$$\begin{tikzcd} K \arrow[rr, bend left, "Q"] \arrow[dr, "L"'] & & \Psi \arrow[dl, "T"] \\ & G & \end{tikzcd}$$

Next, we give a few examples of positive self-decomposable distributions:

Example 2.1 *(i)* Let X be *Gamma*(c, α) distributed with density f given by $f(x) = \frac{\alpha^c}{\Gamma(c)} x^{c-1} e^{-\alpha x} \mathbf{1}_{\{x>0\}}$, c, $\alpha > 0$. The ch.f. and canonical function are given by $\psi(t) = (1 - \alpha^{-1} it)^{-c}$ and $k(x) = ce^{-\alpha x}$ respectively.

(ii) Let X be an α-stable distribution with $\alpha \in (0,1)$. Then X has support $[0,\infty)$ if and only if its ch.f. equals

$$\psi(t) = \exp\left(-|t|^\alpha \left[1 - i\tan\left(\frac{\pi\alpha}{2}\right)\mathrm{Sgn}(t)\right]\right).$$

Its corresponding canonical function is given by $k(x) = c_\alpha x^{-\alpha}$, where $c_\alpha = \alpha/(\Gamma(1-\alpha)\cos(\pi\alpha/2))$. Note that $c_{1/2} = 1/\sqrt{2\pi}$. The density function of X permits a known closed form expression in terms of elementary functions only if $\alpha = 1/2$. In this case $f(x) = \frac{1}{\sqrt{2\pi}} x^{-3/2} e^{-1/(2x)} \mathbf{1}_{\{x>0\}}$. The probability distribution with this density is called the Lévy distribution.

If Z has a standard normal distribution, then W defined by $W = 1/Z^2$ if $Z \neq 0$ and $W = 0$ otherwise, has a Lévy distribution.

(iii) The Inverse Gaussian distribution with parameters δ and γ has probability density function

$$f(x) = \frac{1}{\sqrt{2\pi}} \delta e^{\delta\gamma} x^{-3/2} \exp(-(\delta^2 x^{-1} + \gamma^2 x)/2)\mathbf{1}_{\{x>0\}}, \quad \delta > 0,\, \gamma \geq 0.$$

See for example Barndorff-Nielsen and Shephard (2001a). Its canonical function is given by $k(x) = \frac{1}{\sqrt{2\pi}} \delta x^{-1/2} \exp(-\gamma^2 x/2)\mathbf{1}_{\{x>0\}}$. The case $(\delta, \gamma) = (1, 0)$ corresponds to the Lévy distribution.

2.2 Lévy driven Ornstein-Uhlenbeck processes

In this section we discuss some properties of Lévy driven OU-processes. We can assume that the driving Lévy process $Z = (Z_t,\, t \geq 0)$ has right-continuous sample paths, with existing left-hand limits. It is easily verified that a (strong) solution $X = (X_t,\, t \geq 0)$ to the equation (1.1) is given by

$$X_t = e^{-\lambda t} X_0 + \int_{(0,t]} e^{-\lambda(t-s)} dZ(\lambda s), \quad t \geq 0. \tag{2.3}$$

Up to indistinguishability, this solution is unique (Sato (1999), section 17). Furthermore, since X is given as a stochastic integral with respect to a càdlàg semi-martingale, the

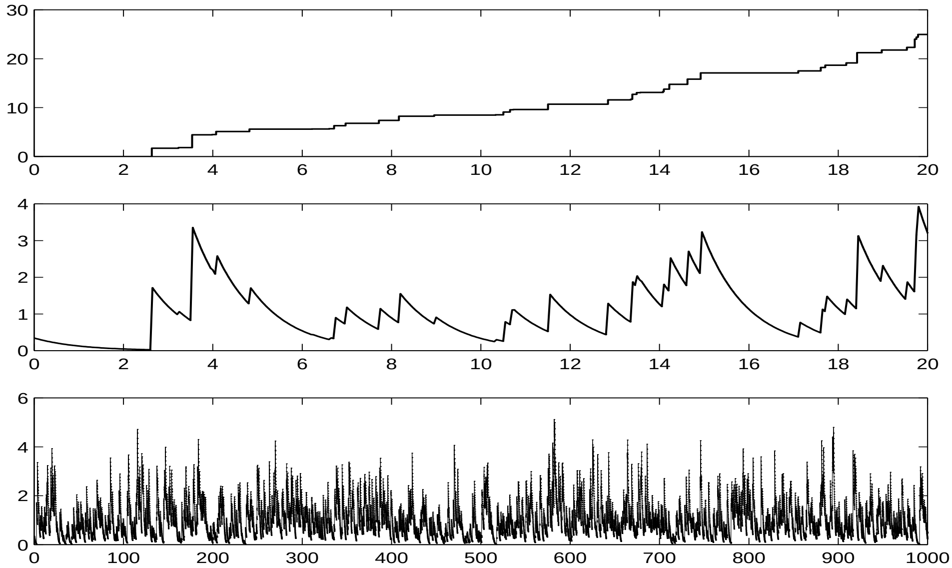

Figure 1: Upper figure: simulation of the BDLP (compound Poisson process of intensity 2 with exponential jumps of expectation 1/2). Middle figure: corresponding OU-process with Gamma$(2, 2)$ marginal distribution. Lower figure: OU-process on longer time-horizon.

OU-process $(X_t,\, t \geq 0)$ can be assumed càdlàg itself. The stochastic integral in (2.3) can be interpreted as a pathwise Lebesgue-Stieltjes integral, since almost surely the paths of Z are of finite variation on each interval $(0, t]$, $t \in (0, \infty)$ (Sato (1999), theorem 21.9). Figure 1 shows a simulation of an OU-process with Gamma$(2, 2)$ marginal distribution.

Denote by $(\mathcal{F}_t^0)_{t \geq 0}$ the natural filtration of (X_t). That is, $(\mathcal{F}_t^0) = \sigma(X_u,\, u \in [0, t])$. As noted in section 2 of Shiga (1990), $(X_t, \mathcal{F}_t^0)$ is a temporally homogeneous Markov process. Denote by $(E, \mathcal{E})$ the state space of X, where $\mathcal{E}$ is the Borel σ-field on E. We take $E = [0, \infty)$. The transition kernel of (X_t), denoted by $P_t(x, B)$ $(x \in E,\, B \in \mathcal{E})$, has characteristic function (Sato (1999), lemma 17.1)

$$\int e^{izy} P_t(x, dy) = \exp\left(ize^{-\lambda t} x + \lambda \int_0^t g(e^{\lambda(u-t)} z) du \right), \quad z \in \mathbb{R}\,, \tag{2.4}$$

where g is the cumulant of $Z(1)$.

Let $b\mathcal{E}$ denote the space of bounded $\mathcal{E}$-measurable functions. The transition kernel induces an operator $P_t : b\mathcal{E} \to b\mathcal{E}$ by

$$P_t f(x) := \int f(y) P_t(x, dy) = \int f(e^{-\lambda t} x + y) P_t(0, dy) . \tag{2.5}$$

The second equality follows directly from the explicit solution (2.3). We call P_t the transition operator. Let $C_0(E)$ denote the space of continuous functions on E vanishing at infinity (i.e. $\forall \varepsilon > 0$ there exists a compact subset K of E such that $|f| \leq \varepsilon$ on $E \setminus K$).

Proposition 2.2 *The transition operator of the OU-process is of Feller-type. That is,*

(i) $P_t C_0(E) \subseteq C_0(E)$ *for all* $t \geq 0$*,*

(ii) $\forall f \in C_0(E)$*,* $\forall x \in E$*,* $\lim_{t \downarrow 0} P_t f(x) = f(x)$*.*

For general notions concerning Markov processes of Feller type we refer to chapter 3 in Revuz and Yor (1999).

Proof Let $f \in C_0(E)$, whence f is bounded. If $x_n \to x$ in E, then $f(e^{-\lambda t} x_n + y) \to f(e^{-\lambda t} x + y)$ in $\mathbb{R}$, by the continuity of f, for any $y \in \mathbb{R}$. By dominated convergence, $P_t f(x_n) \to P_t f(x)$, as $n \to \infty$. Hence, $P_t f$ is continuous. Again by dominated convergence, $P_t f(x) \to 0$, as $x \to \infty$.

For the second part, by dominated convergence $\int_0^t g(e^{\lambda(u-t)} z) du = \int_0^t g(e^{-\lambda u} z) du \to 0$, as $t \downarrow 0$. Here we use the continuity of the cumulant g and $g(0) = 0$. Then it follows from (2.4) that

$$\lim_{t \downarrow 0} \int e^{izy} P_t(x, dy) = e^{izx} .$$

Thus $P_t(x, \cdot)$ converges weakly to $\varepsilon_x(\cdot)$ (Dirac measure at x):

$$\lim_{t \downarrow 0} \int f(y) P_t(x, dy) = \int f(y) \varepsilon_x(dy) = f(x) , \quad \forall f \in C_b(E) .$$

Here $C_b(E)$ denotes the class of bounded, continuous functions on E. The result follows since $C_0(E) \subseteq C_b(E)$. □

The Feller property of (X_t) implies (X_t) is a *Borel right Markov process* (see the definitions in chapter 9 of Getoor (1975)). We will need this result in section 3.

Since P_t is Feller, (X_t) satisfies the strong Markov property (Revuz and Yor (1999), theorem III.3.1). In order to state a useful form of the latter property, we define a *canonical OU-process* on the space $\Omega = D[0,\infty)$, by setting $X_t(\omega) = \omega(t)$, for $\omega \in \Omega$ (here $D[0,\infty)$ denotes the space of càdlàg functions on $[0,\infty)$, equipped with its σ-algebra generated by the cylinder sets). By the Feller property, this process exists (Revuz and Yor (1999), theorem III.2.7). Let ν be a probability measure on $(E, \mathcal{E})$ and denote by P_ν the distribution of the canonical OU-process on $D[0,\infty)$ with initial distribution ν. For $t \in [0,\infty)$, we define the shift maps $\theta_t : \Omega \to \Omega$ by $\theta_t(\omega(\cdot)) = \omega(\cdot + t)$.

Next, we enlarge the filtration by including certain null sets. Denote by $\mathcal{F}_\infty^\nu$ the completion of $\mathcal{F}_\infty^0 = \sigma(\mathcal{F}_t^0, t \geq 0)$ with respect to P_ν. Let $(\mathcal{F}_t^\nu)$ be the filtration obtained by adding to each $\mathcal{F}_t^0$ all the P_ν-negligible sets of $\mathcal{F}_\infty^\nu$. Finally, set $\mathcal{F}_t = \bigcap_\nu \mathcal{F}_t^\nu$ and $\mathcal{F}_\infty = \bigcap_\nu \mathcal{F}_\infty^\nu$, where the intersection is over all initial probability measures ν on $(E, \mathcal{E})$. In the special case of Feller processes, it can be shown that the filtration $(\mathcal{F}_t)$ obtained in this way is automatically right continuous (thus, it satisfies the "usual hypotheses"). See proposition III.2.10 in Revuz and Yor (1999). Moreover, (X_t) is still Markov with respect to this completed filtration (Revuz and Yor (1999), proposition III.2.14). The strong Markov property can now be formulated as follows. Let Z be an $\mathcal{F}_\infty$-measurable and positive (or bounded) random variable. Let T be an $\mathcal{F}_t$-stopping time. Then for any initial measure ν,

$$E_\nu(Z \circ \theta_T | \mathcal{F}_T) = E_{X_T}(Z)\,, \quad P_\nu - \text{a.s. on } \{T < \infty\}\,. \tag{2.6}$$

Here $\mathcal{F}_T = \{A \in \mathcal{F} : A \bigcap \{T \leq t\} \in \mathcal{F}_t, \forall t \geq 0\}$. The expectation on the right-hand side is interpreted as $E_x Z$, evaluated at $x = X_T$.

In section 3 we will apply the strong Markov property to random times such as $\sigma_A := \inf\{t \geq 0 : X_t \in A\}$ with $A \in \mathcal{E}$. By theorem III.2.17 in Revuz and Yor (1999), σ_A is an $(\mathcal{F}_t)$-stopping time.

The theorem below gives a condition in terms of the process Z (called (A)), such that there exists a stationary solution to (1.1). Moreover, it shows that under (A), the marginal distribution of this stationary solution is self-decomposable with canonical function determined by the Lévy measure of the underlying process Z.

Theorem 2.3 *Suppose Z is an increasing Lévy process with Lévy measure ρ (which is by definition the Lévy measure of $Z(1)$). Suppose ρ satisfies the integrability condition*

$$(A) \qquad \int_2^\infty \log x \rho(dx) < \infty\,,$$

then $P_t(x,\cdot)$ converges weakly to a limit distribution π as $t \to \infty$ for each $x \in E$ and each $\lambda > 0$. Moreover, π is self-decomposable with canonical function $k(x) = \rho(x,\infty)\mathbf{1}_{(0,\infty)}(x)$. Furthermore, π is the unique invariant probability distribution of X.

For a proof, see Sato (1999), theorem 17.5 and corollary 17.9. Theorem 24.10*(iii)* in Sato (1999) implies that π has support $[0,\infty)$.

We end this section with two examples of Lévy driven OU-processes. These examples are closely related to the ones given in examples (2.1)*(i)* and (2.1)*(iii)*.

Example 2.4 *(i)* Let $(X_t,\, t \geq 0)$ be the OU-process with $\pi = Gamma(c,\alpha)$. From the previous theorem and example 2.1*(i)* it follows that the BDLP $(Z_t,\, t \geq 0)$ has Lévy measure ρ satisfying $\rho(dx) = c\,\alpha e^{-\alpha x} dx$ (for $x > 0$). Since $\int_0^\infty \rho(dx) < \infty$, Z is a compound Poisson process. By examining the characteristic function of $Z(1)$, we see that the process Z can be represented as $Z_t = \sum_{i=1}^{N_t} Y_i$, where $(N_t,\, t \geq 0)$ is a Poisson process of intensity c, and $Y_1, Y_2, \ldots$ is a sequence of independent random variables, each having an exponential distribution with parameter α. Figure 1 corresponds to the case $c = \alpha = 2$.

(ii) Let $(X_t,\, t \geq 0)$ be the OU-process with $\pi = IG(\delta,\gamma)$. Similar as under *(i)* we obtain for the Lévy measure ρ of the BDLP Z the following expression

$$\rho(dx) = \left(\frac{\delta}{2\sqrt{2\pi}} \frac{1}{x\sqrt{x}} e^{-\gamma^2 x/2} + \frac{\delta\gamma^2}{2\sqrt{2\pi}} \frac{1}{\sqrt{x}} e^{-\gamma^2 x/2} \right) dx\,, \quad x > 0\,.$$

Write $\rho = \rho^{(1)} + \rho^{(2)}$. Then $(Z_t\,, t \geq 0)$ can be constructed as the sum of two independent Lévy processes $Z^{(1)}$ and $Z^{(2)}$, where $Z^{(i)}$ has Lévy measure $\rho^{(i)}$

$(i = 1, 2)$. It is easily seen that $Z^{(1)}(1) \sim IG(\delta/2, \gamma)$. Note that $\int_0^\infty \rho^{(2)}(dx) = \int_0^\infty \frac{\delta\gamma^2}{2\sqrt{2\pi}} \frac{1}{\sqrt{x}} e^{-\gamma^2 x/2} dx < \infty$, so that $Z^{(2)}$ is a compound Poisson process. Some calculations show that we can construct $Z^{(2)}$ as $Z_t^{(2)} = \frac{1}{\gamma^2} \sum_{i=1}^{N_t} W_i^2$, where $(N_t,\, t \geq 0)$ is a Poisson process of intensity $\delta\gamma/2$, and $W_1, W_2, \ldots$ is a sequence of independent standard normal random variables.

Since $\int_0^\infty \rho(dx) = \infty$, this OU-process is a process of infinite activity: it has infinitely many jumps in bounded time intervals.

3 A condition for the OU-process to be β-mixing

Let $(X_t, t \geq 0)$ be a stationary Lévy driven OU-process. The following theorem is the main result of this section.

Theorem 3.1 *If condition (A) of theorem 2.3 holds, then the Ornstein-Uhlenbeck process (X_t) is β-mixing.*

This result will be used in section 7 to obtain consistency proofs for some estimators, that will be defined in the next section. For the remainder of this section we will assume (A) holds. Theorem 2.3 then implies that there exists a unique invariant probability measure π_0.

By proposition 1 in Davydov (1973), the *β-mixing coefficients* for a stationary continuous-time Markov process X are given by

$$\beta_X(t) = \int_E \pi(dx) \|P_t(x, \cdot) - \pi(\cdot)\|_{TV}\,, \quad t > 0\,.$$

Here, $\|\cdot\|_{TV}$ denotes the total variation norm and π the initial distribution. The process is said to be β-mixing if $\beta_X(t) \to 0$, as $t \to \infty$. The analogous definitions for the discrete-time case are obvious. Dominated convergence implies that the following condition is sufficient for (X_t) to be β-mixing:

$$\lim_{t\to\infty} \|P_t(x, \cdot) - \pi(\cdot)\|_{TV} = 0\,, \quad \forall\, x \in E\,. \tag{3.1}$$

That is, it suffices to prove that the transition probabilities converge in total-variation to the invariant distribution for each initial state $x \in E$. The next theorem, taken from Meyn and Tweedie (1993a) (theorem 6.1), can be used to verify this condition.

Theorem 3.2 *Suppose that (X_t) is positive Harris recurrent with invariant probability distribution π. Then (3.1) holds if and only if some skeleton chain is φ-irreducible.*

In the remainder of this section we first prove that the 1-skeleton-chain, obtained from (X_t), is φ-irreducible (corollary 3.6). Subsequently we show that (X_t) is positive Harris recurrent (lemma 3.7). By an application of theorem 3.2, theorem 3.1 then follows immediately.

We start with some definitions from the general theory of stability of continuous time Markov processes. These correspond to the ones used in theorem 3.2. See for more details e.g. Meyn and Tweedie (1993a). Recall from section 2 that P_ν denotes the distribution of the OU-process with initial distribution ν. We write P_x in case ν is Dirac mass at x. For a measurable set A we let

$$\sigma_A = \inf\{t \geq 0 \mid X_t \in A\}, \qquad \eta_A = \int_0^\infty \mathbf{1}_{\{X_t \in A\}} dt\,.$$

Thus, σ_A denotes the first *hitting time* of the set A and η_A denotes the *time spent* in A by the process X. A Markov process is called *φ-irreducible* if for some nonzero σ-finite measure φ,

$$\varphi(A) > 0 \quad \Longrightarrow \quad E_x(\eta_A) > 0, \quad \forall x \in E\,,\ A \in \mathcal{E}\,.$$

The Markov process X is called *Harris recurrent* if for some nonzero σ-finite measure φ,

$$\varphi(A) > 0 \quad \Longrightarrow \quad P_x(\eta_A = \infty) = 1, \quad \forall x \in E\,,\ A \in \mathcal{E}\,.$$

If X is a Borel right Markov process, then this condition can be shown to be equivalent to (Kaspi and Mandelbaum (1994)): for some nonzero σ-finite measure ψ,

$$\psi(A) > 0 \quad \Longrightarrow \quad P_x(\sigma_A < \infty) = 1, \quad \forall x \in E\,,\ A \in \mathcal{E}\,. \tag{3.2}$$

The latter condition is generally more easily verifiable. The process is called *positive Harris recurrent* if it is Harris recurrent and admits an invariant *probability* measure.

The Δ-*skeleton* is defined as the Markov chain obtained by sampling the original process X_t at deterministic time points $\Delta, 2\Delta, \dots$ (the observation scheme *(i)* coincides with this concept). Abusing notation slightly, we shall from now on denote the continuous time process by (X_t) and its Δ-skeleton by (X_n) (thus, $X_n \equiv X_{n\Delta}$). The next proposition says that the 1-skeleton obtained from X constitutes a first order auto-regressive time series, with infinitely divisible noise terms.

Proposition 3.3 *Consider observation scheme (i) with $\Delta = 1$ and denote the observations by $X_0, X_1, \dots$. Then the chain satisfies the first order auto-regressive relation*

$$X_n = e^{-\lambda} X_{n-1} + W_n(\lambda)\,, \quad n \geq 1\,, \tag{3.3}$$

where $(W_n(\lambda))_n$ is an i.i.d. sequence of random variables distributed as

$$W_\lambda := \int_0^1 e^{\lambda(u-1)} dZ(\lambda u)\,.$$

Moreover, W_λ is infinitely divisible with Lévy measure κ given by

$$\kappa(B) = \int_B w^{-1} \rho(w, e^\lambda w] dw\,, \quad B \in \mathcal{E}\,. \tag{3.4}$$

The proof is given in the appendix.

Remark 3.4 Since

$$e^{-\lambda} Z(\lambda) \leq \int_0^1 e^{\lambda(u-1)} dZ(\lambda u) \leq Z(\lambda)\,,$$

W_λ has the same tail behaviour as $Z(\lambda)$. In particular, if $Z(1)$ has infinite expectation, so does W_λ.

Now we will show (X_n) is φ-irreducible. For the discrete time case this means that there exists a nonzero σ-finite measure φ, such that for all $B \in \mathcal{E}$ with $\varphi(B) > 0$, $\sum_{n=1}^\infty P_n(x, B) > 0$, for all $x \in E$.

Lemma 3.5 *Let P^{W_λ} be the distribution function of W_λ. Then P^{W_λ} has an absolutely continuous component w.r.t. Lebesgue measure.*

Proof It follows from proposition 3.3 that P^{W_λ} is infinitely divisible with Lévy measure κ. From (3.4), we see that κ is absolutely continuous with respect to Lebesgue measure.

First consider the case $\kappa[0,\infty) < \infty$. Then P^{W_λ} is compound Poisson, and hence (see equation 27.1 in Sato (1999)),

$$P^{W_\lambda}(\cdot) = e^{-\kappa[0,\infty)} \left(\delta_{\{0\}}(\cdot) + \sum_{k=1}^{\infty} \frac{\kappa^{*k}(\cdot)}{k!} \right), \tag{3.5}$$

where δ_0 means Dirac measure at 0 and $*$ denotes the convolution operator. Since the convolution of two non-zero finite measures σ_1 and σ_2 is absolutely continuous if either of them is absolutely continuous (Sato (1999), lemma 27.1), it follows from the absolute continuity of κ that the second term on the right-hand-side of (3.5) constitutes the absolutely continuous part of P^{W_λ}.

Next consider the case $\kappa[0,\infty) = \infty$. Define for each $n = 1, 2, \ldots$, $\kappa_n(B) := \kappa(B \cap (1/n, \infty)$ for Borel sets B in $(0,\infty)$. Set $c_n = \kappa_n[0,\infty)$. Then $c_n < \infty$ and κ_n is absolutely continuous. Let $P_n^{W_\lambda}$ be the distribution corresponding to κ_n. As in the previous case we have

$$P_n^{W_\lambda}(\cdot) = e^{-c_n} \left(\delta_{\{0\}}(\cdot) + \sum_{k=1}^{\infty} \frac{\kappa_n^{*k}(\cdot)}{k!} \right),$$

and $P_n^{W_\lambda}$ has an absolutely continuous component w.r.t. Lebesgue measure. Since P^{W_λ} contains $P_n^{W_\lambda}$ as a convolution factor, it follows that P^{W_λ} has an absolutely continuous component w.r.t. Lebesgue measure. $\square$

Proposition 6.3.5 in Meyn and Tweedie (1993) asserts that (X_n) as defined in (3.3) is φ-irreducible if the common distribution of the innovation-sequence $(W_n(\lambda))$ has an absolutely continuous component w.r.t. Lebesgue measure. Using the previous lemma, we therefore obtain φ-irreducibility of (X_n).

Corollary 3.6 *The 1-skeleton chain (X_n) is φ-irreducible.*

Lemma 3.7 *Under (A), (X_t) is positive Harris-recurrent.*

Proof Let $\sigma_a = \inf\{t \geq 0 \ : \ X_t = a\}$. We will prove $P_x(\sigma_a < \infty) = 1$, for all $x, a \in E$. Then condition (3.2) is satisfied for any nonzero measure ψ on E.

First, we consider the case $x \geq a$. Since we assume (A), lemma 9.1 from the appendix applies:

$$\int_0^1 \frac{dz}{z} \exp\left(- \int_z^1 \frac{\lambda_\rho(y)}{y} dy \right) = +\infty \, . \tag{3.6}$$

Here λ_ρ is given as in (9.4). Theorem 3.3 in Shiga (1990) now asserts that $P_x(\sigma_a < \infty) = 1$. for every $x \geq a > 0$.

Next, suppose $x < a$. Let (X_n) be the skeleton chain obtained from (X_t). Define $\tau_a = \inf\{n \geq 0 \ : \ X_n \geq a\}$, then for each $m \in \mathbb{N}$,

$$\begin{aligned}
P_x(\tau_a > m) &= P_x(X_1 < a, \ldots, X_m < a) \\
&= P_0(X_1 + e^{-\lambda} x < a, \ldots, X_m + e^{-\lambda m} x < a) \\
&\leq P_0(X_1 < a, \ldots, X_m < a) \\
&= P_0(W_1 < a, \ldots, e^{-\lambda} X_{m-1} + W_m < a) \\
&\leq \mathbb{P}(W_1 < a, \ldots, W_m < a) = [\mathbb{P}(W_\lambda < a)]^m \in [0,1) \, .
\end{aligned}$$

The last assertion holds since the support of any non-degenerate infinitely divisible random variable is unbounded (Sato (1999), theorem 24.3). From this, it follows that

$$P_x(\tau_a < \infty) \geq \lim_{m \to \infty} (1 - [\mathbb{P}(W_\lambda < a)]^m) = 1 \, .$$

It is easy to see that $\{\tau_a + \sigma_a \circ \theta_{\tau_a} < \infty\} \subseteq \{\sigma_a < \infty\}$. Hence,

$$\begin{aligned}
P_x(\sigma_a < \infty) &\geq P_x(\tau_a + \sigma_a \circ \theta_{\tau_a} < \infty) = E_x\{E_x(\mathbf{1}_{\{\tau_a + \sigma_a \circ \theta_{\tau_a} < \infty\}} \mid \mathcal{F}_{\tau_a})\} \\
&= E_x\{\mathbf{1}_{\{\tau_a < \infty\}} E_x(\mathbf{1}_{\{\sigma_a \circ \theta_{\tau_a} < \infty\}} \mid \mathcal{F}_{\tau_a})\} = E_x\{E_{X_{\tau_a}} \mathbf{1}_{\{\sigma_a < \infty\}}\} = 1 \, .
\end{aligned}$$

The second inequality holds since $\{\tau_a + \sigma_a \circ \theta_{\tau_a} < \infty\} = \{\tau_a < \infty\} \bigcap \{\sigma_a \circ \theta_{\tau_a} < \infty\}$. The third equality follows from the strong Markov property, as formulated in (2.6). The last equality follows from the case $x \geq a$.

Hence, for all $x \in E$, we have proved that $P_x(\sigma_a < \infty) = 1$. Thus (X_t) is Harris-recurrent.

By theorem 2.3, the invariant measure of a Lévy driven OU-process is a probability measure, which shows (X_t) is *positive* Harris-recurrent. $\square$

Remark 3.8 The β-mixing property of general (multi-dimensional) OU-processes is also treated in Masuda (2004), section 4. There it is assumed that the OU-process is strictly stationary, and moreover that $\int |x|^\alpha \pi(dx) < \infty$, for some $\alpha > 0$. The latter assumption is stronger than our assumption (A), but also yields the stronger conclusion that $\beta_X(t) = O(e^{-at})$, as $t \to \infty$, for some $a > 0$ (i.e. the process (X_t) is *geometrically ergodic*). It seems hard to extend the argument in Masuda (2004) under assumption (A).

4 Definition of a cumulant M-estimator

Let π_0 be the unique invariant probability distribution of X. Any reference to the true underlying distribution will be 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.

To estimate k_0, based on discrete time observations from X, we first define a preliminary estimator $\tilde{\psi}_n$ for ψ_0. In the sequel, we choose $\tilde{\psi}_n$ such that either

$$\text{for each } n,\ \tilde{\psi}_n \text{ is a ch.f. and } \forall\, t \in \mathbb{R} \quad \tilde{\psi}_n(t) \xrightarrow{\text{a.s.}} \psi_0(t)\,,\ \text{ as } n \to \infty\,, \tag{4.1}$$

or

$$\text{for each } n,\ \tilde{\psi}_n \text{ is a ch.f. and } \forall\, t \in \mathbb{R} \quad \tilde{\psi}_n(t) \xrightarrow{\text{p}} \psi_0(t)\,,\ \text{ as } n \to \infty\,. \tag{4.2}$$

We will show in section 5 that any preliminary estimator satisfying this condition will yield a consistent estimator for k_0. A natural preliminary estimator is the *empirical characteristic function* (e.c.f.). We will return to possible choices for $\tilde{\psi}_n$ in section 7.

Given any preliminary estimator $\tilde{\psi}_n$ for ψ_0, a first idea to construct an estimator for k_0 would be to minimize some distance between $Q(k)$ and $\tilde{\psi}_n$ over all canonical functions

$k \in K$. For example, if we let w be a positive (Lebesgue)-integrable compactly supported weight-function, we could take a weighted L^2 distance and define an estimator by

$$\hat{k}_n = \operatorname*{argmin}_{k \in K} \int |[Q(k)](t) - \tilde{\psi}_n(t)|^2 w(t) dt \,.$$

Apart from the issue whether this estimator is well defined, one disadvantage of this estimating method is that the objective function is non-convex (convexity being desirable from a computational point of view). This problem can be avoided by comparing cumulants. We will see below that $\tilde{\psi}_n$ is non-vanishing on S_w for sufficiently large n and thus admits a distinguished logarithm there. Then the idea is to define an estimator $\hat{k}_n$ as

$$\hat{k}_n = \operatorname*{argmin}_{k \in K} \int |[L(k)](t) - \tilde{g}_n(t)|^2 w(t) dt \,.$$

We call this estimator a *cumulant-M-estimator*. Next, we will make this idea more precise.

Let w be a non-negative integrable weight-function with compact support, denoted by S_w. Assume w is nonzero in a neighborhood of the origin and even. Define the space of square integrable functions w.r.t. $w(t)dt$ by

$$L^2(w) := \left\{ f : \mathbb{R} \to \mathbb{C} \;\middle|\; f \text{ is (Lebesgue) measurable and } \int |f(t)|^2 w(t) dt < \infty \right\},$$

where we identify functions which are a.e. equal with respect to $w(t)dt$. We define an inner-product $\langle \cdot, \cdot \rangle_w$ on $L^2(w)$ by

$$\langle f, g \rangle_w = \Re \int f(t) \overline{g(t)} w(t) dt \,,$$

where the bar over g denotes complex conjugation and $\Re$ the operation of taking the real part of an element of $\mathbb{C}$. For $g \in L^2(w)$ define a norm by $\|g\|_w = \sqrt{\langle g, g \rangle_w}$. The space $(L^2(w), \langle \cdot, \cdot \rangle_w)$ is a Hilbert-space. For the remaining part of the paper, we assume n is large enough such that $\tilde{g}_n$ exists on S_w.

Next, we define an estimator for $g_0 = T(\psi_0)$ as the minimizer of

$$\Gamma_n(g) := \|g - T\tilde{\psi}_n\|_w^2 = \int |g(t) - T\tilde{\psi}_n(t)|^2 w(t) dt$$

over an appropriate subset of G, which we consider as a subspace of $L^2(w)$. It is a standard fact from Hilbert-space theory that every non-empty, closed, convex set in $L^2(w)$ contains a unique element of smallest norm. We will use this result to establish existence and uniqueness of our estimator.

Since Γ_n is a squared norm in a Hilbert space, we only need to specify an appropriate subset of G. For this purpose, we first derive some properties of the mapping L, as defined in section 2.

Lemma 4.1 *The mapping $L : K \to G$ is continuous, onto and one-to-one.*

Proof Let $\{k_n\}$ be a sequence in K converging to $k_0 \in K$, i.e. $\|k_n - k_0\|_\mu \to 0$, as $n \to \infty$.

For $t \in S_w$,

$$\begin{aligned} |L(k_n)(t) - L(k_0)(t)| &= \Big|\int_0^\infty (e^{itx} - 1)\frac{k_n(x) - k_0(x)}{x} dx\Big| \\ &\le |t| \int_0^1 |k_n(x) - k_0(x)| dx + 2\int_1^\infty x^{-1}|k_n(x) - k_0(x)| dx \\ &\le \max\{|t|, 2\}\|k_n - k_0\|_\mu \,, \end{aligned}$$

where we use the inequality $|e^{ix} - 1| \le \min\{|x|, 2\}$. Thus $L(k_n) \to L(k_0)$ uniformly on S_w which implies $\|L(k_n) - L(k_0)\|_w \to 0$ $(n \to \infty)$. Hence, L is continuous.

The surjectivity is trivial by the definition of G. If $g_1, g_2 \in G$ and $\|g_1 - g_2\|_w = 0$, then (by continuity of elements in G), $g_1 = g_2$ on S_w. Then also $\psi_1 := e^{g_1} = e^{g_2} := \psi_2$ on S_w. Lemma 4.2 below implies $\psi_1 = \psi_2$ on $\mathbb{R}$. Since Q is one-to-one we must have $k_1 = k_2$. □

The following lemma extends the uniqueness theorem for characteristic functions. A proof can be found in Loéve (1977), chapter 4.

Lemma 4.2 *Let X be a positive random variable with characteristic function ψ. If ψ_M is the restriction of ψ to an interval $(-M, M)$, then ψ_M determines ψ.*

The set

$$G' := \big\{g : \mathbb{R} \to \mathbb{C} \;:\; g(t) = \beta_0 it + \int_0^\infty \frac{e^{itx} - 1}{x} k(x) dx,\ \beta_0 \ge 0, k \in K\big\},$$

is closed under uniform convergence on compact sets containing the origin. To see this: let S be such a compact set. If $\{g_n\}_n \in G'$ is such that $\sup_{t\in S} |g_n(t) - g(t)| \to 0$ for some g, then $\sup_{t\in S} |\psi_n(t) - \psi(t)| \to 0$ and then (by the same argument as in the proof of Lévy's continuity theorem) the random variables corresponding to $\{\psi_n\}$ are uniformly tight. Denote these r.v. by $\{X_n\}$. By Prohorov's theorem, there exists a subsequence n_l such that X_{n_l} converges weakly to a random variable X^*. Since X_n is a positive SD random variable, and the class of positive SD random variables is closed under weak convergence, X^* is positive SD. Let g^* be the cumulant of X^*. Then $g^* \in G'$ and $\sup_{t\in S} |g_{n_l}(t) - g^*(t)| \to 0$. Together with the continuity of g and g^* on S, this implies $g^* = g$ on S. Hence $g = g^* \in G'$.

However, the set G is not closed under uniform convergence on compact sets containing the origin. Let S again be such a set and define a sequence $\{k_n\}_{n\geq 1} \in K$ by $k_n(x) = n\mathbf{1}_{[0,1/n)}(x)$, then for each $t \in \mathbb{R}$,

$$g_n(t) = [L(k_n)](t) = n\int_0^{1/n} \frac{e^{itx} - 1}{x} dx \to it\,, \quad \text{as } n \to \infty\,.$$

Let $g(t) = it$, then since each g_n and g are uniformly continuous on the compact set S, we have $\sup_{t\in S} |g_n(t) - g(t)| \to 0$. However $g \notin G$, since g can only correspond to a point mass at one. Returning to the set G', we see that this example is the canonical example that can preclude closedness of G.

In view of dominated convergence, this counter example also shows why the set G is not closed in $L^2(w)$. To obtain an appropriate *closed* subset of G, we first define a set of envelope functions in K. Pick for each $R > 0$ a function $k_R \in K$ such that $\|k_R\|_\mu \leq R$ (for example $k_R(x) = R/(4\sqrt{x})$). The collection $\{k_R,\, R > 0\}$ defines a set of envelope functions. Now let

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

and put $G_R = L(K_R)$, i.e. G_R is the image of K_R under L.

Lemma 4.3 *Let $R > 0$,*

(i) K_R is a compact, convex subset of $L^1(\mu)$.

(ii) G_R is a compact, convex subset of $L^2(w)$.

Proof (i): Convexity of K_R is obvious.

Let $\{k_n\}$ be a sequence in K_R. Since each k_n is bounded on all strictly positive rational points, we can use a diagonalization argument to extract a subsequence n_j from n such that the sequence k_{n_j} converges to some function $\bar{k}$ on all strictly positive rationals. For $x \in (0, \infty)$ define

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

This function is (by its definition) decreasing and right-continuous and satisfies $\tilde{k} \leq k_R$ on $(0, \infty)$. Thus $\tilde{k} \in K_R$. Furthermore, k_{n_j} converges pointwise to $\tilde{k}$ at all continuity points of $\tilde{k}$. Since the number of discontinuity points of $\tilde{k}$ is at most countable, k_{n_j} converges to $\tilde{k}$ a.e. on $(0, \infty)$. Moreover, since $k_{n_j} \leq k_R$ on $(0, \infty)$ and $k_R \in L^1(\mu)$, Lebesgue's dominated convergence theorem applies: $\|k_{n_j} - \tilde{k}\|_\mu \to 0$, as $n_j \to \infty$. Hence, K_R is sequentially compact.

(ii): G_R is compact since it is the image of the compact set K_R under the continuous mapping L. Convexity of G_R follows from convexity of K_R. □

Corollary 4.4 *The inverse operator of L, $L^{-1} : G_R \to K_R$ is continuous.*

Proof This is a standard result from topology, see e.g. corollary 9.12 in Jameson (1974). □

Since we want to define our objective function in terms of the canonical function, one last step has to be made. Since Γ_n has a unique minimizer over G_R and to each G_R belongs a unique member of K_R, there exists a unique minimizer of $\Gamma_n \circ L$ (which we will from now on write as $\Gamma_n L$) over K_R. More precisely:

Theorem 4.5 *Let $\hat{g}_n = \operatorname{argmin}_{g \in G_R} \Gamma_n(g)$ (which is by now known to exist uniquely). Then $\hat{k}_n = \operatorname{argmin}_{k \in K_R} [\Gamma_n L](k)$ exists. Moreover, $\hat{k}_n = L^{-1}(\hat{g}_n)$ and $\hat{k}_n$ is unique.*

Proof Since $L : K_R \to G_R$ is onto and one-to-one, to each $g \in G_R$ corresponds a unique $k \in K_R$ such that $L(k) = g$. Thus

$$\beta := \min_{g \in G_R} \Gamma_n(g) = \min_{k \in K_R} [\Gamma_n L](k)\,.$$

Now define $\hat{k}_n = L^{-1}(\hat{g}_n)$ and choose $k \in K_R$ arbitrary (but $k \neq \hat{k}_n$). Then $\hat{k}_n \in K_R$ and

$$[\Gamma_n L](\hat{k}_n) = \Gamma_n(\hat{g}_n) = \beta < [\Gamma_n L](k)\,,$$

which shows that $\hat{k}_n$ is the unique minimizer of $\Gamma_n L$ over K_R. $\square$

5 Consistency

In this section we discuss consistency of the cumulant M-estimator. We start with two results, which strengthen the pointwise convergence in (4.1) and (4.2) to uniform convergence.

Lemma 5.1 *Let* $(\Omega, \mathcal{U}, \mathbb{P})$ *be a probability space. For each* $\omega \in \Omega$, *suppose* $\varphi_n(\cdot, \omega)$ *(*$n = 1, 2, \ldots$*) and* φ *are characteristic functions such that for each* $t \in \mathbb{R}$, $\varphi_n(t, \cdot) \xrightarrow{\text{p}} \varphi(t)$, *as* $n \to \infty$. *Then, as* $n \to \infty$

$$\sup_{t \in K} |\varphi_n(t, \cdot) - \varphi(t)| \xrightarrow{\text{p}} 0\,, \quad \textit{for every compact set } K \subseteq \mathbb{R}\,.$$

Proof Denote the distribution functions corresponding to $\varphi_n(\cdot, \omega)$ and φ by $F_n(\cdot, \omega)$ and F. The functions $x \mapsto e^{itx}$ for $t \in K$ are uniformly bounded and equicontinuous. Therefore (by the Arzelà-Ascoli theorem), if $F_n(\cdot, \omega) \xrightarrow{\text{w}} F$ for some ω along some subsequence, then $\sup_{t \in K} |\varphi_n(t, \cdot) - \varphi(t)| \to 0$ for this ω and subsequence. It follows that it suffices to show that for every subsequence of $\{n\}$ there exists a further subsequence $\{n'\}$ and a set $A \in \mathcal{U}$ with $\mathbb{P}(A) = 1$ such that $F_{n'}(\cdot, \omega) \xrightarrow{\text{w}} F$, $\forall\, \omega \in A$, along the subsequence.

By assumption: for every t there exists a subsequence $\{n\}$ such that $\varphi_n(t, \omega) \xrightarrow{\text{a.s.}} \varphi(t)$. Denote $\mathbb{Q} = \{q_1, q_2, \ldots\}$. There exists a subsequence $\{n^{(1)}\}$ of $\{n\}$ and a set

$A^{(1)} \in \mathcal{U}$ with $\mathbb{P}(A^{(1)}) = 1$ such that $\varphi_{n^{(1)}}(q_1, \omega) \to \varphi(q_1)$, for all $\omega \in A^{(1)}$. There exists a subsequence $\{n^{(2)}\}$ of $\{n^{(1)}\}$ and a set $A^{(2)} \in \mathcal{U}$ with $\mathbb{P}(A^{(2)}) = 1$ such that $\varphi_{n^{(2)}}(q_2, \omega) \to \varphi(q_2)$, for all $\omega \in A^{(2)}$. Proceed iteratively in this way. Consider the diagonal sequence, obtained by $n_i := n_i^{(i)}$, and set $A = \cap_{i=1}^{\infty} A^{(i)}$, then $\mathbb{P}(A) = 1$ and

$$\varphi_n(q, \omega) \to \varphi(q), \quad \forall\, q \in \mathbb{Q}, \quad \forall \omega \in A. \tag{5.1}$$

For every $\delta > 0$,

$$\int_{|x|>2/\delta} F_n(dx, \omega) \leq \frac{1}{2\delta} \int_{-\delta}^{\delta} |1 - \varphi_n(t, \omega)| dt =: a_n(\delta, \omega), \tag{5.2}$$

by a well-known inequality (see for instance Chung (2001), chapter 6.3). Furthermore, with $a(\delta) := \frac{1}{2\delta} \int_{-\delta}^{\delta} |1 - \varphi(t)| dt$, by Fubini's theorem

$$\mathbb{E}|a_n(\delta, \cdot) - a(\delta)| \leq \frac{1}{2\delta} \int_{-\delta}^{\delta} \mathbb{E}\big||1 - \varphi_n(t)| - |1 - \varphi(t)|\big| dt \to 0, \quad n \to \infty,$$

by dominated convergence and the assumed convergence in probability. Thus, for every $\delta > 0$ there exists a further subsequence $\{n\}$ and a set $B \in \mathcal{U}$ with $\mathbb{P}(B) = 1$ such that $a_n(\delta, \omega) \to a(\delta)$ for all $\omega \in B$. By a diagonalization scheme we can find a further subsequence $\{n\}$ and a set $C \in \mathcal{U}$ with $\mathbb{P}(C) = 1$ such that

$$\lim_{n\to\infty} |a_n(\delta, \omega) - a(\delta)| = 0, \quad \forall\, \delta \in \mathbb{Q} \cap (0, \infty), \quad \forall\, \omega \in C.$$

Combined with (5.2) this shows that

$$\limsup_{n\to\infty} \int_{\{|x|>2/\delta\}} F_n(dx, \omega) \leq a(\delta), \quad \forall\, \delta \in \mathbb{Q} \cap (0, \infty), \quad \forall\, \omega \in C,$$

the limsup taken over the subsequence. Because $a(\delta) \downarrow 0$ as $\delta \downarrow 0$ we see that $\{F_n(\cdot, \omega)\}_{n=1}^{\infty}$ is tight for all $\omega \in C$.

If G is a limit point of $F_n(\cdot, \omega)$, then by (5.1)

$$\int e^{itx} dG(x) = \lim_{n\to\infty} \int e^{itx} F_n(dx, \omega) = \int e^{itx} dF(x), \quad \forall\, t \in \mathbb{Q}, \quad \forall\, \omega \in A.$$

Hence $F = G$, and it follows that $\{F_n(\cdot, \omega)\}_n$ has only one limit point, whence $F_n(\cdot, \omega) \xrightarrow{\mathrm{w}} F$, for all $\omega \in A \cap C$, along the subsequence. $\square$

Lemma 5.2 *Let $(\Omega, \mathcal{U}, \mathbb{P})$ be a probability space. For each $\omega \in \Omega$, suppose $\varphi_n(\cdot, \omega)$ ($n = 1, 2, \ldots$) and φ are characteristic functions such that for each $t \in \mathbb{R}$, $\varphi_n(t, \cdot) \xrightarrow{\text{a.s.}} \varphi(t)$, as $n \to \infty$. Then, as $n \to \infty$*

$$\sup_{t \in K} |\varphi_n(t, \cdot) - \varphi(t)| \xrightarrow{\text{a.s.}} 0\,, \quad \textit{for every compact set } K \subseteq \mathbb{R}\,.$$

Proof It suffices to show that there exists an $A \in \mathcal{U}$ with $\mathbb{P}(A) = 1$ such that $F_n(\cdot, \omega) \xrightarrow{\text{w}} F$, for all $\omega \in A$.

With a_n and a as in the proof of the previous lemma

$$\mathbb{E} \sup_{m \geq n} |a_m(\delta, \cdot) - a(\delta)| \leq \frac{1}{2\delta} \int_{-\delta}^{\delta} \mathbb{E} \left(\sup_{m \geq n} \Big| |1 - \varphi_m(t, \cdot)| - |1 - \varphi(t)| \Big| \right) dt \to 0\,, \quad n \to \infty\,.$$

This implies that $|a_n(\delta, \cdot) - a(\delta)| \xrightarrow{\text{a.s.}} 0$ ($n \to \infty$), for all $\delta > 0$. Combined with (5.2) we see

$$\limsup_{n \to \infty} \int_{\{|x| > 2/\delta\}} F_n(dx, \omega) \leq a(\delta)\,, \quad \forall \delta \in \mathbb{Q}\,, \quad \omega \in A_1\,,$$

for some set $A_1 \in \mathcal{U}$ with $\mathbb{P}(A_1) = 1$. Thus, for $\omega \in A_1$ the whole sequence $\{F_n(\cdot, \omega)\}$ is tight.

Let $A_2 \in \mathcal{U}$ be a set of probability one such that $\varphi_n(t, \omega) \to \varphi(t)$, $\forall t \in \mathbb{Q}$ and $\forall \omega \in A_2$. Let $\omega \in A_2$, then (as in the end of the proof of lemma 5.1), $F_n(\cdot, \omega)$ has only F as a limit point.

Hence for all $\omega \in A := A_1 \cap A_2$, $F_n(\cdot, \omega) \xrightarrow{\text{w}} F$. □

Remark 5.3 In case φ_n is the empirical characteristic function of independent random variables with common distribution F, there is a large literature on results as in lemma 5.2. We mention the final result by Csörgö (1983): if $\lim_{n \to \infty} (\log T_n)/n = 0$, then $\sup_{|t| \leq T_n} |\varphi_n(t) - \varphi(t)| \xrightarrow{\text{a.s.}} 0$, as $n \to \infty$, where, $\varphi(t) = \int e^{itx} F(dx)$, ($t \in \mathbb{R}$). The rate $T_n = \exp(o(n))$ is the best possible in general for almost sure convergence.

Now we come to the consistency result, which will be applied in section 7.

Theorem 5.4 *Assume the sequence of preliminary estimators $\tilde{\psi}_n$ satisfies (4.1). If $k_0 \in K_R$ for some $R > 0$, then the cumulant-M-estimator is consistent. That is,*

(i)

$$\|\hat{g}_n - g_0\|_w \to 0 \quad \textit{almost surely, as } n \to \infty\,. \tag{5.3}$$

(ii)

$$\|\hat{k}_n - k_0\|_\mu \to 0 \quad \textit{almost surely, as } n \to \infty\,. \tag{5.4}$$

The same results hold in probability, if we only assume (4.2).

Proof We first prove the statement in case $\tilde{\psi}_n$ converges almost surely to ψ_0.

By lemma 5.2, $\sup_{t\in S_w} |\tilde{\psi}_n(t,\cdot) - \psi_0(t)| \xrightarrow{\text{a.s.}} 0$. Let $A \subseteq \Omega$ be the set of probability one on which the convergence occurs. Fix $\omega \in A$. Since ψ_0 has no zeros, there exists an $\varepsilon > 0$ such that $\inf_{t\in S_w} |\psi_0(t)| > 2\varepsilon$. For this ε there exists an $N = N(\varepsilon, \omega) \in \mathbb{N}$ such that $\sup_{t\in S_w} |\tilde{\psi}_n(t,\omega) - \psi_0(t)| < \varepsilon$ for all $n \geq N$. Hence for all $n \geq N$ and for all $t \in S_w$, $|\tilde{\psi}_n(t,\omega)| \geq |\psi_0(t)| - |\tilde{\psi}(t,\omega) - \psi_0(t)| \geq \varepsilon > 0$.

For $n \geq N$ we can define $\tilde{g}_n(\omega) = T\tilde{\psi}_n(\omega)$ on S_w. Theorem 7.6.3 in Chung (2001) implies that 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. By dominated convergence $\lim_{n\to\infty}\|\tilde{g}_n(\omega) - g_0\|_w = 0$.

Since $\hat{g}_n(\cdot,\omega)$ minimizes Γ_n over G_R, we have

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

as n tends to infinity. By corollary 4.4 this implies

$$\|\hat{k}_n(\cdot,\omega) - k_0\|_\mu = \|L^{-1}(\hat{g}_n(\cdot,\omega)) - L^{-1}(g_0)\|_\mu \to 0, \quad n \to \infty\,.$$

Hence, for all ω in a set with probability one, $\lim_{n\to\infty}\|\hat{k}_n(\cdot,\omega) - k_0\|_\mu \to 0$.

Next, we prove the corresponding statement for convergence in probability. By lemma 5.1 in the appendix $Y_n := \sup_{t\in S_w} |\tilde{\psi}_n(t,\cdot) - \psi_0(t)| \xrightarrow{\text{p}} 0$, as $n \to \infty$.

The following characterization of convergence in probability holds: $Y_n \xrightarrow{\text{p}} Y$ if and only if each subsequence of (Y_n) possesses a further subsequence that converges almost surely to Y.

Let (n_k) be an arbitrary increasing sequence of natural numbers, then $Y_{n_k} \xrightarrow{\text{p}} 0$. Then there exists a subsequence (n_m) of (n_k) such that $Y_{n_m} \xrightarrow{\text{a.s.}} 0$. Now we can apply the statement of the theorem for almost sure convergence, this gives $\|\hat{k}_{n_m} - k_0\|_\mu \xrightarrow{\text{a.s.}} 0$. This in turn shows that $\|\hat{k}_n - k_0\|_\mu \xrightarrow{\text{p}} 0$. □

Corollary 5.5 *Assume (4.1). Denote the distribution function corresponding to $\hat{\psi}_n(\cdot, \omega)$ by $\hat{F}_n(\cdot, \omega)$. Then for all ω in a set of probability one,*

$$\|\hat{F}_n(\cdot, \omega) - F_0(\cdot)\|_\infty \to 0\,, \quad n \to \infty\,.$$

Here $\|\cdot\|_\infty$ denotes the supremum norm. If we only assume (4.2), then

$$\|\hat{F}_n - F_0\|_\infty \xrightarrow{\text{p}} 0\,, \quad n \to \infty\,.$$

Proof First assume (4.1). Theorem 5.4 implies $\|\hat{k}_n - k_0\|_\mu \xrightarrow{\text{a.s.}} 0$, as $n \to \infty$. Fix an arbitrary ω of the set on which the convergence takes place. From the proof of lemma 4.1, we obtain that $\hat{g}_n(\cdot, \omega)$ converges uniformly on compacta to g_0. Then also $\hat{\psi}_n(\cdot, \omega)$ converges uniformly on compacta to ψ_0. By the continuity theorem (Chung (2001), section 6.3), $\hat{F}_n(\cdot, \omega) \xrightarrow{\text{w}} F_0(\cdot)$. Since F_0 is continuous, this is equivalent to $\|\hat{F}_n(\cdot, \omega) - F_0(\cdot)\|_\infty \to 0$, as $n \to \infty$.

The statement for convergence in probability follows by arguing along subsequences, as in the proof of theorem 5.4. □

Theorem 5.4 involves only functional analytic properties of various operators and sets. To fulfill the probabilistic assumption that the sequence of preliminary estimators satisfies a law of large numbers, we can use the β-mixing result from section 3.

6 Computing the cumulant M-estimator

For numerical purposes we will approximate the convex cone K by a finite-dimensional subset. For $N \geq 1$, let $0 < \theta_1 < \theta_2 < \ldots < \theta_N$ be a fixed set of positive numbers and set

$\Theta = \{\theta_1, \ldots, \theta_N\}$. For example, we can take an equidistant grid with grid points $\theta_j = jh$ $(1 \le j \le N)$, where h is the mesh-width. Define "basis functions" by

$$u_\theta(x) := \mathbf{1}_{[0,\theta)}(x), \quad x \ge 0,$$
$$z_\theta(t) = [Lu_\theta](t) = \int_0^{\theta t} \frac{e^{iu} - 1}{u} du, \quad t \in \mathbb{R},$$

and set $\mathcal{U}_\Theta := \{u_\theta, \, \theta \in \Theta\}$. Let K_Θ be the convex cone generated by $\mathcal{U}_\Theta$, i.e.

$$K_\Theta := \left\{ k \in K \;\middle|\; k = \sum_{i=1}^N \alpha_i u_{\theta_i}, \; \alpha_i \in [0, \infty), \; 1 \le i \le N \right\}.$$

Define a *sieved estimator* by

$$\check{k}_n = \operatorname*{argmin}_{k \in K_\Theta} \Gamma_n L(k) = \operatorname*{argmin}_{\alpha_1 \ge 0, \ldots, \alpha_N \ge 0} \Big\| \sum_{i=1}^N \alpha_i z_{\theta_i} - \tilde{g}_n \Big\|_w^2 . \tag{6.1}$$

Since the set $\{\underline{x} \, : \, \underline{x} = (x_1, \ldots, x_N), \, x_i \ge 0 \text{ for all } 1 \le i \le N\}$ is a closed convex subset of $\mathbb{R}^N$ and $\Gamma_n L$ is a continuous mapping, we have

Theorem 6.1 *The sieved estimator $\check{k}_n$ is uniquely defined.*

Note that in this case we do not need conditions in terms of envelope functions, as in section 4.

Next, we study the problem of computing $\check{k}_n$ numerically. Since each $k \in K_\Theta$ is a finite positive mixture of basis-functions $u_\theta \in \mathcal{U}_\Theta$, our minimization problem fits precisely in the setup of Groeneboom et al. (2003). We will follow the approach adopted there to solve (6.1).

6.1 The support reduction algorithm

Define the directional derivative of $\Gamma_n L$ at $k_1 \in K$ in the direction of $k_2 \in K$ by

$$D_{\Gamma_n L}(k_2; k_1) := \lim_{\varepsilon \to 0} \varepsilon^{-1} ([\Gamma_n L](k_1 + \varepsilon k_2) - [\Gamma_n L](k_1)) .$$

This quantity exists (possibly infinite), since $\Gamma_n L$ is a convex functional on K (Γ_n, as an L^2-distance on a Hilbert-space, is a strictly convex functional on G, and L satisfies $L(k_1 + k_2) = L(k_1) + L(k_2)$).

In Groeneboom et al. (2003) it is shown that under conditions that are satisfied here the following characterization of $\check{k}_n$ holds: Write $\check{k}_n = \sum_{j \in J} \alpha_j u_{\theta_j}$, where $J := \{j \in \{1, \ldots, N\},\, \alpha_j > 0\}$. then

$$\check{k}_n \text{ minimizes } \Gamma_n L \text{ over } K_\Theta \quad \iff \quad D_{\Gamma_n L}(u_{\theta_j}; \check{k}_n) \begin{cases} \geq 0 & \forall\, j \in \{1, \ldots, N\} \\ = 0 & \forall\, j \in J \end{cases} . \tag{6.2}$$

This result forms the basis for the *support-reduction algorithm*, which is an iterative algorithm for solving (6.1). We discuss this algorithm briefly. For additional details we refer to section 3 of Groeneboom et al. (2003).

Suppose at each iteration we are given a "current iterate" $k^J \in K_\Theta$ which can be written as

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

(J refers to the index set of positive α-weights). Relation (6.2) gives a criterion to check whether k^J is optimal. As we will shortly see, each iterate k^J will satisfy the equality part of (6.2): $D_{\Gamma_n L}(u_{\theta_j}; k^J) = 0$, for all $j \in J$. This fact, together with (6.2) implies that if k^J is not optimal, then there is an $i \in \{1, \ldots, N\} \setminus J$ with $D_{\Gamma_n L}(u_{\theta_i}, k^J) < 0$. Thus u_{θ_i} provides a direction of descent for $\Gamma_n L$. In that case the algorithm prescribes two steps that have to be carried out:

Step (i). Determine a direction of descent for $\Gamma_n L$. Let

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

then $\Theta_<$ is non-empty. From $\Theta_<$ we choose a direction of descent. Suppose θ_{j^*} is this direction. (A particular choice is the direction of steepest descent, in which case $\theta_{j^*} := \operatorname{argmin}_{\theta \in \Theta_<} D_{\Gamma_n L}(u_\theta, k^J)$. This boils down to finding a minimum element in a vector of length at most N. Below, we give an alternative choice.)

Step (ii). Let the new iterate be given by

$$k^{J^*} = \sum_{j \in J^*} \beta_j u_{\theta_j}\,, \quad J^* := J \cup \{j^*\}\,,$$

where $\{\beta_j\,, j \in J^*\}$ are (yet unknown) weights. The second step consists of first minimizing $\Gamma_n L(k^{J^*})$ with respect to $\{\beta_j\,, j \in J^*\}$, without positivity constraints. In our situation this is a (usually low-dimensional) quadratic unconstrained optimization problem.

If $\min\{\beta_j\;, j \in J^*\} \geq 0$, then $k^{J^*} \in K_\Theta$ and k^{J^*} satisfies the equality part of (6.2). In that case, we check the inequality part of (6.2) and possibly return to step *(i)*. Else, we perform a support-reduction step. Since it can be shown that always $\beta_{j^*} > 0$, we can make a move from k^J towards k^{J^*} and stay within the cone K_Θ initially. As a next iterate, we take $k := k^J + \hat{c}(k^{J^*} - k^J)$, where

$$\begin{aligned}
\hat{c} &= \max\{c \in [0,1]\;:\; k^J + c(k^{J^*} - k^J) \in K_\Theta\} \\
&= \max\{c \in [0,1]\;:\; \sum_{j \in J} [c\beta_j + (1-c)\alpha_j]\, u_{\theta_j} + c\beta_{j^*} u_{\theta_{j^*}} \in K_\Theta\} \\
&= \max\{c \in [0,1]\;:\; c\beta_j + (1-c)\alpha_j \geq 0\,,\ \text{for all } \beta_j\ (j \in J) \text{ with } \beta_j < 0\} \\
&= \min\{\alpha_j/(\alpha_j - \beta_j)\;, j \in J \text{ for which } \beta_j < 0\}\,. \qquad (6.3)
\end{aligned}$$

Then $k \in K_\Theta$. Let j^{**} be the index for which the minimum in (6.3) is attained, i.e. for which $\hat{c}\beta_{j^{**}} + (1-\hat{c})\alpha_{j^{**}} = 0$. Define $J^{**} := J^* \setminus \{j^{**}\}$, then k is supported on $\{\theta_j\,, j \in J^{**}\}$. That is, in the new iterate, the support point $\theta_{j^{**}}$ is removed. Next, set $k^{J^{**}} = \sum_{j \in J^{**}} \gamma_j u_{\theta_j}$ and compute optimal weights γ_j. If all weights γ_j are non-negative, the equality part of (6.2) is satisfied and we can check the inequality part of (6.2) and possibly return to step *(i)*. Else, a new support-reduction step can be performed, since all weights of k are positive. In the end, our iterate k will satisfy the equality part of (6.2).

To start the algorithm, we fix a starting value $\theta^{(0)} \in \Theta$. Then we determine the function $cu_{\theta^{(0)}}$ minimizing $\Gamma_n L$ as a function of $c > 0$. Once the algorithm has been initialized it starts iteratively adding and removing support points, while in between computing optimal weights.

In theorem 3.1 in Groeneboom et al. (2003) conditions are given to guarantee that the sequence of iterates $\{k^{(i)}\}_i$ (generated by the support-reduction algorithm) indeed converges to the solution of our minimization problem. Since these conditions are met in our case, we have

$$(\Gamma_n L)(k^{(i)}) \downarrow (\Gamma_n L)(\check{k}_n)\,, \quad \text{as } i \to \infty\,.$$

6.2 Implementation details

We now work out the actual computations involved, when implementing the algorithm. Suppose $k = \sum_{j=1}^m \alpha_j u_{\theta_j}$.

Step (i). Given the "current iterate" k, we aim to add a function u_θ which provides a direction of descent for $\Gamma_n L$. By linearity of L,

$$\begin{aligned}[\Gamma_n L](k+\varepsilon u_\theta) - [\Gamma_n L](k) &= \|L(k+\varepsilon u_\theta) - \tilde{g}_n\|_w^2 - \|Lk - \tilde{g}_n\|_w^2 \\ &= \varepsilon c_1(\theta, k) + \frac{1}{2}\varepsilon^2 c_2(\theta)\,, \end{aligned} \tag{6.4}$$

where $c_2(\theta) = 2\|Lu_\theta\|_w^2 = 2\|z_\theta\|_w^2 > 0$ and

$$c_1(\theta, k) = 2\langle Lk - \tilde{g}_n, Lu_\theta\rangle_w = 2\Big\langle \sum_{j=1}^m \alpha_j z_{\theta_j} - \tilde{g}_n, z_\theta \Big\rangle_w.$$

In order to find a direction of descent, we can pick any $\theta \in \Theta$ for which $c_1(\theta, k) < 0$. However, since the right-hand side of (6.4) is quadratic in ε, it can be minimized explicitly (and we choose to do so). If $c_1(\theta, k) < 0$, then

$$\operatorname*{argmin}_{\varepsilon > 0}(\varepsilon c_1(\theta, k) + \varepsilon^2 c_2(\theta)) = -\frac{c_1(\theta, k)}{c_2(\theta)} =: \hat{\varepsilon}_\theta\,.$$

Minimizing $[\Gamma_n L](k + \hat{\varepsilon}_\theta u_\theta)$ over all points $\theta \in \Theta$ with $c_1(\theta, k) < 0$ gives

$$\hat{\theta} = \operatorname*{argmin}_{\{\theta\in\Theta\,:\,c_1(\theta,k)<0\}} -\frac{c_1(\theta,k)^2}{2c_2(\theta)} = \operatorname*{argmin}_{\theta\in\Theta} \frac{c_1(\theta,k)}{\sqrt{c_2(\theta)}}\,.$$

Step (ii). Given a set of support points, we compute optimal weights. This is a standard least squares problem, that is solved by the normal equations. In our setup, these are obtained by differentiating $[\Gamma_n L](k)$ with respect to α_j ($j \in \{1, \ldots, m\}$) and setting the partial derivatives equal to zero. This gives the system $A\underline{\alpha} = \underline{b}$, where

$$A_{i,j} = \langle z_{\theta_i}, z_{\theta_j}\rangle_w, \quad i,\, j = 1, \ldots, m\,, \tag{6.5}$$

and

$$b_i = \langle z_{\theta_i}, \tilde{g}_n \rangle_w, \quad i = 1, \dots, m\,.$$

The matrix A is easily seen to be symmetric. By the next lemma, A is non-singular, whence the system $A\underline{\alpha} = \underline{b}$ has a unique solution.

Lemma 6.2 *The matrix A, as defined in (6.5) is non-singular.*

Proof Denote by $\underline{a}_j$ the j-th column of A. Let $h_1, \dots, h_m \in \mathbb{R}$. We aim to show that if $\sum_{i=1}^m h_i \underline{a}_i = \underline{0}$, then all h_j are zero. Now $\sum_{i=1}^m h_i \underline{a}_i = (\langle z_{\theta_1}, \varphi \rangle_w, \dots, \langle z_{\theta_m}, \varphi \rangle_w)^T$, where $\varphi \in L^2(w)$ is given by $\varphi := \sum_{i=1}^m h_j z_{\theta_j}$. Thus if $\sum_{i=1}^m h_i \underline{a}_i = \underline{0}$, then $\varphi \perp$ $span(z_{\theta_1}, \dots, z_{\theta_m})$ in $L^2(w)$. Since $\varphi \in span(z_{\theta_1}, \dots, z_{\theta_m})$, we must have $\varphi = 0$ a.e. w.r.t. Lebesgue measure on S_w. By continuity of $t \mapsto z_\theta(t)$, $\varphi = 0$ on S_w.

Now $\varphi = \sum_{i=1}^m h_i z_{\theta_i} = L(\sum_{i=1}^m h_i u_{\theta_i}) = 0$. If, for $k \in K$, $L(k) = 0$, then $k \equiv 0$. Therefore, $\sum_{i=1}^m h_i u_{\theta_i} \equiv 0$, which can only be true if all h_i are equal to zero. $\square$

In Tucker (1967), section 4.3 an explicit way to calculate the imaginary part of $\tilde{g}_n$ is given.

An estimator $\check{f}_n$ of the density function can be obtained by inverting the ch.f. $Q(\check{k}_n)$. For the density plots in figures 2 and 3, we used the method of Schorr (1975).

7 Applications and examples

We consider two observation schemes for (X_t):

(i) Observe (X_t) on a regularly spaced grid with fixed mesh-width Δ. Write $X_{k\Delta}$ for the observations $(k = 0, 1, \dots)$.

(ii) The same, but now suppose that the mesh-width Δ_n decreases as n increases. This gives, for each n, observations $(X_0, X_{\Delta_n}, X_{2\Delta_n}, \dots)$.

7.1 Data from the OU-process. Observation scheme (i)

Suppose $X_0, X_\Delta, \ldots, X_{(n-1)\Delta}$ are n observations from the stationary OU-process (X_t). Let F_0 denote the marginal distribution of $X_{i\Delta}$ $(0 \le i \le n-1)$. By theorem 2.3, F_0 is positive, self-decomposable and characterized by a function k_0 in K. As a preliminary estimator for $\psi_0 = Q(k_0)$ we propose the *empirical characteristic function* (e.c.f.), defined by

$$\tilde{\psi}_n(t) := \int e^{itx} d\mathbb{F}_n(x) = \frac{1}{n}\sum_{j=0}^{n-1} e^{itX_{j\Delta}}, \quad t \in \mathbb{R}\,.$$

Here, $\mathbb{F}_n$ denotes the empirical distribution function of $X_0, \ldots, X_{(n-1)\Delta}$. By theorem 3.1 the process (X_t) is β-mixing. This implies (X_n) is β-mixing. This in turn implies that (X_n) is *ergodic* (Genon-Catalot et al. (2000)). An application of Birkhoff's ergodic theorem (Krengel (1985), p. 9-10) gives, for $t \in \mathbb{R}$

$$\tilde{\psi}_n(t) \xrightarrow{\text{a.s.}} \int e^{itx} dF_0(x) = \psi_0(t)\,,$$

as n tends to infinity. Consistency of $\hat{k}_n$ now follows directly upon an application of theorem 5.4.

So far, the weight-function is fixed in advance of the estimation procedure. The choice of this function has been more or less arbitrary. Roughly, the larger the number of observations, the larger one would like to take S_w. This suggests a data-adaptive choice for w (or at least for its support). Numerical experiments indicate that such a choice can improve the numerical results obtained so far. Therefore, we have implemented a bootstrap procedure to determine the right-end-point of S_w, which we denote by t_n^*. This procedure runs as follows: sample a large number of times (say N) n observations out of the empirical distribution function $\hat{F}_n$. For the i-th set of observations, let $g_n^{(i)}$ denote its corresponding empirical cumulant function. Then approximate $E_0|\tilde{g}_n(t) - g_0(t)|$ by the average $U_n(t) := \frac{1}{N}\sum_{i=1}^N |g_n^{(i)}(t) - \tilde{g}_n(t)|$ for each t in a (large) interval $[0, M]$. Finally, take some threshold $\eta > 0$ and define $t_n^* := \inf\{t \ge 0 \;:\; U_n(t) > \eta\}$.

Next, we apply the support reduction algorithm of section 6 with $w(\cdot) = \mathbf{1}_{[-t_n^*, t_n^*]}(\cdot)$ $(\eta = 0.1)$. Figure 2 shows some simulation results in case π_0 is Gamma$(3, 2)$. We

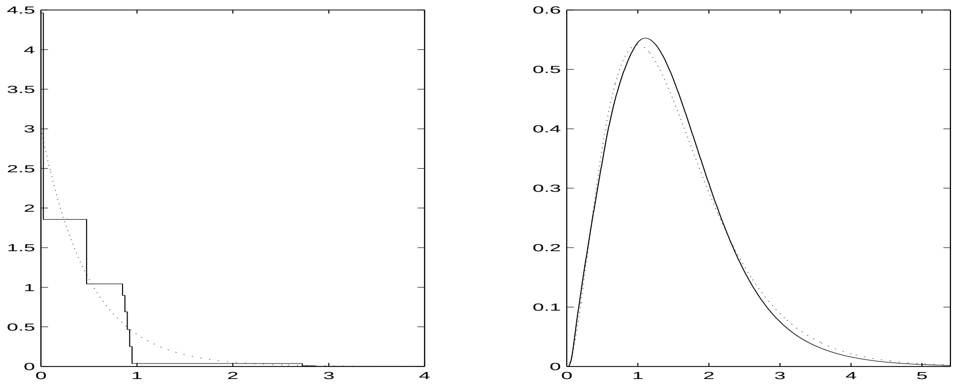

Figure 2: Gamma(3, 2) distribution, $n = 1000$. Left figure: estimated (solid) and true (dotted) canonical function. Right figure: estimated (solid) and true (dotted) density function.

simulated the OU-process on the interval $[0, 1000]$ and took observations at time instants $0, 1, \ldots$ (i.e. we observe the 1-skeleton).

Although the estimate for k_0 is quite inaccurate, the estimate for the density function shows a much better fit. The density plot is obtained by inversion of the characteristic function, corresponding to the estimated canonical function. See the last remark of section 6.

7.2 Data from the OU-process. Observation scheme (ii)

For each $n \geq 1$, denote the observations by $X_0, X_{\Delta_n}, X_{2\Delta_n}, \ldots, X_{(n-1)\Delta_n}$. We will now show that the e.c.f. based on these observations converges pointwise in probability to ψ_0.

Define for each fixed $u \in \mathbb{R}$ the continuous process $(Y_t^u,\, t \geq 0)$ by

$$Y_t^u = e^{iuX_t} - \mathbb{E}e^{iuX_t}\,.$$

Denote by $(Y_{k\Delta_n}^u)_k$ the discretely sampled process, obtained from $(Y_t^u,\, t \geq 0)$ by observation scheme *(ii)*. Thus,

$$Y_{k\Delta_n}^u = e^{iuX_{k\Delta_n}} - \mathbb{E}e^{iuX_{k\Delta_n}}\,, \quad k = 0, \ldots, n-1\,.$$

For a certain stochastic process $(U_s\,, s \geq 0)$ the α-mixing "numbers" are defined by

$$\alpha_U(h) = 2 \sup_t \sup_{\substack{A \in \sigma(U_s,\, s \leq t) \\ B \in \sigma(U_s,\, s \geq t+h)}} |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)|\,, \quad h > 0\,.$$

The process $(U_s\,, s \geq 0)$ is called α-mixing if $\alpha_U(h) \to 0$ as $h \to \infty$. As shown in Genon-Catalot et al. (2000), β-mixing is a stronger property than α-mixing. In fact, for any process $(U_s\,, s \geq 0)$ we have $\alpha_U(t) \leq \beta_U(t)$ $(t > 0)$.

Lemma 7.1 *Suppose that $\Delta_n \to 0$ and $n\Delta_n \to \infty$, as $n \to \infty$. Then, for $u \in \mathbb{R}$,*

$$\frac{1}{n} \sum_{k=0}^{n-1} Y^u_{k\Delta_n} \xrightarrow{p} 0\,, \quad n \to \infty\,.$$

Proof Let $u \in \mathbb{R}$ arbitrary. Denote the α-mixing numbers of $(|Y^u_t|\,, t \geq 0)$ by $\alpha_{|Y^u|}$, and similarly for $(|Y^u_{k\Delta_n}|)_k$ by $\alpha_{|Y^{u,n}|}$. Since $\sigma(|Y^u_t|,\, t \in T) \subseteq \sigma(X_t,\, t \in T)$ for any interval $T \subseteq [0, \infty)$, the definition of the α-mixing numbers implies that any $h > 0$, $\alpha_{|Y^u|}(h) \leq \alpha_X(h)$. In the same way one can verify that for $j \in \mathbb{N}$, $\alpha_{|Y^{u,n}|}(j) \leq \alpha_{|Y^u|}(j\Delta_n)$. Combining these inequalities gives: for $j \in \mathbb{N}$,

$$\alpha_{|Y^{u,n}|}(j) \leq \alpha_{|Y^u|}(j\Delta_n) \leq \alpha_X(j\Delta_n) \leq \beta_X(j\Delta_n)\,, \tag{7.1}$$

where the last inequality follows from the remark just before the lemma.

Lemma (9.2) implies the following inequality holds: for each $h \in \mathbb{N}$

$$\mathbb{P}\Big(\frac{1}{n} \sum_{k=0}^{n-1} |Y^u_{k\Delta_n}| \geq 2\varepsilon\Big) \leq \frac{2h}{n\varepsilon^2} \int_0^1 Q^2(1-w)dw + \frac{2}{\varepsilon} \int_0^{\alpha_{|Y^{u,n}|}(h)} Q(1-w)dw\,, \tag{7.2}$$

where $Q = F^{-1}_{|Y_1|}$. Since $\mathbb{P}(|Y_1| \leq y) = 1$ if $y \geq 2$, we have $Q(u) \leq 2$ for all $u \in (0,1)$. Hence, for all $\varepsilon > 0$,

$$\begin{aligned}
\mathbb{P}\Big(\Big|\frac{1}{n} \sum_{k=0}^{n-1} Y^u_{k\Delta_n}\Big| \geq 2\varepsilon\Big) \quad &\leq \quad \mathbb{P}\Big(\frac{1}{n} \sum_{k=0}^{n-1} |Y^u_{k\Delta_n}| \geq 2\varepsilon\Big) \\
&\leq \quad \frac{8h}{n\varepsilon^2} + \frac{4}{\varepsilon}\alpha_{|Y^{u,n}|}(h) \leq \frac{8h}{n\varepsilon^2} + \frac{4}{\varepsilon}\beta_X(h\Delta_n)\,, \qquad (7.3)
\end{aligned}$$

where the last inequality follows from (7.1).

Take $h = h_n = \sqrt{\frac{n}{\Delta_n}}$, then $h_n/n \to 0$ and $h_n\Delta_n \to \infty$ $(n \to \infty)$. Hence both terms in (7.3) can be made arbitrarily small by letting $n \to \infty$. $\square$

If we define $\tilde{\psi}_n(u) = \frac{1}{n}\sum_{k=0}^{n-1} e^{iuX_{k\Delta_n}}$, then the above lemma shows that $\tilde{\psi}_n(u) \xrightarrow{\mathrm{p}} \psi_0(u)$ for each $u \in \mathbb{R}$. An application of theorem 5.4 gives $\|\hat{k}_n - k_0\|_\mu \xrightarrow{\mathrm{p}} 0$ as $n \to \infty$, proving consistency.

7.3 Estimating a positive self-decomposable density from i.i.d. data

As a sidestep we consider the problem of estimating a positive self-decomposable density from i.i.d. data. Let $X_1, \ldots, X_n$ be independent random variables with common distribution function F_0. As before, F_0 is characterized by k_0 in K. As a preliminary estimator for ψ_0 we take again the e.c.f. Since $\mathbb{F}_n$ converges weakly to F_0 a.s., it follows that $\tilde{\psi}_n$ converges pointwise almost surely to ψ_0, as n tends to infinity. Consistency of $\hat{k}_n$ now follows from theorem 5.4.

Let f_0 denote the density of F_0. We remark that a general closed form expression for the density function f_0 in terms of k_0 is not known, This hampers the use of maximum likelihood techniques for estimating a self-decomposable density, based on i.i.d. data. However, given $\hat{k}_n$, we can calculate $\hat{\psi}_n = Q(\hat{k}_n)$, and then numerically invert this function to obtain a non-parametric estimator $\hat{F}_n$ for F_0. In contrast to the empirical distribution function, our estimator $\hat{F}_n$ is guaranteed to be of the correct type (i.e. self-decomposable). Figure 3 shows plots for the canonical function and the density function, in case π_0 follows a IG$(2, 1)$-distribution.

Alternative preliminary estimators are also possible. For example, suppose we know, in addition to the assumptions already made, that the density of F_0 is decreasing. Then we can take as a preliminary estimator the Grenander estimator $F_{n,Gren}$, which is defined as the least concave majorant of the empirical distribution function $\mathbb{F}_n$. Using similar arguments as before, we can show that the estimator for k based on $F_{n,Gren}$ is consistent. As another example, one could also take the maximum likelihood estimator for a unimodal density as a preliminary estimator for f_0. This makes sense, since every self-decomposable density is unimodal (Sato (1999), theorem 53.1).

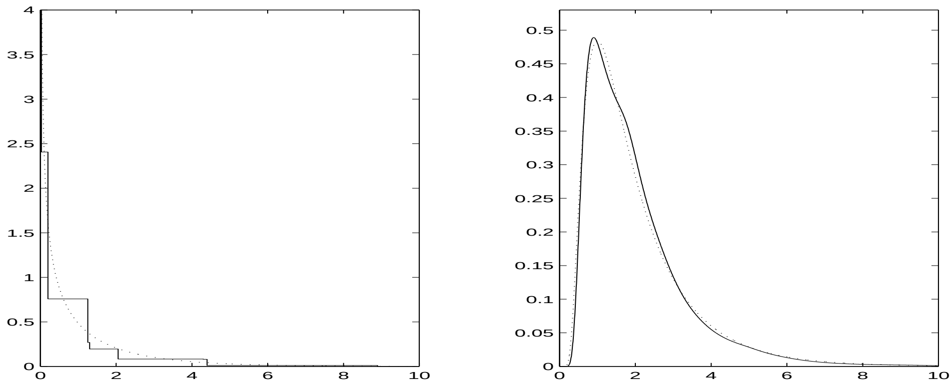

Figure 3: Inverse Gaussian$(2, 1)$ distribution, $n = 1000$. Left figure: estimated (solid) and true (dotted) canonical function. Right figure: estimated (solid) and true (dotted) density function.

8 Estimation of the intensity parameter λ

Suppose $X_0, X_\Delta, \dots, X_{n\Delta}$ are discrete-time observations from the stationary OU-process according to observation scheme *(i)*, for some fixed $\Delta > 0$. In this section, we define an estimator for λ. For ease of notation we write $X_i = X_{i\Delta}$ $(i = 0, \dots, n)$. From the proof of proposition 3.3 we know that for each $n \geq 1$, $X_n = e^{-\lambda} X_{n-1} + W_n(\lambda)$. Here $\{W_n(\lambda)\}_{n\geq 1}$ is a sequence of independent random variables with infinitely divisible distribution function Γ_λ. Since $(X_n,\, n \geq 0)$ is stationary, $X_0 \sim \pi_0$, where π_0 has Lévy density $x \mapsto \rho(x, \infty)/x$.

Let $\theta = e^{-\lambda}$ and denote the true parameter by θ_0. Since $W_n(\lambda) \geq 0$ for each $n \geq 1$, we easily obtain the bound $\theta_0 \leq \min_{n\geq 1} \frac{X_n}{X_{n-1}}$. Define the estimator

$$\hat{\theta}_n = \min_{1\leq k\leq n} \frac{X_k}{X_{k-1}} .$$

Then $\hat{\theta}_n(\omega) \geq \theta_0$, for each ω. Hence $\hat{\theta}_n$ is always biased upwards. However, we have

Lemma 8.1 *The estimator $\hat{\theta}_n$ is consistent: $\hat{\theta}_n \xrightarrow{\mathrm{p}} \theta_0$, as n tends to infinity.*

Proof Let $\varepsilon > 0$. Since

$$\begin{aligned}
\{\hat{\theta}_n \le \theta_0 + \varepsilon\} &= \left\{\exists\, k \in \{1,\dots,n\} \text{ such that } \frac{X_k}{X_{k-1}} \le \theta_0 + \varepsilon\right\} \\
&= \left\{\exists\, k \in \{1,\dots,n\} \text{ such that } \frac{\theta_0 X_{k-1} + W_k(\lambda)}{X_{k-1}} \le \theta_0 + \varepsilon\right\} \\
&= \left\{\exists\, k \in \{1,\dots,n\} \text{ such that } W_k(\lambda) \le \varepsilon X_{k-1}\right\} := A_{n,\varepsilon}\,,
\end{aligned}$$

we have $\mathbb{P}(|\hat{\theta}_n - \theta_0| \le \varepsilon) = \mathbb{P}(A_{n,\varepsilon})$. We aim to show that for each $\varepsilon > 0$, $\lim_{n\to\infty} \mathbb{P}(A^c_{n,\varepsilon}) = 0$. Define $N_n := \sum_{k=1}^n \mathbf{1}_{\{X_{k-1}>1\}}$, then

$$\begin{aligned}
\mathbb{P}(A^c_{n,\varepsilon}) &= \mathbb{P}(W_k(\lambda) > \varepsilon X_{k-1}\,,\ \forall\, k \in \{1,\dots,n\}) \\
&= \sum_{j=0}^{n} \mathbb{P}(W_k(\lambda) > \varepsilon X_{k-1}\ ,\ \forall\, k \in \{1,\dots,n\} \,|\, N_n = j)\, \mathbb{P}(N_n = j) \\
&\le \sum_{j=0}^{n} (\mathbb{P}(W_1(\lambda) > \varepsilon))^j\, \mathbb{P}(N_n = j)\,,
\end{aligned}$$

where the last inequality holds since $\{W_k(\lambda)\}_{k\ge1}$ is an i.i.d. sequence. Since $W_1(\lambda)$ has support $[0,\infty)$ (Sato (1999), corollary 24.8), $\mathbb{P}(W_1(\lambda) > \varepsilon) := \alpha_\varepsilon \in [0,1)$. This gives

$$\mathbb{P}(A^c_{n,\varepsilon}) \le \sum_{j=0}^{\infty} \alpha_\varepsilon^j\, \mathbb{P}(N_n = j)\,.$$

By dominated convergence, $\lim_{n\to\infty} \mathbb{P}(A^c_{n,\varepsilon}) \le \sum_{j=0}^{\infty} \alpha_\varepsilon^j [\lim_{n\to\infty} \mathbb{P}(N_n = j)]$. We are done, once we have proved that $\lim_{n\to\infty} \mathbb{P}(N_n = j) = 0$.

We claim $N_n \xrightarrow{\text{a.s.}} \infty$, as $n \to \infty$. From section 3 we know that $\lim_{n\to\infty} \|P^n(x,\cdot) - \pi_0\|_{TV} = 0$, for all $x \in E$. By proposition 6.3 in Nummelin (1984), this implies that the chain $(X_n)_n$ is positive Harris-recurrent and a-periodic. Now Harris recurrence implies that the set $(1,\infty)$ is visited infinitely many times by $(X_n)_n$, almost surely. Therefore, the claim holds and we conclude $\mathbb{P}(A^c_{n,\varepsilon}) \to 0$. □

By the continuous mapping theorem we have

Corollary 8.2 *Define $\hat{\lambda}_n = -\log \hat{\theta}_n$, then $\hat{\lambda}_n \xrightarrow{\text{P}} \lambda_0$, as $n \to \infty$, where λ_0 denotes the true value of λ.*

In case all innovations $W_n(\lambda)$ are exponentially distributed, $\hat{\theta}_n$ equals the maximum likelihood estimator for the model. A detailed asymptotic analysis for this model is given in Nielsen and Shephard (2003).

9 Appendix

Proof of proposition 3.3 The solution of the OU-equation is given in (2.3). If we discretize the expression for this solution we obtain

$$X_n = e^{-\lambda} X_{n-1} + \int_0^1 e^{\lambda(u-1)} dZ(\lambda(u+n-1)), \quad n \geq 1.$$

Since Z has stationary and independent increments we can write

$$X_n = e^{-\lambda} X_{n-1} + W_n(\lambda), \tag{9.1}$$

where $(W_n(\lambda))_n$ is an i.i.d. sequence of r.v. distributed as W_λ.

Next, we show that the distribution of $(\tilde{X}_t)$, defined by

$$\tilde{X}_t := \int_0^t e^{-\lambda(t-s)} dZ(\lambda s) \tag{9.2}$$

is infinitely divisible for each $t \geq 0$. Since $W_\lambda \stackrel{d}{=} \tilde{X}_1$ we then obtain infinite divisibility for the noise variables. Note that $(\tilde{X}_t)$ is simply the OU-process with initial condition $X(0) = 0$,

Similar as in equation (17.3) in Sato (1999) we have the following relation between the characteristic function of $\tilde{X}$ and $T(\psi_{Z(1)})$ (the cumulant of $Z(1)$).

$$\mathbb{E} e^{iz\tilde{X}_t} = \exp\left(\lambda \int_0^t T(\psi_{Z(1)})(e^{-\lambda(t-u)} z) du\right).$$

Since we assume Z has Lévy measure ρ (that is, $T(\psi_{Z(1)})(u) = \int_0^\infty (e^{iux} - 1)\rho(dx)$), we have

$$
\begin{aligned}
\log \mathbb{E} e^{iz\tilde{X}_t} &= \lambda \int_0^t \int_0^\infty \left(e^{ie^{-\lambda(t-u)}zx} - 1 \right) \rho(dx) du \\
&= \lambda \int_0^\infty \int_0^t \left(e^{ie^{-\lambda(t-u)}zx} - 1 \right) du \rho(dx) = \int_0^\infty \int_{e^{-\lambda t}x}^x (e^{iwz} - 1) w^{-1} dw \rho(dx) \\
&= \int_0^\infty (e^{iwz} - 1) w^{-1} dw \int_w^{e^{\lambda t} w} \rho(dx) = \int_0^\infty (e^{iwz} - 1) \kappa_t(dw) \,.
\end{aligned}
$$

Here

$$
\kappa_t(B) = \int_B w^{-1} \rho(w, e^{\lambda t} w] dw \,, \quad B \in \mathcal{E} \,.
$$

Hence if we let $\kappa := \kappa_1$, the Lévy measure has the form as given in (3.4).

It remains to be shown that κ_t satisfies $\int_0^\infty (1 \wedge x) \kappa_t(dx) < \infty$ for each $t > 0$. This follows from

$$
\int_0^1 x \kappa_t(dx) = \int_0^1 \rho(x, e^{\lambda t} x] dx \leq \int_0^{e^{\lambda t}} y \rho(dy) < \infty \,,
$$

and

$$
\begin{aligned}
\int_1^\infty \kappa_t(dx) &= \kappa_t(1, \infty) = \int_1^\infty \int_{(1 \vee e^{-\lambda t} y)}^y \frac{1}{w} dw \rho(dy) \\
&= \int_1^\infty \log \left(\frac{y}{(1 \vee e^{-\lambda t} y)} \right) \rho(dy) < \infty \\
&= \int_1^{e^{\lambda t}} \log y \rho(dy) + \int_{e^{\lambda t}}^\infty \lambda t \rho(dy) < \infty.
\end{aligned}
$$

□

Lemma 9.1 *Under* (A),

$$
I := \int_0^1 \frac{dz}{z} \exp \left(- \int_z^1 \frac{\lambda_\rho(y)}{y} dy \right) = +\infty \,, \tag{9.3}
$$

where

$$
\lambda_\rho(y) = \int_0^\infty (1 - e^{-yx}) \rho(dx) \,. \tag{9.4}
$$

Proof Let $y \in (0,1)$. Since $1-e^{-u} \le \min(u,1)$ for $u>0$ we get

$$\begin{aligned}
\lambda_\rho(y) &= \int_0^1 \ldots + \int_1^{1/\sqrt{y}} \ldots + \int_{1/\sqrt{y}}^\infty (1-e^{-yx})\rho(dx) \\
&\le y\int_0^1 x\rho(dx) + \int_1^{1/\sqrt{y}} \frac{y}{\sqrt{y}}\rho(dx) + \int_{1/\sqrt{y}}^\infty \frac{1-e^{-yx}}{\log x}\log x\rho(dx) \\
&\le c_1 y + c_2\sqrt{y} - \frac{2}{\log y}\int_{1/\sqrt{y}}^\infty \log x\rho(dx)\,,
\end{aligned}$$

where $c_1 = \int_0^1 x\rho(dx)$ and $c_2 = \rho(1,\infty)$.

Choose $\alpha \in (0,1)$ such that $c_3 := 2\int_{1/\sqrt{\alpha}}^\infty \log x\rho(dx) < 1$, which is possible by (A). Since $y \mapsto \int_{1/\sqrt{y}}^\infty \log x\rho(dx)$ is increasing on $(0,1)$, we have

$$\lambda_\rho(y) \le c_1 y + c_2\sqrt{y} - c_3/\log y\,, \qquad \text{if } y \in (0,\alpha)\,.$$

For $y \in (\alpha,1)$, we have the simple estimate $\lambda_\rho(y) \le c_1 y + c_2$. If $z \in (0,\alpha)$, then

$$\begin{aligned}
\int_z^1 \frac{\lambda_\rho(y)}{y}dy &= \int_z^\alpha \frac{\lambda_\rho(y)}{y}dy + \int_\alpha^1 \frac{\lambda_\rho(y)}{y}dy \\
&\le c_1(\alpha - z) + 2c_2(\sqrt{\alpha}-\sqrt{z}) - c_3\int_z^\alpha \frac{1}{y\log y}dy + c_1(1-\alpha) - c_2\log\alpha \\
&= K_\alpha - c_1 z - 2c_2\sqrt{z} + c_3\log(-\log z)\,,
\end{aligned}$$

where

$$K_\alpha = c_1 + c_2(2\sqrt{\alpha} - \log(\alpha)) - c_3\log(-\log\alpha) \in \mathbb{R}\,.$$

Using this inequality we get

$$I \ge \int_0^\alpha \frac{dz}{z}\exp\left(-\int_z^1 \frac{\lambda_\rho(y)}{y}dy\right) \ge e^{-K_\alpha}\int_0^\alpha e^{c_1 z + 2c_2\sqrt{z}}(-\log z)^{-c_3}\frac{dz}{z}\,.$$

The last integral exceeds

$$\int_0^\alpha \frac{1}{z(-\log z)^{c_3}}dz = \int_{-\log\alpha}^\infty \frac{1}{u^{c_3}}du = \infty\,,$$

since α was chosen such that $c_3 < 1$. $\square$

The statement and proof of the following lemma is similar to theorem 3.2 in Rio (2000).

Lemma 9.2 *For any mean zero time series X_t with α-mixing numbers $\alpha(h)$, every $x > 0$ and every $h, n \in I\!N$, with $Q_t = F^{-1}_{|X_t|}$,*

$$\mathbb{P}(\bar{X}_n \geq 2x) \leq \frac{2}{nx^2} \int_0^1 \frac{h}{n} \sum_{t=1}^{n} Q_t^2(1-u)\, du + \frac{2}{x} \int_0^{\alpha(h)} \frac{1}{n} \sum_{t=1}^{n} Q_t(1-u)\, du. \tag{9.5}$$

Acknowledgement: We thank O.E. Barndorff-Nielsen and L. Korsholm for suggesting to us the problem of estimating the Lévy density of Z nonparametrically, based on data from X. L. Korsholm was involved in the very early stages of this research and we thank him for sharing his ideas with us. We also would like to thank the associate editor and two referees for their helpful remarks, which improved the paper substantially,

References

Barndorff-Nielsen, O.E. and Shephard, N. (2001) *Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial economics*, Journal of the Royal Statistical Society, Series B, Vol. **63**, pp. 167-241.

Barndorff-Nielsen, O.E. and Shephard, N. (2001a) *Modelling by Lévy processes for Financial Econometrics*, in "Lévy Processes — Theory and Applications" (eds. Barndorff-Nielsen, O.E., Mikosch T. and Resnick, S), Birkhauser,

Basawa, I.V. and Brockwell, P.J. (1982) *Nonparametric estimation for nondecreasing Lévy processes*, Journal of the Royal Statistical Society, Series B, Vol. **44**, pp. 262-269.

Chung, K.L. (2001) *A course in probability theory*, Academic Press, 3rd Edition.

Çinlar, E. and Pinsky, M. (1971) *A stochastic integral in storage theory*, Zeit. Wahrsch. Verw. Gebiete **17**, 227-240.

Csörgö, S. (1983) *On how long interval is the empirical characteristic function uniformly consistent?*, Acta Sci. Math. (Szeged) **45**, pp. 141-149.

Davydov, Yu.A. (1973) *Mixing condtions for Markov chains*, Theory Probab. Appl. **18**, pp. 312-328.

Feuerverger, A. and McDunnough, P. (1981) *On the efficiency of empirical characteristic function procedures*, Journal of the Royal Statistical Society, Series B, Vol. **43**, pp. 20-27.

Genon-Catalot, V., Jeantheau, T. and Laredo, C. (2000) *Stochastic volatility models as hidden Markov models and statistical applications*, Bernoulli **6**, pp. 1051-1079.

Getoor, R.K. (1975) *Markov processes: Ray processes and right processes*, Lecture notes in mathematics 440, Springer.

Groeneboom, P., Jongbloed, G. and Wellner, J.A. (2003) *The support reduction algorithm for computing nonparametric function estimates in mixture models*, submitted.

Harn, K. van and Steutel, F.W. (2004) *Infinite divisibility of probability distributions on the real lihne*, Marcel Dekker New York.

Hughett, P. (1998) *Error bounds for numerical inversion of a probability characteristic function*, SIAM J. Numer. Anal. **35** No. 4 pp. 1368-1392.

Jameson, G.J.O. (1974) *Topology and Normed Spaces*, Chapman and Hall.

Jongbloed, G. and Van der Meulen, F.H. (2004) *Parametric estimation for OU-processes driven by a subordinator*, Preprint, `http://www.math.vu.nl/~meulen`.

Kaspi, H. and Mandelbaum, A. (1994) *On Harris recurrence in continuous time*, Mathematics of operations research, Vol **19** No. 1.

Knight, J.L. and Satchell, S.E. (1997) *The cumulant generating function method: implementation and asymptotic efficiency*, Econometric Theory **13**, pp 170-184.

Krengel, U. (1985) *Ergodic theorems*, Berlin: De Gruyter.

Loève, M. (1977) *Probability Theory I.* Springer, 4th edition.

Luong, A. and Thompson, M.E. (1987) *Minimum-distance methods based on quadratic distances for transforms*, Canad. J. Stat. **15** pp. 239-251.

Masuda, H. (2004) *On multidimensional Ornstein-Uhlenbeck processes driven by a general Lévy process*, Bernoulli **10**, p. 97-120.

MEYN, S.P. AND TWEEDIE, R.L. (1993) *Markov Chains and Stochastic Stability*, Springer.

MEYN, S.P. AND TWEEDIE, R.L. (1993a) *Stability of Markovian processes II: Continuous-time processes and sampled chains*, Adv. Appl. Prob. **25**, pp. 487-517.

NIELSEN, B. AND SHEPHARD, N. (2003) *Likelihood analysis of a first order autoregressive model with exponential innovations*, Journal of Time Series Analysis **24**, p. 337-344.

NUMMELIN, E. (1984) *General irreducible Markov chains and non-negative operators*, Cambridge.

PAPASPILIOPOULOS, O., DELLAPORTAS, P. AND ROBERTS, G. (2004) *Bayesian inference for non-Gaussian Ornstein-Uhlenbeck stochastic volatility processes*, Journal of Royal Statistical Society, Series B, Vol. **68**, to appear

REVUZ, D. AND YOR, M. (1999) *Continuous martingales and Browmian Motion*, Springer, 3rd Edition.

RIO, E. (2000) *Théorie asymptotique des processus aléatoires faiblement dependants*, In: Collection Mathématiques & Applications, vol. **31**, Springer.

RUBIN, H. AND TUCKER, H.G. (1959) *Estimating the parameters of a differential process*, Ann. Math. Statist. **30**, pp. 641-658.

RUDIN, W. (1987) *Real and comples analysis*, McGraw-Hill, 3rd Edition.

SATO, K.I. (1999) *Lévy processes and infinitely divisible distributions.* Cambridge University Press.

SCHORR, B. (1975) *Numerical inversion of a class of characteristic functions*, BIT **15**, pp. 94-102.

SHIGA, T. (1990) *A recurrence criterion for Markov processes of Ornstein-Uhlenbeck type*, Prob. Theory Related Fields **85**, pp. 425-447.

TUCKER, H.G. (1967) *A Graduate Course in Probability.* Academic Press.