Method of Moment Estimation in Time-Changed Lévy Models

Jan Kallsen[*]   Johannes Muhle-Karbe[†]

Abstract

This paper introduces a method of moment estimator for the time-changed Lévy processes proposed by Carr, Geman, Madan and Yor (2003). By establishing that the returns sequence is strongly mixing with exponentially decreasing rate, we prove consistency and asymptotic normality of the resulting estimators. In addition, we fit parametrized versions of the model to real data and examine the quality of our estimators by performing a simulation study. Finally, we also show how to estimate the current level of volatility.

Key words: time-changed Lévy models, moment estimator, mixing property, volatility estimation

Mathematics Subject Classification (2010): 62M05, 62P05, 60G99

1   Introduction

Most financial time series exhibit certain distinct features, usually called *stylized facts*:

1. *gain/loss asymmetry*, i.e., returns are negatively skewed,

2. *heavy tails* of the returns compared to the normal distribution,

3. *conditionally heavy tails*, i.e., heavy tails even after correcting for volatility clustering,

4. *absence of autocorrelation* of asset returns, but *volatility clustering*, i.e., significant autocorrelation of the squared returns,

5. *leverage effect*, i.e., negative crosscorrelation between returns and squared returns.

[*]Mathematisches Seminar, Christian-Albrechts-Universität zu Kiel, Westring 383, D–24118 Kiel, Germany, (e-mail: kallsen@math.uni-kiel.de).

[†]ETH Zürich, Departement Mathematik, Rämistrasse 101, CH–8092 Zürich, Switzerland, (e-mail: johannes.muhle-karbe@math.ethz.ch).

1

Consequently, there exists a growing literature on different models trying to recapture these empirical observations. In continuous time the first three characteristics are typically tackled by allowing for jumps in the asset price, whereas the last two are usually accounted for by introducing some kind of stochastic volatility (cf. [7, 26] for an overview).

For applications in Mathematical Finance, finding a suitable statistical model for the data under consideration is of course only one part of the story. Indeed, one prefers models that are able to explain at least some of the stylized facts, but at the same time one needs enough mathematical structure to allow for the solution of financial problems. These requirements are fit surprisingly well by the class of *affine stochastic volatility models* (cf. [19] for an overview). Since the stochastic volatility y and the logarithmized asset price X are modelled as a bivariate affine process in these models, the joint conditional characteristic function can be computed by solving some generalized Riccati equations, as shown in great generality by [9]. This opens the door to explicit solutions of diverse financial problems dealing with, e.g., optimal investment (cf., e.g., [4, 21, 23]) and hedging of derivatives (see, e.g., [8, 15, 20, 23]). In this paper, we introduce an estimation algorithm for the subclass of time-changed Lévy models introduced by [5]. In these the asset price is modelled as $S_t = S_0 \exp(X_t)$ with $S_0 > 0$ and

$$X_t = \mu t + B_{\int_0^t y_s ds}, \tag{1.1}$$

where $\mu \in \mathbb{R}$ and B denotes a Lévy process, whereas the *activity process* y is assumed to be strictly positive, stationary and independent of B. These models can capture several stylized facts observed in the data, nevertheless they are quite tractable from an analytical point of view.

When performing statistical estimation, it is typically assumed that the time series under consideration is mean adjusted, i.e., μ is set equal to 0 and B is assumed to be a martingale in Equation (1.1). By [3], it is straightforward to estimate μ from the mean adjustment if B is a martingale, since different values for μ do not change any of the higher centered moments or the second order dependence structure. If on the other hand, we do not require B to be a martingale, the situation becomes more involved (cf. [3]).

For applications in Mathematical Finance, the situation is completely different though. Here, many problems can only be solved if the parameter μ is set equal to zero, thus requiring a non-martingale B to model the drift of the asset under consideration (cf., e.g., [20, 21, 23] for examples when this condition is necessary).

Statistical estimation of stochastic volatility models typically falls into one of the following two broad categories:

1. *Simulation based techniques*: See, e.g., [1, 6, 10] and the references therein for applications to affine jump-diffusion models, which correspond to choosing B in Equation (1.1) to be the sum of a standard Brownian motion and a compound Poisson process. These approaches could also be used in the more general setup considered here. However, they require lengthy computations and are tedious to implement for the non-specialist. Furthermore, consistency and asymptotic normality are typically only

assured under regularity conditions that are not easily checked in concrete models (cf. [12] for more details).

2. *Approaches using exact formulas for moments of the model*: [3] calculate the moments and second order dependence structure of model (1.1), exactly in the case where B is a martingale and approximately for frequent observations in the general case. They proceed to construct a quasi-maximum likelihood (QML) estimator in the case where B is a martingale with symmetric marginal distributions and note that it would also be possible to argue approximately otherwise. Again easy-to-check regularity conditions ensuring good asymptotic properties are missing. Furthermore, QML estimation involves nonlinear minimization and is also not robust with respect to model misspecification.

This last drawback is avoided by performing a direct (generalized-) method of moment estimation, matching theoretical moments of the model to the corresponding empirical moments of the data. For affine jump-diffusions, this approach has been considered by [18] in the case where B is a martingale. They use the first four moments of the returns as well as some autocorrelations of the squared returns to exemplarily estimate the Heston model. However, asymptotic results are once more only obtained subject to regularity conditions (cf. [12]) that may be difficult to check in concrete models. On the contrary, [14], who use a similar moment based approach for the COGARCH model, only impose conditions on the parameters of the model that are easily verified for a concrete specification.

The aim of the present study is fourfold. First, we extend the method of moments algorithms used by [14, 18] to the setup considered here (which encompasses pure jump driving processes of infinite activity like the Normal Inverse Gaussian (henceforth NIG) process, for example), drawing on the results of [3]. In particular, we consider the case where B is possibly skewed and not necessarily a martingale. No simulation is required and all estimators are given explicitly, which makes straightforward implementation for diverse models possible. Inspired by [14], we then present exact asymptotic results if B is assumed to be a martingale and approximate asymptotic results if this assumption is dropped, only imposing conditions that are easily verified in concrete models. Thirdly, we analyze the small sample behavior of our estimation algorithms by fitting parametrized versions of the models to real data and performing simulation studies with the parameters obtained in this way. Finally, we also show how to estimate the current level of volatility by using a Kalman filter (if B is a martingale) respectively an extended Kalman filter (for general B).

The remainder of this article is organized as follows. In Section 2, we introduce the model and supply the formulas for its moments obtained by [3]. Subsequently, we deal with estimation in the case where B is a martingale. In Section 4, we then turn to estimating the model if the martingale assumption on B is dropped.

For a Lévy process B, we denote by ψ^B the corresponding *Lévy exponent*, i.e., the continuous function $\psi^B : i\mathbb{R} \to \mathbb{C}$ such that $\mathrm{E}(\exp(iuB_t)) = \exp(t\psi^B(iu))$ for $t \geq 0$

and $u \in \mathbb{R}$. Moreover, we use the shorthand notations $\mathbb{N}^* := \{1, 2, 3, \ldots\}$ as well as $\lfloor x \rfloor := \max\{n \in \mathbb{N}^* : n \leq x\}$ for $x \in \mathbb{R}_+$, and $\mathbb{R}_{++} := (0, \infty)$.

2 Moments and second-order dependence structure of time-changed Lévy models

Denote by $S = S_0 \exp(X)$ some asset price process with initial value $S_0 > 0$. We consider the time changed Lévy-models proposed by [5], where the *return process* X is modelled as

$$X_t = \mu t + B_{\int_0^t y_s ds}. \tag{2.2}$$

Here, $\mu \in \mathbb{R}$ and B denotes a real-valued Lévy process, whereas the *activity process* y is assumed to be strictly positive, stationary, and independent of B.

Example 2.1 If y is chosen to be a Lévy-driven OU process, i.e.,

$$dy_t = -\lambda y_{t-} dt + dZ_{\lambda t}, \quad y_0 > 0,$$

for $\lambda > 0$ and an increasing Lévy process Z independent of B, Equation (2.2) leads to a generalization of the model proposed by Barndorff-Nielsen and Shephard ([2], henceforth BNS). Similarly, one obtains a generalization of the Heston model without correlation, if one instead uses a strictly positive square-root process

$$dy_t = \lambda(\eta - y_t)dt + \sigma dZ_t, \quad y_0 > 0,$$

for a Wiener process Z and constants $\lambda, \eta, \sigma > 0$ satisfying $2\lambda\eta > \sigma^2$. For these specifications, it is shown in [19] that the process (y, X) is *affine* in the sense of [9].

To use the generalized method of moments for parameter estimation, one needs to calculate sufficiently many moments of the model under consideration. For time-changed Lévy models this has been done by [3] by conditioning on the time-change $\int_0^{\cdot} y_s ds$. More specifically, let $\Delta > 0$ be some grid size and define the discrete increments $X_{(n)}$ of the log-price X as

$$X_{(n)} := X_{n\Delta} - X_{(n-1)\Delta}, \quad n \in \mathbb{N}^*. \tag{2.3}$$

[3] relate the moments and dependence structure of $(X_{(n)})_{n \in \mathbb{N}^*}$ to the moments and dependence structure of y as well as the *cumulants* of B, defined as

$$c_n := \frac{\partial^n}{\partial u^n} \psi^B(u)\big|_{u=0}, \quad n \in \mathbb{N}^*.$$

Summing up results from [3], the following holds.

4

Theorem 2.2 *Let B be a Lévy process with $c_4 < \infty$ and suppose y is stationary with $E(y_t^4) < \infty$, $E(y_t) =: \xi$ and $\mathrm{Var}(y_t) =: \omega^2$ for all $t \in \mathbb{R}_+$. Let r_y be the autocorrelation function of y and*

$$r_y^{**}(t) := \int_0^t \int_0^v r_y(u)du\, dv.$$

Then, if $\mu = 0$, the following holds:

$$E(X_{(n)}) = c_1 \Delta \xi,$$
$$E(X_{(n)}^2) = c_2 \Delta \xi + c_1^2 \left(2\omega^2 r_y^{**}(\Delta) + (\Delta \xi)^2\right),$$
$$E(X_{(n)}^3) = c_3 \Delta \xi + 3c_1 c_2 \left(2\omega^2 r_y^{**}(\Delta) + (\Delta \xi)^2\right) + c_1^3 E(Y_\Delta^3),$$
$$E(X_{(n)}^4) = c_4 \Delta \xi + (4c_1 c_3 + 3c_2^2)\left(2\omega^2 r_y^{**}(\Delta) + (\Delta \xi)^2\right) + 6c_1^2 c_2 E(Y_\Delta^3) + c_1^4 E(Y_\Delta^4),$$

where $Y = \int_0^{\cdot} y_s ds$ and, for $s \in \mathbb{N}^*$,

$$\mathrm{Cov}(X_{(n)}, X_{(n+s)}) = c_1^2 \mathrm{Cov}(y_{n\Delta}, y_{(n+s)\Delta}),$$
$$\mathrm{Cov}(X_{(n)}, X_{(n+s)}^2) = c_1 c_2 \mathrm{Cov}(y_{n\Delta}, y_{(n+s)\Delta}) + c_1^3 \mathrm{Cov}(y_{n\Delta}, y_{(n+s)\Delta}^2),$$
$$\mathrm{Cov}(X_{(n)}^2, X_{(n+s)}^2) = c_2^2 \mathrm{Cov}(y_{n\Delta}, y_{(n+s)\Delta}) + c_1^2 c_2 \mathrm{Cov}(y_{n\Delta}^2, y_{(n+s)\Delta})$$
$$+ c_2 c_1^2 \mathrm{Cov}(y_{n\Delta}, y_{(n+s)\Delta}^2) + c_1^4 \mathrm{Cov}(y_{n\Delta}^2, y_{(n+s)\Delta}^2).$$

Moreover,

$$\mathrm{Cov}(y_{n\Delta}, y_{(n+s)\Delta}) = \omega^2 (r_y^{**}((s+1)\Delta) - 2r_y^{**}(s\Delta) + r_y^{**}((s-1)\Delta)).$$

PROOF. See [3, Propositions 2 and 5]. $\qquad\square$

Example 2.3 If y is either a stationary OU process or a stationary square-root process, it follows from [7, Chapter 15] that the autocorrelation function r_y of y is given by

$$r_y(u) = e^{-\lambda u}, \quad u \in \mathbb{R}_+.$$

Consequently, by [2, Example 4], we have

$$r_y^{**}(u) = \frac{1}{\lambda^2}\left(e^{-\lambda u} - 1 + \lambda u\right), \quad u \in \mathbb{R}_+.$$

3 Moment estimation if B is a martingale

We now use Theorem 2.2 to set up a generalized method of moments estimator, extending similar approaches used by [18] and [14] to estimate affine jump diffusion models and the COGARCH(1,1) model, respectively. This is done subject to the following assumptions:

(A1)   For time horizon $T > 0$ and grid size $\Delta > 0$, we have equally spaced observations $X_{j\Delta}$, $j = 0, ..., \lfloor T/\Delta \rfloor$ and returns $X_{(j)} = X_{j\Delta} - X_{(j-1)\Delta}$, $j = 1, ..., \lfloor T/\Delta \rfloor$.

(A2)   The cumulants c_j of B satisfy $c_1 = 0$, $c_2 = 1$ and $c_4 < \infty$.

(A3) y is a stationary OU or square-root process with mean reversion $\lambda > 0$, mean $\xi > 0$ and variance $\omega^2 \in (0, \infty)$.

Remark 3.1 $c_4 < \infty$ holds for most Lévy processes used in the literature, e.g., Variance Gamma (VG) and Normal Inverse Gaussian (NIG) processes. The normalization $c_2 = 1$ just leads to a rescaling of the time change and therefore can be assumed without leading to a loss of generality in the model. The final parameter restriction $c_1 = 0$ is equivalent to B being a martingale. It is commonly made in the literature (see, e.g., [2, 14, 24]), because it drastically simplifies the moment and dependence structure of the model (cf. Theorem 2.2). We will discuss the case $c_1 \neq 0$ in Section 4 below.

For given $\Delta > 0$, denote by $m_{i,\Delta}$ and $\mu_{i,\Delta}$, $i \in \mathbb{N}^*$, the i-th uncentered and centered moments of $X_{(n)}$, respectively. Furthermore, let $\gamma_\Delta(s) := \mathrm{Cov}(X_{(n)}^2, X_{(n+s)}^2)$ for $n, s \in \mathbb{N}^*$ and define $\gamma_{\Delta,d} := (\gamma_\Delta(1), \ldots, \gamma_\Delta(d))$ for $d \in \mathbb{N}^*$. Given (A1)-(A3), Theorem 2.2 then reads as follows.

Corollary 3.2 *Assume (A1)-(A3) hold. Then we have $m_{1,\Delta} = \mu\Delta$ as well as*

$$\mu_{2,\Delta} = \Delta\xi, \quad \mu_{3,\Delta} = c_3\Delta\xi, \quad \mu_{4,\Delta} = c_4\Delta\xi + 6\frac{\omega^2}{\lambda^2}(e^{-\lambda\Delta} - 1 + \lambda\Delta) + 3(\Delta\xi)^2,$$

$$\gamma_\Delta(s) = \omega^2 \frac{\left(1 - e^{-\lambda\Delta}\right)^2}{\lambda^2} e^{-\lambda\Delta(s-1)}, \quad s \in \mathbb{N}^*.$$

3.1 The estimation procedure

We begin by showing that the unknown model parameters $\mu, c_3, c_4, \lambda, \xi, \omega^2$ are uniquely determined as a continuously differentiable function of the first four moments of the returns as well as the autocovariance function of the squared returns.

Proposition 3.3 *Suppose (A1)-(A3) are satisfied and let $k_\Delta, p > 0$ such that, for $s \in \mathbb{N}^*$,*

$$\gamma_\Delta(s) = k_\Delta e^{-p\Delta(s-1)}.$$

Then $\mu, c_3, c_4, \lambda, \xi, \omega^2$ are uniquely determined by $m_{1,\Delta}, \mu_{2,\Delta}, \mu_{3,\Delta}, \mu_{4,\Delta}, k_\Delta, p$ as

$$(\mu, c_3, c_4, \lambda, \xi, \omega^2) = H_\Delta(m_{1,\Delta}, \mu_{2,\Delta}, \mu_{3,\Delta}, \mu_{4,\Delta}, k_\Delta, p)$$

with $H_\Delta : \mathbb{R} \times \mathbb{R}_{++} \times \mathbb{R}^3 \times \mathbb{R}_{++} \to \mathbb{R}^6$ defined as

$$H_\Delta(m_1, \mu_2, \mu_3, \mu_4, k, p))$$
$$:= \left(\frac{m_1}{\Delta}, \frac{\mu_3}{\mu_2}, \frac{\mu_4}{\mu_2} - 3\mu_2 - \frac{6k_\Delta(e^{-p\Delta} - 1 + p\Delta)}{\mu_2(1 - e^{-p\Delta})^2}, p, \frac{\mu_2}{\Delta}, \frac{p^2 k}{(1 - e^{-p\Delta})^2} \right).$$

Furthermore, H_Δ is continuously differentiable in $(m_{1,\Delta}, \mu_{2,\Delta}, \mu_{3,\Delta}, \mu_{4,\Delta}, k_\Delta, p)$.

PROOF. This follows immediately from Theorem 2.2 and Corollary 3.2 above. □

Proposition 3.3 motivates the following estimation algorithm, which estimates μ, c_3, c_4, λ, ξ, ω^2 by matching the first four moments of the model to the corresponding empirical moments and fitting the logarithmized autocovariance function of the model to its empirical counterpart via linear regression.

6

Algorithm 3.4 *1. Calculate the moment estimators*

$$\widehat{m}_{1,\Delta,T} := \frac{1}{\lfloor T/\Delta \rfloor} \sum_{j=1}^{\lfloor T/\Delta \rfloor} X_{(j)},$$

$$\widehat{\mu}_{i,\Delta,T} := \frac{1}{\lfloor T/\Delta \rfloor} \sum_{j=1}^{\lfloor T/\Delta \rfloor} (X_{(j)} - \widehat{m}_{1,\Delta,T})^i, \quad i = 2, 3, 4,$$

and for $d \geq 2$ the empirical autocovariances $\widehat{\gamma}_{\Delta,T,d} := (\widehat{\gamma}_{\Delta,T}(1), \ldots, \widehat{\gamma}_{\Delta,T}(d))$ as

$$\widehat{\gamma}_{\Delta,T}(s) := \frac{1}{\lfloor T/\Delta \rfloor} \sum_{j=1}^{\lfloor T/\Delta \rfloor - s} \left(X_{(j)}^2 - \widehat{\mu}_{2,\Delta,T} \right) \left(X_{(j+s)}^2 - \widehat{\mu}_{2,\Delta,T} \right),$$

for $s = 1, \ldots, d$.

2. *For fixed $d \geq 2$ define the mapping $K_\Delta : \mathbb{R}_{++}^d \times \mathbb{R}^2 \to \mathbb{R}$ by*

$$K_\Delta((\gamma_1, \ldots, \gamma_d), k, p) := \sum_{s=1}^{d} \left(\log(\gamma_s) - \log(k) + p\Delta s \right)^2,$$

and compute the least squares estimator

$$(\widehat{k}_\Delta(\widehat{\gamma}_{\Delta,T,d}), \widehat{p}_\Delta(\widehat{\gamma}_{\Delta,T,d})) := \operatorname*{arg\,min}_{(k,p)\in\mathbb{R}^2} K_\Delta(\widehat{\gamma}_{\Delta,T,d}, k, p),$$

which is given by

$$\widehat{p}_\Delta(\widehat{\gamma}_{\Delta,T,d}) = -\frac{\sum_{s=1}^{d} \left(\log(\widehat{\gamma}_{\Delta,T,d}(s)) - \overline{\log(\widehat{\gamma}_{\Delta,T,d})} \right) \left(s - \frac{d+1}{2} \right)}{\Delta \sum_{s=1}^{d} (s - \frac{d+1}{2})^2},$$

$$\widehat{k}_\Delta(\widehat{\gamma}_{\Delta,T,d}) = \exp\left(\overline{\log(\widehat{\gamma}_{\Delta,T,d})} + \Delta \frac{d+1}{2} \widehat{p}_\Delta(\widehat{\gamma}_{\Delta,T,d}) \right),$$

with $\overline{\log(\widehat{\gamma}_{\Delta,T,d})} := \frac{1}{d} \sum_{s=1}^{d} \log(\widehat{\gamma}_{\Delta,T,d}(s))$.

3. *Define the mapping $J_\Delta : \mathbb{R}^{d+4} \to \mathbb{R}^6$ by*

$$J_\Delta(m_1, \mu_2, \mu_3, \mu_4, \gamma)$$
$$:= \begin{cases} H_\Delta(m_1, \mu_2, \mu_3, \mu_4, \widehat{k}_\Delta(\gamma), \widehat{p}_\Delta(\gamma)) & \text{if } \mu_2, \gamma, \widehat{p}_\Delta(\gamma) > 0, \\ (0, 0, 0, 0, 0, 0) & \text{otherwise,} \end{cases}$$

and compute the estimator

$$(\widehat{\mu}_{\Delta,T}, \widehat{c}_{3,\Delta,T}, \widehat{c}_{4,\Delta,T}, \widehat{\lambda}_{\Delta,T}, \widehat{\xi}_{\Delta,T}, \widehat{\omega^2}_{\Delta,T})$$
$$:= J_\Delta(\widehat{m}_{1,\Delta,T}, \widehat{\mu}_{2,\Delta,T}, \widehat{\mu}_{3,\Delta,T}, \widehat{\mu}_{4,\Delta,T}, \widehat{\gamma}_{\Delta,T,d}).$$

Remark 3.5 By 3.2 and Assumption (A3), we have $\mu_{2,\Delta} > 0$ as well as $p > 0$ and $\gamma_\Delta(s) > 0$ for all $s \in \mathbb{N}^*$. However, the corresponding estimators are not necessarily strictly positive. But we will show in Corollary 3.10 below that all estimators are strongly consistent, which implies that all estimators will be almost surely well defined for sufficiently large samples.

Similarly, $\widehat{c}_{4,\Delta,T} < 0$ is possible depending on the data. On the other hand, we have $c_4 = 0$ if B is chosen as a Brownian motion as well as $c_4 > 0$ for all other Lévy process B with jumps. Hence we take $\widehat{c}_{4,\Delta,T} < 0$ as a strong indication that the data is too light tailed to be suitably modeled by the class of (semi-) heavy tailed models considered here.

Remark 3.6 If one considers the special case where B is chosen to be a Brownian motion, i.e., in the BNS or Heston model, we have $c_3 = c_4 = 0$. Hence one can still use Algorithm 3.4 above by simply neglecting the moments of order 3 and 4 and setting $\widehat{c}_{3,\Delta,T} = \widehat{c}_{4,\Delta,T} = 0$. All asymptotic considerations in Section 3.2 below remain true.

Remark 3.7 As in [14], we fit the model to the logarithms of the empirical autocovariances rather than the covariances themselves, because this leads to a linear regression and allows to compute the least squares estimator explicitly. Using the empirical covariances as proposed by [2], one is lead to a nonlinear least squares problem. Consequently, the existence of a unique solution, which depends on the model parameters in a continuously differentiable way, is no longer obvious and can only be guaranteed under additional assumptions (cf. [12]). Nevertheless, this approach seems to work fine in practice and is the natural choice when considering *superpositions* of OU processes of the form $y = \sum_{j=1}^m y^{(j)}$, where $y^{(j)}$, $j = 1, \ldots, m$ denote independent stationary OU processes. If each $y^{(j)}$ has mean reversion λ_j and $\mathrm{IG}(w_j a, b)$ or $\Gamma(w_j a, b)$ marginals with $\sum_{j=1}^m w_j = 1$, we have

$$\gamma_\Delta(s) = \omega^2 \sum_{j=1}^m \frac{w_j}{\lambda_j^2} \left(1 - e^{-\lambda_j \Delta}\right)^2 e^{-\lambda_j \Delta(s-1)}, \quad s \in \mathbb{N}^*,$$

which can be used to fit the parameters $\lambda_j, w_j, j = 1, \ldots, m$ to the empirical autocovariances via a nonlinear least squares regression.

3.2 Asymptotic properties of the estimator

Since all estimators in Algorithm 3.4 are continuously differentiable functions of empirical moments, strong consistency and asymptotic normality will follow from ergodicity of the process $(X_{(n)})_{n \in \mathbb{N}^*}$. For stochastic volatility models with stock prices driven by Brownian motion, it has been shown independently by [11] and [28] that the return sequence $(X_{(n)})_{n \in \mathbb{N}^*}$ is α-mixing (and hence ergodic), if y is α-mixing and further that the mixing coeffients for returns are smaller than or equal to the mixing coefficients of y. An inspection of the arguments in [11] shows that this remains true for time-changed Lévy models.

Theorem 3.8 *Suppose the process y is strictly stationary and α-mixing with mixing coefficients $(\alpha_y(k))_{k \in \mathbb{R}_+}$. Then $(X_{(n)})_{n \in \mathbb{N}^*}$ is also strictly stationary and α-mixing with mixing coefficients $(\alpha_X(n))_{n \in \mathbb{N}^*}$ satisfying*

$$\alpha_X(n) \leq \alpha_y(n), \quad \forall n \in \mathbb{N}^*.$$

In particular, $(X_{(n)})_{n \in \mathbb{N}^*}$ is ergodic and if y is α-mixing with exponentially decreasing rate, then $(X_{(n)})_{n \in \mathbb{N}^*}$ is α-mixing with exponentially decreasing rate, too.

PROOF. We generalize the arguments of [11, Sections 3.1, 3.2] to time-changed Lévy models. In view of [11, Proposition 3.1] it is enough to check the prerequisites of [11, Definition 3.1]. The first property of [11, Definition 3.1] follows as in [11, Theorem 3.1] if the space of continuous functions and its Borel σ-algebra associated with the uniform topology are replaced with the Skorokhod space $\mathbb{D}$ and its Borel σ-algebra $\mathscr{D}$ associated with the Skorokhod topology (cf. [17, Chapter VI and in particular VI.1.14] for more details), because the mapping

$$T : \mathbb{D} \to \mathbb{R}^2; \quad (f(t))_{t \in \mathbb{R}_+} \mapsto \left(\int_0^\Delta f(s)ds, f(\Delta) \right)$$

is $\mathscr{D}$-$\mathscr{B}(\mathbb{R})$ measurable. The other two properties of [11, Definition 3.1] follow literally as in [11, Theorem 3.1] by applying [17, II.4.15], because X has independent increments on $[\![0, n\Delta]\!]$ conditional on $\sigma(y_s, s \leq n\Delta)$. $\qquad\square$

Theorem 3.8 is often applicable due to the following well known result.

Lemma 3.9 *Let y be a strictly positive square-root process or an OU process such that $\mathrm{E}(|y_t|^p) < \infty$ for some $p > 0$. Then y is α-mixing with exponentially decreasing rate.*

PROOF. The first part of the assertion can be found, e.g., in [11, Section 2.6]. The second follows from [22, Theorem 4.3]. $\qquad\square$

By Birkhoff's ergodic theorem (cf. [27, Theorem V.3.1]) all moments estimators in Algorithm 3.4 are strongly consistent.

Corollary 3.10 *Assuming that (A1)-(A3) hold, we have, for $T \to \infty$,*

$$\widehat{m}_{1,\Delta,T} \overset{a.s.}{\to} m_{1,\Delta}, \quad \widehat{\mu}_{i,\Delta,T} \overset{a.s.}{\to} \mu_{i,\Delta}, \quad i = 2,3,4, \quad \widehat{\gamma}_{\Delta,T}(s) \overset{a.s.}{\to} \gamma_\Delta(s), \quad s = 1, \ldots, d.$$

Next we turn to asymptotic normality, which can be obtained by applying a central limit theorem for strongly mixing processes under the following additional assumption.

(A4) $\quad \mathrm{E}(X_{(n)}^{8+\varepsilon}) < \infty$ for some $\varepsilon > 0$.

Remark 3.11 Since $\mathrm{E}(B_1) = 0$, Condition (A4) holds if $\mathrm{E}(B_1^{10}) < \infty$ and $\mathrm{E}(|y_1|^5) < \infty$, since this implies $\mathrm{E}(X_t^{10}) < \infty$ and hence $\mathrm{E}(X_{(n)}^{10}) < \infty$. This can be seen by conditioning on the time-change $\int_0^\cdot y_s ds$ and using [25, Proposition 2.5].

Lemma 3.12 *Let (A1)-(A4) be satisfied. Then, for $T \to \infty$,*

$$\sqrt{\left\lfloor \frac{T}{\Delta} \right\rfloor} \Big((\widehat{m}_{1,\Delta,T}, \widehat{\mu}_{2,\Delta,T}, \widehat{\mu}_{3,\Delta,T}, \widehat{\mu}_{4,\Delta,T}, \widehat{\gamma}_{\Delta,T,d}) - (m_{1,\Delta}, \mu_{2,\Delta}, \mu_{3,\Delta}, \mu_{4,\Delta}, \gamma_{\Delta,d}) \Big)$$
$$\xrightarrow{d} N_{d+4}(0, \Sigma),$$

where the covariance matrix Σ has components

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

with

$$G_n := \Big(X_{(n)}, (X_{(n)} - m_{1,\Delta})^2, (X_{(n)} - m_{1,\Delta})^3, (X_{(n)} - m_{1,\Delta})^4,$$
$$(X_{(n)}^2 - \mu_{2,\Delta})(X_{(n+1)}^2 - \mu_{2,\Delta}), \dots, (X_{(n)}^2 - \mu_{2,\Delta})(X_{(n+d)}^2 - \mu_{2,\Delta}) \Big).$$

PROOF. Since $(X_{(n)})_{n \in \mathbb{N}^*}$ is strongly mixing with exponentially decreasing rate, the claim follows from the Ibragimov central limit theorem for strongly mixing processes (cf. [16, Theorem 18.5.3]) along the lines of the proof of [14, Proposition 3.7]. $\square$

Summing up, we have the following result.

Theorem 3.13 *Assume (A1)-(A3) hold. Then, for $T \to \infty$,*

$$(\widehat{\mu}_{\Delta,T}, \widehat{c}_{3,\Delta,T}, \widehat{c}_{4,\Delta,T}, \widehat{\lambda}_{\Delta,T}, \widehat{\xi}_{\Delta,T}, \widehat{\omega^2}_{\Delta,T}) \xrightarrow{\text{a.s.}} (\mu, c_3, c_4, \lambda, \xi, \omega^2).$$

If additionally (A4) holds, then, for $T \to \infty$,

$$\sqrt{\left\lfloor \frac{T}{\Delta} \right\rfloor} \Big((\widehat{\mu}_{\Delta,T}, \widehat{c}_{3,\Delta,T}, \widehat{c}_{4,\Delta,T}, \widehat{\lambda}_{\Delta,T}, \widehat{\xi}_{\Delta,T}, \widehat{\omega}^2_{\Delta,T}) - (\mu, c_3, c_4, \lambda, \xi, \omega^2) \Big)$$
$$\xrightarrow{\text{d}} \nabla J_\Delta(m_{1,\Delta}, \mu_{2,\Delta}, \mu_{3,\Delta}, \mu_{4,\Delta}, \gamma_{\Delta,d}) N_{d+4}(0, \Sigma),$$

where Σ is defined as in Lemma 3.12.

PROOF. The strong consistency follows from Corollary 3.10 by the continuous mapping theorem (cf. [29, Theorem 2.3]) and the asymptotic normality is a consequence of Lemma 3.12 and the delta method (cf. [29, Theorem 3.1]), because J_Δ is continuously differentiable in $(m_{1,\Delta}, \mu_{2,\Delta}, \mu_{3,\Delta}, \mu_{4,\Delta}, \gamma_{\Delta,d})$. $\square$

3.3 Estimation results for real data

Using Algorithm 3.4 proposed above, we now fit the time-changed Lévy model to real data. As in, e.g., [1, 6, 10] we consider a long time series of daily returns, since this provides rich information about the conditional and unconditional distribution of the returns while

allowing us to sidestep the seasonality issues inherent in high frequency data, which are beyond our scope here. We use a daily time series of the German industrial index *DAX* spanning from the 14th of June in 1988 to the 10th of April in 2008 (i.e. $T = 20$, $\Delta = 1/250$ and $T/\Delta = 5000$ returns). Following [14], we use $d \approx \sqrt{\lfloor T/\Delta \rfloor}$, i.e., $d = 70$ for $T = 20$ and $\Delta = 1/250$. The results are shown in Table 1.

$\widehat{\mu}_{1/250,20}$	$\widehat{c}_{1,1/250,20}$	$\widehat{c}_{3,1/250,20}$	$\widehat{c}_{4,1/250,20}$	$\widehat{\lambda}_{1/250,20}$	$\widehat{\xi}_{1/250,20}$	$\widehat{\omega}^2_{1/250,20}$
0.0894	0	-0.00549	0.000445	2.54	0.0485	0.00277

Table 1: Estimation results based on Algorithm 3.4 with $d = 70$.

The fitted model accounts for the skewness of -0.3943 and the kurtosis of 8.8210 exhibited by our data set. The empirical autocorrelation functions and their theoretical counterparts are shown in Figure 1, indicating that the second-order dependency structure is fit quite well, too.
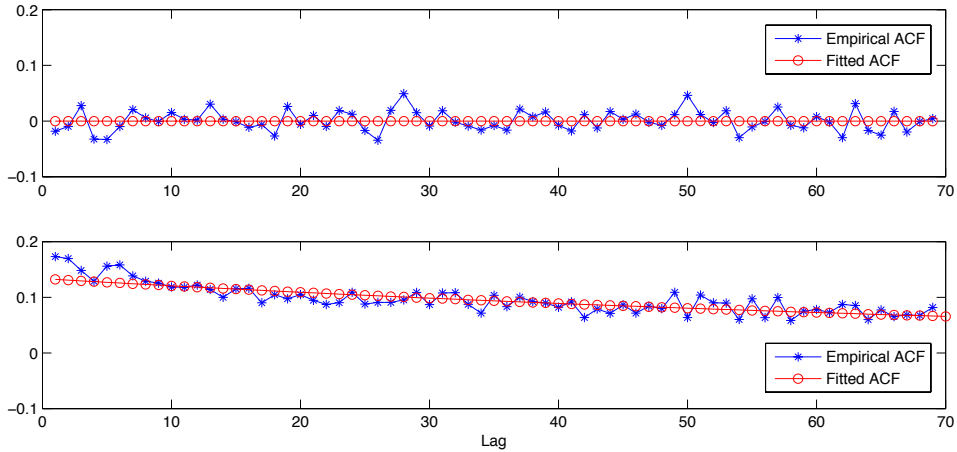

Figure 1: Empirical and fitted autocorrelation functions of the log returns (first) and the squared log returns (second)

Remark 3.14 Many applications in Mathematical Finance require a model for the stock price discounted by a bond $S_t^0 = e^{rt}$ with constant interest rate r. If we use the average 0.0456 of the 6-month EURIBOR from its inception as a proxy for r and estimate the parameters of the discounted model using Algorithm 3.4, we obtain the results shown in Table 2. Only the estimate of μ changes, since all other estimators use centered moments.

$\widehat{\mu}_{1/250,20}$	$\widehat{c}_{1,1/250,20}$	$\widehat{c}_{3,1/250,20}$	$\widehat{c}_{4,1/250,20}$	$\widehat{\lambda}_{1/250,20}$	$\widehat{\xi}_{1/250,20}$	$\widehat{\omega}^2_{1/250,20}$
0.0438	0	-0.00549	0.000445	2.54	0.0485	0.00277

Table 2: Estimation results for the discounted stock price with Algorithm 3.4.

Remark 3.15 As discussed in Remark 3.7 above, it is also possible to consider superpositions of OU processes and fit them to the empirical autocovariances. Using the MATLAB nonlinear least squares routine `lsqnonlin`, this approach yields the following set of parameter estimates for the superposition of two independent OU-processes with mean reversion λ_j, mean $w_j\xi$ and variance $w_j\omega^2$, $j = 1, 2$:

$$\widehat{\xi} = 0.0485, \quad \widehat{\omega}^2 = 0.00402, \quad \widehat{w}_1 = 0.446, \quad \widehat{\lambda}_1 = 32.5, \quad \widehat{w}_2 = 0.554, \quad \widehat{\lambda}_2 = 1.38.$$

The corresponding fitted autocorrelation function for the superposition of two OU processes is shown alongside its counterpart for one OU process in Figure 2. Clearly, the fit is im-
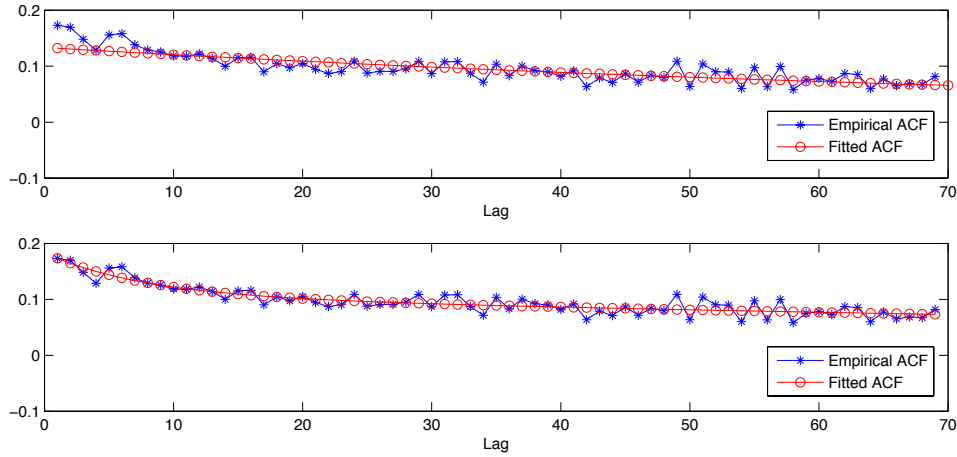

Figure 2: Empirical and fitted autocorrelation functions of the squared log returns for a superposition of one (first) and two (second) OU processes.

proved considerably for short lags, although the overall effect is not too big for our daily data. If one moves to more highly frequent data, however, several OU processes become indispensable to model dependencies on different time scales.

So far these results are really of semiparametric nature, since we have not specified the processes B and y yet. We now present some parametric examples.

Example 3.16 (IG-OU process, Gamma-OU process) Suppose y follows a stationary IG-OU process (cf., e.g., [26]) with $IG(a, b)$ marginals. Then $a = \sqrt{\xi^3/\omega^2}$ and $b = \sqrt{\xi/\omega^2}$, hence strongly consistent and asymptotically normal estimators are given by $\widehat{a}_{1/250,20} = 0.203$ and $\widehat{b}_{1/250,20} = 4.1835$. If y follows a stationary Gamma-OU process (cf., e.g., [26]) with $\Gamma(a, b)$ marginals, the corresponding estimators are $\widehat{a}_{1/250,20} = 0.8483$ and $\widehat{b}_{1/250,20} = 17.5013$.

Example 3.17 (Square-root process) Assume y is given by a strictly stationary and positive square-root process. Then y_t follows a $\Gamma(2\lambda\eta/\sigma^2, 2\lambda/\sigma^2)$-law by, e.g., [7, Section 15.1.2]. This implies that $\eta = \xi$ and $\sigma^2 = 2\lambda\omega^2/\xi$. Hence consistent and asymptotically normal estimators are given by $\widehat{\eta}_{1/250,20} = 0.0485$ and $\widehat{\sigma}_{1/250,20} = 0.5386$.

Note that the stationary distributions for the square-root and the Gamma-OU process coincide, whereas only the first two moments are identical for the IG-OU process.

Example 3.18 (BNS model) In the BNS model, B is chosen to be a Brownian motion with drift. In this case, $c_1 = 0$ and $c_2 = 1$ imply that B is a standard Brownian motion.

If the BNS model is estimated using Algorithm 3.4, the third and fourth moments of the model are not fitted to the data. More specifically, Theorem 2.2 yields that the fitted BNS model has skewness 0 and kurtosis 6.52 compared with the values -0.39 and 8.82 observed in our data set. This shows that in our setup, stochastic volatility without jumps in the asset price cannot explain the skewness in the data and can only account for a part of the heavy tails. To show the full flexibility of the class of models considered here, we now consider a Lévy process B with jumps, namely the NIG process, which is a popular model for stock prices itself.

Example 3.19 (NIG process) Let B be a NIG process with Lévy exponent

$$\psi^B(u) = u\delta + \vartheta \left(\sqrt{\alpha^2 - \beta^2} - \sqrt{\alpha^2 - (\beta + u)^2} \right),$$

where $\delta \in \mathbb{R}$, $\alpha, \vartheta > 0$ and $\beta \in (-\alpha, \alpha)$. Then by [26, Section 5.3.8], Conditions (A1)-(A4) are satisfied for $\vartheta = (\alpha^2 - \beta^2)^{3/2}\alpha^{-2}$ and $\delta = -\beta(\alpha^2 - \beta^2)\alpha^{-2}$ and solving for $\alpha, \beta, \delta, \vartheta$ we obtain the following estimators, which are strongly consistent and asymptotically normal by Theorem 3.13 above:

$$\widehat{\beta}_{\Delta,T} := \frac{\widehat{c}_{3,\Delta,T}}{\widehat{c}_{4,\Delta,T} - 5\widehat{c}_{3,\Delta,T}^2/3}, \quad \widehat{\alpha}_{\Delta,T} := \sqrt{\widehat{\beta}_{\Delta,T}^2 + 3\widehat{\beta}_{\Delta,T}/\widehat{c}_{3,\Delta,T}},$$

$$\widehat{\vartheta}_{\Delta,T} := \frac{(\widehat{\alpha}_{\Delta,T}^2 - \widehat{\beta}_{\Delta,T}^2)^{3/2}}{\widehat{\alpha}_{\Delta,,T}^2}, \quad \widehat{\delta}_{\Delta,T} := \frac{-\widehat{\vartheta}_{\Delta,T}\widehat{\beta}_{\Delta,T}}{\sqrt{\widehat{\alpha}_{\Delta,T}^2 - \widehat{\beta}_{\Delta,T}^2}}.$$

For our data set, this yields

$$\widehat{\beta}_{1/250,20} = -13.9, \quad \widehat{\alpha}_{1/250,20} = 88.3, \quad \widehat{\vartheta}_{1/250,20} = 85.0, \quad \widehat{\delta}_{1/250,20} = 13.6.$$

3.4 Simulation study

To investigate the small sample behavior of our estimation algorithm, we now assume that X is given by a NIG-IG-OU process, i.e., y is chosen to be a stationary IG-OU process with mean reversion λ and marginal $IG(\sqrt{\xi^3/\omega^2}, \sqrt{\xi/\omega^2})$ distributions, whereas the Lévy process B is assumed to be an NIG process.

As for parameters, we use the estimates obtained from our daily DAX time series in Examples 3.16 and 3.19 above. Sample paths of an NIG-IG-OU process can easily be simulated using algorithms found in [26, Sections 8.4.5 and 8.4.7]. We simulate 1000 samples of equidistant observations of returns $X_{(n)}$ for $\Delta = 1/250$ and $T = 20$ and $T = 40$, where we

first work on a finer grid with 80 steps per day and then only use the returns on the original grid to reduce discretization errors. The results are shown in Table 3. As above, we have chosen $d \approx \sqrt{\lfloor T/\Delta \rfloor}$, i.e., $d = 70$ for $T = 20$ and $d = 100$ for $T = 40$ here. Note that we measure the estimation error relative to the true values of the parameters in order to account for the different sizes of the parameters.

	μ	c_3	c_4	λ	ξ	ω^2
True Value	0.0894	-0.00549	0.000445	2.54	0.0485	0.00277
$T = 20$	$\widehat{\mu}_{1/250,T}$	$\widehat{c}_{3,1/250,T}$	$\widehat{c}_{4,1/250,T}$	$\widehat{\lambda}_{1/250,T}$	$\widehat{\xi}_{1/250,T}$	$\widehat{\omega}^2_{1/250,T}$
Mean	0.0856	-0.0543	0.000454	3.01	0.0478	0.00250
AAPE	0.427	0.354	0.307	0.339	0.174	0.454
$T = 40$	$\widehat{\mu}_{1/250,T}$	$\widehat{c}_{3,1/250,T}$	$\widehat{c}_{4,1/250,T}$	$\widehat{\lambda}_{1/250,T}$	$\widehat{\xi}_{1/250,T}$	$\widehat{\omega}^2_{1/250,T}$
Mean	0.0910	-0.00547	0.000450	2.82	0.0484	0.00272
AAPE	0.311	0.271	0.231	0.242	0.125	0.347

Table 3: Estimated mean and average absolute percentage error for the parameters $\widehat{\mu}_{\Delta,T}$, $\widehat{c}_{3,\Delta,T}$, $\widehat{c}_{4,\Delta,T}$, $\widehat{\lambda}_{\Delta,T}$, $\widehat{\xi}_{\Delta,T}$ and $\widehat{\omega}^2_{\Delta,T}$ estimated with Algorithm 3.4.

The estimators seem to be fairly consistent for the sample size under consideration, the only notable exception being the mean reversion parameter λ which is markedly biased to the right. Moving from $T = 20$ to $T = 40$ we observe that the mean absolute errors decrease by factors of roughly $\sqrt{2}$ as would be expected from the Ibragimov central limit theorem.

3.5 Estimation of the current level of volatility

The current value of the activity process y is needed in many applications in Mathematical Finance, e.g., for portfolio optimization or hedging of derivatives. Since it cannot be observed directly, it has to be filtered from the given returns. Assuming y follows an OU process and $c_1 = 0$, we can proceed along the lines of [2, Section 5.4.3], and obtain a linear state space representation which allows to use the Kalman filter (cf. [13] for more details), to provide a best linear (based on $X_{(n)}$ and $X^2_{(n)}$) predictor of y. More specifically, it follows from Corollary 3.2 and [2, Section 5.4.3] that a linear state space representation of $(X_{(n)}, X^2_{(n)})$ is given by

$$\begin{pmatrix} X_{(n)} \\ X^2_{(n)} \end{pmatrix} = \begin{pmatrix} \mu\Delta \\ \mu^2\Delta^2 \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ \lambda^{-1} & 0 \end{pmatrix} \begin{pmatrix} \lambda(Y_{n\Delta} - Y_{(n-1)\Delta}) \\ y_{n\Delta} \end{pmatrix} + u_n,$$

where the vector martingale difference sequence u_n satisfies

$$\mathrm{Var}(u_{1n}) = \Delta\xi, \quad \mathrm{Cov}(u_{1n}, u_{2n}) = 2\mu\Delta^2\xi + c_3\Delta\xi,$$

$$\mathrm{Var}(u_{2n}) = 4\mu^2\Delta^3\xi + 4\frac{\omega^2}{\lambda^2}\left(e^{-\lambda\Delta} - 1 + \lambda\Delta\right) + 2\xi^2\Delta^2 + c_4\Delta\xi + 4\mu\Delta^2 c_3\xi,$$

and

$$\begin{pmatrix} \lambda(Y_{(n+1)\Delta} - Y_{n\Delta}) \\ y_{(n+1)\Delta} \end{pmatrix} = \begin{pmatrix} 0 & 1 - e^{-\lambda\Delta} \\ 0 & e^{-\lambda\Delta} \end{pmatrix} \begin{pmatrix} \lambda(Y_{n\Delta} - Y_{(n-1)\Delta}) \\ y_{n\Delta} \end{pmatrix} + w_n,$$

with IID noise w_n (uncorrelated with u_n) satisfying

$$E(w_n) = \xi \begin{pmatrix} e^{-\lambda\Delta} - 1 + \lambda\Delta \\ 1 - e^{-\lambda\Delta} \end{pmatrix},$$

$$\text{Var}(w_n) = 2\omega^2 \begin{pmatrix} \lambda\Delta - 2(1 - e^{-\lambda\Delta}) + \frac{1}{2}(1 - e^{-2\lambda\Delta}) & \frac{1}{2}(1 - e^{-\lambda\Delta})^2 \\ \frac{1}{2}(1 - e^{-\lambda\Delta})^2 & \frac{1}{2}(1 - e^{-2\lambda\Delta}) \end{pmatrix}.$$

In Figure 3 we show the results of applying the Kalman filter to the simulated returns, suggesting it is possible to obtain decent estimates of the volatility in this way.
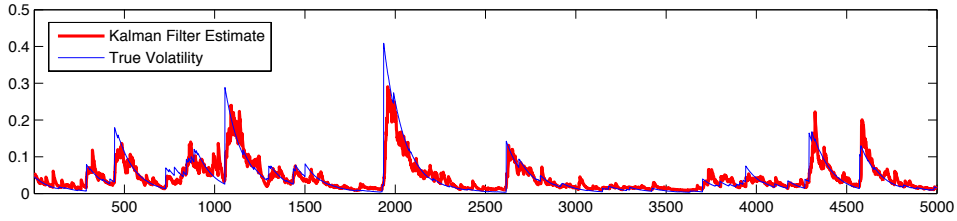

Figure 3: Sample paths of an IG-OU process (thin line) with parameters as in Example 3.16 and the Kalman filter estimate (thick line) obtained from the corresponding NIG-IG-OU process with parameters as in Examples 3.19, 3.16.

4 Moment estimation for arbitrary B

We now consider the case where $\mu = 0$ and the Lévy process B is not necessarily assumed to be a martingale, i.e., $c_1 \neq 0$. Estimation is done subject to the following assumptions:

(B1) For time horizon $T > 0$ and grid size $\Delta > 0$ we have equally spaced observations $X_{j\Delta}$, $j = 0, ..., \lfloor T/\Delta \rfloor$ and returns $X_{(j)} = X_{j\Delta} - X_{(j-1)\Delta}$, $j = 1, ..., \lfloor T/\Delta \rfloor$.

(B2) $\mu = 0$ and the cumulants of B satisfy $c_2 = 1$ and $c_4 < \infty$.

(B3) y is a stationary OU or CIR process with mean reversion $\lambda > 0$, mean $\xi > 0$, variance $\omega^2 > 0$ and existing fourth moments.

(B4) $E(X_{(n)}^{8+\varepsilon}) < \infty$ for some $\varepsilon > 0$.

Remark 4.1 Note that if $c_1 = E(B_1)$ does not vanish anymore, one now has to to require $E(B_1^{10}) < \infty$ and $E(y_1^{10}) < \infty$ in order to ensure that (B4) holds. Like in Remark 3.11, this follows from [25, Proposition 2.5] by conditioning on the time change.

As above, for given grid size $\Delta > 0$ we write $\mu_{i,\Delta}$ and $m_{i,\Delta}$ for the i-th centered and uncentered moment of $X_{(n)}$, set $\gamma_\Delta(s) := \text{Cov}(X_{(n)}, X_{(n+s)})$ for $s \in \mathbb{N}^*$ and define $\gamma_{\Delta,d_\Delta} := (\gamma_\Delta(1), \ldots, \gamma_\Delta(d_\Delta))$ for $d_\Delta \in \mathbb{N}^*$.

4.1 Approximate moments

The key to the estimation algorithms proposed below is the following observation by [3].

Lemma 4.2 *Assume that (B1)-(B3) hold. Then for $\Delta \downarrow 0$,*

$$m_{3,\Delta} = c_3 \Delta \xi + 3c_1 \left(\frac{2\omega^2}{\lambda^2}(e^{-\lambda\Delta} - 1 + \lambda\Delta) + \Delta^2\xi^2 \right) + O(\Delta^3),$$

$$m_{4,\Delta} = c_4 \Delta \xi + (3 + 2c_1 c_3) \left(\frac{2\omega^2}{\lambda^2}(e^{-\lambda\Delta} - 1 + \lambda\Delta) + \Delta^2\xi^2 \right) + O(\Delta^3),$$

as well as , for $s \in \mathbb{N}^*$ and $\Delta \downarrow 0$,

$$\gamma_\Delta(s) = \omega^2 \frac{(1 - e^{-\lambda\Delta})^2}{\lambda^2} e^{-\lambda\Delta(s-1)} + O(\Delta^3) = \omega^2 \Delta^2 e^{-\lambda\Delta(s-1)} + O(\Delta^3).$$

PROOF. This is shown in [3, Propositions 4 and 5]. $\qquad\square$

4.2 The estimation procedure

By neglecting all terms of order Δ^3 or higher in Lemma 4.2, we obtain the following approximations of the model parameters by moments of the returns and the autocovariance function of the squared returns.

Lemma 4.3 *Assume (B1)-(B3) hold and let $k, p \in \mathbb{R}_{++}$ be constants such that, for fixed $D \in \mathbb{N}^*$ and $\Delta \downarrow 0$,*

$$\gamma_\Delta(s) = k\Delta^2 e^{-p\Delta(s-1)} + O(\Delta^3), \quad s \in \left\{ 1, \ldots, \left\lfloor \frac{D}{\sqrt{\Delta}} \right\rfloor + 1 \right\}. \tag{4.4}$$

Then, for sufficiently small Δ, there exists a largest solution $x_\Delta > 0$ to

$$0 = \mu_{2,\Delta} x^2 - \Delta x^3 - m_{1,\Delta}^2 k,$$

and we have, for $\Delta \downarrow 0$,

$$\lambda = p + O(\sqrt{\Delta}/D), \quad \omega^2 = k + O(\Delta), \quad \xi = x_\Delta + O(\Delta^2), \quad c_1 = \frac{m_{1,\Delta}}{\Delta x_\Delta} + O(\Delta^2),$$

$$c_3 = \frac{m_{3,\Delta}}{\Delta x_\Delta} - 3m_{1,\Delta} \left(1 + \frac{k}{x_\Delta^2} \right) + O(\Delta^2),$$

$$c_4 = \frac{m_{4,\Delta}}{\Delta x_\Delta} - \left\{ \frac{3\Delta}{x_\Delta} + \frac{2m_{1,\Delta}}{x_\Delta^2} \left(\frac{m_{3,\Delta}}{\Delta x_\Delta} - 3m_{1,\Delta} \left(1 + \frac{k}{x_\Delta^2} \right) \right) \right\} (x_\Delta^2 + k) + O(\Delta^2).$$

PROOF. This follows from Theorem 2.2 and Lemma 4.2 using some technical but straightforward arguments, cf. [23, Lemma 3.25] for more details. $\qquad\square$

Lemma 4.3 motivates the following estimation algorithm. In view of Theorem 4.6 below, all estimators will again be almost surely well-defined for sufficiently small Δ and sufficiently large samples.

16

Algorithm 4.4   *1. Calculate the moment estimators*

$$\widehat{m}_{i,\Delta,T} := \frac{1}{\lfloor T/\Delta \rfloor} \sum_{j=1}^{\lfloor T/\Delta \rfloor} X_{(n)}^i, \quad i = 1,2,3,4,$$

as well as, for fixed $D \in \mathbb{N}^*$ and $d_\Delta := \lfloor D/\sqrt{\Delta} \rfloor + 1$, the empirical autocovariances $\widehat{\gamma}_{\Delta,T,d_\Delta} := (\widehat{\gamma}_{\Delta,T}(1), \ldots, \widehat{\gamma}_{\Delta,T}(d_\Delta))$, as

$$\widehat{\gamma}_{\Delta,T}(s) := \frac{1}{\lfloor T/\Delta \rfloor} \sum_{j=1}^{\lfloor T/\Delta \rfloor - s} \left(X_{(j)}^2 - \widehat{m}_{2,\Delta,T} \right) \left(X_{(j+s)}^2 - \widehat{m}_{2,\Delta,T} \right),$$

for $s = 1, \ldots, d_\Delta$.

2. Define the mapping $K_\Delta : \mathbb{R}_{++}^d \times \mathbb{R}^2 \to \mathbb{R}$ by

$$K_\Delta((\gamma_1, \ldots, \gamma_d), k, p) := \sum_{s=1}^{d_\Delta} \left(\log(\gamma_s) - \log(\Delta^2 k) + p\Delta s \right)^2,$$

and compute the least square estimator

$$\left(\widehat{k}_\Delta(\widehat{\gamma}_{\Delta,T,d_\Delta}), \widehat{p}_\Delta(\widehat{\gamma}_{\Delta,T,d_\Delta}) \right) := \operatorname*{arg\,min}_{(k,p) \in \mathbb{R}^2} K_\Delta(\widehat{\gamma}_{\Delta,T,d_\Delta}, k, p),$$

which is given by

$$\widehat{p}_\Delta(\widehat{\gamma}_{\Delta,T,d_\Delta}) = -\frac{\sum_{s=1}^{d_\Delta} \left(\log(\widehat{\gamma}_{\Delta,T,d_\Delta}(s)) - \overline{\log(\widehat{\gamma}_{\Delta,T,d_\Delta})} \right) (s - \frac{d_\Delta+1}{2})}{\Delta \sum_{s=1}^{d_\Delta} (s - \frac{d_\Delta+1}{2})^2)},$$

$$\widehat{k}_\Delta(\widehat{\gamma}_{\Delta,T,d_\Delta}) = \Delta^{-2} \exp\left(\overline{\log(\widehat{\gamma}_{\Delta,T,d_\Delta})} + \Delta \frac{d_\Delta+1}{2} \widehat{p}_\Delta(\widehat{\gamma}_{\Delta,T,d_\Delta}) \right),$$

with $\overline{\log(\widehat{\gamma}_{\Delta,T,d_\Delta})} := \frac{1}{d_\Delta} \sum_{s=1}^{d_\Delta} \log(\widehat{\gamma}_{\Delta,T,d_\Delta}(s))$.

3. Compute

$$\widehat{x}_\Delta(\widehat{m}_{1,\Delta,T}, \widehat{m}_{2,\Delta,T}, \widehat{\gamma}_{\Delta,T})$$
$$:= \max\left\{ x \in \mathbb{R} : \widehat{\mu}_{2,\Delta,T} x^2 - \Delta x^3 - \widehat{m}_{1,\Delta,T}^2 \widehat{k}_\Delta(\widehat{\gamma}_{\Delta,T,d_\Delta}) = 0 \right\}.$$

4. Define the mapping $H_\Delta : \mathbb{R}_{++} \times \mathbb{R}^4 \times \mathbb{R}_{++} \to \mathbb{R}^6$ by

$$H_\Delta(x, m_1, m_3, m_4, k, p) := \left(\frac{m_1}{\Delta x}, \frac{m_3}{\Delta x} - 3m_1 \left(1 - \frac{k}{x^2} \right), \right.$$

$$\left. \frac{m_4}{\Delta x} - \left\{ \frac{3\Delta}{x} + \frac{2m_1}{x^2} \left(\frac{m_3}{\Delta x} - 3m_1 \left(1 - \frac{k}{x^2} \right) \right) \right\} (x^2 + k), p, x, k \right).$$

17

5. *Define the mapping $J_\Delta : \mathbb{R}^4 \times \mathbb{R}_+^{d_\Delta} \to \mathbb{R}^6$ by*

$$J_\Delta(m_1, m_2, m_3, m_4, \gamma)$$
$$:= \begin{cases} H(\widehat{x}_\Delta(m_1, m_2, \gamma), m_1, m_3, m_4, \widehat{k}_\Delta(\gamma), \widehat{p}_\Delta(\gamma)) & \text{if } \gamma, \widehat{x}_\Delta(m_1, m_2, \gamma), \widehat{p}_\Delta(\gamma) > 0, \\ (0, 0, 0, 0, 0, 0) & \text{otherwise}, \end{cases}$$

and compute the estimator

$$(\widehat{c}_{1,\Delta,T}, \widehat{c}_{3,\Delta,T}, \widehat{c}_{4,\Delta,T}, \widehat{\lambda}_{\Delta,T}, \widehat{\xi}_{\Delta,T}, \widehat{\omega}^2_{\Delta,T})$$
$$= J_\Delta(\widehat{m}_{1,\Delta,T}, \widehat{m}_{2,\Delta,T}, \widehat{m}_{3,\Delta,T}, \widehat{m}_{4,\Delta,T}, \widehat{\gamma}_{\Delta,T,d_\Delta}).$$

Remark 4.5 1. Note that the mapping J_Δ is continuously differentiable in the true parameter values $(m_{1,\Delta}, m_{2,\Delta}, m_{3,\Delta}, m_{4,\Delta}, \gamma_{\Delta,d_\Delta})$, because the implicit function theorem shows that $\widehat{x}_\Delta$ is continuously differentiable in $(m_{1,\Delta}, m_{2,\Delta}, \gamma_{\Delta,d_\Delta})$.

2. As above, $\widehat{c}_{4,\Delta,T} < 0$ is possible depending on the data, which we once again take as a strong indication that the data is too light tailed to be suitably modelled by the class of (semi-) heavy tailed models considered here.

4.3 Asymptotic properties of the estimator

For the construction of the estimation algorithms in Section 4.2, we had to resort to approximate moments with an error term vanishing only as $\Delta \downarrow 0$. Consequently, strong consistency and asymptotic normality only hold up to this error term as well.

Theorem 4.6 *Define $\widehat{c}_{1,\Delta,T}$, $\widehat{c}_{3,\Delta,T}$, $\widehat{c}_{4,\Delta,T}$, $\widehat{\lambda}_{\Delta,T}$, $\widehat{\xi}_{\Delta,T}$, $\widehat{\omega^2}_{\Delta,T}$ as in Algorithm 4.4 and assume (B1)-(B3) hold. Then we have*

$$\lim_{T\to\infty} \left((\widehat{c}_{1,\Delta,T}, \widehat{c}_{3,\Delta,T}, \widehat{c}_{4,\Delta,T}, \widehat{\lambda}_{\Delta,T}, \widehat{\xi}_{\Delta,T}, \widehat{\omega^2}_{\Delta,T}) - \left((c_1, c_3, c_4, \lambda, \xi, \omega^2) + \varepsilon_\Delta \right) \right) \stackrel{\text{a.s.}}{=} 0,$$

and, if additionally (B4) holds,

$$\sqrt{\lfloor T/\Delta \rfloor} \left((\widehat{c}_{1,\Delta,T}, \widehat{c}_{3,\Delta,T}, \widehat{c}_{4,\Delta,T}, \widehat{\lambda}_{\Delta,T}, \widehat{\xi}_{\Delta,T}, \widehat{\omega^2}_{\Delta,T}) - \left((c_1, c_3, c_4, \lambda, \xi, \omega^2) + \varepsilon_\Delta \right) \right)$$
$$\stackrel{\text{d}}{\to} \nabla J_\Delta(m_1, m_2, m_3, m_4, \gamma_{\Delta,d_\Delta}) N_{d+4}(0, \Sigma),$$

as $T \to \infty$, where

$$\varepsilon_\Delta = (O(\Delta^2), O(\Delta^2), O(\Delta^2), O(\sqrt{\Delta}), O(\Delta^2), O(\Delta)) \quad \text{for } \Delta \downarrow 0,$$

and the covariance matrix Σ has components

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

for

$$G_n := \left(X_{(n)}, X^2_{(n)}, X^3_{(n)}, X^4_{(n)}, \right.$$
$$\left. (X^2_{(n)} - m_{2,\Delta})(X^2_{(n+1)} - m_{2,\Delta}), \dots, (X^2_{(n)} - m_{2,\Delta})(X^2_{(n+d)} - m_{2,\Delta}) \right).$$

PROOF. Set

$$\varepsilon_\Delta := (c_1, c_3, c_4, \lambda, \xi, \omega^2) - J_\Delta(m_{1,\Delta}, m_{2,\Delta}, m_{3,\Delta}, m_{4,\Delta}, \gamma_{\Delta, d_\Delta}),$$

where $d_\Delta = \lfloor \sqrt{D/\Delta} \rfloor$ for some $D \in \mathbb{R}_+$. By Lemma 4.3 and the definition of J_Δ in Algorithm 4.4, we have

$$\varepsilon_\Delta = (O(\Delta^2), O(\Delta^2), O(\Delta^2), O(\sqrt{\Delta}), O(\Delta^2), O(\Delta)) \quad \text{for } \Delta \downarrow 0.$$

Notice that the proof of Theorem 3.8 also holds in the present setup. Hence, for fixed $\Delta > 0$, the series $(X_{(n)})_{n \in \mathbb{N}^*}$ is ergodic and Birkoff's ergodic theorem yields that for $T \to \infty$, we have

$$\widehat{m}_{i,\Delta,T} \overset{\text{a.s.}}{\to} m_{i,\Delta}, \quad i = 1, 2, 3, 4, \qquad \widehat{\gamma}_{\Delta,T,d_\Delta} \overset{\text{a.s.}}{\to} \gamma_{\Delta, d_\Delta}.$$

By the continuous mapping theorem (cf. [29], Theorem 2.3), this implies

$$J_\Delta(\widehat{m}_{1,\Delta,T}, \widehat{m}_{2,\Delta,T}, \widehat{m}_{3,\Delta,T}, \widehat{m}_{4,\Delta,T}, \widehat{\gamma}_{\Delta,T,d_\Delta})$$
$$\overset{\text{a.s.}}{\to} J_\Delta(\widehat{m}_{1,\Delta}, \widehat{m}_{2,\Delta}, \widehat{m}_{3,\Delta}, \widehat{m}_{4,\Delta}, \widehat{\gamma}_{\Delta,d_\Delta}) \quad \text{as } T \to \infty,$$

since J_Δ is continuous in $(m_{1,\Delta}, m_{2,\Delta}, m_{3,\Delta}, m_{4,\Delta}, \gamma_{\Delta, d_\Delta})$. This shows the first statement. The second follows from the Ibragimov central limit theorem by an application of the delta method as in the proofs of Lemma 3.12 and Theorem 3.13. □

4.4 Estimation results for real data

We now apply Algorithm 4.4 to the same set of daily DAX data used in Section 3 above. The results are shown in Table 4. As in Remark 3.14, one can again discount by a constant deterministic interest rate $r = 0.0456$ first and then apply the estimation Algorithm 4.4. Since uncentered moments are used for the estimation of all parameters in Algorithm 4.4, all parameters are potentially affected by this. However, the results shown in Table 5 suggest that the effect is quite small for all parameters except for the drift c_1.

$\widehat{\mu}$	$\widehat{c}_{1,1/250,20}$	$\widehat{c}_{3,1/250,20}$	$\widehat{c}_{4,1/250,20}$	$\widehat{\lambda}_{1/250,20}$	$\widehat{\xi}_{1/250,20}$	$\widehat{\omega}^2_{1/250,20}$
0	1.85	-0.00675	0.000448	2.54	0.0485	0.00277

Table 4: Estimation results based on Algorithm 4.4.

$\widehat{\mu}$	$\widehat{c}_{1,1/250,20}$	$\widehat{c}_{3,1/250,20}$	$\widehat{c}_{4,1/250,20}$	$\widehat{\lambda}_{1/250,20}$	$\widehat{\xi}_{1/250,20}$	$\widehat{\omega}^2_{1/250,20}$
0	0.904	-0.00610	0.000444	2.54	0.0485	0.00278

Table 5: Estimation results for the discounted stock price based on Algorithm 4.4.

Example 4.7 (IG-OU process, Gamma-OU process, square-root process) Since the estimators $\widehat{\lambda}, \widehat{\xi}, \widehat{\omega}^2$ are virtually unchanged compared to Section 3, the resulting estimators for the stationary distribution of y also practically coincide with the respective values from Examples 3.16 and 3.17.

Example 4.8 (BNS model) If B is given by a Brownian motion with drift $\delta \in \mathbb{R}$ and volatility $\sigma \in \mathbb{R}_+$, we have $\delta = c_1$ and $\sigma^2 = c_2$. Consequently, $\sigma = 1$ and the estimator $\widehat{\delta}_{1/250,20} = \widehat{c}_{1,1/250,20} = 1.85$ is approximately consistent and asymptotically normal for small Δ. If one considers data discounted with the constant deterministic interest rate $r = 0.0456$, the corresponding estimator is given by $\widehat{\delta}_{1/250,20} = 0.904$.

Example 4.9 (NIG process) Suppose that B is given by an NIG process. Plugging $c_2 = 1$ and the estimates for c_1, c_3, c_4 given in Table 4 above into

$$\beta = \frac{c_3}{c_4 - 5c_3^2/3}, \quad \alpha = \sqrt{\beta^2 + 3\beta/c_3}, \quad \vartheta = \frac{(\alpha^2 - \beta^2)^{3/2}}{\alpha^2}, \quad \delta = c_1 - \frac{\vartheta\beta}{\sqrt{\alpha^2 - \beta^2}},$$

yields estimators $\widehat{\beta}_{1/250,20}, \widehat{\alpha}_{1/250,20}, \widehat{\vartheta}_{1/250,20}, \widehat{\delta}_{1/250,20}$ for the parameters $\beta, \alpha, \vartheta, \delta$ of the NIG process, which are approximately consistent and asymptotically normal for small Δ:

$$\widehat{\beta}_{1/250,20} = -18.2, \quad \widehat{\alpha}_{1/250,20} = 91.7, \quad \widehat{\vartheta}_{1/250,20} = 86.3, \quad \widehat{\delta}_{1/250,20} = 19.3.$$

For discounted data, we obtain

$$\widehat{\beta}_{1/250,20} = -16.0, \quad \widehat{\alpha}_{1/250,20} = 90.1, \quad \widehat{\vartheta}_{1/250,20} = 85.9, \quad \widehat{\delta}_{1/250,20} = 16.5.$$

4.5 Simulation study

We now investigate the performance of Algorithm 4.4 by performing the same simulation study as for Algorithm 3.4 in Section 3.4 above. As for parameters we use the values given in Examples 4.7 and 4.9, respectively. The results are shown in Table 6. Comparing the results with Table 3, we find that the use of the approximate moments seems to entail virtually no loss in the quality of the estimators for our daily data. This suggests that the approximation errors resulting from the use of the approximate moment are rather small compared to the variance of the estimators.

	c_1	c_3	c_4	λ	ξ	ω^2
True Value	1.84	-0.00675	0.000448	2.54	0.0485	0.00278
$T = 20$	$\widehat{c}_{1,1/250,T}$	$\widehat{c}_{3,1/250,T}$	$\widehat{c}_{4,1/250,T}$	$\widehat{\lambda}_{1/250,T}$	$\widehat{\xi}_{1/250,T}$	$\widehat{\omega}^2_{1/250,T}$
Mean	1.90	-0.00667	0.000452	3.00	0.0482	0.00253
AAPE	0.463	0.278	0.335	0.347	0.176	0.467
$T = 40$	$\widehat{c}_{1,1/250,T}$	$\widehat{c}_{3,1/250,T}$	$\widehat{c}_{4,1/250,T}$	$\widehat{\lambda}_{1/250,T}$	$\widehat{\xi}_{1/250,T}$	$\widehat{\omega}^2_{1/250,T}$
Mean	1.84	-0.00673	0.000445	2.83	0.0486	0.00264
AAPE	0.318	0.189	0.270	0.243	0.123	0.357

Table 6: Estimated mean and average absolute percentage error for the parameters $\widehat{c}_{1,\Delta,T}$, $\widehat{c}_{3,\Delta,T}, \widehat{c}_{4,\Delta,T}, \widehat{\lambda}_{\Delta,T}, \widehat{\xi}_{\Delta,T}$ and $\widehat{\omega}^2_{\Delta,T}$.

4.6 Computation of the approximation error

The results of our simulation studies suggest that the errors resulting from the use of approximate moments are quite small. However, it is generally difficult to quantify them without resorting to large scale Monte-Carlo simulations. For affine models however, it is sometimes possible to explicitly calculate the joint characteristic function of the returns $X_{(n)}$ and $X_{(n+s)}$ for $n, s \in \mathbb{N}^*$. Differentiation and evaluation at zero via MATLAB's symbolic toolbox then lead to exact formulas for moments and autocovariances. These equations do not yield any favorable estimation algorithms, because they are extremely complicated and hideously nonlinear. However, they can comfortably be used for an a posteriori error estimation. We have the following general result from [19]:

Lemma 4.10 *Let y be an OU-process driven by a subordinator Z. Then for $n, s \in \mathbb{N}^*$, the joint characteristic function of the returns $X_{(n)}$ and $X_{(n+s)}$ is given by*

$$\mathrm{E}\left(e^{iu_1 X_{(n)}+iu_2 X_{(n+s)}}\right) = \mathrm{E}\left(e^{\Psi^1(\Delta, \Psi^1((s-1)\Delta, \Psi^1(\Delta, 0, iu_2), 0), iu_1)y_{(n-1)\Delta}}\right)$$
$$\times e^{\Psi^0(\Delta, 0, iu_2)+\Psi^0((s-1)\Delta, \Psi^1(\Delta, 0, iu_2), 0)+\Psi^0(\Delta, \Psi^1((s-1)\Delta, \Psi^1(\Delta, 0, iu_2), 0), iu_1)},$$

where

$$\Psi^1(t, u_1, u_2) := u_1 e^{-\lambda t} + \frac{1-e^{-\lambda t}}{\lambda}\psi^B(u_2),$$

$$\Psi^0(t, u_1, u_2) := \int_0^t \psi^Z(\Psi^1(s, u_1, u_2))ds,$$

and ψ^B resp. ψ^Z denote the Lévy exponents of B and Z, respectively.

PROOF. Follows from [19, Corollaries 3.2 and 3.1]. $\qquad\qquad\square$

For a Gamma-OU process y, all terms can be determined explicitly. A similar closed-form expression can also be obtained if y is chosen to be a square-root process, cf. [5].

Corollary 4.11 *Let y be a Gamma-OU process with stationary $\Gamma(\xi^2/\omega^2, \xi/\omega^2)$-distribution. Then for $s \in \mathbb{N}^*$, we have*

$$\mathrm{E}\left(e^{iu_1 X_{(n)}+iu_2 X_{(n+s)}}\right) = \left(1 - \frac{\omega^2}{\xi}\Psi^1(\Delta, \Psi^1((s-1)\Delta, \Psi^1(\Delta, 0, iu_2), 0), iu_1)\right)^{-\xi^2/\omega^2}$$
$$\times e^{\Psi^0(\Delta, 0, iu_2)+\Psi^0((s-1)\Delta, \Psi^1(\Delta, 0, iu_2), 0)+\Psi^0(\Delta, \Psi^1((s-1)\Delta, \Psi^1(\Delta, 0, iu_2), 0), iu_1)},$$

where

$$\Psi^1(t, u_1, u_2) := u_1 e^{-\lambda t} + \frac{1-e^{-\lambda t}}{\lambda}\psi_B(u_2),$$

$$\Psi^0(t, u_1, u_2) := \frac{\frac{\xi^2}{\omega^2}\left(\frac{\xi}{\omega^2}\log\left(\frac{\xi/\omega^2-\Psi^1(t, u_1, u_2)}{\xi/\omega^2-u_1}\right)+\psi_B(u_2)t\right)}{\xi/\omega^2 - \psi_B(u_2)/\lambda}.$$

Here $\log$ *denotes the* distinguished logarithm *in the sense of [25, Lemma 7.6].*

PROOF. Since Ψ^1 is $\mathbb{C}_-$-valued by [9, Propositions 6.1 and 6.4], the first formula follows from Lemma 4.10 by inserting the analytic continuation of the characteristic function of the $\Gamma(\xi^2/\omega^2, \xi/\omega^2)$-distribution to $\mathbb{C}_-$. By, e.g., [26, Section 7.1.1] we have $\psi^Z(u) = (u\lambda\xi^2/\omega^2)/(\xi/\omega^2 - u)$ for the stationary Gamma-OU process. Substitution into Lemma 4.10 and integration using partial fractions yield the assertion. $\qquad\square$

The results of using MATLAB's symbolic toolbox to differentiate and evaluate the characteristic function given in Corollary 4.11 are given in Table 7. Clearly, the first four moments are still fit very well despite the approximation errors involved.

	$m_{1,1/250}$	$m_{2,1/250}$	$m_{3,1/250}$	$m_{4,1/250}$
Data	3.5777×10^{-4},	1.9400×10^{-4}	-8.5630×10^{-7},	3.3018×10^{-7}
Model	3.5777×10^{-4}	1.9404×10^{-4}	-8.5680×10^{-7}	3.2683×10^{-7}
APE	$< 10^{-11}\%$	$< 0.02\%$	$< 0.06\%$	$< 1.02\%$

Table 7: Empirical moments of data, exact theoretical moments of the model fitted with Algorithm 4.4, and the corresponding absolute percentage errors.

We can also compute the exact autocorrelation and crosscorrelation functions of the returns and squared returns. They are plotted together with the corresponding approximations and their empirical counterparts in Figure 4. Again, the approximation errors involved turn out to be negligible compared to the variance of the corresponding estimators. Furthermore, it is clearly visible that while the positive autocorrelation of the returns and the positive crosscorrelation between the returns and the squared returns are of course negative features of the model from a theoretical point of view, the size of these effects is very small. Hence we can conclude that the second-order structure of the data is still fit satisfactorily for practical purposes.

4.7   Estimation of the current level of volatility

We now propose an approach to estimate the current level of volatility in the case $c_1 \neq 0$. Assuming $\mu = 0$ and y follows an OU process, [2, Section 5.4.3] and Theorem 2.2 yield the following state-space representation of $(X_{(n)}, X_{(n)}^2)$:

$$\begin{pmatrix} X_{(n)} \\ X_{(n)}^2 \end{pmatrix} = \begin{pmatrix} c_1(Y_{n\Delta} - Y_{(n-1)\Delta}) \\ (Y_{n\Delta} - Y_{(n-1)\Delta}) + c_1^2(Y_{n\Delta} - Y_{(n-1)\Delta})^2 \end{pmatrix} + u_n,$$

where the vector martingale difference sequence u_n satisfies, for $\Delta \downarrow 0$,

$$\mathrm{Var}(u_{1n}) = \Delta\xi, \quad \mathrm{Cov}(u_{1n}, u_{2n}) = c_3\Delta\xi + 2c_1\Delta^2(\omega^2 + \xi^2) + O(\Delta^3),$$
$$\mathrm{Var}(u_{2n}) = c_4\Delta\xi + (4c_1c_3 + 2)\Delta^2(\omega^2 + \xi^2) + O(\Delta^3),$$

and

$$\begin{pmatrix} \lambda(Y_{(n+1)\Delta} - Y_{n\Delta}) \\ y_{(n+1)\Delta} \end{pmatrix} = \begin{pmatrix} 0 & 1 - e^{-\lambda\Delta} \\ 0 & e^{-\lambda\Delta} \end{pmatrix} \begin{pmatrix} \lambda(Y_{n\Delta} - Y_{(n-1)\Delta}) \\ y_{n\Delta} \end{pmatrix} + w_n,$$
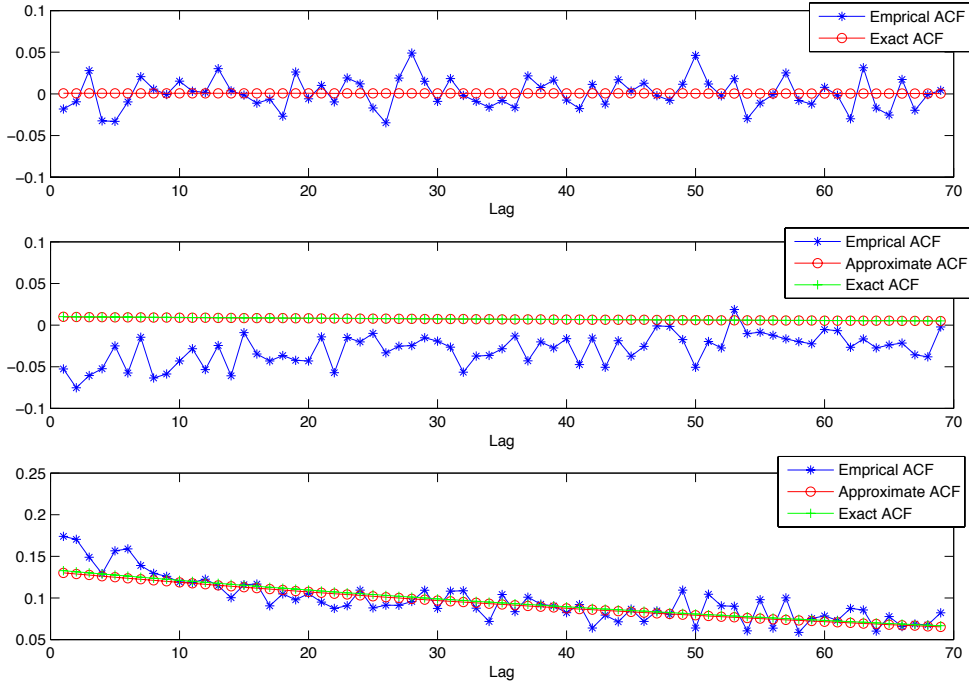
Figure 4: Empirical, approximate and exact autocorrelation functions of the log returns (first), crosscorrelation function of the returns and the squared returns (second) autocorrelation function of the squared log returns (third).

with IID noise w_n (uncorrelated with u_n) satisfying

$$\mathrm{E}(w_n) = \xi \begin{pmatrix} e^{-\lambda\Delta} - 1 + \lambda\Delta \\ 1 - e^{-\lambda\Delta} \end{pmatrix},$$

$$\mathrm{Var}(w_n) = 2\omega^2 \begin{pmatrix} \lambda\Delta - 2(1 - e^{-\lambda\Delta}) + \frac{1}{2}(1 - e^{-2\lambda\Delta}) & \frac{1}{2}(1 - e^{-\lambda\Delta})^2 \\ \frac{1}{2}(1 - e^{-\lambda\Delta})^2 & \frac{1}{2}(1 - e^{-2\lambda\Delta}) \end{pmatrix}.$$

While the nonlinearity of this representation prohibits the use of the Kalman filter, it is still possible to use the extended Kalman filter by neglecting terms of order $O(\Delta^3)$ or higher once again. Despite the approximations involved, the results shown in Figure 5 suggest that it is still possible to obtain decent estimates of the volatility in this way.
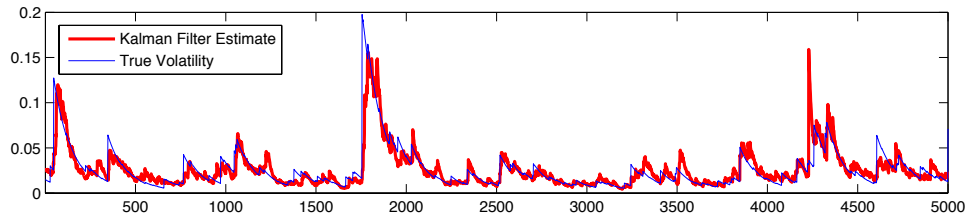
Figure 5: Sample paths of IG-OU process (thin line) with parameters as in Examples 4.7 and the approximate extended Kalman filter estimate (thick line) obtained from the corresponding NIG-IG-OU process with parameters as in Examples 4.7, 4.9.

References

[1] T. Andersen, L. Benzoni, and J. Lund. An empirical investigation of continuous-time equity return models. *Journal of Finance*, 57:1239–1284, 2002.

[2] O. Barndorff-Nielsen and N. Shephard. Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial economics. *Journal of the Royal Statistical Society, Series B*, 63:167–241, 2001.

[3] O. Barndorff-Nielsen and N. Shephard. Impact of jumps on returns and realised variances: Econometric analysis of time-deformed Lévy processes. *Journal of Econometrics*, 131:217–252, 2006.

[4] F. Benth, K. Karlsen, and K. Reikvam. Merton's portfolio optimization problem in a Black and Scholes market with non-Gaussian stochastic volatility of Ornstein-Uhlenbeck type. *Mathematical Finance*, 13:215–244, 2003.

[5] P. Carr, H. Geman, D. Madan, and M. Yor. Stochastic volatility for Lévy processes. *Mathematical Finance*, 13:345–382, 2003.

[6] M. Chernov, R. Gallant, E. Ghysels, and G. Tauchen. Alternative models for stock price dynamics. *Journal of Econometrics*, 116:225–257, 2003.

[7] R. Cont and P. Tankov. *Financial Modelling with Jump Processes*. Chapman & Hall/CRC, Boca Raton, 2004.

[8] R. Cont, P. Tankov, and E. Voltchkova. Hedging with options in models with jumps. In *Stochastic analysis and applications: The Abel Symposium 2005*, pages 197–217. Springer, Berlin, 2007.

[9] D. Duffie, D. Filipović, and W. Schachermayer. Affine processes and applications in finance. *The Annals of Applied Probability*, 13:984–1053, 2003.

[10] B. Eraker, M. Johannes, and N. Polson. The impact of jumps in volatility and returns. *The Journal of Finance*, 58:1269–1300, 2003.

[11] V. Genon-Catalot, T. Jeantheau, and C. Larédo. Stochastic volatility models as hidden Markov models and statistical applications. *Bernoulli*, 6:1051–1079, 2000.

[12] L. Hansen. Large sample properties of generalized method of moments estimators. *Econometrica*, 50:1029–1054, 1982.

[13] A. Harvey. *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press, Cambridge, 1989.

[14] S. Haug, C. Klüppelberg, A. Lindner, and M. Zapp. Method of moment estimation in the COGARCH(1,1) model. *The Econometrics Journal*, 10:320–341, 2007.

[15] F. Hubalek and C. Sgarra. Quadratic hedging for the Bates model. *International Journal of Theoretical and Applied Finance*, 10:873–885, 2007.

[16] I. Ibragimov and Y. Linnik. *Independent and stationary sequences of random variables*. Wolters-Noordhoff Publishing, Groningen, 1971.

[17] J. Jacod and A. Shiryaev. *Limit theorems for stochastic processes*, volume 288 of *Grundlehren der Mathematischen Wissenschaften*. Springer-Verlag, Berlin, second edition, 2003.

[18] G. Jiang and J. Knight. Estimation of continuous-time processes via the empirical characteristic function. *Journal of Business & Economic Statistics*, 20:198–212, 2002.

[19] J. Kallsen. A didactic note on affine stochastic volatility models. In Yu. Kabanov, R. Liptser, and J. Stoyanov, editors, *From Stochastic Calculus to Mathematical Finance*, pages 343–368. Springer, Berlin, 2006.

[20] J. Kallsen and R. Vierthauer. Quadratic hedging in affine stochastic volatility models. *Review of Derivatives Research*, 12:3–27, 2009.

[21] H. Kraft. Optimal portfolios and Heston's stochastic volatility model: An explicit solution for power utility. *Quantitative Finance*, 5:303–313, 2005.

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

[23] J. Muhle-Karbe. *On Utility-Based Investment, Pricing and Hedging in Incomplete Markets*. Dissertation Technische Universität München, 2009.

[24] C. Pigorsch and R. Stelzer. A multivariate Ornstein-Uhlenbeck type stochastic volatility model. Preprint, 2008.

[25] K. Sato. *Lévy Processes and Infinitely Divisible Distributions*. Cambridge University Press, Cambridge, 1999.

[26] W. Schoutens. *Lévy Processes in Finance*. Wiley, New York, 2003.

[27] A. Shiryaev. *Probability*. Springer, New York, second edition, 1995.

[28] M. Sørensen. Prediction-based estimating functions. *The Econometrics Journal*, 3:123–147, 2000.

[29] A. van der Vaart. *Asymptotic statistics*, volume 3 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 1998.