October 11, 2012

PARAMETER ESTIMATION FOR THE DRIFT OF A TIME-INHOMOGENEOUS JUMP DIFFUSION PROCESS

BRICE FRANKE AND THOMAS KOTT

Abstract. This work deals with parameter estimation for the drift of jump diffusion processes which are driven by a Lévy process and whose drift term is linear in the parameter. In contrast to the commonly used maximum likelihood estimator, our proposed estimator has the practical advantage that its calculation does not require the evaluation of the continuous part of the sample path. In the important case of an Ornstein-Uhlenbeck-type jump diffusion, which is a widely used model, we prove consistency and asymptotic normality.

1. Introduction

In statistical inference for time-continuous stochastic processes, parameter estimators that are based on the observation of the entire time-continuous path are natural objects to study: These estimators have often a closed-form representation in terms of stochastic integrals such that large sample results like consistency and asymptotic normality may be obtained by using techniques from stochastic analysis, see Section 5. In many situations continuous time estimators can be fairly approximated by their discrete time versions, see Lemma 3.1 for a representation.

There exists a large number of publications on drift parameter estimation for time-continuously observed diffusion processes. Maximum likelihood estimation is thereby, as well as in many other fields of statistical inference, the most commonly used estimation method. For a continuous diffusion process, which is an important model in many applied fields, with stochastic differential

$$dX_t = f(t, X_t, \theta)dt + dB_t, \quad 0 \leq t \leq T, \tag{1}$$

where $(B_t)_{t\geq 0}$ denotes Brownian motion and θ the unknown parameter, maximum likelihood estimation is based on Girsanov's theorem which provides an expression of the likelihood function. The resulting maximum likelihood estimator requires the computation of integrals of the form

$$\int_0^t f(t, X_t, \theta)dX_t. \tag{2}$$

Asymptotic properties of maximum likelihood estimates from time-continuous realizations of the process in (1) can be found e.g. in [3] and [17]. Given time-continuous observations $\{X_t, 0 \leq t \leq T\}$ the integral in (2) is approximated by an Itô sum using time-discrete increments of the sample path.

Especially in mathematical finance, an important generalization of model (1) is the jump diffusion process allowing for the possibility of discontinuities and a wide variety of marginal

Key words and phrases. Time-inhomogeneous diffusion process, time-continuous sample, least squares estimation, maximum likelihood

Acknowledgement. This work was partly supported by the Collaborative Research Center 'Statistical modeling of non-linear dynamic processes' (SFB 823) of the German Research Foundation.

distributions, see [2] and [6] for some applications. The jump diffusion process is defined as the solution to

$$dX_t = f(t, X_t, \theta)dt + dL_t, \quad 0 \leq t \leq T, \tag{3}$$

where $(L_t)_{t\geq 0}$ is a homogeneous Lévy process, and the application of the maximum likelihood approach to the Girsanov density yields estimators that are based on

$$\int_0^T f(t, X_t, \theta) dX_t^{0,c}$$

where $X_t^{0,c}$ is the continuous local martingale part of the process (see section 3 for the exact definition). This integral cannot be computed without further ado since the process $(X_t^c)_{t\geq 0}$ is not observed separately in practice. Large sample results on the maximum likelihood estimator for jump diffusion processes are derived in [21] and [22].

In this treatment we present an alternative continuous-time estimator for the drift parameter θ of the jump diffusion process given in (3) where the drift term is linear in the parameter, that means that the process solves

$$dX_t = f(t, X_t)\theta\, dt + dL_t. \tag{4}$$

Our estimator is derived by making use of the least squares method. In detail, we first regard the discretized version of the stochastic differential equation (4) and apply ordinary least squares estimation. In doing so, we obtain a time-discrete estimator. Thereafter, we take the limit as the discretization step Δt goes toward zero and get thereby a continuous time estimator of the drift parameter. Note that the resulting least squares estimator does not coincide with the trajectory fitting estimator (see Section 2.2.3 in [18]) which is sometimes referred to as time-continuous least squares estimator as well.

The crucial point of this work is the fact that, unlike the maximum likelihood estimator, our estimator requires the computation of integrals of the form (2) which can be calculated from the given data and which do not require further investigation determining the continuous part of the sample path.

In Sections 5, 6 and 7 we prove strong consistency and asymptotic normality of our time-continuous least squares estimator for a time-inhomogeneous, mean-reverting Ornstein-Uhlenbeck process of the form (4) provided with a periodic drift. The case of a continuous driving process in this Ornstein-Uhlenbeck model is studied by the authors in [7]. Note that mean reversion, periodicity and the occurrence of jumps are meaningful properties of, for example, energy commodity and particularly electricity data, cf. [11]. In the case of the ordinary Ornstein-Uhlenbeck process with jumps, [12] studies the large sample behavior of the time-continuous trajectory fitting estimator, various time-discrete estimation techniques are investigated in [4],[8], [13] and [23].

2. Description of the model

Suppose we are given a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, P)$ where $\mathcal{F}_0$ contains all P-null sets of $\mathcal{F}$ and where $\mathcal{F}_t$ is right continuous, i.e. $\mathcal{F}_t = \cap_{u>t}\mathcal{F}_u$. Let $(L_t)_{t\geq 0}$ be a Lévy process, that is an $\mathcal{F}_t$-adapted process which is continuous in probability and which has independent and stationary increments. Throughout this work we want to consider its unique càdlàg (right-continuous with left limits) modification. The famous Lévy-Itô decomposition

([1], Theorem 2.4.16 on p. 108) gives the path-wise representation

$$L_t = bt + \sigma B_t + \int_0^t \int_{|x|<1} x\tilde{q}_L(dt,dx) + \int_0^t \int_{|x|\geq 1} xq_L(dt,dx) \tag{5}$$

where $(B_t)_{t\geq 0}$ is a (standard) Brownian motion, $b \in \mathbb{R}$, $\sigma > 0$ and where q_L denotes the Poisson random measure associated with $(L_t)_{t\geq 0}$ while $\tilde{q}_L$ is the corresponding compensated measure. In detail, q_L is a random measure on $\mathbb{R}_+ \times (\mathbb{R}\backslash\{0\})$ defined by

$$q_L(t,A) = \#\{0 \leq s \leq t : \Delta L_s \in A\} = \sum_{0\leq s\leq t} \mathbf{1}_A(\Delta L_s)$$

for all Borel sets $A \in \mathbb{R}\backslash\{0\}$. Thereby, we use the notation $\Delta L_s = L_s - L_{s-}$, $L_{s-} = \lim_{\varepsilon\to 0} L_{s-\varepsilon}$ and $\mathbf{1}_A$ for the indicator function of the set A. Further, the compensated Poisson random measure is given by $\tilde{q}_L(dt,dx) = q_L(dt,dx) - dt\nu(dx)$ where ν is the Lévy measure associated with $(L_t)_{t\geq 0}$ satisfying $\int_{\mathbb{R}\backslash\{0\}}(x^2 \wedge 1)\nu(dx) < \infty$. It holds that $b = E(L_1) - \int_{|x|\geq 1} x\nu(dx)$. We also point out that the Poisson random measure $q_L(dt,dx)$ and the Brownian motion B representing L are independent.

The model of interest is the jump diffusion process $(X_t)_{t\geq 0}$ solving the stochastic differential equation

$$dX_t = f(t,X_t)\theta\, dt + dL_t, \quad X_0 = X_*, \tag{6}$$

where

$$f(t,x) = (f_1(t,x),\ldots,f_p(t,x)), \quad p \in \mathbb{N},$$

and where each $f_i(t,x)$ is a known, real-valued function on $[0,\infty)\times\mathbb{R}$. Further, let the random variable X_* be independent of the Lévy process and $E(X_*^2) < \infty$. The drift parameter $\theta \in \mathbb{R}^p$ is unknown and has to be estimated. We require that the distribution of X_* does not depend on θ otherwise the Radon-Nikodym derivative given in Proposition 4.1 would contain an additional factor, see [18] (p. 37) for details.

Note that equation (6) is a short form of the integral equation

$$X_t - X_* = \int_0^t f(s,X_s)\theta\, ds + L_t.$$

We implicitly assume that the well-known Lipschitz and linear growth conditions on the drift function f are satisfied (see [15], Theorem III.2.32 on p. 145) such that a unique càdlàg solution to (6) exists.

3. Least squares estimator

In this section we introduce a least squares estimator for the drift parameter θ. The derivation is based on a discretization of the stochastic differential equation (6) to which the ordinary least squares approach is applied. Taking the limit as the refinement improves yields a time-continuous estimator.

The stochastic differential equation (6) can be discretized on a time interval $[0,T]$ to the difference equation

$$X_{(i+1)\Delta t} - X_{i\Delta t} = f(i\Delta t, X_{i\Delta t})\theta\, \Delta t + (L_{(i+1)\Delta t} - L_{i\Delta t}), \quad i = 0,1,\ldots,N, \tag{7}$$

where $N = \lfloor T/\Delta t\rfloor - 1$ and where $\Delta t > 0$ denotes the constant time increment. Here $\lfloor x\rfloor$ denotes the integer part of x. The structure of (7) is similar to that of the classical linear model. Even though the conditions of the Gauss-Markov Theorem for linear models are not fulfilled we want to apply least squares estimation which is based on the minimization of the functional

$$g : \theta \mapsto \sum_{i=0}^{N} \left(X_{(i+1)\Delta t} - X_{i\Delta t} - f(i\Delta t, X_{i\Delta t})\theta \, \Delta t \right)^2 .$$

Lemma 3.1. *The solution vector $\tilde{\theta}_{T,\Delta t}$ to the minimization problem $g(\theta) \to \min$ is given by*

$$\tilde{\theta}_{T,\Delta t} = Q_{T,\Delta t}^{-1} R_{T,\Delta t}$$

where $Q_{T,\Delta t} = \left(\sum_{i=0}^{N} f_j(i\Delta t, X_{i\Delta t}) f_k(i\Delta t, X_{i\Delta t}) \Delta t \right)_{1 \leq j,k \leq p} \in \mathbb{R}^{p \times p}$ and

$$R_{T,\Delta t} = \left(\sum_{i=0}^{N} f_1(i\Delta t, X_{i\Delta t})(X_{(i+1)\Delta t} - X_{i\Delta t}), \ldots, \sum_{i=0}^{N} f_p(i\Delta t, X_{i\Delta t})(X_{(i+1)\Delta t} - X_{i\Delta t}) \right)^t \in \mathbb{R}^p.$$

Proof. By general theory of least squares estimation in linear models, the solution to the minimization problem $g(\theta) \to \min$ is given by

$$\theta_{\min} = (A^{\mathrm{T}} A)^{-1} A \mathrm{T} D \tag{8}$$

where

$$A = \Delta t \begin{pmatrix} f_1(0, X_0) & \ldots & f_p(0, X_0) \\ f_1(\Delta t, X_{\Delta t}) & \ldots & f_p(\Delta t, X_{\Delta t}) \\ \vdots & \ddots & \vdots \\ f_1(N\Delta t, X_{N\Delta t}) & \ldots & f_p(N\Delta t, X_{N\Delta t}) \end{pmatrix}, \quad D = \begin{pmatrix} X_{\Delta t} - X_0 \\ X_{2\Delta t} - X_{\Delta t} \\ \vdots \\ X_{(N+1)\Delta t} - X_{N\Delta t} \end{pmatrix}.$$

Hence, the products in equation (8) can be calculated to be

$$A^{\mathrm{T}} D = \Delta t R_{T,\Delta t}$$

and

$$A^{\mathrm{T}} A = \Delta t Q_{T,\Delta t}.$$

Thus we get $\tilde{\theta}_{T,\Delta t} = (A^{\mathrm{T}} A)^{-1} A^{\mathrm{T}} D = Q_{T,\Delta t}^{-1} R_{T,\Delta t}$. □

Now a continuous-time estimator can be derived from the least squares estimator by considering $\Delta t \to 0$. Note that any càdlàg function can be uniformly approximated on finite intervals by a sequence of step functions since it has countably many discontinuities on finite intervals. Hence, it is Riemann-integrable. This justifies the following convergence of the entries of $Q_{T,\Delta t}$ (as $\Delta t \to 0$):

$$\sum_{i=0}^{N} f_l(i\Delta t, X_{i\Delta t}) f_m(i\Delta t, X_{i\Delta t}) \Delta t \to \int_0^T f_l(t, X_t) f_m(t, X_t) dt \tag{9}$$

since $f_j(t, X_t)$ has càdlàg paths (because X_t has càdlàg paths) and the left-hand side of (9) is a Riemann sum. Regarding the entries of $R_{T,\Delta t}$ it holds that

$$\sum_{i=0}^{N} f_j(i\Delta t, X_{i\Delta t}) \cdot (X_{(i+1)\Delta t} - X_{i\Delta t}) \to \int_0^T f_j(t, X_{t-}) dX_t$$

uniformly on compacts in probability since X_t is a semi-martingale with càdlàg paths, see [20] (Theorem II.21 on p. 64). We have thus proved the following proposition.

Proposition 3.2. *As $\Delta t \to 0$, the least squares estimator $\tilde{\theta}_{T,\Delta t}$ converges in probability to $\hat{\theta}_T = Q_T^{-1} R_T$ where $Q_T = \left(\int_0^T f_j(t, X_t) f_k(t, X_t) dt\right)_{1 \le j,k \le p} \in \mathbb{R}^{p \times p}$ and*

$$R_T = \left(\int_0^T f_1(t, X_{t-}) dX_t, \dots, \int_0^T f_p(t, X_{t-}) dX_t\right)^{\mathrm{T}} \in \mathbb{R}^p.$$

We call the estimator $\hat{\theta}_T$ continuous-time least squares estimator.

Remark 1. We implicitly assumed that Q_T is invertible. This condition holds for many reasonable models, like for jump diffusions of Ornstein-Uhlenbeck form. However, in the case of a singular matrix Q_T, one has to find solutions $\gamma \in \mathbb{R}^p$ to the normal equations

$$Q_T \gamma = R_T$$

and make further constraints in order to determine a proper estimator. Note that we have the same possible ambiguity of the solution vector in the case of maximum likelihood estimation which is presented in the next section.

Remark 2. The referee for this article suggested another least square analogy which also leads to the same estimator: in a discrete linear model $y_i = f_i \theta + \epsilon_i;\ i = 1, ..., n$ the least square estimator also has to minimize the expression $-2 \sum y_i f_i \theta + \sum (f_i \theta)^2$. In our case the equation leads to the continuous linear model $dX_t = f_t \theta dt + dL_t$ and it is then manifest to look for an estimator, which minimize the expression

$$-2 \int_0^T f_t \theta dX_t + \int_0^T \theta^{\mathrm{T}} f_t^{\mathrm{T}} f_t \theta dt.$$

The resulting minimizer turns out to have the expression $\hat{\theta}_T = Q_T^{-1} \int_0^T f_t^{\mathrm{T}} dX_t$, where Q_T^{-1} denotes the inverse of the matrix $Q_T := \int_0^T f_t^{\mathrm{T}} f_t dt$.

4. Maximum likelihood estimation

We want to demonstrate the practical advantage of the continuous-time least squares estimator introduced in the previous section over the maximum likelihood estimator. An extensive study of maximum likelihood methods for jump-type processes has been presented in [22]. Let $D[0, T]$ denote the space of càdlàg functions from $[0, T]$ to $\mathbb{R}$. Denote by P_X and P_L the measures induced on $D[0, T]$ by the process $(X_t)_{0 \le t \le T}$ solving equation (6) and by the Lévy process $(L_t)_{0 \le t \le T}$, respectively. Note, that the law P_X implicitly depends on the parameter θ. Under the law P_X the canonical process $(\xi_t)_{t \ge 0}$ on $D[0, T]$ is a jump-diffusion with the same characteristics as the solution of the stochastic differential equation (6) and the process

$$\xi_t^{\theta,c} := \xi_t - \xi_0 - \sum_{s \le t} \mathbf{1}_{\{|\Delta \xi_s| \ge 1\}} \Delta \xi_s - \int_0^t \theta f(s, \xi_s) ds + t \int_{\{|y| > 1\}} y \nu(dy) - \int_0^t \int_{\{|y| \le 1\}} y (\mu - \nu)(ds, dy)$$

is a continious local martingale. Here the random measure μ is defined on $[0, T] \times (\mathbb{R}/\{0\})$ through

$$\mu(t, A) := \sum_{0 \le s \le t} \mathbf{1}_A(\Delta \xi_s).$$

In particular under P_L the process

$$\xi_t^{0,c} := \xi_t - \xi_0 - \sum_{s\le t} \mathbf{1}_{\{|\Delta\xi_s|\ge 1\}}\Delta\xi_s + t\int_{\{|y|>1\}} y\nu(dy) - \int_0^t\int_{\{|y|\le 1\}} y(\mu-\nu)(ds,dy)$$

is a continious local martingale (see [22] p.73). We mention that (see [22] p.80)

$$\xi_t^{\theta,c} = \xi_t^{0,c} - \int_0^t f(s,\xi_s)\theta ds.$$

The following proposition is an adaptation of the results from [22] to our situation. Like in [22] we will assume that the function f is bounded on $[0,n]\times\{x:|x|\le n\}$ for all $n\in\mathbb{N}$.

Proposition 4.1. *The measures P_X and P_L are equivalent on $D[0,T]$ and one has P_L-almost surely*

$$\frac{dP_X}{dP_L}(\xi) = \exp\Big(\int_0^T f(t,\xi_{t-})\theta\sigma^{-1}d\xi_t^{0,c} - \frac{1}{2}\int_0^T (f(t,\xi_{t-})\theta)^2\sigma^{-1}dt\Big).$$

Proof. Clearly one has that $P_X(A_T(\theta)<\infty) = P_L(A_T(\theta)<\infty) = 1$ since for every θ the random variable

$$A_T(\theta) := \int_0^T (f(t,\xi_t)\theta)^2 dt$$

is bounded. Furthermore the condition C from [22] is clearly satisfied and thus the formula on p.77 from [22] yields the particular form of the Radon Nikodym derivative. □

Remark 3. In our model specified in (6), the Radon-Nikodym derivative has a simpler form than the one in [22] because $X^{(T)}(\omega)$ and $L^{(T)}(\omega)$ exhibit the same jumps since

$$\Delta X_t = \lim_{\varepsilon\to 0}(X_t - X_{t-\varepsilon}) = \lim_{\varepsilon\to 0}\int_{t-\varepsilon}^t dX_t = \lim_{\varepsilon\to 0}\int_{t-\varepsilon}^t f(t,X_t)\theta\,dt + \lim_{\varepsilon\to 0}(L_t - L_{t-\varepsilon}) = \Delta L_t.$$

Hence, it holds for the point process q_X associated with $(X_t)_{t\ge 0}$ that

$$q_X(t,A) = q_L(t,A)$$

for all Borel sets $A\in\mathbb{R}\backslash\{0\}$ and all t. Consequently, the change of measure from P_L to P_X does not change the 'weights' of the discontinuities. So, in our framework, the density of P_X with respect to P_L does not include any term that accounts for the jumps.

Remark 4. Similar expressions for the Radon-Nikodym derivative $\frac{dP_X}{dP_L}$ have been described in [5] for financial models with jumps.

We will use the notation $X^{(T)}$ to denote the observed trajectory of the process X during the time-interval $[0,T]$. Based on X^T we define the continuous trajectory $X^{0,c} := F(X^T)$ where

$$F: D[0,T]\to C[0,T]; \xi\mapsto \xi^{0,c}.$$

The maximum likelihood estimator $\check\theta_T$ is defined by

$$\check\theta_T := \arg\max_\theta \frac{dP_X}{dP_L}(X^{(T)}),$$

where the Radon-Nikodym density dP_X/dP_L obviously depends on the parameter θ. Consistency and asymptotic normality for this estimator in quite general jump diffusion models have been studied in [22] (see p.84). In our model the maximum likelihood estimator has an explicit representation.

Proposition 4.2. *The maximum likelihood estimator $\check{\theta}_T$ is given by*

$$\check{\theta}_T = Q_T^{-1}\tilde{R}_T$$

where $Q_T = \left(\int_0^T f_j(t,X_t)f_k(t,X_t)dt\right)_{1\le j,k\le p} \in \mathbb{R}^{p\times p}$ *and*

$$\tilde{R}_T = \left(\int_0^T f_1(t,X_{t-})\,dX_t^{0,c},\ldots,\int_0^T f_p(t,X_{t-})\,dX_t^{0,c}\right)^{\mathrm{T}} \in \mathbb{R}^p.$$

Proof. Suppose we observe a sample path $X^{(T)} = \{X_t, 0 \le t \le T\}$ of the process with stochastic differential given in (6). The partial derivatives of the logarithm of the Radon Nikodym derivative in Proposition 4.1 are of the form

$$\frac{\partial}{\partial\theta_i}\ln\left(\frac{dP_X}{dP_L}(X^{(T)})\right) = \int_0^T \frac{\partial}{\partial\theta_i}f(t,X_{t-})\theta\sigma^{-1}dX_t^{0,c} - \int_0^T f(t,X_t)\theta\frac{\partial}{\partial\theta_i}f(t,X_t)\theta\sigma^{-1}dt, \tag{10}$$

$i = 1,\ldots,p$, and the single derivatives of the linear drift function are given by $\frac{\partial}{\partial\theta_i}f(t,X_t)\theta = f_i(t,X_t)$. Setting the derivatives in (10) equal zero results in a system of equations

$$\int_0^T f_i(t,X_{t-})dX_t^{0,c} - \int_0^T f(t,X_t)\check{\theta}_T f_i(t,X_t)dt = 0, \quad i = 1,\ldots,p,$$

which can be written as

$$\tilde{R}_T - Q_T\check{\theta}_T = 0$$

This proves our claim. □

Note that the expression for the maximum likelihood estimator $\check{\theta}_T$ is similar to the least squares estimator $\hat{\theta}_T$ given in Proposition 3.2. The discrepancy lies in the vectors $\tilde{R}_T$ and R_T. The entries of the latter are of the form $\int_0^T f_i(t,X_{t-})dX_t$ and can be calculated in practice without any difficulty. If time-discrete observations are available these integrals can be approximated by sums. In contrast to that, the integrals $\int_0^T f_i(t,X_{t-})dX_t^{0,c}$ in $\tilde{R}_T$ cannot be computed without further investigation due to the integrator which is the continuous part of the sample path. In practice, a discontinuous path is observed such that the continuous part of this path has to be determined by means of further techniques detaching discontinuities. This is a challenging issue unless the paths of the Lévy process have a finite number of jumps along the time interval $[0,T]$. In the case of time-discrete observations the always arising problem is to distinguish the jumps from the continuous points since the entire time-discrete sample looks discontinuous.

Remark 5. In the case of an ordinary diffusion process without jumps, that is the process solving

$$dX_t = f(t,X_t)\theta\,dt + dB_t, \quad X_0 = X_*,$$

where $(B_t)_{t\ge0}$ is a Brownian motion, the Radon-Nikodym derivative in Proposition 4.1 takes the form

$$\frac{dP_X}{dP_B}(X^{(T)}) = \exp\left(\int_0^T f(t,X_t)\theta\,dX_t - \frac{1}{2}\int_0^T (f(t,X_t)\theta)^2dt\right)$$

for data $X^{(T)}$. This expression is in accordance with the famous Girsanov Theorem, see [19] (Theorem 7.6, p. 246). The first integral is computed with respect to the entire path since there do not occur any discontinuities in this model. Note that the derivation of the maximum likelihood estimator goes in line with the considerations given above, i.e. differentiating and solving the resulting system of equations, such that $\check{\theta}_T = Q_T^{-1} R_T = \hat{\theta}_T$. That means, in this continuous diffusion model, the maximum likelihood method provides the same estimator as our least squares methodology presented in the previous section.

5. Consistency and asymptotic normality of the least squares estimator

In order to substantiate the convenience of the least squares estimator introduced in Section 3 we show (strong) consistency of this estimator for a concrete jump diffusion model. In the quite general setup with regard to the drift function f in (6) consistency requires general conditions on the convergence of the matrix Q_T in Proposition 3.2 which are not helpful for the application in concrete models.

Let us consider the time-inhomogeneous, mean-reverting Ornstein-Uhlenbeck process with jumps which we define as the solution to

$$dX_t = \Phi(t, X_t)\theta \, dt + dL_t, \quad X_0 = X_*, \tag{11}$$

where

$$\Phi(t, x) = (\varphi_1(t), \dots, \varphi_p(t), -x), \quad p \in \mathbb{N},$$

with known, real-valued functions $\varphi_1, \dots, \varphi_p$ and $E[X_*^2] < \infty$, X_* being independent of the Lévy process. We denote the parameter vector by $\theta = (\mu_1, \dots \mu_p, \alpha)^{\mathrm{T}} \in \mathbb{R}^p \times \mathbb{R}_+$. We assume that the drift function Φ is periodic in t, i.e.

$$\Phi(t + \ell, x) = \Phi(t, x) \quad \text{for all } x$$

where ℓ is a known period. Seasonality in the drift is a quite frequent phenomenon in applications, e.g. in commodity prices or temperature modeling. This assumed periodicity leads to the requirement $\varphi_j(t + \ell) = \varphi_j(t)$. Due to Gram-Schmidt orthogonalization, we may assume without loss of generality that $\varphi_1(t), \dots, \varphi_p(t)$ form an orthonormal system in $L_2([0, \ell], \frac{1}{\ell} d\lambda)$, that means that

$$\int_0^\ell \varphi_j(t)\varphi_k(t) dt = \begin{cases} 1, & j = k \\ 0, & j \neq k. \end{cases} \tag{12}$$

Henceforth, we will assume that we observe an integral multiple of the period length, i.e. that

$$T = N \, \ell \tag{13}$$

for some integer N. Moreover, we will assume without loss of generality that $\ell = 1$.

The driving process $(L_t)_{t \geq 0}$ is the right-continuous modification of a Lévy process of the form as described in Section 2, that is, as usual, a stochastic process that is continuous in probability with stationary and independent increments. For this model, we additionally require

$$\mathbb{E}\left[L_t^2\right] < \infty \tag{14}$$

for all t which is equivalent to the requirement $\int_{|x|>1} x^2 \mu(dx) < \infty$ where μ denotes the Lévy measure. Further, we assume that the measure ν is symmetric; i.e.: $\nu(A) = \nu(-A)$ for all Borel sets in $\mathbb{R} \backslash \{0\}$. This results in $b = 0$ in representation (5) and

$$\mathbb{E}[L_1] = 0 \tag{15}$$

implying that $(L_t)_{t \geq 0}$ is a square integrable martingale (see [14]).

Note that under the assumptions (12) and (13) the matrix Q_T in Proposition 3.2 simplifies to

$$Q_T = \begin{pmatrix} T\, I_p & -a_T \\ -a_T^{\mathrm{T}} & b_T \end{pmatrix} \tag{16}$$

where $a_T = \left(\int_0^T \varphi_1(t) X_t \, dt, \ldots, \int_0^T \varphi_p(t) X_t \, dt \right)^{\mathrm{T}}$, $b_T = \int_0^T X_t^2 dt$ and where I_p denotes the $(p \times p)$-identity matrix.

We have the following strong consistency result. Its proof is postponed to Section 6.

Theorem 1. *Let $\{X_t, 0 \le t \le T\}$ be observations of the periodic Ornstein-Uhlenbeck process introduced above satisfying (12), (13), (14) and (15). Then the least squares estimator given in Proposition 3.2 is consistent, i.e.*

$$\hat{\theta}_T \to \theta, \ \textit{almost surely,}$$

as $T \to \infty$.

We are also interested in the distributional asymptotic of the least squares estimator $\hat{\theta}_T$. For this we introduce the following function

$$\tilde{h}(t) = e^{-\alpha t} \sum_{j=1}^{p} \mu_j \int_{-\infty}^{t} e^{\alpha s} \varphi_j(s) ds \tag{17}$$

and the $(p+1) \times (p+1)$-matrix

$$C = \begin{pmatrix} I_p + \gamma \Lambda \Lambda^{\mathrm{T}} & -\gamma \Lambda \\ -\gamma \Lambda^{\mathrm{T}} & \gamma \end{pmatrix}$$

with

$$\Lambda := (\Lambda_1, ..., \Lambda_p)^{\mathrm{T}}$$

$$\gamma := \left(\int_0^1 \left(\tilde{h}(t) \right)^2 dt + \frac{c^2}{2\alpha} - \sum_{i=1}^{p} \Lambda_i^2 \right)^{-1}$$

where

$$\Lambda_i = \int_0^1 \varphi_i(t) \tilde{h}(t) dt, \quad i = 1, ..., p$$

and

$$c^2 := \sigma^2 + \int_{\mathbb{R}^d \setminus \{0\}} y^2 \nu(dy).$$

With those definitions we have the following asymptotic normality result. Its proof can be found in Section 7.

Theorem 2. *Let $\{X_t, 0 \le t \le T\}$ be observations of the periodic Ornstein-Uhlenbeck process introduced above satisfying (12), (13), (14) and (15). Then the least squares estimator given in Proposition 3.2 is asymptotically normal; i.e.:*

$$\sqrt{T} c^{-1} \left(\hat{\theta}_T - \theta \right) \xrightarrow{\mathcal{D}} N(0, C)$$

where C is defined in (5).

Here, $N(0, C)$ denotes a normally distributed random vector with zero-mean and covariance matrix C.

Remark 6. In the particular case where the driving process is Brownian motion, asymptotic normality of the least squares estimator is proved in [7] (Theorem 2). In that continuous model, least squares and maximum likelihood give the same estimator, see Remark 4.

6. Proof of consistency

The following representation of the least squares estimator is essential for the proof:

Proposition 6.1. *The least squares estimator $\hat{\theta}_T$ can be written as*

$$\hat{\theta}_T = \theta + Q_T^{-1} S_T \tag{18}$$

where

$$S_T = \begin{pmatrix} \int_0^T \varphi_1(t) dL_t \\ \vdots \\ \int_0^T \varphi_p(t) dL_t \\ -\int_0^T X_{t-} dL_t \end{pmatrix} \tag{19}$$

and where Q_T is given in (16).

Proof. By definition, we have $\hat{\theta}_T = Q_T^{-1} R_T$ where Q_T is given in (16) and

$$R_T = \begin{pmatrix} \int_0^T \varphi_1(t) dX_t \\ \vdots \\ \int_0^T \varphi_p(t) dX_t \\ -\int_0^T X_{t-} dX_t \end{pmatrix}$$

in the model considered here, see Proposition 3.2. Due to the stochastic differential equation (11) which is generating the data and which can be written as

$$dX_t = \Big(\sum_{j=1}^{p} \mu_j \varphi_j(t) - \alpha X_t \Big) dt + dL_t$$

the stochastic integrals in R_T are seen to be

$$\begin{aligned} \int_0^T \varphi_i(t) dX_t &= \sum_{j=1}^{p} \mu_j \int_0^T \varphi_i(t) \varphi_j(t) dt - \alpha \int_0^T \varphi_i(t) X_t dt + \int_0^T \varphi_i(t) dL_t, \quad i = 1, \dots, p, \\ \int_0^T X_{t-} dX_t &= \sum_{j=1}^{p} \mu_j \int_0^T \varphi_j(t) X_t dt - \alpha \int_0^T X_t^2 dt + \int_0^T X_{t-} dL_t. \end{aligned}$$

Observe that the set $\{t \in [0, T] : X_t \neq X_{t-}\}$ is countable and has thus zero-mass with respect to dt. Hence, we can conclude from these representations combined with (12) and (13) that

$$R_T = \begin{pmatrix} T\, I_p & -a_T \\ -a_T^{\mathrm{T}} & b_T \end{pmatrix} \theta + S_T$$

and it follows that $Q_T^{-1} R_T = \theta + Q_T^{-1} S_T$. □

Due to representation (18) the aim in the sequel is to show that

$$Q_T^{-1} S_T = \left(T Q_T^{-1}\right)\left(\frac{1}{T} S_T\right)$$

converges to zero almost surely as $T \to \infty$. Therefor, we will prove that $T Q_T^{-1}$ converges to a finite limit and that $\frac{1}{T} S_T$ tends to zero, almost surely respectively. Both of these results require some auxiliary results.

Lemma 6.2. *The solution to the stochastic differential equation (11) is explicitly given by*

$$X_t = e^{-\alpha t} X_0 + h(t) + Z_t \tag{20}$$

where

$$h(t) = e^{-\alpha t} \sum_{i=1}^{p} \mu_i \int_0^t e^{\alpha s} \varphi_i(s) ds$$

and

$$Z_t = e^{-\alpha t} \int_0^t e^{\alpha s} dL_s.$$

Proof. This is a direct consequence of the Itô lemma applied to the function $F(t, x) = e^{\alpha t} x$. See [20] (Theorem II.32, p. 71) or [16] (Theorem 7.6.1 and 7.6.4, p. 111 and 113) for reference to the general Itô formula. □

Remark 7. In the sequel we will use the following version of the well known Itô isometry for Levy driven stochastic integrals (see [14] p.56):

$$\begin{aligned}
\mathbb{E}\left[\left(\int_0^T f(X_t) dL_t\right)^2\right] &= \mathbb{E}\left[\left(\int_0^T f(X_t)\sigma dB_t\right)^2\right] + \mathbb{E}\left[\left(\int_0^T \int_{\mathbb{R}\setminus\{0\}} f(X_t) y q_L(dy, dt)\right)^2\right] \\
&= \sigma^2 \mathbb{E}\left[\int_0^T (f(X_t))^2 dt\right] + \mathbb{E}\left[\int_0^T \int_{\mathbb{R}\setminus\{0\}} (f(X_t))^2 y^2 \nu(dy) dt\right] \\
&= c^2 \mathbb{E}\left[\int_0^T (f(X_t))^2 dt\right].
\end{aligned}$$

Due to the time-dependence of h and Z in (20) the process $(X_t)_{t \geq 0}$ is not stationary in the ordinary sense such that the ergodic theorem is not directly applicable. We will introduce a solution to the stochastic differential equation (11) with time index $t \in \mathbb{R}$ instead of $t \geq 0$. By interpreting this process as a sequence of path-valued random variables we prove stationarity and ergodicity of this sequence.

Define the process

$$\tilde{X}_t = \tilde{h}(t) + \tilde{Z}_t \tag{21}$$

where $\tilde{h} : [0, \infty) \to \mathbb{R}$ is defined by

$$\tilde{h}(t) = e^{-\alpha t} \sum_{j=1}^{p} \mu_j \int_{-\infty}^t e^{\alpha s} \varphi_j(s) ds \tag{22}$$

and

$$\tilde{Z}_t = e^{-\alpha t} \int_{-\infty}^t e^{\alpha s} d\tilde{L}_s \tag{23}$$

whereby

$$\tilde{L}_s := L_s \mathbf{1}_{\{s\geq 0\}}(s) + \bar{L}_s \mathbf{1}_{\{s<0\}}(s)$$

by taking $\bar{L}_s$, when $s < 0$, to be an independent copy of $-L_{-(s-)}$ (see p. 214 in [1]). Constructed in this way, the process $(\tilde{L}_t)_{t\in\mathbb{R}}$ is a continuation of $(L_t)_{t\geq 0}$ to $\mathbb{R}$ such that $(\tilde{L}_t)_{t\in\mathbb{R}}$ is also a Lévy process with càdlàg paths. We know that each function φ_j is periodic with period one and its restriction to $[0,1]$ belongs to $L_1([0,1],\lambda)$. This implies that the function $\tilde{h}$ is well defined since we have

$$\Big| \int_{-\infty}^{0} e^{\alpha s}\varphi_j(s)ds \Big| \leq \sum_{j=0}^{\infty} \int_{j}^{j+1} e^{-\alpha s}|\varphi_j(s)|ds \leq \sum_{j=0}^{\infty} \int_0^1 e^{-\alpha j \nu}|\varphi_j(s)|ds < \infty.$$

We also mention that $\tilde{Z}$ is well defined, since we can use Remark 7 to prove that its second moments exists:

$$\mathbb{E}\left[\Big(\int_{-\infty}^{t} e^{\alpha s} d\tilde{L}_s \Big)^2 \right] = c^2 \mathbb{E}\left[\int_{-\infty}^{t} e^{2\alpha s} ds \right] < \infty.$$

In order to prove our main results we will need the following ergodicity result:

Lemma 6.3. *The sequence $(W_k)_{k\in\mathbb{N}}$ of $D[0,1]$-valued random variables defined by*

$$W_k(s) := \tilde{X}_{k-1+s}, \quad 0 \leq s \leq 1,$$

is stationary and ergodic.

Proof. Let $\tilde{h}_0$ be the restriction of the function $\tilde{h}$ to $[0,1]$. Since $\tilde{h}$ is periodic, we have the decomposition

$$\begin{aligned} W_k(t) &= \tilde{h}(k-1+t) + e^{-\alpha(k-1+t)} \int_{-\infty}^{k-1+t} e^{\alpha s} d\tilde{L}_s \\ &= \tilde{h}_0(t) + e^{-\alpha(k-1+t)} \int_{k-1}^{k-1+t} e^{\alpha s} d\tilde{L}_s + \sum_{l=-\infty}^{k-1} e^{-\alpha(k-1+t)} \int_{l-1}^{l} e^{\alpha s} d\tilde{L}_s. \end{aligned}$$

The time shifted Lévy martingale $\tilde{L}_s^{(l)} := \tilde{L}_{s+l}$ yields

$$\begin{aligned} W_k(t) &= \tilde{h}_0(t) + e^{-\alpha t} \int_0^t e^{\alpha s} d\tilde{L}_s^{(k-1)} + \sum_{l=-\infty}^{k-1} e^{-\alpha(k-l+t)} \int_0^1 e^{\alpha s} d\tilde{L}_s^{(l-1)} \\ &= \tilde{h}_0(t) + e^{-\alpha t} \int_0^t e^{\alpha s} d\tilde{L}_s^{(k-1)} + \sum_{j=-\infty}^{0} e^{-\alpha(1+t-j)} \int_0^1 e^{\alpha s} d\tilde{L}_s^{(j+k-2)}. \end{aligned}$$

Consequently, we can write

$$W_k(\cdot) = \tilde{h}_0(\cdot) + F_0(Y_{k-1}) + \sum_{j=-\infty}^{0} e^{\alpha(j-1)} F(Y_{j+k-2})$$

by using the almost surely defined functionals

$$F_0 : D[0,1] \to D[0,1];\ \omega \mapsto \left(t \mapsto e^{-\alpha t} \int_0^t e^{\alpha s} d\omega(s) \right),$$

$$F : D[0,1] \to D[0,1];\ \omega \mapsto \left(t \mapsto e^{-\alpha t} \int_0^1 e^{\alpha s} d\omega(s) \right)$$

and the $D[0,1]$-valued random variables

$$Y_l : s \mapsto (\tilde{L}_s^{(l)} - \tilde{L}_0^{(l)}),\ 0 \leq s < 1.$$

The sequence $(Y_l)_{l\in\mathbb{Z}}$ consists of independent and identically distributed random variables. This implies that $(W_k)_{k\in\mathbb{N}}$ is stationary and ergodic since each element of this sequence can be represented as a measurable function $G : (D[0,1])^{\mathbb{N}} \to D[0,1]$ of elements of the iid sequence $(Y_l)_{l\in\mathbb{Z}}$, i.e.

$$W_k = G(Y_{k-1}, Y_{k-2}, \ldots).$$

□

Lemma 6.4. *As $t \to \infty$, one has*

$$|\tilde{X}_t - X_t| \to 0, \quad \textit{almost surely.}$$

Proof. We have

$$\begin{aligned} |\tilde{X}_t - X_t| &\leq e^{-\alpha t}|X_0| + |\tilde{h}(t) - h(t)| + |\tilde{Z}_t - Z_t| \\ &\leq e^{-\alpha t}|X_0| + e^{-\alpha t}\sum_{i=1}^{p}\mu_i \int_{-\infty}^{0} e^{\alpha s}|\varphi_i(s)|ds + \left|e^{-\alpha t}\int_{-\infty}^{0} e^{\alpha s} d\tilde{L}_s\right|. \end{aligned}$$

Obviously, the first two terms on the right-hand side converge toward zero as $t \to \infty$. Further, we can use Remark 7 to see

$$\mathbb{E}\left[\left(\int_{-\infty}^{0} e^{\alpha s} d\tilde{L}_s\right)^2\right] = c^2\mathbb{E}\left[\int_{-\infty}^{0} e^{2\alpha s} ds\right] < \infty.$$

Hence we have shown that $E\left[\int_{-\infty}^{0} e^{\alpha s} d\tilde{L}_s\right]^2 < \infty$ which implies that $\left|\int_{-\infty}^{0} e^{\alpha s} d\tilde{L}_s\right| < \infty$ almost surely. It follows that

$$e^{-\alpha t}\left|\int_{-\infty}^{0} e^{\alpha s} d\tilde{L}_s\right| \to 0 \tag{24}$$

as $t \to \infty$.

□

Let us now turn to the matrix Q_T. Due to its simplified form in this model, see representation (16), its inverse can be explicitly computed.

Lemma 6.5. *The inverse of the matrix Q_T given in (16) can be computed to be*

$$Q_T^{-1} = \frac{1}{T}\begin{pmatrix} I_p + \gamma_T\Lambda_T\Lambda_T^{\mathrm{T}} & -\gamma_T\Lambda_T \\ -\gamma_T\Lambda_T^{\mathrm{T}} & \gamma_T \end{pmatrix} \tag{25}$$

where $\Lambda_T = (\Lambda_{T,1}, \ldots, \Lambda_{T,p})^{\mathrm{T}} = \frac{1}{T}a_T$, see (16), and

$$\gamma_T = \left(\frac{1}{T}\int_0^T X_t^2 dt - \sum_{i=1}^{p}\Lambda_{T,i}^2\right)^{-1}.$$

Proof. This can be proved by using the Frobenius matrix inversion formula which can be found in [10] (p. 73), or by directly multiplying the above expression for Q_T^{-1} with

$$Q_T = T \begin{pmatrix} I_p & -\Lambda_T \\ -\Lambda_T^{\mathrm{T}} & \frac{1}{T}\int_0^T X_t^2 dt \end{pmatrix}.$$

□

Remark 8. Note that the Frobenius matrix inversion formula holds if and only if the entries of the matrix on the right hand side of (25) are well-defined. We will see in the proof of Proposition 6.6 that the limit of $\frac{1}{T}Q_T^{-1}$ is well defined since we show that the limit of γ_T denoted by γ is greater than zero. Consequently, $\frac{1}{T}Q_T^{-1}$ exists almost surely if T is sufficiently large.

Proposition 6.6. *As $T \to \infty$, we have*

$$T\, Q_T^{-1} \to C, \qquad \textit{almost surely},$$

where C is the $(p+1)\times(p+1)$ matrix

$$C = \begin{pmatrix} I_p + \gamma\Lambda\Lambda^{\mathrm{T}} & -\gamma\Lambda \\ -\gamma\Lambda^{\mathrm{T}} & \gamma \end{pmatrix}$$

whereby $\Lambda = (\Lambda_1, \ldots, \Lambda_p)^{\mathrm{T}}$ and

$$\begin{aligned} \Lambda_i &= \int_0^1 \varphi_i(t)\tilde{h}(t)dt, \quad i = 1, \ldots, p, \\ \gamma &= \left(\int_0^1 (\tilde{h}(t))^2 dt + \frac{c^2}{2\alpha} - \sum_{i=1}^p \Lambda_i^2 \right)^{-1}. \end{aligned}$$

The function $\tilde{h}$ and the random variable $\tilde{Z}_t$ are specified in (22) and (23).

Proof. Consider the entries of the vector Λ_T first, i.e. $\frac{1}{T}\int_0^T X_t\varphi_j(t)dt$. From Lemma 6.4 we may conclude that

$$\frac{1}{T}\int_0^T \tilde{X}_t\varphi_j(t)dt - \frac{1}{T}\int_0^T X_t\varphi_j(t)dt \to 0,$$

almost surely. Since $(\tilde{X}_{k-1+s})_{k\in\mathbb{N}}$ is stationary and ergodic by Lemma 6.3, the ergodic theorem justifies

$$\frac{1}{T}\int_0^T \tilde{X}_t\varphi_j(t)dt = \frac{1}{T}\sum_{k=1}^T \int_{k-1}^k \tilde{X}_t\varphi_j(t)dt \to \mathbb{E}\left[\int_0^1 \tilde{X}_t\varphi_j(t)\,dt\right] = \int_0^1 \tilde{h}(t)\varphi_j(t)dt,$$

almost surely. Thus we have established convergence of $\Lambda_{T,j}$, $1 \le j \le p$. For the asymptotic behavior of γ_T, it suffices to investigate $\frac{1}{T}\int_0^T X_t^2 dt$.

It holds by (24) that

$$\left|\frac{1}{T}\int_0^T (Z_t - \tilde{Z}_t)\,dt\right| \le \frac{1}{T}\int_0^T \left|Z_t - \tilde{Z}_t\right| dt = \frac{1}{T}\int_0^T e^{-\alpha t}\left|\int_{-\infty}^0 e^{\alpha s} d\tilde{L}_s\right| dt \to 0,$$

almost surely, as $T \to \infty$. The ergodic theorem gives

$$\frac{1}{T}\int_0^T \tilde{Z}_t\,dt \to \mathbb{E}\left[\tilde{Z}_0\right] = 0,$$

compare the proof of Lemma 6.3 and we may conclude that

$$\limsup_{T\to\infty} \frac{1}{T}\int_0^T Z_t\,dt < \infty. \tag{26}$$

Observe that $h(t)$ is bounded and $X_0 < \infty$ almost surely. These facts combined with (26) and representation (20) justify

$$\limsup_{T\to\infty} \frac{1}{T}\int_0^T X_t\,dt = \limsup_{T\to\infty} \frac{1}{T}\int_0^T (e^{-\alpha t}X_0 + h(t) + Z_t)\,dt < \infty, \tag{27}$$

almost surely. It follows from (27) and Lemma 6.4 that

$$\frac{1}{T}\int_0^T \tilde{X}_t^2 dt - \frac{1}{T}\int_0^T X_t^2 dt = \frac{1}{T}\int_0^T (\tilde{X}_t + X_t)(\tilde{X}_t - X_t)dt \to 0.$$

Consequently, again by the ergodic theorem, we get

$$\begin{aligned}
\frac{1}{T}\int_0^T \tilde{X}_t^2 dt \quad = \quad \frac{1}{T}\sum_{k=1}^{T}\int_{k-1}^{k} \tilde{X}_t^2 dt \quad &\longrightarrow \quad \mathbb{E}\left[\int_0^1 \tilde{X}_t^2 dt\right] \\
&= \quad \mathbb{E}\left[\int_0^1 (\tilde{h}(t) + \tilde{Z}_t)^2 dt\right] \\
&= \quad \mathbb{E}\left[\int_0^1 (\tilde{h}(t))^2 dt + 2\int_0^1 \tilde{h}(t)\tilde{Z}_t dt + \int_0^1 \tilde{Z}_t^2 dt\right] \\
&= \quad \int_0^1 (\tilde{h}(t))^2 dt + E\left[\tilde{Z}_0^2\right] \\
&= \quad \int_0^1 (\tilde{h}(t))^2 dt + \frac{c^2}{2\alpha}.
\end{aligned}$$

By Bessel's inequality (see [25] p.33), we have

$$\sum_{i=1}^{p} \Lambda_i^2 \le \int_0^1 (\tilde{h}(t))^2 dt$$

and thus

$$\int_0^1 (\tilde{h}(t))^2 dt + \frac{c^2}{2\alpha} - \sum_{i=1}^{p} \Lambda_i^2 \ge \frac{c^2}{2\alpha} > 0.$$

□

Lemma 6.7. *The term $\frac{1}{\sqrt{T}}S_T$ is bounded in L^2.*

Proof. Note that $\frac{1}{\sqrt{T}}\int_0^T \varphi_i(t)\,dL_t$ is L^2-bounded since by Remark 7 one has

$$\mathbb{E}\left[\left(\frac{1}{\sqrt{T}}\int_0^T \varphi_i(t)\,dL_t\right)^2\right] = c^2\mathbb{E}\left[\frac{1}{T}\int_0^T \varphi_i(t)^2\,dt\right] = \frac{c}{T}\int_0^T \varphi_i(t)^2\,dt < \infty.$$

For the last entry of $\frac{1}{\sqrt{T}}S_T$ we use Remark 7 to prove that

$$\begin{aligned}&\mathbb{E}\left[\left(\frac{1}{\sqrt{T}}\int_0^T X_t\,dL_t\right)^2\right] = c^2\mathbb{E}\left[\frac{1}{\sqrt{T}}\int_0^T X_t^2\,dt\right]\\ &= \frac{c^2}{T}\mathbb{E}\left[\int_0^T \left(2e^{-\alpha t}X_0 h(t) + 2e^{-\alpha t}X_0 Z_t + e^{-2\alpha t}X_0^2 + 2h(t)Z_t + h(t)^2 + Z_t^2\right)\,dt\right].\end{aligned}$$

Since Z_t is a zero-mean random variable the expectation of the second and fourth term is zero. Moreover, $\mathbb{E}\left[Z_t^2\right] = \frac{c^2}{2\alpha}(1-e^{-2\alpha t}) < \infty$ such that

$$\sup_{T\geq 0}\frac{1}{T}\mathbb{E}\left[\int_0^T Z_t^2 dt\right] < \infty.$$

Further, the function h is bounded and $\mathbb{E}(X_0^2) < \infty$ resulting in

$$\sup_{T\geq 0}\frac{1}{T}\mathbb{E}\left[\int_0^T e^{-\alpha t}X_0 h(t)dt\right] < \infty$$

and

$$\sup_{T\geq 0}\frac{1}{T}\int_0^T h(t)^2 dt < \infty.$$

□

Proposition 6.8. *As $T \to \infty$, we have*

$$\lim_{T\to\infty}\frac{1}{T}S_T = 0, \text{ almost surely.}$$

Proof. Observe that S_T is a martingale since the Lévy process is a martingale due to condition (15). By Lemma 6.7, $\frac{1}{\sqrt{T}}S_T$ is L^2-bounded. Doob's maximal inequality for time-discontinuous sub-martingales, see Theorem 2.1.5 in [1] (p. 74), provides for any $\epsilon > 0$ that

$$\begin{aligned}\mathbb{P}\left(\sup_{2^k\leq T\leq 2^{k+1}}\frac{1}{T}|S_T| \geq \epsilon\right) &\leq \mathbb{P}\left(\sup_{2^k\leq T\leq 2^{k+1}}|S_T| \geq \epsilon 2^k\right)\\ &\leq \frac{4}{\epsilon^2 2^{2k}}\mathbb{E}\left[|S_{2^{k+1}}|^2\right] = O(2^{-k}).\end{aligned}$$

Applying the Borel-Cantelli theorem, we obtain $\limsup_{T\to\infty}\frac{1}{T}|S_T| \leq \epsilon$, almost surely, and thus we have shown that $S_T/T \to 0$. □

Remark 9. The referee suggested another proof for Proposition 6.8: we have seen in Lemma 6.7 that the components S^j of the martingale S satisfy $\sup_{T\geq 0}\frac{1}{T}E[(S_T^j)^2] < \infty$ and moreover one has $\langle S^j\rangle_T \sim T$ a.s.. An application of the martingale stability theorem then proves Proposition 6.8. See [24] (Theorem 3.3.1) for the martingale stability theorem.

Proof of Theorem 1. This follows directly from Proposition 6.6 and Proposition 6.8. □

7. Proof of asymptotic normality

We define the following $(p+1) \times (p+1)$-matrix

$$\Sigma := \begin{pmatrix} I_p & \Lambda \\ \Lambda^{\mathrm{T}} & \omega \end{pmatrix} \tag{28}$$

where

$$\Lambda_i := \int_0^1 \varphi_i(t)\tilde{h}(t)dt, \quad i = 1, \dots, p$$

$$\omega := \int_0^1 (\tilde{h}(t))^2 dt + \frac{c^2}{2\alpha}$$

and where the function $\tilde{h}(t)$ is defined in (17).

We know that from Proposition 6.1 that

$$\hat{\theta}_T = \theta + Q_T^{-1} R_T$$

with

$$R_T = \begin{pmatrix} \int_0^T \varphi_1(t)dL_t \\ \vdots \\ \int_0^T \varphi_p(t)dL_t \\ -\int_0^T X_{t-}dL_t \end{pmatrix}$$

Thus we have

$$\sqrt{T}\left(\hat{\theta}_T - \theta\right) = \left(TQ_T^{-1}\right)\frac{1}{\sqrt{T}}R_T.$$

We further know from Proposition 6.6 that as $T \to \infty$

$$TQ_T^{-1} \longrightarrow C = \Sigma^{-1} \quad \text{almost surely.}$$

Thus the following proposition is sufficient to prove Theorem 2

Proposition 7.1. *We have that*

$$\frac{1}{\sqrt{T}}R_T \xrightarrow{\mathcal{D}} N(0, \Sigma).$$

Proof. Let $\mathcal{F}_t^{(n)} := \sigma(L_s; s \leq nt)$. We use a general limit theorem presented in [9] to prove the weak convergence of the sequence of vector-valued martingales

$$R_t^{(n)} := \left(\frac{1}{\sqrt{n}}\int_0^{nt} \varphi_1(s)dL_s, ..., \frac{1}{\sqrt{n}}\int_0^{nt} \varphi_p(s)dL_s, \frac{1}{\sqrt{n}}\int_0^{nt} X_s dL_s\right)$$

toward a Brownian motion. We define the matrix valued processes

$$A_t^{(n)} := \begin{pmatrix} \Phi_n(t) & \Psi_n(t) \\ \Psi_n^{\mathrm{T}}(t) & \rho_n(t) \end{pmatrix}$$

with

$$\Phi_n^{ij}(t) := \frac{c^2}{n}\int_0^{nt} \varphi_i(s)\varphi_j(s)ds,$$

$$\Psi_n^i(t) := -\frac{c^2}{n}\int_0^{nt} \varphi_i(s)X_s ds,$$

$$\rho_n(t) := \frac{c^2}{n} \int_0^{nt} X_s^2 ds.$$

For two square-integrable $\mathcal{F}_t$-martingales $M_t; t \geq 0$ and $N_t, t \geq 0$ there exists a $\mathcal{F}_t$-adapted process noted $\langle M, N \rangle_t, t \geq 0$ such that the process

$$M_t \cdot N_t - \langle M, N \rangle_t; \ t \geq 0$$

is a $\mathcal{F}_t$-martingalesee ([14] p.53) . Since one has that $\langle L, L \rangle_t = c^2 t$ (use formula (3.9) on p.62 in [14]) it follows that

$$\left\langle \int \varphi_i dL, \int \varphi_j dL \right\rangle = \int \varphi_i \varphi_j d\langle L, L \rangle_s = c^2 \int \varphi_i \varphi_j ds$$

$$\left\langle \int \varphi_i dL, \int X dL \right\rangle = \int \varphi_i X d\langle L, L \rangle_s = c^2 \int \varphi_i X_s ds$$

$$\left\langle \int X dL, \int X dL \right\rangle = \int X^2 d\langle L, L \rangle_s = c^2 \int X_s^2 ds.$$

This implies that the processes

$$\begin{aligned} M_n^j(t) &:= \frac{1}{n} \int_0^{nt} \varphi_i(s) dL_s \int_0^{nt} \varphi_j(s) dL_s - \Phi_n^{ij}(t) \\ M_n(t) &:= \frac{1}{n} \int_0^{nt} \varphi_i(s) dL_s \int_0^{nt} X_s dL_s - \Psi_n^i(t) \\ M_n(t) &:= \frac{1}{n} \left(\int_0^{nt} X_s dL_s \right)^2 - \rho_n(t) \end{aligned}$$

are $\mathcal{F}_t^{(n)}$-martingales. Using Lemma 6.3 and Lemma 6.4 like in the proof of Proposition 6.6 we obtain that for $n \to \infty$

$$\Phi_n^{ij} \longrightarrow tc^2 \int_0^1 \varphi_i(s) \varphi_j(s) ds$$

and

$$\begin{aligned} \Psi_n^i(t) \longrightarrow -tc^2 \mathbb{E} \left[\int_0^1 \varphi_i(s) \tilde{X}_s d \right] &= -tc^2 \mathbb{E} \left[\int_0^1 \varphi_i(s) (\tilde{Z}_s + \tilde{h}(s)) ds \right] \\ &= -tc^2 \int_0^1 \varphi_i(s) \tilde{h}(s) ds. \end{aligned}$$

Further,

$$\begin{aligned} \rho_n(t) \longrightarrow -tc^2 \mathbb{E} \left[\int_0^1 \tilde{X}_s^2 ds \right] &= tc^2 \mathbb{E} \left[\int_0^1 (\tilde{Z}_s + \tilde{h}(s))^2 ds \right] \\ &= tc^2 \left(\mathbb{E} \left[\tilde{Z}_1^2 \right] + \int_0^1 (\tilde{h}(s))^2 ds \right). \end{aligned}$$

Thus the matrices $A^{(n)}(t)$ converge for $n \uparrow \infty$ toward the matrix $t\Sigma$, which is the covariance matrix of a Brownian motion process $W_t^{\Sigma}; t \geq 0$. We note that the process $A^{(n)}(t)$ has no jumps; i.e.:

$$\mathbb{E} \left[\sup_{0 \leq t \leq T} \left| A_t^{(n)} - A_{t-}^{(n)} \right| \right] = 0$$

Moreover, we have that

$$
\begin{aligned}
& \mathbb{E}\left[\sup_{0\leq t\leq T}\frac{1}{\sqrt{n}}\left|\int_0^{nt} X_s dL_s - \int_0^{nt-} X_s dL_s\right|^2\right] \\
\leq\; & \epsilon + \mathbb{E}\left[\left(\int_0^{nt}\int_{\mathbb{R}\setminus B_{\sqrt{n}\epsilon}(0)}\left|\frac{1}{\sqrt{n}}X(s)y\right| q_L(dy,ds)\right)^2\right] \\
=\; & \epsilon + \mathbb{E}\left[\frac{1}{n}\int_0^{nt}\int_{\mathbb{R}\setminus B_{\sqrt{n}\epsilon}(0)} (X(s)y)^2\,\nu(dy)ds\right] \\
=\; & \epsilon + \mathbb{E}\left[\frac{1}{n}\int_0^{nt}(X(s))^2 ds\int_{\mathbb{R}\setminus\{0\}} \mathbf{1}_{B^c_{\sqrt{n}\epsilon}(0)}(X_s y)y^2\nu(dy)\right] \\
=\; & \epsilon + \mathbb{E}\left[X_s^2\int_{\mathbb{R}\setminus\{0\}} \mathbf{1}_{B^c_{\sqrt{n}\epsilon}(0)}(X_s y)y^2\nu(dy)\right].
\end{aligned}
$$

By Lebesgue's theorem on dominated convergence the last expectation converges toward zero as $n\uparrow\infty$ since one has $\mathbb{P}$-almost surely

$$
\int_{\mathbb{R}\setminus\{0\}} \mathbf{1}_{B^c_{\sqrt{n}\epsilon}(0)}(X_s y)y^2\nu(dy) \longrightarrow 0 \quad \text{as } n\to\infty.
$$

The same argument can be applied for the other components of the process $R^{(n)}$. We thus have that

$$
\mathbb{E}\left[\sup_{0\leq t\leq T}\left|R_t^{(n)} - R_{t-}^{(n)}\right|^2\right] \longrightarrow 0.
$$

An application of the limit theorem for martingales in [9] shows that $R^{(n)}$ converges in distribution toward a Wiener process with covariance matrix Σ. □

Remark 10. In order to prove the distributional convergence of the processes $R^{(n)}$ toward W^{Σ} one could also use the more general results from [15].

Acknowledgement. This research grew from some discussions with Herold Dehling from Ruhr-Universität Bochum. Those discussions lead to the idea of introducing the least square estimator through discretization in section 3. We would like to thank Jeannette Woerner, TU Dortmund, and Rogemar Mamon, University of Western Ontario, for helpful advice on change of measure for jump diffusion processes. We are further grateful to Mathias Vetter, Ruhr-Universität Bochum, for his comments that improved the manuscript. We also thank the referee for many valuable comments which helped us to improve the paper.

References

[1] D. Applebaum (2004): *Lévy Processes and Stochastic Calculus.* Cambridge University Press.

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

[3] J. P. N. Bishwal (2008): *Parameter Estimation in Stochastic Differential Equations.* Springer-Verlag.

[4] P. J. Brockwell, R. A. Davis and Y. Yang (2007): Estimation for nonnegative Lévy-driven Ornstein-Uhlenbeck processes. *Journal of Applied Probability* **44**, 977–989.

[5] T. Chan (1999): Pricing contingent claims on stocks driven by Lévy processes. *The Annals of Applied Probability* **9**, 504–528.

[6] R. Cont and P. Tankov (2004): *Financial Modelling with Jump Processes*. Financial Mathematics Series, Chapman & Hall/CRC.
[7] H. Dehling, B. Franke and T. Kott (2011): Drift estimation for a periodic mean reversion process. *Statistical Inference for Stochastic Processes* **13**, 175-192.
[8] A. Diop and A. F. Yode (2010): Minimum distance parameter estimation for Ornstein-Uhlenbeck processes driven by Lévy process. *Statistics and Probability Letters* **80**, 122–127.
[9] S. Ethier and T. Kurtz (1986): *Markov Processes: Characterization and Convergence*. J. Wiley.
[10] F. R. Gantmacher (1986): *Matrizentheorie*. VEB Deutscher Verlag der Wissenschaften.
[11] H. Geman (2005): *Commodities and Commodity Derivatives.* J. Wiley.
[12] Y. Hu and H. Long (2007): Parameter estimation for Ornstein-Uhlenbeck processes driven by α-stable Lévy motions. *Communications on Stochastic Analysis* **1**, 175–192.
[13] Y. Hu and H. Long (2009): Least squares estimator for Ornstein-Uhlenbeck processes driven by α-stable Lévy motions. *Stochastic Processes and their Applications* **119**, 2465–2480.
[14] N. Ikeda and S. Watanabe (1989): *Stochastic Differential Equations and Diffusion Processes*, North-Holland.
[15] J. Jacod and A. N. Shiryaev (1987): *Limit Theorems for Stochastic Processes*. Grundlehren der mathematischen Wissenschaft **288**. Springer-Verlag.
[16] H. H. Kuo (2006): *Introduction to Stochastic Integration*. Springer-Verlag.
[17] Y. A. Kutoyants (1984): *Parameter Estimation for Stochastic Processes*. Heldermann Verlag.
[18] Y. A. Kutoyants (2004): *Statistical Inference for Ergodic Diffusion Processes*. Springer-Verlag.
[19] R. S. Lipster and A. N. Shiryayev (1977): *Statistics of Random Processes I.*, Springer-Verlag.
[20] P. Protter (2004): *Stochastic Integration and Differential Equations*. Springer-Verlag.
[21] M. Sørensen (1989): A note on the existence of a consistent maximum likelihood estimator for diffusions with jumps. *Markov Processes and Control Theory*, 229–234. Akademie-Verlag.
[22] M. Sørensen (1991): Likelihood methods for diffusions with jumps. *Statistical Inference in Stochastic Processes*, 67–105. Marcel Dekker, Inc.
[23] K. Spiliopoulos (2009): Method of moments estimation of Ornstein-Uhlenbeck processes driven by general Lévy process. *Publications de L'Institut de Statistique de l'Université de Paris* **53**, 3–17.
[24] W. Stout (1974): *Almost sure convergence*. Academic Press.
[25] H. Weyl: (1950) *The Theory of Groups and Quantum Mechanics*, Dover Publications, Inc.

Brice Franke, Département de Mathématique, Université de Bretagne Occidentale, 6 avenue Le Gorgeu, 29200 Brest, France
E-mail address: `brice.franke@univ-brest.fr`

Thomas Kott, Fakultät für Mathematik, Ruhr-Universität Bochum, 44780 Bochum, Germany
E-mail address: `thomas.kott@ruhr-uni-bochum.de`