Electronic Journal of Statistics
Vol. 13 (2019) 4508–4526
ISSN: 1935-7524
https://doi.org/10.1214/19-EJS1631

On parameter estimation of hidden ergodic Ornstein-Uhlenbeck process

Yury A. Kutoyants

Le Mans University, Le Mans, France and Tomsk State University, Tomsk, Russia
e-mail: kutoyants@univ-lemans.fr

Abstract: We consider the problem of parameter estimation for the partially observed linear stochastic differential equation. We assume that the unobserved Ornstein-Uhlenbeck process depends on some unknown parameter and estimate the unobserved process and the unknown parameter simultaneously. We construct the One-step MLE-process for the estimator of the parameter and describe its large sample asymptotic properties, including consistency and asymptotic normality. Using the Kalman-Bucy filtering equations we construct recurrent estimators of the state and the parameter.

MSC 2010 subject classifications: Primary 62M05, 62M20; secondary 62F12.
Keywords and phrases: Partially observed linear system, parameter estimation, hidden process, Kalman-Bucy filtration, ergodic process, One-step MLE-process.

Received March 2019.

1. Introduction

We are given a partially observed linear system, defined by the equations

$$\mathrm{d}X_t = aY_t\,\mathrm{d}t + \sigma\,\mathrm{d}W_t, \qquad X_0 = 0, \tag{1.1}$$

$$\mathrm{d}Y_t = -fY_t\,\mathrm{d}t + b\,\mathrm{d}V_t, \qquad Y_0 = \xi, \tag{1.2}$$

where $a \neq 0, \sigma \neq 0, b \neq 0$ and $f > 0$ are constants, $W^T = (W_t, 0 \le t \le T)$ and $V^T = (V_t, 0 \le t \le T)$ are two independent Wiener processes. The random variable $\xi \sim \mathcal{N}\left(0, d^2\right)$ is independent of W^T and V^T.

The system (1.1)-(1.2) is defined by the four parameters a, f, b, σ^2. Recall that the parameter σ^2 can be estimated without error by continuous time observations X^T as follows. By the Itô formula we can write

$$X_t^2 = 2\int_0^t X_s\,\mathrm{d}X_s + \sigma^2 t.$$

Hence, for any $t \in (0, T]$, we have the estimator

$$\hat\sigma_t^2 = t^{-1}X_t^2 - 2t^{-1}\int_0^t X_s\,\mathrm{d}X_s = \sigma^2,$$

and this estimator equals the true value. Therefore we consider only the estimation of the three other parameters f, b and a. Note that the consistent estimation of the three-dimensional parameter $\vartheta = (a, b, f)$ or two-dimensional parameter $\vartheta = (a, b)$ is impossible. The heuristic explication of this is given at the end of this section.

The observations are $X^T = (X_t, 0 \le t \le T)$ and the Ornstein-Uhlenbeck process Y^T is unobservable (hidden), i.e., we have partially observed linear model of observations.

We consider estimation of the one-dimensional parameters f, b and a separately given the continuous time observations X^T. The unknown parameter will be denoted by ϑ and we will assume that $\vartheta \in \Theta = (\alpha, \beta)$ for some constants $\alpha < \beta$. In all the cases the set Θ does not contain 0. Thus we are faced with three different problems: $\vartheta = f$, $\vartheta = b$ and $\vartheta = a$. In each problem we propose a two-step construction of asymptotically efficient estimator-process of recurrent nature. First we propose a preliminary consistent estimator ϑ_{T^δ} based on the observations $X^{T^\delta} = \left(X_t, 0 \le t \le T^\delta\right)$ with $\delta \in (1/2, 1)$. Then this estimator is used for construction of One-step MLE-process, which has recurrent structure. In the last section we discuss the possibilities of the joint estimation of two dimensional parameters $\vartheta = (f, b)$ and $\vartheta = (f, a)$.

Equations (1.1)-(1.2) is a prototypical model in the Kalman-Bucy filtering theory, which provides a closed form system of equations for the conditional expectation $m\left(t\right) = \mathbf{E}\left(Y_t | X_s, 0 \le s \le t\right)$ ([1], [9], [18]). The statistical problems for discretely observed hidden Markov processes were studied by many authors (see [2], [3], [6], [7] and the references therein). However, the literature on continuous time models is limited. For the results in continuous time setup, we refer the interested reader to [13] (linear and non linear partially observed systems with small noise), [6] (continuous-time hidden Markov models estimation), [4] and [11] (hidden telegraph process observed in the white Gaussian noise).

In the present paper we are particularly interested in the asymptotic behavior of the maximum likelihood estimator (MLE) $\hat\vartheta_T$ in the *large sample* asymptotic regime, i.e., when $T \to \infty$. The statistical problems for such observation models have been widely studied, motivated by the importance of the Kalman-Bucy filtering in engineering applications.

Let us now recall the definitions of the MLE in the case $\vartheta = f$, when the other two parameters a and b are known. As the parameters of the model take finite values and $\sigma^2 > 0$, the measures $\left\{\mathbf{P}_\vartheta^{(T)}, \vartheta \in \Theta\right\}$ induced by the observations (1.1) on the space of continuous functions on $[0, T]$ are equivalent. The likelihood ratio function ([18]) is given by the expression

$$L\left(\vartheta, X^T\right) = \exp\left\{\int_0^T \frac{a\, m\left(\vartheta, t\right)}{\sigma^2}\, \mathrm{d}X_t - \int_0^T \frac{a^2 m\left(\vartheta, t\right)^2}{2\sigma^2}\, \mathrm{d}t\right\}, \quad \vartheta \in \Theta. \tag{1.3}$$

Then the MLE $\hat\vartheta_T$ is defined by the equation

$$L(\hat\vartheta_T, X^T) = \sup_{\vartheta \in \Theta} L\left(\vartheta, X^T\right). \tag{1.4}$$

This means that to calculate $\hat{\vartheta}_T$ we need the values of the family of stochastic processes $(m(\vartheta,t), 0 \leq t \leq T), \vartheta \in \Theta$. The random process $m(\vartheta,\cdot)$ is solution of the Kalman-Bucy filtering equations (see [1], [9], [18])

$$\begin{aligned} \mathrm{d}m(\vartheta,t) &= -\vartheta\, m(\vartheta,t)\,\mathrm{d}t + \frac{\gamma(\vartheta,t)\,a}{\sigma^2}\left[\mathrm{d}X_t - a\, m(\vartheta,t)\,\mathrm{d}t\right] \\ &= -\left[\vartheta + \frac{\gamma(\vartheta,t)\,a^2}{\sigma^2}\right] m(\vartheta,t)\,\mathrm{d}t + \frac{\gamma(\vartheta,t)\,a}{\sigma^2}\mathrm{d}X_t, \end{aligned} \tag{1.5}$$

where $m(\vartheta,0) = \mathbf{E}_\vartheta(\xi|X_0) = 0$. The function $\gamma(\vartheta,t) = \mathbf{E}_\vartheta(m(\vartheta,t) - Y_t)^2$ is the solution of the Ricatti equation

$$\frac{\partial\gamma(\vartheta,t)}{\partial t} = -2\vartheta\,\gamma(\vartheta,t) - \frac{\gamma(\vartheta,t)^2\,a^2}{\sigma^2} + b^2, \qquad \gamma(\vartheta,0) = d^2. \tag{1.6}$$

Due to importance of this model in many applied problems, much engineering literature is concerned with identification of this model.

The behavior of the MLE was studied at least in three asymptotics:

- *Small noise in both equations* $\sigma = b = \varepsilon \to 0$ (T is fixed) [12], [13]

$$\frac{\hat{\vartheta}_\varepsilon - \vartheta}{\varepsilon} \Longrightarrow \mathcal{N}\left(0, \mathrm{I}_1(\vartheta)^{-1}\right).$$

- *Large sample* $T \to \infty$ (σ and b are fixed) [14]

$$\sqrt{T}\left(\hat{\vartheta}_T - \vartheta\right) \Longrightarrow \mathcal{N}\left(0, \mathrm{I}_2(\vartheta)^{-1}\right).$$

- *Small noise in observation* only, $\sigma = \varepsilon \to 0$ (T and b are fixed) [16]

$$\frac{\hat{\vartheta}_\varepsilon - \vartheta}{\sqrt{\varepsilon}} \Longrightarrow \mathcal{N}\left(0, \mathrm{I}_3(\vartheta)^{-1}\right).$$

In all three cases $\mathrm{I}_i(\vartheta)$ are corresponding Fisher informations. It was also shown that the polynomial moments of the scaled estimation error converge and the MLE is asymptotically efficient.

Let us remind that the simultaneous consistent estimation of the parameters a and b is impossible. According to (1.3) the Kullback-Leibler distance between the measures with different $\vartheta = (a,b)$ is

$$D_{K-L}(\vartheta,\vartheta_0) = \frac{1}{2\sigma^2}\int_0^T \mathbf{E}_{\vartheta_0}\left[am(\vartheta,t) - a_0 m(\vartheta_0,t)\right]^2 \mathrm{d}t.$$

Remark that if we denote $\tilde{m}(\vartheta,t) = am(\vartheta,t)$ and $\tilde{\gamma}(\vartheta,t) = a^2\gamma(\vartheta,t)$, then we obtain equations

$$\mathrm{d}\tilde{m}(\vartheta,t) = -\left[f + \frac{\tilde{\gamma}(\vartheta,t)}{\sigma^2}\right]\tilde{m}(\vartheta,t)\,\mathrm{d}t + \frac{\tilde{\gamma}(\vartheta,t)}{\sigma^2}\tilde{m}(\vartheta_0,t)\,\mathrm{d}t + \frac{\tilde{\gamma}(\vartheta,t)}{\sigma}\mathrm{d}\bar{W}_t,$$

$$
\begin{aligned}
{\rm d}\tilde m\left(\vartheta_0,t\right)&=-f\tilde m\left(\vartheta_0,t\right){\rm d}t+\frac{\tilde \gamma\left(\vartheta_0,t\right)}{\sigma}{\rm d}\bar W_t,\\
\frac{\partial \tilde\gamma\left(\vartheta,t\right)}{\partial t}&=-2f\,\tilde\gamma\left(\vartheta,t\right)-\frac{\tilde\gamma\left(\vartheta,t\right)^2}{\sigma^2}+a^2b^2,\qquad \tilde\gamma\left(\vartheta,0\right)=d^2a^2,\\
\frac{\partial \tilde\gamma\left(\vartheta_0,t\right)}{\partial t}&=-2f\,\tilde\gamma\left(\vartheta_0,t\right)-\frac{\tilde\gamma\left(\vartheta_0,t\right)^2}{\sigma^2}+a_0^2b_0^2,\qquad \tilde\gamma\left(\vartheta_0,0\right)=d^2a_0^2.
\end{aligned}
$$

Here $\bar W_t$ is innovation Wiener process (see Theorems 6.2 and 7.11 in [18]). We see that except the initial values the equations for $\tilde m\left(\cdot,\cdot\right)$ do not depend on ϑ and the equations for $\tilde\gamma\left(\cdot,\cdot\right)$ depend on the products a^2b^2 and $a_0^2b_0^2$ only. Therefore the invariat distributions of $\tilde m\left(\vartheta,\cdot\right)$ and $\tilde m\left(\vartheta_0,\cdot\right)$ depend on these products and the limit Kullback-Leibler distance

$$
T^{-1}D_{K-L}\left(\vartheta,\vartheta_0\right)\longrightarrow B\left(ab,a_0b_0\right).
$$

Hence the consistent estimation of the parameter $\vartheta=(a,b)$ is impossible.

It is evident that the numerical calculation of the MLE $\hat\vartheta_T$ according to (1.3)-(1.6) is quite a difficult problem. The goal of this work is to suggest the new estimator, called One-step MLE-process $\vartheta_t^\star,\tau\leq t\leq T$, which has two advantages. First, its numerical calculation is much more simple than that of the MLE and, second, this estimator has a recurrent structure and can be used for the joint estimation of the hidden process Y_t and the parameter ϑ. Similar One-step MLE's and Multi-step MLE-processes, introduced in [15], have been applied in the problem of parameter estimation of the hidden telegraph process [11], parameter estimation in diffusion processes by the discrete time observations [10], in the problem of frequency estimation [8], intensity parameter estimation for inhomogeneous Poisson processes [5], parameter estimation for the Markov sequences [17].

2. Preliminary estimator

Following [11] One-step MLE process will be constructed in two steps. First we introduce a consistent and asymptotically normal preliminary estimator and then this estimator is used to define One-step MLE-process. Preliminary estimator is constructed using an asymptotically negligible amount of the observations $X^K=\left(X_t,0\leq t\leq K\right)$, where $K=T^\delta,\delta\in(1/2,1)$.

Suppose that $\vartheta=f$ and introduce the statistic $\mathbb{S}_K$ and the function $\Phi\left(\vartheta\right),\vartheta\in\Theta$:

$$
\mathbb{S}_K=\frac{1}{K}\sum_{k=1}^{K}\left[X_k-X_{k-1}\right]^2,\qquad\qquad \Phi\left(\vartheta\right)=\frac{a^2b^2}{\vartheta^3}\left[e^{-\vartheta}-1+\vartheta\right]+\sigma^2.
$$

In the cases $\vartheta=b$ and $\vartheta=a$ the counterparts of the latter function are

$$
\Phi_*\left(\vartheta\right)=\frac{a^2\vartheta^2}{f^3}\left[e^{-f}-1+f\right]+\sigma^2,\qquad \hat\Phi\left(\vartheta\right)=\frac{\vartheta^2b^2}{f^3}\left[e^{-f}-1+f\right]+\sigma^2
$$

respectively.

In this section we consider the case $\vartheta = f$ only. Therefore

$$\mathrm{d}X_t = aY_t\,\mathrm{d}t + \sigma\,\mathrm{d}W_t, \qquad X_0 = 0, \tag{2.1}$$
$$\mathrm{d}Y_t = -\vartheta Y_t\,\mathrm{d}t + b\,\mathrm{d}V_t, \qquad Y_0 = \xi. \tag{2.2}$$

Note that the function $\Phi(\vartheta), \alpha < \vartheta < \beta$ is strictly decreasing. Define the preliminary estimator $\bar\vartheta_K$, base on the observations X^K:

$$\bar\vartheta_K = \vartheta_K^* \mathbb{1}_{\{\mathcal{A}_K\}} + \alpha \mathbb{1}_{\{\mathcal{A}_K^-\}} + \beta \mathbb{1}_{\{\mathcal{A}_K^+\}}.$$

Here ϑ_K^* is the root of equation $\Phi(\vartheta_K^*) = \mathbb{S}_K$ and $\mathcal{A}_K, \mathcal{A}_K^-, \mathcal{A}_K^+$ are the sets

$$\mathcal{A}_K = \{\omega : \Phi(\beta) < \mathbb{S}_K < \Phi(\alpha)\}, \quad \mathcal{A}_K^- = \{\omega : \mathbb{S}_K \geq \Phi(\alpha)\},$$
$$\mathcal{A}_K^+ = \{\omega : \mathbb{S}_K \leq \Phi(\beta)\}.$$

The asymptotic behavior of $\bar\vartheta_K$ as $K \to \infty$ is described in the following proposition.

Proposition 1. *The estimator $\bar\vartheta_K$ is consistent in L_2 sense, uniformly on compacts $\left[\bar\alpha, \bar\beta\right] \subset \Theta$ and*

$$\sup_{\vartheta_0 \in \left[\bar\alpha, \bar\beta\right]} \mathbf{E}_{\vartheta_0} \left|\bar\vartheta_K - \vartheta_0\right|^2 \leq \frac{C}{K} \tag{2.3}$$

with some constant $C > 0$.

Proof. We have

$$\mathbf{E}_{\vartheta_0} \left[\bar\vartheta_K - \vartheta_0\right]^2 = \mathbf{E}_{\vartheta_0} \left[\vartheta_K^* - \vartheta_0\right]^2 \mathbb{1}_{\{\mathcal{A}_K\}} + (\vartheta_0 - \alpha)^2 \mathbf{P}_{\vartheta_0}\left(\mathcal{A}_K^-\right) + (\beta - \vartheta_0)^2 \mathbf{P}_{\vartheta_0}\left(\mathcal{A}_K^+\right).$$

For the probabilities we have the estimates

$$\mathbf{P}_{\vartheta_0}\left(\mathcal{A}_K^-\right) = \mathbf{P}_{\vartheta_0}\left(\mathbb{S}_K - \Phi(\vartheta_0) \geq \Phi(\alpha) - \Phi(\vartheta_0)\right) \leq \frac{\mathbf{E}_{\vartheta_0}\left[\mathbb{S}_K - \Phi(\vartheta_0)\right]^2}{\left|\Phi(\alpha) - \Phi(\vartheta_0)\right|^2},$$
$$\mathbf{P}_{\vartheta_0}\left(\mathcal{A}_K^+\right) \leq \frac{\mathbf{E}_{\vartheta_0}\left|\mathbb{S}_K - \Phi(\vartheta_0)\right|^2}{\left[\Phi(\beta) - \Phi(\vartheta_0)\right]^2}.$$

Therefore we have to study the asymptotics of the statistic $\mathbb{S}_K$ as $K \to \infty$:

$$\mathbb{S}_K = \frac{1}{K}\sum_{k=1}^K \left[X_k - X_{k-1}\right]^2 = \frac{1}{K}\sum_{k=1}^K \left[\int_{k-1}^k \mathrm{d}X_s\right]^2$$
$$= \frac{a^2}{K}\sum_{k=1}^K \eta_k^2 + \frac{2a\sigma}{K}\sum_{k=1}^K \eta_k\left[W_k - W_{k-1}\right] + \frac{\sigma^2}{K}\sum_{k=1}^K \left[W_k - W_{k-1}\right]^2,$$

where

$$\eta_k = \int_{k-1}^{k} Y_t \, \mathrm{d}t.$$

We have

$$\mathbf{E}_{\vartheta_0} \mathbb{S}_K = \frac{a^2}{K} \sum_{k=1}^{K} \mathbf{E}_{\vartheta_0} \eta_k^2 + \sigma^2$$

because $Y^T = (Y_t, 0 \le t \le T)$ and $W^T = (W_t, 0 \le t \le T)$ are independent.

The process Y^T can be written as

$$Y_t = \xi e^{-\vartheta_0 t} + b \int_0^t e^{-\vartheta_0 (t-r)} \mathrm{d}V_r.$$

Hence

$$\begin{aligned}
\mathbf{E}_{\vartheta_0} Y_t Y_s &= \mathbf{E}_{\vartheta_0} \xi^2 e^{-\vartheta_0 (t+s)} + b^2 e^{-\vartheta_0 (t+s)} \int_0^{t \wedge s} e^{2\vartheta_0 r} \mathrm{d}r \\
&= \left[d^2 - \frac{b^2}{2\vartheta_0} \right] e^{-\vartheta_0 (t+s)} + \frac{b^2}{2\vartheta_0} e^{-\vartheta_0 |t-s|}
\end{aligned}$$

and

$$\begin{aligned}
\mathbf{E}_{\vartheta_0} \eta_k^2 &= \int_{k-1}^{k} \int_{k-1}^{k} \mathbf{E}_{\vartheta_0} Y_t Y_s \, \mathrm{d}s\mathrm{d}t \\
&= \left[d^2 - \frac{b^2}{2\vartheta_0} \right] \left(\int_{k-1}^{k} e^{-\vartheta_0 t} \mathrm{d}t \right)^2 + \frac{b^2}{2\vartheta_0} \int_{k-1}^{k} \int_{k-1}^{k} e^{-\vartheta_0 |t-s|} \mathrm{d}s\mathrm{d}t \\
&= \left[\frac{d^2}{\vartheta_0^2} - \frac{b^2}{2\vartheta_0^3} \right] \left[e^{\vartheta_0} - 1 \right]^2 e^{-2\vartheta_0 k} + \frac{b^2}{{\vartheta_0}^3} \left[e^{-\vartheta_0} - 1 + \vartheta_0 \right].
\end{aligned}$$

Therefore

$$\begin{aligned}
\mathbf{E}_{\vartheta_0} \mathbb{S}_K &= \left[e^{\vartheta_0} - 1 \right]^2 \frac{a^2 \left[2\vartheta_0 d^2 - b^2 \right]}{2\vartheta_0^3 K} \sum_{k=1}^{K} e^{-2\vartheta_0 k} + \frac{a^2 b^2}{{\vartheta_0}^3} \left[e^{-\vartheta_0} - 1 + \vartheta_0 \right] + \sigma^2 \\
&= \frac{a^2 b^2}{{\vartheta_0}^3} \left[e^{-\vartheta_0} - 1 + \vartheta_0 \right] + \sigma^2 + r_K, \qquad |r_K| \le \frac{C}{K}.
\end{aligned}$$

Using similar calculations we obtain the estimate

$$\mathbf{E}_{\vartheta_0} \eta_k \eta_m \le C \, e^{-\vartheta_0 |k-m|},$$

which allows us to prove the law of large numbers: for $K \to \infty$ we have convergence in mean square

$$\mathbf{E}_{\vartheta_0} \left(\mathbb{S}_K - \mathbf{E}_{\vartheta_0} \mathbb{S}_K \right)^2 \le \frac{C}{K}, \qquad \mathbb{S}_K \longrightarrow \Phi\left(\vartheta_0\right) = \frac{a^2 b^2}{{\vartheta_0}^3} \left[e^{-\vartheta_0} - 1 + \vartheta_0 \right] + \sigma^2$$

and

$$\mathbf{E}_{\vartheta_0}\left(\mathbb{S}_K-\Phi\left(\vartheta_0\right)\right)^2\leq 2\mathbf{E}_{\vartheta_0}\left(\mathbb{S}_K-\mathbf{E}_{\vartheta_0}\mathbb{S}_K\right)^2+2\left(\mathbf{E}_{\vartheta_0}\mathbb{S}_K-\Phi\left(\vartheta_0\right)\right)^2\leq\frac{C}{K}.$$

Hence

$$\sup_{\bar\alpha<\vartheta_0\leq\bar\beta}\left[\mathbf{P}_{\vartheta_0}\left(\mathcal{A}_K^-\right)+\mathbf{P}_{\vartheta_0}\left(\mathcal{A}_K^+\right)\right]\leq\frac{C}{K}.$$

The function $\Phi\left(\vartheta\right),\alpha<\vartheta<\beta$ is strictly decreasing. If we denote its inverse function as $\Psi\left(\phi\right)=\Phi^{-1}\left(\phi\right),\Phi\left(\beta\right)<\phi<\Phi\left(\alpha\right)$, then we have

$$\Psi'\left(\phi\right)=\frac{1}{\Phi'\left(\vartheta\right)},\qquad\text{for}\qquad\phi=\Phi\left(\vartheta\right)$$

and

$$\sup_{\vartheta\in\Theta}\left|\Psi'\left(\Phi\left(\vartheta\right)\right)\right|=\left(\inf_{\vartheta\in\Theta}\left|\Phi'\left(\vartheta\right)\right|\right)^{-1}=\left|\Phi'\left(\beta\right)\right|^{-1}\equiv c_*>0.$$

We can write

$$\begin{aligned}\mathbf{E}_{\vartheta_0}\left[\vartheta_K^*-\vartheta_0\right]^2\mathbb{1}_{\{\mathcal{A}_K\}}&=\mathbf{E}_{\vartheta_0}\left[\Psi\left(\mathbb{S}_K\right)-\Psi\left(\Phi\left(\vartheta_0\right)\right)\right]^2\mathbb{1}_{\{\mathcal{A}_K\}}\\&\leq c_*^2\mathbf{E}_{\vartheta_0}\left[\mathbb{S}_K-\Phi\left(\vartheta_0\right)\right]^2\leq\frac{Cc_*^2}{K}\longrightarrow 0\end{aligned}$$

as $K\to\infty$.

If we put $K=T^\delta$, then

$$\sup_{\bar\alpha\leq\vartheta_0\leq\bar\beta}\mathbf{E}_{\vartheta_0}\left[\bar\vartheta_{T^\delta}-\vartheta_0\right]^2\leq C\,T^{-\delta}.\tag{2.4}$$

$\square$

3. One-step MLE-process. Case $\vartheta=f$

Suppose that the unknown parameter is $\vartheta=f$ and we have the model (2.1)-(2.2), where the process X^T is observable and the Ornstein-Uhlenbeck process Y^T is "hidden". We realize the asymptotically efficient estimation of the parameter $\vartheta\in\Theta$ in two steps. First we calculate the preliminary estimator $\bar\vartheta_{T^\delta}$ and then using this estimator we construct the One-step MLE-process.

Recall that the equation (1.6) has explicit solution (see Lemma 1 in the Section 1.8, [1])

$$\gamma\left(\vartheta,t\right)=e^{-2r\left(\vartheta\right)t}\left[\frac{1}{d^2-\gamma\left(\vartheta\right)}+\frac{a^2}{2r\left(\vartheta\right)\sigma^2}\left(1-e^{-2r\left(\vartheta\right)t}\right)\right]^{-1}+\gamma\left(\vartheta\right).$$

Here $\gamma_0 = d^2$,

$$r(\vartheta) = \left(\vartheta^2 + \frac{b^2 a^2}{\sigma^2}\right)^{1/2}, \qquad \gamma(\vartheta) = \frac{\vartheta\sigma^2}{a^2}\left(\sqrt{1 + \frac{b^2 a^2}{\vartheta^2\sigma^2}} - 1\right).$$

Therefore we have exponential convergence of $\gamma(\vartheta, t)$ to the stationary solution $\gamma(\vartheta)$

$$|\gamma(\vartheta, t) - \gamma(\vartheta)| \leq C\, e^{-2r(\vartheta)t}.$$

To simplify the exposition we suppose that $d^2 = \gamma(\vartheta)$; then we have $\gamma(\vartheta, t) = \gamma(\vartheta)$. The case with an arbitrary d^2 requires cumbersome calculations, but the main results remain intact.

The equation for $m(\vartheta, t)$ in this case is

$$\mathrm{d}m(\vartheta, t) = -\left[\vartheta + \frac{\gamma(\vartheta)\, a^2}{\sigma^2}\right] m(\vartheta, t)\,\mathrm{d}t + \frac{\gamma(\vartheta)\, a}{\sigma^2}\mathrm{d}X_t.$$

Denote $m_t = m(\vartheta_0, t)$ and $\gamma(\vartheta_0) = \gamma_*$, where ϑ_0 is the true value. Then for the process $m_t, 0 \leq t \leq T$ we obtain the equation

$$\mathrm{d}m_t = -\vartheta_0 m_t \mathrm{d}t + \frac{\gamma_* a}{\sigma}\mathrm{d}\bar{W}_t, \quad m_0 \sim \mathcal{N}(0, \gamma_*), \quad 0 \leq t \leq T. \tag{3.1}$$

Here we used once more the theorem 7.11 in [18].

$$\mathrm{d}X_t = a m_t\,\mathrm{d}t + \sigma\,\mathrm{d}\bar{W}_t, \qquad X_0 = 0, \qquad 0 \leq t \leq T.$$

The *innovation* Wiener process $\bar{W}_t$ is defined by this equation and m_0 is independent on $\bar{W}_t, 0 \leq t \leq T$. With probability 1, the random process $m(\vartheta, t)$ has continuous derivatives w.r.t. ϑ and derivative processes $\dot{m}(\vartheta, t), \ddot{m}(\vartheta, t)$ satisfy the equations

$$\begin{aligned}
\mathrm{d}\dot{m}(\vartheta, t) = &-\left[\vartheta + \frac{\gamma(\vartheta)\, a^2}{\sigma^2}\right] \dot{m}(\vartheta, t)\,\mathrm{d}t + \frac{\dot{\gamma}(\vartheta)\, a}{\sigma^2}\mathrm{d}X_t \\
&- \left[1 + \frac{\dot{\gamma}(\vartheta)\, a^2}{\sigma^2}\right] m(\vartheta, t)\,\mathrm{d}t, \qquad (3.2)\\
\mathrm{d}\ddot{m}(\vartheta, t) = &-\left[\vartheta + \frac{\gamma(\vartheta)\, a^2}{\sigma^2}\right] \ddot{m}(\vartheta, t)\,\mathrm{d}t + \frac{\ddot{\gamma}(\vartheta)\, a}{\sigma^2}\mathrm{d}X_t \\
&- 2\left[1 + \frac{\dot{\gamma}(\vartheta)\, a^2}{\sigma^2}\right] \dot{m}(\vartheta, t)\,\mathrm{d}t - \frac{\ddot{\gamma}(\vartheta)\, a^2}{\sigma^2} m(\vartheta, t)\,\mathrm{d}t. \qquad (3.3)
\end{aligned}$$

The Fisher information for this model of observations is (see, e.g., Section 3.1.2 [14])

$$\mathrm{I}(\vartheta) = \frac{1}{2\vartheta} - \frac{2\dot{r}(\vartheta)}{r(\vartheta) + \vartheta} + \frac{\dot{r}(\vartheta)^2}{2r(\vartheta)}.$$

Note that $\mathrm{I}(\vartheta)$ has continuous bounded derivatives and is uniformly in $\vartheta \in \Theta$ separated from zero.

According to [15] the One-step MLE-process $\vartheta_t^\star, T^\delta < t \leq T$ is introduced as follows

$$\vartheta_t^\star = \bar\vartheta_{T^\delta} + \frac{a}{\sigma^2 t \mathrm{I}\left(\bar\vartheta_{T^\delta}\right)} \int_{T^\delta}^{t} \dot m(\bar\vartheta_{T^\delta}, s) \left[\mathrm{d}X_s - am(\bar\vartheta_{T^\delta}, s)\mathrm{d}s\right]. \tag{3.4}$$

Let us change the variables $t = \tau T$ and denote $\vartheta_{\tau T}^\star = \vartheta_T^\star(\tau), T^{\delta-1} < \tau \leq 1$.

Theorem 1. *One-step MLE-process $\vartheta_T^\star(\tau), T^{\delta-1} < \tau \leq 1$ with $\delta \in (1/2, 1)$ is consistent: for any $\nu > 0$ and any $\tau \in (0, 1]$*

$$\lim_{T\to\infty} -\mathbf{P}_{\vartheta_0}\left(\left|\vartheta_T^\star(\tau) - \vartheta_0\right| > \nu\right) = 0,$$

and asymptotically normal

$$\sqrt{\tau T}\left(\vartheta_T^\star(\tau) - \vartheta_0\right) \Longrightarrow \mathcal{N}\left(0, \mathrm{I}(\vartheta_0)^{-1}\right).$$

Proof. Consider the difference

$$\begin{aligned}\sqrt{\tau T}\left(\vartheta_T^\star(\tau) - \vartheta_0\right) &= \sqrt{\tau T}\left(\bar\vartheta_{T^\delta} - \vartheta_0\right) \\ &\quad + \frac{a}{\sigma\sqrt{\tau T}\mathrm{I}\left(\bar\vartheta_{T^\delta}\right)} \int_{T^\delta}^{\tau T} \dot m(\bar\vartheta_{T^\delta}, s)\mathrm{d}\bar W_s \\ &\quad + \frac{a^2}{\sigma^2\sqrt{\tau T}\mathrm{I}\left(\bar\vartheta_{T^\delta}\right)} \int_{T^\delta}^{\tau T} \dot m\left(\bar\vartheta_{T^\delta}, s\right)\left[m_s - m(\bar\vartheta_{T^\delta}, s)\right]\mathrm{d}s. \end{aligned}\tag{3.5}$$

Note that as it follows from the equations (3.2)-(3.3), the Gaussian processes $\dot m(\vartheta, t)$ and $\ddot m(\vartheta_0, t)$ have bounded variances and therefore for any $p > 1$ we have

$$\sup_{\vartheta\in\Theta} \mathbf{E}_{\vartheta_0}\left|\dot m(\vartheta, t)\right|^p \leq C, \qquad \sup_{\vartheta\in\Theta} \mathbf{E}_{\vartheta_0}\left|\ddot m(\vartheta, t)\right|^p \leq C,$$

where the constants do not depend on t. We can write

$$\begin{aligned}\dot m(\bar\vartheta_{T^\delta}, s) &= \dot m(\vartheta_0, s) + \dot m(\bar\vartheta_{T^\delta}, s) - \dot m(\vartheta_0, s) \\ &= \dot m(\vartheta_0, s) + \left(\vartheta_0 - \bar\vartheta_{T^\delta}\right)\ddot m(\tilde\vartheta, s) = \dot m(\vartheta_0, s) + O\left(T^{-\delta/2}\right)\end{aligned}$$

because

$$\left(\mathbf{E}_{\vartheta_0}\left|\left(\vartheta_0 - \bar\vartheta_{T^\delta}\right)\ddot m(\tilde\vartheta, s)\right|\right)^2 \leq \mathbf{E}_{\vartheta_0}\left(\vartheta_0 - \bar\vartheta_{T^\delta}\right)^2 \mathbf{E}_{\vartheta_0}\ddot m(\tilde\vartheta, s)^2 \leq \frac{C}{T^\delta}.$$

Here $\left|\tilde\vartheta - \bar\vartheta_{T^\delta}\right| \leq \left|\vartheta_0 - \bar\vartheta_{T^\delta}\right|$ and $O\left(T^{-\delta/2}\right)$ means that $T^{\delta/2}O\left(T^{-\delta/2}\right)$ is bounded in probability.

Further, for the Fisher information we have

$$\left|\frac{1}{\mathrm{I}\left(\bar{\vartheta}_{T^\delta}\right)}-\frac{1}{\mathrm{I}\left(\vartheta_0\right)}\right|=\frac{\left|\mathrm{I}\left(\bar{\vartheta}_{T^\delta}\right)-\mathrm{I}\left(\vartheta_0\right)\right|}{\mathrm{I}\left(\bar{\vartheta}_{T^\delta}\right)\mathrm{I}\left(\vartheta_0\right)}\leq C\left|\bar{\vartheta}_{T^\delta}-\vartheta_0\right|=O\left(T^{-\delta/2}\right).$$

This allows us to write

$$\begin{aligned}\Delta_T&=\frac{a}{\sigma\sqrt{\tau T}\mathrm{I}\left(\bar{\vartheta}_{T^\delta}\right)}\int_{T^\delta}^{\tau T}\dot{m}(\bar{\vartheta}_{T^\delta},s)\,\mathrm{d}\bar{W}_s\\&=\frac{a}{\sigma\mathrm{I}\left(\vartheta_0\right)\sqrt{\tau T-T^\delta}}\int_{T^\delta}^{\tau T}\dot{m}(\vartheta_0,s)\,\mathrm{d}\bar{W}_s\left(1+o\left(1\right)\right),\end{aligned}$$

where $o\left(1\right)$ is small in probability. By the law of large numbers

$$\frac{a^2}{\sigma^2\tau T}\int_{T^\delta}^{\tau T}\dot{m}(\vartheta_0,s)^2\,\mathrm{d}s\longrightarrow\mathrm{I}\left(\vartheta_0\right)$$

and therefore by the central limit theorem

$$\frac{a}{\sigma\mathrm{I}\left(\vartheta_0\right)\sqrt{\tau T}}\int_{T^\delta}^{\tau T}\dot{m}(\vartheta_0,s)\,\mathrm{d}\bar{W}_s\Longrightarrow\mathcal{N}\left(0,\mathrm{I}\left(\vartheta_0\right)^{-1}\right).$$

The similar arguments allow us to write

$$\begin{aligned}&\int_{T^\delta}^{\tau T}\dot{m}\left(\bar{\vartheta}_{T^\delta},s\right)\left[m\left(\vartheta_0,s\right)-m(\bar{\vartheta}_{T^\delta},s)\right]\mathrm{d}s\\&\qquad=-(\bar{\vartheta}_{T^\delta}-\vartheta_0)\int_{T^\delta}^{\tau T}\dot{m}\left(\bar{\vartheta}_{T^\delta},s\right)\dot{m}(\tilde{\vartheta},s)\mathrm{d}s\\&\qquad=-(\bar{\vartheta}_{T^\delta}-\vartheta_0)\int_{T^\delta}^{\tau T}\dot{m}\left(\vartheta_0,s\right)^2\mathrm{d}s\left(1+O\left(T^{-\delta/2}\right)\right).\end{aligned}$$

Recall that $\mathbf{E}_{\vartheta_0}\dot{m}(\vartheta_0,s)^2=\sigma^2a^{-2}\mathrm{I}\left(\vartheta_0\right)\left(1+o\left(1\right)\right)$. Therefore

$$\frac{1}{\tau T}\int_{T^\delta}^{\tau T}\dot{m}\left(\vartheta_0,s\right)^2\mathrm{d}s-\sigma^2a^{-2}\mathrm{I}\left(\vartheta_0\right)=\frac{1}{\sqrt{\tau T}}A\left(\tau T\right)+o\left(1\right),$$

where the integral (see, e.g., Proposition 1.23 in [14])

$$A\left(\tau T\right)=\frac{1}{\sqrt{\tau T}}\int_{T^\delta}^{\tau T}\left[\dot{m}\left(\vartheta_0,s\right)^2-\mathbf{E}_{\vartheta_0}\dot{m}\left(\vartheta_0,s\right)^2\right]\mathrm{d}s\Longrightarrow\mathcal{N}\left(0,D\left(\vartheta_0\right)\right).$$

Hence we obtained the representation

$$\begin{aligned}&\frac{a^2}{\sigma^2\sqrt{\tau T}\mathrm{I}\left(\bar{\vartheta}_{T^\delta}\right)}\int_{T^\delta}^{\tau T}\dot{m}\left(\bar{\vartheta}_{T^\delta},s\right)\left[m_s-m(\bar{\vartheta}_{T^\delta},s)\right]\mathrm{d}s\\&\qquad=-\sqrt{\tau T}\left(\bar{\vartheta}_{T^\delta}-\vartheta_0\right)\left(1+O\left(T^{-\delta/2}\right)\right).\end{aligned}$$

Substitution of this relation into the initial representation (3.5) yields the final expression

$$\sqrt{\tau T}\left(\vartheta^{\star}\left(\tau\right)-\vartheta_0\right)=\Delta_T+\sqrt{\tau T}\left(\bar\vartheta_{T^\delta}-\vartheta_0\right)O\left(T^{-\delta/2}\right)=\Delta_T+O\left(T^{\frac12-\delta}\right)$$
$$\Longrightarrow\mathcal{N}\left(0,\mathrm{I}\left(\vartheta_0\right)^{-1}\right),$$

since $\delta\in(1/2,1)$.

Note that the process $\vartheta_t^\star, T^\delta<t\leq T$ can be written in recurrent form

$$\mathrm{d}\vartheta_t^\star=-\frac{\vartheta_t^\star-\bar\vartheta_{T^\delta}}{t}\mathrm{d}t+\frac{a\dot m(\bar\vartheta_{T^\delta},t)}{\sigma^2t\mathrm{I}(\bar\vartheta_{T^\delta})}\left[\mathrm{d}X_t-am(\bar\vartheta_{T^\delta},t)\mathrm{d}t\right] \tag{3.6}$$

and we can introduce the adaptive filtering equations as follows

$$\mathrm{d}\hat m_t=-\left[\vartheta_t^\star+\frac{\gamma\left(\vartheta_t^\star\right)a^2}{\sigma^2}\right]\hat m_t\mathrm{d}t+\frac{\gamma\left(\vartheta_t^\star\right)a}{\sigma^2}\mathrm{d}X_t,\quad T^\delta<t\leq T, \tag{3.7}$$

$$\gamma\left(\vartheta_t^\star\right)=\frac{\vartheta_t^\star\sigma^2}{a^2}\left(\sqrt{1+\frac{b^2a^2}{\left(\vartheta_t^\star\right)^2\sigma^2}}-1\right) \tag{3.8}$$

with the initial value $\hat m_{T^\delta}=m\left(\bar\vartheta_{T^\delta},T^\delta\right)$.

It will be interesting to see the behavior of the system (3.6)-(3.8) using numerical simulations.

Recall that if we put $\tau=1$, then $\vartheta_T^\star$ is One-step MLE with

$$\sqrt{T}\left(\vartheta_T^\star-\vartheta_0\right)\Longrightarrow\mathcal{N}\left(0,\mathrm{I}\left(\vartheta_0\right)^{-1}\right)$$

studied for ergodic diffusion processes in the Section 2.5 [14]. Therefore the estimator $\vartheta_T^\star$ is asymptotically equivalent to the asymptotically efficient MLE $\hat\vartheta_T$ defined by the equation (1.4). There is essential computational difference between these two estimators. The calculation of $\hat\vartheta_T$ using (1.3)-(1.6) requires solving the differential equations (1.5)-(1.6) for *numerous* values of $\vartheta\in\Theta$, which is computationally inefficient. To construct One-step MLE-process $\vartheta_T^\star$ we have to calculate a simple preliminary estimator $\bar\vartheta_{T^\delta}$ and then to solve the system (1.5)-(1.6) for just one value $\vartheta=\bar\vartheta_{T^\delta}$. The difference between these two approaches becomes even more significant in the case of multidimensional ϑ. □

4. One-step MLE-process. Case $\boldsymbol{\vartheta}=\boldsymbol{b}$

Suppose that the volatility $b=\vartheta$ is the unknown parameter and we have the equations

$$\mathrm{d}X_t=aY_t\,\mathrm{d}t+\sigma\,\mathrm{d}W_t,\qquad X_0=0, \tag{4.1}$$
$$\mathrm{d}Y_t=-fY_t\,\mathrm{d}t+\vartheta\,\mathrm{d}V_t,\qquad Y_0=\xi. \tag{4.2}$$

As before all parameters a, σ, ϑ do not vanish and $f > 0$. The volatility $\vartheta \in (\alpha, \beta)$ with $\alpha > 0$ and the function

$$\Phi_*(\vartheta) = \frac{a^2\vartheta^2}{f^3}\left[e^{-f} - 1 + f\right] + \sigma^2, \qquad \alpha < \vartheta < \beta$$

is strictly increasing.

The statistic $\mathbb{S}_K$, with the new notations, converges to this function

$$\mathbb{S}_K \longrightarrow \Phi_*(\vartheta_0) \qquad \text{as} \qquad K \to \infty.$$

Therefore we have the explicit expression for the preliminary estimator

$$\bar\vartheta_K = \vartheta_K^* \mathbb{1}_{\{\mathcal{B}_K\}} + \alpha \mathbb{1}_{\{\mathcal{B}_K^-\}} + \beta \mathbb{1}_{\{\mathcal{B}_K^+\}},$$

where on $\mathcal{B}_K$

$$\vartheta_K^* = \left(\frac{f^3\left(\mathbb{S}_K - \sigma^2\right)}{a^2\left[e^{-f} - 1 + f\right]}\right)^{1/2}.$$

Here the sets $\mathcal{B}^{\pm}$ are defined by the similar relations

$$\begin{aligned}
&\mathcal{B}_K^- = \{\omega : \mathbb{S}_K \leq \Phi_*(\alpha)\}, \qquad \mathcal{B}_K^+ = \{\omega : \mathbb{S}_K \geq \Phi_*(\beta)\}, \\
&\mathcal{B}_K = \{\omega : \mathbb{S}_K \in (\Phi_*(\alpha), \Phi_*(\beta))\}.
\end{aligned}$$

As before, we have the consistency

$$\bar\vartheta_K \longrightarrow \vartheta_0 \qquad \text{as} \qquad K \to \infty$$

and

$$\mathbf{E}_{\vartheta_0}\left|\bar\vartheta_K - \vartheta_0\right|^2 \leq \frac{C}{K}.$$

We need the equation for $\dot m(\vartheta, t)$ and expression for Fisher information

$$\mathrm{I}(\vartheta_0) = \sigma^{-2}a^2\mathbf{E}_{\vartheta_0}\dot m(\vartheta_0, t)^2$$

in this case. The filtering equations in the stationary regime are

$$\begin{aligned}
\mathrm{d}m(\vartheta, t) &= -\left[f + \frac{\gamma_*(\vartheta)\,a^2}{\sigma^2}\right]m(\vartheta, t)\,\mathrm{d}t + \frac{\gamma_*(\vartheta)\,a}{\sigma^2}\mathrm{d}X_t, \qquad m(\vartheta, 0) = \xi, \\
\gamma_*(\vartheta) &= \frac{f\sigma^2}{a^2}\left(\sqrt{1 + \frac{\vartheta^2a^2}{f^2\sigma^2}} - 1\right), \qquad\qquad \xi \sim \mathcal{N}(0, \gamma_*(\vartheta)).
\end{aligned}$$

Therefore

$$\mathrm{d}\dot m(\vartheta, t) = -\left[f + \frac{\gamma_*(\vartheta)\,a^2}{\sigma^2}\right]\dot m(\vartheta, t)\,\mathrm{d}t + \frac{\dot\gamma_*(\vartheta)\,a}{\sigma^2}\left[\mathrm{d}X_t - am(\vartheta, t)\,\mathrm{d}t\right].$$

For $\vartheta = \vartheta_0$

$$\mathrm{d}m\left(\vartheta_0,t\right) = -fm\left(\vartheta_0,t\right)\mathrm{d}t + \frac{\gamma_*\left(\vartheta_0\right)a}{\sigma}\mathrm{d}\bar W_t, \qquad m\left(\vartheta,0\right)\sim\mathcal{N}\left(0,\gamma_*\left(\vartheta_0\right)\right),$$

$$\mathrm{d}\dot m\left(\vartheta_0,t\right) = -A\left(\vartheta_0\right)\dot m\left(\vartheta_0,t\right)\mathrm{d}t + \frac{\vartheta_0 a}{\sigma A\left(\vartheta_0\right)}\mathrm{d}\bar W_t,\ \dot m\left(\vartheta_0,0\right)\sim\mathcal{N}\left(0,q\left(\vartheta_0\right)\right),$$

where

$$A\left(\vartheta_0\right) = f + \frac{\gamma_*\left(\vartheta_0\right)a^2}{\sigma^2} = \sqrt{f^2+\frac{\vartheta_0^2a^2}{\sigma^2}}, \qquad q\left(\vartheta_0\right) = \frac{\vartheta_0^2a^2}{\sigma^2A\left(\vartheta_0\right)^3}.$$

Since

$$\dot m\left(\vartheta_0,t\right) = \dot m\left(\vartheta_0,0\right)e^{-At} + \int_0^t e^{-A(t-s)}\frac{\dot\gamma_*\left(\vartheta_0\right)a}{\sigma}\mathrm{d}\bar W_s$$

we obtain

$$\mathbf{E}_{\vartheta_0}\dot m\left(\vartheta_0,t\right)^2 = \frac{\vartheta_0^2a^2}{2\sigma^2A\left(\vartheta_0\right)^3}.$$

Therefore the Fisher information is

$$\mathrm{I}\left(\vartheta\right) = \frac{\vartheta_0^2a^4}{2\sigma^4A\left(\vartheta_0\right)^3}.$$

Now we can write the One-step MLE-process $\vartheta_t^\star, T^\delta < t \leq T$ as follows

$$\vartheta_t^\star = \bar\vartheta_{T^\delta} + \frac{a}{\sigma^2\left(t-T^\delta\right)\mathrm{I}\left(\bar\vartheta_{T^\delta}\right)}\int_{T^\delta}^t \dot m(\bar\vartheta_{T^\delta},s)\left[\mathrm{d}X_s - am(\bar\vartheta_{T^\delta},s)\mathrm{d}s\right]. \tag{4.3}$$

If we change the variables $t=\tau T$ and denote $\vartheta_{\tau T}^\star = \vartheta_T^\star\left(\tau\right), T^{\delta-1} < \tau \leq 1$, then we obtain the same assertions as in the Theorem 1:

Proposition 2. *One-step MLE-process* $\vartheta_T^\star = \left(\vartheta_T^\star\left(\tau\right), T^{\delta-1} < \tau \leq 1\right)$ *with* $\delta \in (1/2,1)$ *is consistent: for any* $\nu > 0$ *and any* $\tau \in (0,1]$

$$\lim_{T\to\infty} -\mathbf{P}_{\vartheta_0}\left(\left|\vartheta_T^\star\left(\tau\right)-\vartheta_0\right| > \nu\right) = 0,$$

and asymptotically normal

$$\sqrt{\tau T}\left(\vartheta_T^\star\left(\tau\right)-\vartheta_0\right) \Longrightarrow \mathcal{N}\left(0,\mathrm{I}\left(\vartheta_0\right)^{-1}\right).$$

Proof. Similarly to (3.6), we have exactly the same representation for the estimator $\vartheta_t^\star$ as in (3.4), with the only difference in the forms of $\dot m\left(\vartheta,t\right)$ and $\mathrm{I}\left(\vartheta\right)$. Thus the previous proof works in this case as well. □

It is possible to write the system of recurrent equations as in (3.6)-(3.8).

5. One-step MLE-process. Case $\vartheta = a$

It is clear that the suggested estimation approach also works for the partially observed system

$$\mathrm{d}X_t = \vartheta Y_t\,\mathrm{d}t + \sigma\,\mathrm{d}W_t, \qquad X_0 = 0, \tag{5.1}$$

$$\mathrm{d}Y_t = -fY_t\,\mathrm{d}t + b\,\mathrm{d}V_t, \qquad Y_0 = \xi, \tag{5.2}$$

where the unknown parameter is the drift $\vartheta = a$.

The function

$$\hat{\Phi}(\vartheta) = \frac{b^2\vartheta^2}{f^3}\left[e^{-f} - 1 + f\right] + \sigma^2, \qquad \alpha < \vartheta < \beta$$

is strictly increasing and the corresponding preliminary estimator

$$\bar{\vartheta}_{T^\delta} = \frac{f^3\left(\mathbb{S}_{T^\delta} - \sigma^2\right)}{b^2\left(e^f - 1 + f\right)}\mathbb{1}_{\{\mathcal{C}_{T^\delta}\}} + \alpha\mathbb{1}_{\left\{\mathcal{C}^-_{T^\delta}\right\}} + \beta\mathbb{1}_{\left\{\mathcal{C}^+_{T^\delta}\right\}},$$

where

$$\mathcal{C}^-_{T^\delta} = \left\{\omega : \mathbb{S}_{T^\delta} \leq \hat{\Phi}(\alpha)\right\}, \qquad \mathcal{C}^+_{T^\delta} = \left\{\omega : \mathbb{S}_{T^\delta} \geq \hat{\Phi}(\beta)\right\},$$
$$\mathcal{C}_{T^\delta} = \left\{\omega : \hat{\Phi}(\alpha) < \mathbb{S}_{T^\delta} < \hat{\Phi}(\beta)\right\}$$

admits the same asymptotic properties as in the preceding section.

The filtering equations are

$$\mathrm{d}m(\vartheta,t) = -\left[f + \frac{\hat{\gamma}(\vartheta)\,\vartheta^2}{\sigma^2}\right]m(\vartheta,t)\,\mathrm{d}t + \frac{\hat{\gamma}(\vartheta)\,\vartheta}{\sigma^2}\mathrm{d}X_t, \qquad m(\vartheta,0) = \xi,$$
$$\hat{\gamma}(\vartheta) = \frac{f\sigma^2}{\vartheta^2}\left(\sqrt{1 + \frac{\vartheta^2 b^2}{f^2\sigma^2}} - 1\right), \qquad \xi \sim \mathcal{N}\left(0, \hat{\gamma}(\vartheta)\right).$$

Therefore

$$\mathrm{d}\dot{m}(\vartheta,t) = -\left[f + \frac{\hat{\gamma}(\vartheta)\,\vartheta^2}{\sigma^2}\right]\dot{m}(\vartheta,t)\,\mathrm{d}t + \frac{\dot{\hat{\gamma}}(\vartheta)\,\vartheta + \hat{\gamma}(\vartheta)}{\sigma^2}\mathrm{d}X_t$$
$$- \frac{\left[\dot{\hat{\gamma}}(\vartheta)\,\vartheta^2 + 2\hat{\gamma}(\vartheta)\,\vartheta\right]}{\sigma^2}m(\vartheta,t)\,\mathrm{d}t.$$

To calculate Fisher information $\mathrm{I}(\vartheta_0) = \sigma^{-2}\mathbf{E}_{\vartheta_0}\left[m(\vartheta_0,t) + \vartheta_0\dot{m}(\vartheta_0,t)\right]^2$ we write the representations

$$m(\vartheta_0,t) = \frac{\hat{\gamma}(\vartheta_0)\,\vartheta_0}{\sigma}\int_0^t e^{-f(t-s)}\mathrm{d}\bar{W}_s + o(1), \qquad A = f + \frac{\hat{\gamma}(\vartheta_0)\,\vartheta_0^2}{\sigma^2},$$
$$\dot{m}(\vartheta_0,t) = \frac{\dot{\hat{\gamma}}(\vartheta_0)\,\vartheta_0 + \hat{\gamma}(\vartheta_0)}{\sigma}\int_0^t e^{-A(t-s)}\mathrm{d}\bar{W}_s$$

$$- \frac{\hat{\gamma}(\vartheta_0)\,\vartheta_0}{\sigma^2} \int_0^t e^{-A(t-s)} m(\vartheta_0, s)\,\mathrm{d}s + o(1)\,.$$

In the last integral we change the order of integration

$$\begin{aligned}
\int_0^t e^{-A(t-s)} m(\vartheta_0, s)\,\mathrm{d}s &= \frac{\hat{\gamma}(\vartheta_0)\,\vartheta_0}{\sigma} \int_0^t e^{-A(t-s)} \int_0^s e^{-f(s-r)}\mathrm{d}\bar{W}_r\,\mathrm{d}s \\
&= \frac{\hat{\gamma}(\vartheta_0)\,\vartheta_0}{\sigma} e^{-At} \int_0^t \left(\int_r^t e^{(A-f)s}\mathrm{d}s \right) e^{fr}\mathrm{d}\bar{W}_r \\
&= \frac{\hat{\gamma}(\vartheta_0)\,\vartheta_0}{\sigma(A-f)} e^{-At} \int_0^t \left(e^{(A-f)t} - e^{(A-f)r} \right) e^{fr}\mathrm{d}\bar{W}_r \\
&= \frac{\sigma}{\vartheta_0} \int_0^t e^{-f(t-r)}\mathrm{d}\bar{W}_r - \frac{\sigma}{\vartheta_0} \int_0^t e^{-A(t-r)}\mathrm{d}\bar{W}_r.
\end{aligned}$$

Hence

$$m(\vartheta_0, t) + \vartheta_0 \dot{m}(\vartheta_0, t) = N(\vartheta_0) \int_0^t e^{-A(t-s)}\mathrm{d}\bar{W}_s + o(1)\,,$$

where we denoted

$$N(\vartheta_0) = \frac{\dot{\hat{\gamma}}(\vartheta_0)\,\vartheta_0^2 + 2\hat{\gamma}(\vartheta_0)\,\vartheta_0}{\sigma}.$$

Therefore the Fisher information in this problem is the function

$$\mathrm{I}(\vartheta_0) = \frac{\left(\dot{\hat{\gamma}}(\vartheta_0)\,\vartheta_0^2 + 2\hat{\gamma}(\vartheta_0)\,\vartheta_0 \right)^2}{2\sigma^2 \left(f\sigma^2 + \hat{\gamma}(\vartheta_0)\,\vartheta_0^2 \right)}.$$

Having the preliminary estimator $\bar{\vartheta}_{T^\delta}$, expression for Fisher information $\mathrm{I}(\vartheta_0)$ and the equation for $\dot{m}(\vartheta, t)$ we can construct the One-step MLE-process

$$\vartheta_t^\star = \bar{\vartheta}_{T^\delta} + \frac{1}{\mathrm{I}\left(\bar{\vartheta}_{T^\delta}\right)} \int_{T^\delta}^t \frac{\left[m(\bar{\vartheta}_{T^\delta}, s) + \bar{\vartheta}_{T^\delta}\dot{m}(\bar{\vartheta}_{T^\delta}, s) \right]}{\sigma^2 t} \left[\mathrm{d}X_s - \bar{\vartheta}_{T^\delta} m(\bar{\vartheta}_{T^\delta}, s)\mathrm{d}s \right].$$

This estimator has the same asymptotic properties: it is consistent and asymptotically normal

$$\sqrt{\tau T}\left(\vartheta_T^\star(\tau) - \vartheta_0\right) \Longrightarrow \mathcal{N}\left(0, \mathrm{I}(\vartheta_0)^{-1}\right).$$

The proof follows the same pattern as in the previous cases.

6. Discussion

The results, presented above, can be developed in several directions by means of already known approaches.

1. It is interesting to find preliminary estimator in the cases of unknown parameters f, b, a. Recall that the estimation of all parameters is impossible. Of course, with one statistic $\mathbb{S}_{T^\delta}$ it is impossible and we need at least two different statistics. Consider the case of two-dimensional parameter $\vartheta = (f, b)$ or $\vartheta = (f, a)$ and two statistics

$$\mathbb{S}_K = \frac{1}{K}\sum_{k=1}^{K}\left[X_k - X_{k-1}\right]^2, \quad \mathbb{R}_K = \frac{1}{K}\sum_{k=1}^{K}\left[X_k - X_{k-1}\right]\left[X_{k-1} - X_{k-2}\right].$$

The limits are

$$\mathbb{S}_K \longrightarrow \Phi(\vartheta) = \frac{a^2b^2}{f^3}\left[e^{-f} - 1 + f\right] + \sigma^2,$$
$$\mathbb{R}_K \longrightarrow \Xi(\vartheta) = \frac{a^2b^2}{2f^3}\left[e^{-f} - 1\right]^2.$$

Therefore

$$\mathbb{Q}_K = \frac{\mathbb{S}_K - \sigma^2}{\mathbb{R}_K} \longrightarrow \frac{2\left[e^{-f} - 1 + f\right]}{\left[e^{-f} - 1\right]^2}.$$

The function

$$\phi(x) = \frac{2\left[e^{-x} - 1 + x\right]}{\left[e^{-x} - 1\right]^2}, \quad x > 0$$

is strictly increasing and $\lim_{x\to 0}\phi(x) = 1$, $\lim_{x\to\infty}\phi(x) = \infty$. Therefore, the parameter f can be estimated with the help of the statistic $\mathbb{Q}_K$:

$$\mathbb{Q}_K = \phi\left(f_K^*\right).$$

Having this estimator the second parameter, say, a or b can be obtained as solution of one of these equations

$$\mathbb{S}_K = \Phi\left(f_K^*, a_K^*\right), \qquad \text{or} \qquad \mathbb{S}_K = \Phi\left(f_K^*, b_K^*\right),$$

with obvious notation. As soon as we have a consistent preliminary estimator, say, $\bar\vartheta_{T^\delta} = \left(f_{T^\delta}^*, b_{T^\delta}^*\right)$ and explicit expression for the information matrix $\mathrm{I}(\vartheta)$, then

$$\vartheta_t^\star = \bar\vartheta_{T^\delta} + t^{-1}\mathrm{I}\left(\bar\vartheta_{T^\delta}\right)^{-1}\int_{T^\delta}^{t}\frac{a\,\dot m\left(\bar\vartheta_{T^\delta}, s\right)}{\sigma^2}\left[\mathrm{d}X_s - am\left(\bar\vartheta_{T^\delta}, s\right)\mathrm{d}s\right].$$

Recall that such processes were studied in [15].
2. The One-step MLE-process has *learning interval* $\left[0, T^\delta\right]$ with $\delta \in (\frac{1}{2}, 1]$. It can be interesting to have such process with *shorter* learning. This can be done with the help of another construction called Two-step MLE-process

introduced in [15]. Let us recall this construction using the model of observation (2.1)-(2.2). The first preliminary estimator $\bar\vartheta_{T^\delta}$ is constructed using the observations $X^{T^\delta}=\left(X_t,0\leq t\leq T^\delta\right)$ with $\delta\in(1/3,1/2]$ (shorter learning interval). The second preliminary *estimator-process* $\vartheta_t^*, T^\delta<t\leq T$ is

$$\vartheta_t^*=\bar\vartheta_{T^\delta}+\frac{a}{\sigma^2 t\mathrm{I}\left(\bar\vartheta_{T^\delta}\right)}\int_{T^\delta}^t \dot m(\bar\vartheta_{T^\delta},s)\left[\mathrm{d}X_s-am(\bar\vartheta_{T^\delta},s)\mathrm{d}s\right].$$

The Two-step MLE-process is

$$\vartheta_t^{\star\star}=\vartheta_t^*+\frac{a}{\sigma^2 t\mathrm{I}\left(\vartheta_t^*\right)}\int_{T^\delta}^t \dot m(\bar\vartheta_{T^\delta},s)\left[\mathrm{d}X_s-am(\vartheta_t^*,s)\mathrm{d}s\right].$$

Following the same arguments as in the proof of Theorem 2 in [15] it can be shown that

$$\sqrt{\tau T}\left(\vartheta_T^{\star\star}\left(\tau\right)-\vartheta_0\right)\Longrightarrow\mathcal{N}\left(0,\mathrm{I}\left(\vartheta_0\right)^{-1}\right),$$

where $\vartheta_T^{\star\star}\left(\tau\right)=\vartheta_{\tau T}^{\star\star}$.
The learning interval $\left[0,T^\delta\right]$ can be made even shorter if $\delta\in(1/4,1/3]$. In this case we use Three-step MLE-process (see details in [15]).

3. Consider the model (2.1)-(2.2) and the estimator-process $\vartheta_T^\star\left(\tau\right),\tau\in[\kappa,1]$, where $\kappa>0$. Let us denote by $\mathcal{P}_T$ the measure induced by the process

$$\zeta_T\left(\tau\right)=\sqrt{T\mathrm{I}\left(\vartheta_0\right)}\left(\vartheta_T^\star\left(\tau\right)-\vartheta_0\right),\kappa\leq\tau\leq 1$$

in the measurable space $\left(\mathscr{C}\left[\kappa,1\right],\mathscr{B}\right)$ of continuous on $[\kappa,1]$ functions. It is possible to verify the weak convergence

$$\mathcal{P}_T\Longrightarrow\mathcal{P}$$

where $\mathcal{P}$ corresponds to the Gaussian process $\zeta\left(\tau\right),\tau\in[\kappa,1]$ with

$$\mathbf{E}_{\vartheta_0}\zeta\left(\tau\right)=0,\qquad\mathbf{E}_{\vartheta_0}\zeta\left(\tau_1\right)\zeta\left(\tau_2\right)=\tau_1\wedge\tau_2,$$

i.e. $\zeta\left(\cdot\right)$ is a Wiener process on the interval $[\kappa,1]$.
The proof in similar situation can be found in [15], Theorem 1. It consists of proving convergence of the finite-dimensional distributions

$$\left(\zeta_T\left(\tau_1\right),\ldots,\zeta_T\left(\tau_k\right)\right)\Longrightarrow\left(\zeta\left(\tau_1\right),\ldots,\zeta\left(\tau_k\right)\right)$$

and the estimate

$$\mathbf{E}_{\vartheta_0}\left|\zeta_T\left(\tau_1\right)-\zeta_T\left(\tau_2\right)\right|^4\leq C\left|\tau_2-\tau_1\right|^2,$$

where the constant $C>0$ does not depend on T. The approach applied in the present work allows us the direct verification these two conditions.

Acknowledgments

I would like to thank P. Chigansky, AE and Reviewer for useful comments.

References

[1] Arato, M. (1983). *Linear Stochastic Systems with Constant Coefficients A Statistical Approach.* Lecture Notes in Control and Inform. Sci., 45. Springer, New York. MR0791212

[2] Bickel, P. J., Ritov, Y. and Rydén, T. (1998). Asymptotic normality of the maximum likelihood estimator for general hidden Markov models. *Ann. Statist.*, 26, 4, 1614–1635. MR1647705

[3] Cappé, O., Moulines, E. and Rydén, T. (2005). *Inference in Hidden Markov Models*. Springer, New York. MR2159833

[4] Chigansky, P. (2009). Maximum likelihood estimation for hidden Markov models in continuous time. *Statist. Inference Stoch. Processes*, 12, 2, 139–163. MR2511676

[5] Dabye, A. S., Gounoung, A. A. and Kutoyants, Yu. A. (2018). Method of moments estimators and multu-step MLE for Poisson processes. *J. Contemp. Math. Analysis*, 53, 4, 187–196. MR3860735

[6] Elliott, R. J., Aggoun, L. and Moor, J. B. (1995). *Hidden Markov Models.* Springer, New York. MR1323178

[7] Ephraim, Y., Mehrav, N. (2002). Hidden Markov processes. *IEEE Trans. Inform. Theory,* 48, 6, 1518–1569. MR1909472

[8] Golubev, G. K. (1984). Fisher's method of scoring in the problem of frequency estimation. *J. of Soviet Math.,* 25, 3, 1125–1139.

[9] Kalman, R. E. and Bucy, R. S. (1961). New results in linear filtering and prediction theory. *Trans. ASME*, 83D, 95–100. MR0234760

[10] Kamatani, K. and Uchida, M. (2015). Hybrid multi-step estimators for stochastic differential equations based on sampled data. *Statist. Inference Stoch. Processes*, 18, 2, 177–204. MR3348584

[11] Khasminskii, R. Z. and Kutoyants, Yu. A. (2018). On parameter estimation of hidden telegraph process. *Bernoulli*, 24, 3, 2064–2090. MR3757523

[12] Kutoyants, Y. A. (1984). *Parameter Estimation for Stochastic Processes.* Heldermann, Berlin. MR0777685

[13] Kutoyants, Y. A. (1994). *Identification of Dynamical Systems with Small Noise.* Kluwer Academic Publisher, Dordrecht. MR1332492

[14] Kutoyants, Yu. A. (2004). *Statistical Inference for Ergodic Diffusion Processes.* Springer, London. MR2144185

[15] Kutoyants, Yu. A. (2017). On the multi-step MLE-process for ergodic diffusion. *Stochastic Process. Appl.*, 127, 2243–2261. MR3652412

[16] Kutoyants, Yu. A. (2019). On parameter estimation of the hidden Ornstein–Uhlenbeck process. *J. Multivar. Analysis*, 169, 1, 248–269. MR3875598

[17] Kutoyants, Yu. A. and Motrunich, A. (2016). On multi-step MLE-process for Markov sequences. *Metrika*, 79, 705–724. MR3518583

[18] Liptser, R. S. and Shiryayev, A. N. (2001). *Statistics of Random Processes, I. General Theory*, 2nd ed. Springer, N.Y. MR1800857