

On parameter estimation of the hidden Gaussian process in perturbed SDE.*

Yury A. Kutoyants^{2,3} and Li Zhou¹

¹*Shandong University, Weihai, China*

²*Le Mans University, Le Mans, France*

³*Tomsk State University, Tomsk, Russia*

Abstract: We present results on parameter estimation of the linear partially observed Gaussian system of stochastic differential equations. We propose new one-step estimators which have the same asymptotic properties as the MLE, but much more simple to calculate, the estimators are so-called “estimator-processes”. The construction of the estimators is based on the equations of Kalman-Bucy filtration and the asymptotic corresponds to the *small noises* in the observations and state (hidden process) equations. We give conditions which provide the consistency, asymptotic normality and asymptotic efficiency of the proposed estimators.

MSC2020 subject classifications: Primary 62M05; secondary 62F12.

Keywords and phrases: Filter system, parameter estimation, small noise asymptotics, one-step MLE-process.

Received June 2020.

Contents

1	Introduction	211
2	Preliminary estimator	213
3	One-step MLE	217
4	One-step MLE-process	220
5	On efficient estimation of $m(\vartheta_0, t)$	223
6	Two examples	228
7	Discussion	231
	Acknowledgements	233
	References	233

1. Introduction

Let us consider the problem of parameter estimation for partially observed linear system. The observed process is:

$$dX_t = f(\vartheta, t)Y_t dt + \varepsilon\sigma(t) dW_t, \quad X_0 = 0, \quad (1.1)$$

*Corresponding author: Li Zhou. Email: zhou.li@sdu.edu.cn

where the *hidden* process Y_t is solution of the equation

$$dY_t = a(\vartheta, t)Y_t dt + \varepsilon b(t) dV_t, \quad Y_0 = y_0 \neq 0. \quad (1.2)$$

Here $\{W_t, 0 \leq t \leq T\}$ and $\{V_t, 0 \leq t \leq T\}$ are independent Wiener processes and the functions $f(\cdot), \sigma(\cdot), a(\cdot), b(\cdot)$ are known. The unknown parameter $\vartheta \in \Theta = (\alpha, \beta)$, for $|\alpha| + |\beta| < \infty$. Therefore we have to estimate ϑ by observations $X^T = (X_t, 0 \leq t \leq T)$.

The problem of the construction of adaptive Kalman filter was treated by many authors in engineering literature, where the models of observations are mainly of discrete time form (see, e.g., [9], [10], [11], [24] and references therein). The identification of continuous time partially observed systems were studied as well by some authors (see, e.g., [1], [2], [4], [5], [15], [17], [19], [20] and references therein). Similar problems for continuous time hidden telegraph process were studied in [3], [14].

In this paper we discuss the asymptotic properties of the estimators in “small noises” case, that is for $\varepsilon \rightarrow 0$. In [15], the well known maximum likelihood estimator (MLE) $\hat{\vartheta}_\varepsilon$ and Bayes estimator (BE) $\tilde{\vartheta}_\varepsilon$ are introduced. Denote the true value of the unknown parameter as $\vartheta_0 \in \Theta$. The construction of these estimators is based on the likelihood ratio function

$$\begin{aligned} L(\vartheta, X^T) &= \frac{dP_\vartheta}{dP_0}(X^T) \\ &= \exp \left\{ \int_0^T \frac{f(\vartheta, t) m(\vartheta, t)}{\varepsilon^2 \sigma(t)^2} dX_t - \int_0^T \frac{f(\vartheta, t)^2 m(\vartheta, t)^2}{2\varepsilon^2 \sigma(t)^2} dt \right\}, \end{aligned}$$

Here $m(\vartheta, t) = \mathbf{E}_\vartheta(Y_t | X_s, 0 \leq s \leq t)$ is the conditional expectation, satisfying the Kalman-Bucy filtration equations

$$\begin{aligned} dm(\vartheta, t) &= a(\vartheta, t) m(\vartheta, t) dt \\ &\quad + \frac{\gamma(\vartheta, t) f(\vartheta, t)}{\varepsilon^2 \sigma(t)^2} [dX_t - f(\vartheta, t) m(\vartheta, t) dt], \end{aligned} \quad (1.3)$$

$$\frac{\partial \gamma(\vartheta, t)}{\partial t} = 2a(\vartheta, t) \gamma(\vartheta, t) - \frac{\gamma(\vartheta, t)^2 f(\vartheta, t)^2}{\varepsilon^2 \sigma(t)^2} + \varepsilon^2 b(t)^2, \quad (1.4)$$

with initial values $m(\vartheta, 0) = y_0$ and $\gamma(\vartheta, 0) = 0$. The Kalman-Bucy filter enables us to have online estimating or track unobservable signals. Thus the equation is linear and the Gaussianity of the observable perturbed SDE yields the asymptotical properties of the estimators. For example, the MLE and BE are given by the following relations

$$L(\hat{\vartheta}_\varepsilon, X^T) = \sup_{\vartheta \in \Theta} L(\vartheta, X^T), \quad \tilde{\vartheta}_\varepsilon = \frac{\int_\Theta \vartheta p(\vartheta) L(\vartheta, X^T) d\vartheta}{\int_\Theta p(\vartheta) L(\vartheta, X^T) d\vartheta}. \quad (1.5)$$

with $p(\vartheta)$ some known prior density function. Denoting $I(\vartheta_0)$ (details on Section 3) the Fisher information, it is shown in [15] that under regularity condi-

tions, the two estimators are consistent, asymptotically normal

$$\frac{\hat{\vartheta}_\varepsilon - \vartheta_0}{\varepsilon} \implies \mathcal{N}\left(0, \mathbf{I}(\vartheta_0)^{-1}\right), \quad \frac{\tilde{\vartheta}_\varepsilon - \vartheta_0}{\varepsilon} \implies \mathcal{N}\left(0, \mathbf{I}(\vartheta_0)^{-1}\right) \quad (1.6)$$

and asymptotically efficient. We take here and after “ \implies ” for “convergence in distribution”.

It is evident that the construction of the MLE and BE according to the relations (1.5) is hard to compute, because we need solutions of the system (1.3)-(1.4) for all $\vartheta \in \Theta$. Therefore numerical realization of these Constructions is difficult.

In the present work our goal is to propose other estimators, which can be much more easily calculated, and have the same asymptotic properties as mentioned in (1.6). Moreover, we construct *estimator-processes*, i.e., estimators which evolve in time. The One-step estimator in general is well known. Such one-step procedure was firstly proposed by Fisher [6]. Then it was used by many authors, see, e.g., [8], [14], [12], [21], [18].

The construction is done in two steps. In Section 2, we construct a preliminary consistent estimator $\bar{\vartheta}_{\tau_\varepsilon}$ by the observation $X^{\tau_\varepsilon} = (X_t, 0 \leq t \leq \tau_\varepsilon)$, where $\tau_\varepsilon = \varepsilon^\delta$, with $\delta \in (0, 1)$. Note that $\tau_\varepsilon \rightarrow 0$ when $\varepsilon \rightarrow 0$. This means that the preliminary estimator depends only on a small part of initial observations, of which the length converges to zero, but slower than ε . In Section 3 and Section 4, we propose the One-step MLE and then the estimator process, in using the preliminary estimator and the score function:

$$\vartheta_{t,\varepsilon}^* = \bar{\vartheta}_{\tau_\varepsilon} + I_{\tau_\varepsilon}(\bar{\vartheta}_{\tau_\varepsilon})^{-1} \int_{\tau_\varepsilon}^t \frac{\dot{M}(\bar{\vartheta}_{\tau_\varepsilon}, s)}{\sigma(s)^2} [dX_s - f(\bar{\vartheta}_{\tau_\varepsilon}, s)m(\bar{\vartheta}_{\tau_\varepsilon}, s)ds]$$

where dot means the derivation w.r.t. ϑ , $m(t, \vartheta)$ is the conditional expectation of Y_t w.r.t. $\{X_s, 0 \leq s \leq t\}$ and $M(\vartheta, s) = f(\vartheta, s)y(\vartheta, s)$, with $y(\vartheta, s)$ defined in Section 3. In Section 5, we propose an efficient estimator for conditional expectation $m(\theta_0, t)$, which is used in the definition of estimator, but the solution could be disturbing. In Section 6, two examples are given and so does a numerical realization.

2. Preliminary estimator

We say that $h(\vartheta, t) \in \mathcal{C}_b$ if the function $h(\vartheta, t), \vartheta \in \Theta, t \in [0, T]$ is bounded; $h(\vartheta, t) \in \mathcal{C}_\vartheta^{(1)}$ or $h(\vartheta, t) \in \mathcal{C}_t^{(1)}$ if the function $h(\cdot, \cdot)$ is continuously differentiable on ϑ or t respectively. We denote the derivatives w.r.t. ϑ as $\dot{h}(\vartheta, t)$ and the derivatives w.r.t. t as $h^t(\vartheta, t)$.

Let us introduce the *Conditions* \mathcal{R} :

\mathcal{R}_1 . The functions $f(\vartheta, t), \sigma(t), a(\vartheta, t), b(t) \in \mathcal{C}_b$.

\mathcal{R}_2 . The functions $f(\vartheta, t), a(\vartheta, t) \in \mathcal{C}_\vartheta^{(1)}$. The function $\dot{f}(\vartheta, t) \in \mathcal{C}_t^{(1)}$.

\mathcal{R}_3 . The function $f(\vartheta, 0)$ and its derivative $\dot{f}(\vartheta, 0)$ are separated from zero:

$$\inf_{\vartheta \in \Theta} |f(\vartheta, 0)| > 0, \quad \inf_{\vartheta \in \Theta} |\dot{f}(\vartheta, 0)| > 0.$$

\mathcal{R}_4 . The initial value $y_0 \neq 0$.

Recall that \mathcal{R}_4 is a necessary condition for existence of consistent estimator for the model of observations (1.1)-(1.2). If $y_0 = 0$, then we can denote $\hat{X}_t = \varepsilon^{-1}X_t$, $\hat{Y}_t = \varepsilon^{-1}Y_t$ and rewrite the system (1.1)-(1.2) as follows

$$\begin{aligned} d\hat{X}_t &= f(\vartheta, t)\hat{Y}_t dt + \sigma(t) dW_t, & \hat{X}_0 &= 0, \\ d\hat{Y}_t &= a(\vartheta, t)\hat{Y}_t dt + b(t) dV_t, & \hat{Y}_0 &= 0. \end{aligned}$$

Hence this system does not depend on ε and there is no consistent estimation (see Khasminskii [13]). The condition which makes main sense here is \mathcal{R}_3 , which propose the identification of the unknown parameter. In proposition 2.1 we consider how to construct the preliminary estimator if this condition is replaced by another one. Without loss of generality we suppose that

$$y_0 > 0, f(\vartheta, 0) > 0 \text{ and } \dot{f}(\vartheta, 0) > 0. \quad (2.1)$$

Let us denote $x_t(\vartheta)$ and $y_t(\vartheta)$ the solutions of the equations (1.1), (1.2) for $\varepsilon = 0$:

$$\begin{aligned} \frac{\partial x_t(\vartheta)}{\partial t} &= f(\vartheta, t)y_t(\vartheta), & x_0(\vartheta) &= 0, \\ \frac{\partial y_t(\vartheta)}{\partial t} &= a(\vartheta, t)y_t(\vartheta), & y_0(\vartheta) &= y_0. \end{aligned}$$

Hence

$$\begin{aligned} x_t(\vartheta) &= y_0 \int_0^t f(\vartheta, s)A(\vartheta, s) ds, \\ y_t(\vartheta) &= y_0 A(\vartheta, t), \quad A(\vartheta, t) = \exp \left\{ \int_0^t a(\vartheta, r) dr \right\}. \end{aligned}$$

The following equalities are easy to calculate:

$$\begin{aligned} Y_t - y_t(\vartheta_0) &= \varepsilon \int_0^t b(s) \frac{A(\vartheta_0, t)}{A(\vartheta_0, s)} dV_s, \\ X_t - x_t(\vartheta_0) &= \varepsilon \int_0^t f(\vartheta_0, s) \int_0^s b(q) \frac{A(\vartheta_0, s)}{A(\vartheta_0, q)} dV_q ds + \varepsilon \int_0^t \sigma(s) dW_s \\ &= \varepsilon \int_0^t b(q) \int_q^t f(\vartheta_0, s) \frac{A(\vartheta_0, s)}{A(\vartheta_0, q)} ds dV_q + \varepsilon \int_0^t \sigma(s) dW_s. \end{aligned}$$

Hence we can write

$$Y_t - y_t(\vartheta_0) = \varepsilon \xi_t, \quad X_t - x_t(\vartheta_0) = \varepsilon \eta_t, \quad 0 \leq t \leq T,$$

where ξ_t and η_t are Gaussian processes with $\mathbf{E}_{\vartheta_0}\xi_t = 0$, $\mathbf{E}_{\vartheta_0}\eta_t = 0$ and for any $p > 0$

$$\mathbf{E}_{\vartheta_0} |\xi_t|^p < C_1 t^{p/2}, \quad \mathbf{E}_{\vartheta_0} |\eta_t|^p < C_2 t^{p/2}. \quad (2.2)$$

The constants $C_1 > 0, C_2 > 0$ do not depend on ϑ_0 and $t \in [0, T]$ (see [23]).

In the vicinity of the point $t = 0$ the function $x_t(\vartheta)$ is monotonically increasing on ϑ according to (2.1), and we have

$$x_t(\alpha) = \inf_{\theta} x_t(\theta), \quad x_t(\beta) = \sup_{\theta} x_t(\theta).$$

Further, following [14] and [20] we put $\tau_\varepsilon = \varepsilon^\delta, \delta > 0$, and introduce three sets

$$\begin{aligned} \mathbb{B}_\varepsilon &= \{x_{\tau_\varepsilon}(\alpha) < X_{\tau_\varepsilon} < x_{\tau_\varepsilon}(\beta)\}, \\ \mathbb{B}_\varepsilon^- &= \{X_{\tau_\varepsilon} \leq x_{\tau_\varepsilon}(\alpha)\}, \quad \mathbb{B}_\varepsilon^+ = \{X_{\tau_\varepsilon} \geq x_{\tau_\varepsilon}(\beta)\}, \end{aligned}$$

we define the estimator

$$\bar{\vartheta}_{\tau_\varepsilon} = \alpha \mathbb{1}_{\{\mathbb{B}_\varepsilon^-\}} + \mu_\varepsilon \mathbb{1}_{\{\mathbb{B}_\varepsilon\}} + \beta \mathbb{1}_{\{\mathbb{B}_\varepsilon^+\}}, \quad (2.3)$$

where μ_ε is the solution of the equation $x_{\tau_\varepsilon}(\mu_\varepsilon) = X_{\tau_\varepsilon}$.

This is what we called preliminary estimator, which will be used in the next section for construction of asymptotically efficient estimator. The definition of this preliminary depends on the solution of the equation $x_{\tau_\varepsilon}(\mu_\varepsilon) = X_{\tau_\varepsilon}$, but only for fixed time τ_ε , which is of small value. This would not bring difficulties in numerical realization, as what we will do in Section 6.

Theorem 2.1. *Suppose that the conditions \mathcal{R} be fulfilled and $\delta \in (0, 2)$, then the estimator $\bar{\vartheta}_{\tau_\varepsilon}$ is uniformly consistent on compacts $\mathbb{K} \in \Theta$, i.e., for any $\nu > 0$ and any compact \mathbb{K}*

$$\sup_{\vartheta_0 \in \mathbb{K}} \mathbf{P}_{\vartheta_0} (|\bar{\vartheta}_{\tau_\varepsilon} - \vartheta_0| > \nu) \longrightarrow 0, \quad \text{as } \varepsilon \rightarrow 0. \quad (2.4)$$

Moreover, for any $p > 0$,

$$\sup_{\vartheta_0 \in \mathbb{K}} \mathbf{E}_{\vartheta_0} |\bar{\vartheta}_{\tau_\varepsilon} - \vartheta_0|^p \leq C \left(\frac{\varepsilon}{\sqrt{\tau_\varepsilon}} \right)^p \longrightarrow 0. \quad (2.5)$$

Proof. We have

$$\begin{aligned} \mathbf{P}_{\vartheta_0} (|\bar{\vartheta}_{\tau_\varepsilon} - \vartheta_0| > \nu) &= \mathbf{P}_{\vartheta_0} (|\bar{\vartheta}_{\tau_\varepsilon} - \vartheta_0| > \nu, \mathbb{B}_\varepsilon^-) + \mathbf{P}_{\vartheta_0} (|\mu_\varepsilon - \vartheta_0| > \nu, \mathbb{B}_\varepsilon) \\ &\quad + \mathbf{P}_{\vartheta_0} (|\bar{\vartheta}_{\tau_\varepsilon} - \vartheta_0| > \nu, \mathbb{B}_\varepsilon^+) \\ &\leq \mathbf{P}_{\vartheta_0} (\mathbb{B}_\varepsilon^-) + \mathbf{P}_{\vartheta_0} (|\mu_\varepsilon - \vartheta_0| > \nu, \mathbb{B}_\varepsilon) + \mathbf{P}_{\vartheta_0} (\mathbb{B}_\varepsilon^+). \end{aligned}$$

Let $[\alpha_K, \beta_K] \subset \Theta$ be such that $\mathbb{K} \subset [\alpha_K, \beta_K]$. We can write

$$\dot{x}_t(\vartheta) = y_0 \int_0^t \left[\dot{f}(\vartheta, s) + f(\vartheta, s) \int_0^s \dot{a}(\vartheta, r) dr \right] A(\vartheta, s) ds.$$

By conditions \mathcal{R}_2 and \mathcal{R}_3 , there exists $\kappa_* > 0$ such that for sufficiently small t we have the estimate

$$\inf_{\vartheta \in \Theta} \dot{x}_t(\vartheta) \geq \kappa_* t.$$

With the help of (2.2), we have

$$\begin{aligned} \sup_{\vartheta_0 \in \mathbb{K}} \mathbf{P}_{\vartheta_0}(\mathbb{B}_\varepsilon^-) &= \sup_{\vartheta_0 \in \mathbb{K}} \mathbf{P}_{\vartheta_0}(X_{\tau_\varepsilon} - x_{\tau_\varepsilon}(\vartheta_0) \leq x_{\tau_\varepsilon}(\alpha) - x_{\tau_\varepsilon}(\vartheta_0)) \\ &\leq \sup_{\vartheta_0 \in \mathbb{K}} \mathbf{P}_{\vartheta_0}(|X_{\tau_\varepsilon} - x_{\tau_\varepsilon}(\vartheta_0)| \geq x_{\tau_\varepsilon}(\vartheta_0) - x_{\tau_\varepsilon}(\alpha)) \\ &\leq \sup_{\vartheta_0 \in \mathbb{K}} \mathbf{P}_{\vartheta_0}(|X_{\tau_\varepsilon} - x_{\tau_\varepsilon}(\vartheta_0)| \geq x_{\tau_\varepsilon}(\alpha_K) - x_{\tau_\varepsilon}(\alpha)) \\ &\leq \sup_{\vartheta_0 \in \mathbb{K}} \mathbf{P}_{\vartheta_0}(|X_{\tau_\varepsilon} - x_{\tau_\varepsilon}(\vartheta_0)| \geq \kappa_*(\alpha_K - \alpha)\tau_\varepsilon) \\ &\leq \sup_{\vartheta_0 \in \mathbb{K}} \frac{\mathbf{E}_{\vartheta_0} |X_{\tau_\varepsilon} - x_{\tau_\varepsilon}(\vartheta_0)|^p}{\kappa_*^p (\alpha_K - \alpha)^p \tau_\varepsilon^p} \leq C \frac{\varepsilon^p}{\tau_\varepsilon^{p/2}}. \end{aligned}$$

Similar estimate could be obtained for the probability $\mathbf{P}_{\vartheta_0}(\mathbb{B}_\varepsilon^+)$. Further

$$\mathbf{P}_{\vartheta_0}(|\mu_\varepsilon - \vartheta_0| > \nu, \mathbb{B}_\varepsilon) = \mathbf{P}_{\vartheta_0} \left(\frac{|X_{\tau_\varepsilon} - x_{\tau_\varepsilon}(\vartheta_0)|}{\dot{x}_{\tau_\varepsilon}(\tilde{\mu}_\varepsilon)} > \nu, \mathbb{B}_\varepsilon \right) \leq C \frac{\varepsilon^p}{\nu^p \tau_\varepsilon^{p/2}},$$

with $\tilde{\mu}_\varepsilon$ certain value between μ_ε and ϑ_0 according to the mean value theorem. For the moments,

$$\begin{aligned} \mathbf{E}_{\vartheta_0} |\bar{\vartheta}_{\tau_\varepsilon} - \vartheta_0|^p &\leq (\vartheta_0 - \alpha)^p \mathbf{P}_{\vartheta_0}(\mathbb{B}_\varepsilon^-) + (\beta - \vartheta_0)^p \mathbf{P}_{\vartheta_0}(\mathbb{B}_\varepsilon^+) \\ &\quad + \mathbf{E}_{\vartheta_0} |\mu_\varepsilon - \vartheta_0|^p \mathbb{1}_{\{\mathbb{B}_\varepsilon\}} \leq C \frac{\varepsilon^p}{\tau_\varepsilon^{p/2}} = C \varepsilon^{(1-\frac{1}{2})\delta} p. \end{aligned}$$

Thus we obtain the convergence of the moments for $\delta \in (0, 2)$.

Case $f(\vartheta, t) = f(t)$.

The condition \mathcal{R}_3 is not fulfilled in this case. We introduce new conditions as follows:

\mathcal{R}'_2 The function $a(\vartheta, t) \in \mathcal{C}_\vartheta^{(1)}$ and $\dot{a}(\vartheta, t) \in \mathcal{C}_t^{(1)}$.

\mathcal{R}'_3 The function $a(\vartheta, t)$ and its derivative $\dot{a}(\vartheta, t)$ are separated from zero

$$\inf_{\vartheta \in \Theta} a(\vartheta, 0) > 0, \quad \inf_{\vartheta \in \Theta} \dot{a}(\vartheta, 0) > 0.$$

Proposition 2.1. *Let the conditions $\mathcal{R}_1, \mathcal{R}'_2, \mathcal{R}'_3, \mathcal{R}_4$ be fulfilled and $\tau_\varepsilon = \varepsilon^\delta$ with $\delta \in (0, 2/3)$. Then the estimator $\bar{\vartheta}_\varepsilon$ is uniformly consistent on compacts $\mathbb{K} \subset \Theta$ and*

$$\mathbf{E}_{\vartheta_0} |\bar{\vartheta}_{\tau_\varepsilon} - \vartheta_0|^2 \leq C \varepsilon^{2-3\delta}. \quad (2.6)$$

Proof. The proof of Theorem 2.1 can be applied here with the only difference in the estimation of $\dot{x}_t(\vartheta)$. We have

$$\dot{x}_t(\vartheta) = y_0 \int_0^t f(s) \int_0^s \dot{a}(\vartheta, r) dr A(\vartheta, s) ds \geq \kappa^* t^2.$$

Hence follow the proof of Theorem 2.1 we obtain the estimate

$$\mathbf{E}_{\vartheta_0} |\bar{\vartheta}_{\tau_\varepsilon} - \vartheta_0|^2 \leq C \frac{\varepsilon^2}{\tau_\varepsilon^3}.$$

3. One-step MLE

Let us re-write the equation (1.3)-(1.4) as follows

$$\begin{aligned} dm(\vartheta, t) &= [a(\vartheta, t) - D(\vartheta, t) f(\vartheta, t)] m(\vartheta, t) dt + D(\vartheta, t) dX_t, \\ \frac{\partial \gamma_*(\vartheta, t)}{\partial t} &= 2a(\vartheta, t) \gamma_*(\vartheta, t) - \frac{\gamma_*(\vartheta, t)^2 f(\vartheta, t)^2}{\sigma(t)^2} + b(t)^2, \quad \gamma_*(\vartheta, 0) = 0, \end{aligned}$$

where $\gamma_*(\vartheta, t) = \gamma(\vartheta, t) / \varepsilon^2$ and $D(\vartheta, t) = \gamma_*(\vartheta, t) f(\vartheta, t) / \sigma(t)^2$.

We introduce the condition

$\tilde{\mathcal{H}}_2$. The functions $f(\vartheta, \cdot), a(\vartheta, \cdot) \in \mathcal{C}_\vartheta^{(2)}$. The function $\dot{f}(\cdot, t) \in \mathcal{C}_t^{(1)}$.

It can be shown that the random process $m(\vartheta, t), 0 \leq t \leq T$ is continuously differentiable w.r.t. ϑ with probability 1 (see, e.g., [16]). The derivative $\dot{m}(\vartheta, t)$ satisfies the equation

$$\begin{aligned} d\dot{m}(\vartheta, t) &= [a(\vartheta, t) - D(\vartheta, t) f(\vartheta, t)] \dot{m}(\vartheta, t) dt + \dot{D}(\vartheta, t) dX_t \\ &\quad + \left[\dot{a}(\vartheta, t) - \dot{D}(\vartheta, t) f(\vartheta, t) - D(\vartheta, t) \dot{f}(\vartheta, t) \right] m(\vartheta, t) dt. \end{aligned} \quad (3.1)$$

Therefore for $\vartheta = \vartheta_0$ and $\varepsilon = 0$, we obtain deterministic function $\dot{y}(\vartheta_0, t) \equiv \dot{m}(\vartheta_0, t)|_{\varepsilon=0}$, which can be written as follows

$$\dot{y}(\vartheta_0, t) = \int_0^t e^{B(\vartheta_0, t, s)} \left[\dot{a}(\vartheta_0, s) - D(\vartheta_0, s) \dot{f}(\vartheta_0, s) \right] y_s(\vartheta_0) ds, \quad (3.2)$$

where

$$B(\vartheta_0, t, s) = \int_s^t [a(\vartheta_0, r) - D(\vartheta_0, r) f(\vartheta_0, r)] dr.$$

Note that $y(\vartheta, t) \neq y_t(\vartheta)$ except at point ϑ_0 , and that $\dot{y}(\vartheta_0, t) \neq \dot{y}_t(\vartheta_0)$.

We define the Fisher information as follows

$$\mathbf{I}_\tau(\vartheta) = \int_\tau^T \left[\frac{\dot{f}(\vartheta, t) y_t(\vartheta) + f(\vartheta, t) \dot{y}(\vartheta, t)}{\sigma(t)} \right]^2 dt, \quad \mathbf{I}(\vartheta) = \mathbf{I}_0(\vartheta)$$

which is obtained as the following limit

$$I(\vartheta) = -\lim_{\varepsilon \rightarrow 0} \varepsilon^2 \partial_{\vartheta}^2 \ln L(\vartheta; X^T).$$

The family of measures of this statistical experiment is locally asymptotically normal, therefore we have the Hajek-Le Cam's lower bound on the mean square risks of all estimators ϑ_{ε} (see, e.g., [16])

$$\lim_{\nu \rightarrow 0} \underline{\lim}_{\varepsilon \rightarrow 0} \sup_{|\vartheta - \vartheta_0| < \nu} \varepsilon^{-2} \mathbf{E}_{\vartheta} (\vartheta_{\varepsilon} - \vartheta)^2 \geq I(\vartheta_0)^{-1}.$$

We call the estimator $\vartheta_{\varepsilon}^*$ *asymptotically efficient* if for all $\vartheta_0 \in \Theta$ we have

$$\lim_{\nu \rightarrow 0} \lim_{\varepsilon \rightarrow 0} \sup_{|\vartheta - \vartheta_0| < \nu} \varepsilon^{-2} \mathbf{E}_{\vartheta} (\vartheta_{\varepsilon}^* - \vartheta)^2 = I(\vartheta_0)^{-1}. \quad (3.3)$$

Let us introduce the estimator

$$\vartheta_{\varepsilon}^* = \bar{\vartheta}_{\tau_{\varepsilon}} + \frac{1}{\mathbf{I}_{\tau_{\varepsilon}}(\bar{\vartheta}_{\tau_{\varepsilon}})} \int_{\tau_{\varepsilon}}^T \frac{\dot{M}(\bar{\vartheta}_{\tau_{\varepsilon}}, t)}{\sigma(t)^2} [dX_t - f(\bar{\vartheta}_{\tau_{\varepsilon}}, t) m(\bar{\vartheta}_{\tau_{\varepsilon}}, t) dt], \quad (3.4)$$

where we denoted $\dot{M}(\vartheta, t) = \dot{f}(\vartheta, t) y_t(\vartheta) + f(\vartheta, t) \dot{y}(\vartheta, t)$.

Theorem 3.1. *Suppose that the conditions $\mathcal{R}_1, \tilde{\mathcal{R}}_2, \mathcal{R}_3, \mathcal{R}_4$ are fulfilled and $\delta \in (0, 1)$, then the One-step MLE $\vartheta_{\varepsilon}^*$ has following properties:*

1. *It is uniformly consistent on compacts $\mathbb{K} \in \Theta$: for any $\nu > 0$*

$$\sup_{\vartheta_0 \in \mathbb{K}} \mathbf{P}_{\vartheta_0} \{|\vartheta_{\varepsilon}^* - \vartheta_0| > \nu\} \rightarrow 0.$$

2. *It is uniformly asymptotically normal*

$$\varepsilon^{-1} (\vartheta_{\varepsilon}^* - \vartheta_0) \implies \zeta \sim \mathcal{N}(0, \mathbf{I}(\vartheta_0)^{-1}).$$

3. *The moments converge: for any $p > 0$*

$$\varepsilon^{-p} \mathbf{E}_{\vartheta_0} |\vartheta_{\varepsilon}^* - \vartheta_0|^p \rightarrow \mathbf{E}_{\vartheta_0} |\zeta|^p.$$

4. *It is asymptotically efficient satisfying (3.3).*

Proof. We have

$$\begin{aligned} \frac{\vartheta_{\varepsilon}^* - \vartheta_0}{\varepsilon} &= \frac{\bar{\vartheta}_{\tau_{\varepsilon}} - \vartheta_0}{\varepsilon} + \mathbf{I}_{\tau_{\varepsilon}}(\bar{\vartheta}_{\tau_{\varepsilon}})^{-1} \int_{\tau_{\varepsilon}}^T \frac{\dot{M}(\bar{\vartheta}_{\tau_{\varepsilon}}, t)}{\sigma(t)} d\bar{W}_t \\ &+ \frac{1}{\varepsilon \mathbf{I}_{\tau_{\varepsilon}}(\bar{\vartheta}_{\tau_{\varepsilon}})} \int_{\tau_{\varepsilon}}^T \frac{\dot{M}(\bar{\vartheta}_{\tau_{\varepsilon}}, t)}{\sigma(t)^2} [f(\vartheta_0, t) m(\vartheta_0, t) - f(\bar{\vartheta}_{\tau_{\varepsilon}}, t) m(\bar{\vartheta}_{\tau_{\varepsilon}}, t)] dt. \end{aligned}$$

Here we used the *innovation representation* [23]

$$dX_t = f(\vartheta_0, t) m(\vartheta_0, t) dt + \varepsilon \sigma(t) d\bar{W}_t, \quad X_0 = 0,$$

the Wiener process \bar{W}_t is defined by this equality as [23]. Note that $\bar{\vartheta}_{\tau_\varepsilon} \rightarrow \vartheta_0$ and $\tau_\varepsilon \rightarrow 0$, there is the convergence uniformly on compacts:

$$\int_{\tau_\varepsilon}^T \frac{\dot{M}(\bar{\vartheta}_{\tau_\varepsilon}, t)^2}{\sigma(t)^2} dt \longrightarrow \int_0^T \frac{\dot{M}(\vartheta_0, t)^2}{\sigma(t)^2} dt = \mathbf{I}(\vartheta_0),$$

and by the central limit theorem we have

$$\Delta_\varepsilon = \frac{1}{\mathbf{I}_{\tau_\varepsilon}(\bar{\vartheta}_{\tau_\varepsilon})} \int_{\tau_\varepsilon}^T \frac{\dot{M}(\bar{\vartheta}_{\tau_\varepsilon}, t)}{\sigma(t)} d\bar{W}_t \Longrightarrow \mathcal{N}(0, \mathbf{I}(\vartheta_0)^{-1}).$$

Recall that $\dot{m}(\vartheta, t) = \dot{y}(\vartheta, t) + O(\varepsilon)$, and that $m(\vartheta, t) = y_t(\vartheta) + O(\varepsilon) = y(\vartheta, t) + O(\varepsilon)$. We have

$$\begin{aligned} & f(\vartheta_0, t)m(\vartheta_0, t) - f(\bar{\vartheta}_{\tau_\varepsilon}, t)m(\bar{\vartheta}_{\tau_\varepsilon}, t) \\ &= [f(\vartheta_0, t) - f(\bar{\vartheta}_{\tau_\varepsilon}, t)] m(\vartheta_0, t) - f(\bar{\vartheta}_{\tau_\varepsilon}, t) [m(\bar{\vartheta}_{\tau_\varepsilon}, t) - m(\vartheta_0, t)] \\ &= (\vartheta_0 - \bar{\vartheta}_{\tau_\varepsilon}) \dot{f}(\tilde{\vartheta}_{\tau_\varepsilon}, t)m(\vartheta_0, t) - (\bar{\vartheta}_{\tau_\varepsilon} - \vartheta_0) \dot{m}(\tilde{\vartheta}_{\tau_\varepsilon}, t)f(\bar{\vartheta}_{\tau_\varepsilon}, t) \\ &= (\vartheta_0 - \bar{\vartheta}_{\tau_\varepsilon}) \left[\dot{f}(\bar{\vartheta}_{\tau_\varepsilon}, t)m(\bar{\vartheta}_{\tau_\varepsilon}, t) + f(\bar{\vartheta}_{\tau_\varepsilon}, t)\dot{m}(\bar{\vartheta}_{\tau_\varepsilon}, t) + \bar{R}_\varepsilon \right] \\ &= (\vartheta_0 - \bar{\vartheta}_{\tau_\varepsilon}) \left[\dot{f}(\bar{\vartheta}_{\tau_\varepsilon}, t)y(\bar{\vartheta}_{\tau_\varepsilon}, t) + f(\bar{\vartheta}_{\tau_\varepsilon}, t)\dot{y}(\bar{\vartheta}_{\tau_\varepsilon}, t) + \tilde{R}_\varepsilon \right]. \end{aligned}$$

where $\tilde{\vartheta}_{\tau_\varepsilon}$ is certain value between $\bar{\vartheta}_{\tau_\varepsilon}$ and ϑ_0 , according to Taylor's formula. Here we denoted $\bar{R}_\varepsilon, \tilde{R}_\varepsilon$, the random variables satisfying the estimates

$$\sup_{\vartheta_0 \in \mathbb{K}} \mathbf{E}_{\vartheta_0} |\bar{R}_\varepsilon|^p \leq C \sup_{\vartheta_0 \in \mathbb{K}} \mathbf{E}_{\vartheta_0} |\bar{\vartheta}_{\tau_\varepsilon} - \vartheta_0|^p \leq C \left(\frac{\varepsilon}{\sqrt{\tau_\varepsilon}} \right)^p.$$

Therefore we obtained the representation

$$\begin{aligned} \frac{\vartheta_\varepsilon^* - \vartheta_0}{\varepsilon} &= \Delta_\varepsilon + \frac{\bar{\vartheta}_{\tau_\varepsilon} - \vartheta_0}{\varepsilon} \left[1 - \frac{1}{\mathbf{I}_{\tau_\varepsilon}(\bar{\vartheta}_{\tau_\varepsilon})} \int_{\tau_\varepsilon}^T \frac{\dot{M}(\bar{\vartheta}_{\tau_\varepsilon}, t)^2}{\sigma(t)^2} dt + R_\varepsilon^* \right] \\ &= \Delta_\varepsilon + \frac{(\bar{\vartheta}_{\tau_\varepsilon} - \vartheta_0) R_\varepsilon^*}{\varepsilon}, \end{aligned}$$

with corresponding R_ε^* . We can write

$$\sup_{\vartheta_0 \in \mathbb{K}} \mathbf{E}_{\vartheta_0} \left| \frac{(\bar{\vartheta}_{\tau_\varepsilon} - \vartheta_0) R_\varepsilon^*}{\varepsilon} \right|^p \leq C \sup_{\vartheta_0 \in \mathbb{K}} \mathbf{E}_{\vartheta_0} \left| \frac{(\bar{\vartheta}_{\tau_\varepsilon} - \vartheta_0)^2}{\varepsilon} \right|^p \leq C \varepsilon^{p(1-\delta)} \rightarrow 0.$$

Hence

$$\frac{\vartheta_\varepsilon^* - \vartheta_0}{\varepsilon} = \Delta_\varepsilon + o(1) \Longrightarrow \mathcal{N}(0, \mathbf{I}(\vartheta_0)^{-1}).$$

We obtain thus the uniform convergence of moments on compacts:

$$\varepsilon^{-2} \mathbf{E}_{\vartheta_0} (\vartheta_\varepsilon^* - \vartheta_0)^2 \longrightarrow \mathbf{I}(\vartheta_0)^{-1}.$$

This uniform convergence gives us the asymptotical efficiency:

$$\sup_{|\vartheta - \vartheta_0| < \nu} \varepsilon^{-2} \mathbf{E}_{\vartheta} (\vartheta_{\varepsilon}^* - \vartheta)^2 \xrightarrow{\varepsilon \rightarrow 0} \sup_{|\vartheta - \vartheta_0| < \nu} \mathbf{I}(\vartheta)^{-1} \xrightarrow{\nu \rightarrow 0} \mathbf{I}(\vartheta_0)^{-1}.$$

Case $f(\vartheta, t) = f(t)$.

The estimator $\bar{\vartheta}_{\tau_{\varepsilon}}$ is defined by the same equation (2.3). The Fisher information is

$$\mathbf{I}_{\tau}(\vartheta) = \int_{\tau}^T \frac{f(t)^2 \dot{y}(\vartheta, t)^2}{\sigma(t)^2} dt, \quad \mathbf{I}(\vartheta) = \mathbf{I}_0(\vartheta).$$

We define the estimator just as (3.4), in putting $\dot{f}(\bar{\vartheta}_{\tau_{\varepsilon}}, t) = 0$:

$$\vartheta_{\varepsilon}^* = \bar{\vartheta}_{\tau_{\varepsilon}} + \frac{1}{\mathbf{I}_{\tau_{\varepsilon}}(\bar{\vartheta}_{\tau_{\varepsilon}})} \int_{\tau_{\varepsilon}}^T \frac{f(t) \dot{y}(\bar{\vartheta}_{\tau_{\varepsilon}}, t)}{\sigma(t)^2} [dX_t - f(t) m(\bar{\vartheta}_{\tau_{\varepsilon}}, t) dt].$$

Let us introduce

\mathcal{R}_2^* . The function $a(\vartheta, \cdot) \in C_{\vartheta}^{(2)}$. The function $\dot{a}(\vartheta, t) \in C_t^{(1)}$.

Proposition 3.1. Suppose that the conditions $\mathcal{R}_1, \mathcal{R}_2^*, \mathcal{R}_3', \mathcal{R}_4$ be fulfilled and $\delta \in (0, 1/3)$, then the One-step MLE $\vartheta_{\varepsilon}^*$ has the properties

1. It is uniformly on compacts consistent: for any $\nu > 0$

$$\sup_{\vartheta_0 \in \mathbb{K}} \mathbf{P}_{\vartheta_0} \{|\vartheta_{\varepsilon}^* - \vartheta_0| > \nu\} \rightarrow 0.$$

2. It is uniformly asymptotically normal

$$\varepsilon^{-1} (\vartheta_{\varepsilon}^* - \vartheta_0) \Rightarrow \mathcal{N}\left(0, \mathbf{I}(\vartheta_0)^{-1}\right),$$

3. It is asymptotically efficient.

Proof. The proof of this proposition is similar as the one given in Theorem 3.1. The only difference is in the estimate:

$$\mathbf{E}_{\vartheta_0} \frac{|(\bar{\vartheta}_{\tau_{\varepsilon}} - \vartheta_0) R_{\varepsilon}|}{\varepsilon} \leq C \varepsilon^{-1} \mathbf{E}_{\vartheta_0} (\bar{\vartheta}_{\tau_{\varepsilon}} - \vartheta_0)^2 \leq C \varepsilon^{-1+2-3\delta} = C \varepsilon^{1-3\delta} \rightarrow 0.$$

4. One-step MLE-process

Let us consider the problem slightly different. We have the model of observations (1.1)-(1.2) with unknown parameter $\vartheta \in \Theta$ and we are interested in the construction of the adaptive Kalman-Bucy filter to approximate $m(\vartheta_0, t) = \mathbf{E}_{\vartheta_0}(Y_t | X_s, 0 \leq s \leq t)$. We can not use $m(\vartheta_{\varepsilon}^*, t)$ because the estimator $\vartheta_{\varepsilon}^*$ depends on the observations $X^T = (X_s, 0 \leq s \leq T)$, where $(X_s, t < s \leq T)$ are the future. Therefore we need an estimator-process $\{\vartheta_{t,\varepsilon}^*, 0 < t \leq T\}$, where the estimator $\vartheta_{t,\varepsilon}^*$ has the following properties

- it is adaptive and depends on $(X_s, 0 \leq s \leq t)$;
- it can be easily calculated;
- it is asymptotically efficient.

We prefer not to use the MLE $\hat{\vartheta}_{t,\varepsilon}$ obtained as solution of the equation

$$L(\hat{\vartheta}_{t,\varepsilon}, X^t) = \sup_{\vartheta \in \Theta} L(\vartheta, X^t)$$

because its calculation for all $t \in (0, T)$ requires the solutions of the filtration equations (1.3)-(1.4). The numerical realization of such procedure would be too complicated.

We propose the One-step MLE $\hat{\vartheta}_{t,\varepsilon}, \tau \leq t \leq T$, called *One-step MLE-process*, defined as follows

$$\hat{\vartheta}_{t,\varepsilon} = \bar{\vartheta}_\tau + \frac{1}{\mathbf{I}_\tau^t(\bar{\vartheta}_\tau)} \int_\tau^t \frac{\dot{M}(\bar{\vartheta}_\tau, s)}{\sigma(s)^2} [dX_s - f(\bar{\vartheta}_\tau, s) m(\bar{\vartheta}_\tau, s) ds], \quad (4.1)$$

where

$$\mathbf{I}_\tau^t(\vartheta) = \int_\tau^t \frac{\dot{M}(\vartheta, s)^2}{\sigma(s)^2} ds, \quad \mathbf{I}^t(\vartheta) = \mathbf{I}_0^t(\vartheta).$$

Here τ is fixed which means we construct the preliminary estimator on observation $\{X_s, 0 \leq s \leq \tau\}$.

Proposition 4.1. *Let the conditions $\mathcal{R}_1, \mathcal{R}_2^*, \mathcal{R}_3', \mathcal{R}_4$ be fulfilled and $\delta \in (0, 1)$. Then the One-step MLE-process $\hat{\vartheta}_{t,\varepsilon}, \tau < t \leq T$ is uniformly consistent, asymptotically normal*

$$\frac{\hat{\vartheta}_{t,\varepsilon} - \vartheta_0}{\varepsilon} \Longrightarrow \mathcal{N}\left(0, \mathbf{I}^t(\vartheta_0)^{-1}\right),$$

and the random process $\eta_\varepsilon(t) = \varepsilon^{-1}(\hat{\vartheta}_{t,\varepsilon} - \vartheta_0), \tau \leq t \leq T$ for any $\tau \in (0, T)$, converges in distribution to the random process $\eta(t)$:

$$\eta_\varepsilon(t) \Longrightarrow \eta(t) = \mathbf{I}^t(\vartheta_0)^{-1} \int_0^t \frac{\dot{M}(\vartheta_0, s)}{\sigma(s)} dw_s, \quad \tau \leq t \leq T,$$

where $\{w_s, 0 \leq s \leq T\}$ is some Wiener process.

Proof. For any $t \in (\tau, T]$, the consistency and asymptotical normality follow from Theorem 3.1. We consider the vector $(\eta_\varepsilon(t_1), \dots, \eta_\varepsilon(t_k))$. The representation obtained in the proof of Theorem 3.1:

$$\begin{aligned} \eta_\varepsilon(t) &= \frac{\bar{\vartheta}_\tau - \vartheta_0}{\varepsilon} + \mathbf{I}_\tau^t(\bar{\vartheta}_\tau)^{-1} \int_\tau^t \frac{\dot{M}(\bar{\vartheta}_\tau, s)}{\sigma(s)} d\bar{W}_s \\ &+ \frac{1}{\varepsilon \mathbf{I}_\tau^t(\bar{\vartheta}_\tau)} \int_\tau^t \frac{\dot{M}(\bar{\vartheta}_\tau, s)}{\sigma(s)^2} [f(\vartheta_0, s)m(\vartheta_0, s) - f(\bar{\vartheta}_\tau, s)m(\bar{\vartheta}_\tau, s)] ds, \end{aligned}$$

allows us to verify the estimate

$$\mathbf{E}_{\vartheta_0} |\eta_\varepsilon(t_1) - \eta_\varepsilon(t_2)|^4 \leq C |t_1 - t_2|^2, \quad (4.2)$$

where the constant $C > 0$ does not depend on ε . The calculations are direct but cumbersome. The convergence of finite-dimensional distributions and estimate (4.2) provide us that the measures induced in the space of continuous functions $\mathcal{C}[\tau, T]$ by the processes $\eta_\varepsilon(\cdot)$ converge weakly to the measure of the process $\eta(\cdot)$ (see details of the proof in the similar situation in [18], [25]).

This weak convergence provides us the relation: for any $\tau \in (0, T)$ and any $\nu > 0$

$$\begin{aligned} & \mathbf{P}_{\vartheta_0} \left(\sup_{\tau \leq t \leq T} \mathbf{I}_\tau^t(\bar{\vartheta}_\tau) \left| \frac{\hat{\vartheta}_{t,\varepsilon} - \vartheta_0}{\varepsilon} \right| > \nu \right) \\ & \quad \rightarrow \mathbf{P}_{\vartheta_0} \left(\sup_{\tau \leq t \leq T} \left| \int_0^t \frac{\dot{M}(\vartheta_0, s)}{\sigma(s)} dw_s \right| > \nu \right) \\ & \quad = \mathbf{P}_{\vartheta_0} \left(\sup_{\Lambda_\tau \leq \lambda \leq \Lambda_T} |W(\lambda)| > \nu \right). \end{aligned}$$

Here $W(\cdot)$ is a Wiener process and

$$\lambda = \int_0^t \frac{\dot{M}(\vartheta_0, s)^2}{\sigma(s)^2} ds, \quad \Lambda_\tau = \int_0^\tau \frac{\dot{M}(\vartheta_0, s)^2}{\sigma(s)^2} ds, \quad \Lambda_T = \int_0^T \frac{\dot{M}(\vartheta_0, s)^2}{\sigma(s)^2} ds.$$

Thus we have the One-step MLE-process is uniformly consistent: for any $\tau \in (0, T)$ and any $\nu > 0$ we have

$$\mathbf{P}_{\vartheta_0} \left(\sup_{\tau \leq t \leq T} |\vartheta_{t,\varepsilon}^* - \vartheta_0| > \nu \right) \rightarrow 0.$$

We remind that the initial time τ could be fixed or depend on ε as $\tau_\varepsilon = \varepsilon^\delta \rightarrow 0$, this means that the length of preliminary observation converges to zero as $\varepsilon \rightarrow 0$, but slower. We define a modified One-step MLE $\vartheta_{t,\varepsilon}^*, \tau_\varepsilon \leq t \leq T$ as follows

$$\vartheta_{t,\varepsilon}^* = \bar{\vartheta}_{\tau_\varepsilon} + \frac{1}{\mathbf{I}_{\tau_\varepsilon}^t(\bar{\vartheta}_{\tau_\varepsilon})} \int_{\tau_\varepsilon}^t \frac{\dot{M}(\bar{\vartheta}_{\tau_\varepsilon}, s)}{\sigma(s)^2} [dX_s - f(\bar{\vartheta}_{\tau_\varepsilon}, s) m(\bar{\vartheta}_{\tau_\varepsilon}, s) ds]. \quad (4.3)$$

Proposition 4.2. *Let the conditions $\mathcal{R}_1, \mathcal{R}_2^*, \mathcal{R}_3', \mathcal{R}_4$ be fulfilled. Then the modified One-step MLE-process $\vartheta_{t,\varepsilon}^*, \tau_\varepsilon \leq t \leq T$ is consistent, asymptotically normal,*

$$\frac{\vartheta_{t,\varepsilon}^* - \vartheta_0}{\varepsilon} \Rightarrow \mathcal{N}(0, \mathbf{I}^t(\vartheta_0)^{-1})$$

and efficient.

Proof. As $\tau_\varepsilon \rightarrow 0$ we have $\tau_\varepsilon < \tau$ for any $\tau > 0$. The result follows from the Theorem 3.1.

The adaptive equations of filtration can be written as follows

$$\begin{aligned} dm_\varepsilon(t) &= a(\vartheta_{t,\varepsilon}^*, t) m_\varepsilon(t) dt \\ &\quad + \frac{\gamma_{*,\varepsilon}(t) f(\vartheta_{t,\varepsilon}^*, t)}{\sigma(t)^2} [dX_t - f(\vartheta_{t,\varepsilon}^*, t) m_\varepsilon(t) dt], \end{aligned} \quad (4.4)$$

$$\frac{\partial \gamma_{*,\varepsilon}(t)}{\partial t} = 2a(\vartheta_{t,\varepsilon}^*, t) \gamma_{*,\varepsilon}(t) - \frac{\gamma_{*,\varepsilon}(t)^2 f(\vartheta_{t,\varepsilon}^*, t)^2}{\sigma(t)^2} + b(t)^2, \quad (4.5)$$

with initial values $m_\varepsilon(\tau_\varepsilon) = m_\varepsilon(\bar{\vartheta}_{\tau_\varepsilon, \varepsilon}, \tau_\varepsilon)$ and $\gamma_{*,\varepsilon}(\tau_\varepsilon) = \gamma(\bar{\vartheta}_{\tau_\varepsilon, \varepsilon}, \tau_\varepsilon)$.

The equations (4.3)-(4.5) give us closed system, which allows us to calculate the estimator and the approximation of the conditional expectation in recurrent form.

5. On efficient estimation of $m(\vartheta_0, t)$

Recall that $m(\vartheta_0, t)$ is the mean squared optimal estimator of Y_t . The random process $m_\varepsilon(t)$, $\tau_\varepsilon \leq t \leq T$ can be considered as an estimator of the random function $m(\vartheta_0, t)$, $\tau_\varepsilon \leq t \leq T$. It is interesting to study the asymptotically efficient estimators in this problem. For any estimator $\bar{m}_\varepsilon(t)$ of the random function $m(\vartheta, t)$, we have the following lower bound on the mean square error.

Theorem 5.1. *Let the conditions $\mathcal{R}_1, \tilde{\mathcal{R}}_2, \mathcal{R}_3, \mathcal{R}_4$ be fulfilled, then for any $\vartheta_0 \in \Theta$, any $t \in (0, T]$ and any estimator $\bar{m}_\varepsilon(t)$ we have*

$$\lim_{\nu \rightarrow 0} \liminf_{\varepsilon \rightarrow 0} \sup_{|\vartheta - \vartheta_0| < \nu} \varepsilon^{-2} \mathbf{E}_\vartheta |\bar{m}_\varepsilon(t) - m(\vartheta, t)|^2 \geq \frac{\dot{y}(\vartheta_0, t)^2}{I^t(\vartheta_0)}. \quad (5.1)$$

Proof. The proof of this inequality follows the main steps of the proof of van Trees inequality in [7] and in [22]. Let us remind the main steps of the proof. For given density function $p(\vartheta)$, $\vartheta_0 - \nu < \vartheta < \vartheta_0 + \nu$ such that $p(\vartheta_0 \pm \nu) = 0$, we introduce the Fisher information

$$I_p = \int_{\vartheta_0 - \nu}^{\vartheta_0 + \nu} \frac{\dot{p}(\vartheta)^2}{p(\vartheta)} d\vartheta < \infty.$$

Then we can write

$$\begin{aligned} \sup_{|\vartheta - \vartheta_0| < \nu} \mathbf{E}_\vartheta |\bar{m}_\varepsilon(t) - m(\vartheta, t)|^2 &\geq \int_{\vartheta_0 - \nu}^{\vartheta_0 + \nu} \mathbf{E}_\vartheta |\bar{m}_\varepsilon(t) - m(\vartheta, t)|^2 p(\vartheta) d\vartheta \\ &:= \mathbb{E} |\bar{m}_\varepsilon(t) - m(\vartheta, t)|^2. \end{aligned}$$

Note that \mathbb{E} is the double mathematical expectation defined by the last equality. Let us denote

$$\begin{aligned} L(\vartheta, \vartheta_0, X^t) &= \frac{dP_\vartheta}{dP_{\vartheta_0}}(X^t) \\ &= \exp \left\{ \int_0^t \frac{f(\vartheta, s) m(\vartheta, s) - f(\vartheta_0, s) m(\vartheta_0, s)}{\varepsilon^2 \sigma(s)^2} dX_s \right. \\ &\quad \left. - \int_0^t \frac{f(\vartheta, s)^2 m(\vartheta, s)^2 - f(\vartheta_0, s)^2 m(\vartheta_0, s)^2}{2\varepsilon^2 \sigma(s)^2} ds \right\}. \end{aligned}$$

We have

$$\begin{aligned} &\int_{\vartheta_0-\nu}^{\vartheta_0+\nu} [\bar{m}_\varepsilon(t) - m(\vartheta, s)] \frac{\partial}{\partial \vartheta} [L(\vartheta, \vartheta_0, X^t) p(\vartheta)] d\vartheta \\ &= (\bar{m}_\varepsilon(t) - m(\vartheta, s)) L(\vartheta, \vartheta_0, X^t) p(\vartheta) \Big|_{\vartheta_0-\nu}^{\vartheta_0+\nu} \\ &\quad + \int_{\vartheta_0-\nu}^{\vartheta_0+\nu} \dot{m}(\vartheta, s) L(\vartheta, \vartheta_0, X^t) p(\vartheta) d\vartheta \\ &= \int_{\vartheta_0-\nu}^{\vartheta_0+\nu} \dot{m}(\vartheta, s) L(\vartheta, \vartheta_0, X^t) p(\vartheta) d\vartheta. \end{aligned}$$

Therefore

$$\begin{aligned} &\mathbf{E}_{\vartheta_0} \int_{\vartheta_0-\nu}^{\vartheta_0+\nu} [\bar{m}_\varepsilon(t) - m(\vartheta, s)] \frac{\partial}{\partial \vartheta} [L(\vartheta, \vartheta_0, X^t) p(\vartheta)] d\vartheta \\ &= \mathbf{E}_{\vartheta_0} \int_{\vartheta_0-\nu}^{\vartheta_0+\nu} \dot{m}(\vartheta, s) L(\vartheta, \vartheta_0, X^t) p(\vartheta) d\vartheta \\ &= \int_{\vartheta_0-\nu}^{\vartheta_0+\nu} \mathbf{E}_\vartheta \dot{m}(\vartheta, s) p(\vartheta) d\vartheta, \end{aligned} \tag{5.2}$$

where we changed the measure $\mathbf{E}_{\vartheta_0} L(\vartheta, \vartheta_0, X^t) = \mathbf{E}_\vartheta$. Recall that

$$\begin{aligned} &\frac{\partial}{\partial \vartheta} \ln L(\vartheta, \vartheta_0, X^t) \\ &= \int_0^t \frac{f(\vartheta, s) \dot{m}(\vartheta, s) + \dot{f}(\vartheta, s) m(\vartheta, s)}{\varepsilon^2 \sigma(s)^2} [dX_s - f(\vartheta, s) m(\vartheta, s) ds]. \end{aligned}$$

Hence

$$\begin{aligned} &\left(\mathbf{E}_{\vartheta_0} \int_{\vartheta_0-\nu}^{\vartheta_0+\nu} [\bar{m}_\varepsilon(t) - m(\vartheta, t)] \frac{\partial}{\partial \vartheta} [L(\vartheta, \vartheta_0, X^t) p(\vartheta)] d\vartheta \right)^2 \\ &= \left(\mathbf{E}_{\vartheta_0} \int_{\vartheta_0-\nu}^{\vartheta_0+\nu} [\bar{m}_\varepsilon(t) - m(\vartheta, t)] \right. \end{aligned}$$

$$\begin{aligned}
& \times \frac{\partial \ln [L(\vartheta, \vartheta_0, X^t) p(\vartheta)]}{\partial \vartheta} L(\vartheta, \vartheta_0, X^t) p(\vartheta) \, d\vartheta \Big)^2 \\
& \leq \mathbf{E}_{\vartheta_0} \int_{\vartheta_0-\nu}^{\vartheta_0+\nu} [\bar{m}_\varepsilon(t) - m(\vartheta, t)]^2 L(\vartheta, \vartheta_0, X^t) p(\vartheta) \, d\vartheta \\
& \quad \times \mathbf{E}_{\vartheta_0} \int_{\vartheta_0-\nu}^{\vartheta_0+\nu} \left[\frac{\partial \ln [L(\vartheta, \vartheta_0, X^t) p(\vartheta)]}{\partial \vartheta} \right]^2 L(\vartheta, \vartheta_0, X^t) p(\vartheta) \, d\vartheta \\
& = \int_{\vartheta_0-\nu}^{\vartheta_0+\nu} \mathbf{E}_\vartheta [\bar{m}_\varepsilon(t) - m(\vartheta, t)]^2 p(\vartheta) \, d\vartheta \\
& \quad \times \left(\int_{\vartheta_0-\nu}^{\vartheta_0+\nu} \mathbf{E}_\vartheta \left[\frac{\partial \ln L(\vartheta, \vartheta_0, X^t)}{\partial \vartheta} \right]^2 p(\vartheta) \, d\vartheta + \mathbf{I}_p \right). \quad (5.3)
\end{aligned}$$

Here we used Cauchy-Schwarz inequality and the property of stochastic integral

$$\begin{aligned}
& \mathbf{E}_\vartheta \frac{\partial \ln L(\vartheta, \vartheta_0, X^t)}{\partial \vartheta} \\
& = \mathbf{E}_\vartheta \int_0^t \frac{\dot{f}(\vartheta, s) m(\vartheta, s) + f(\vartheta, s) \dot{m}(\vartheta, s)}{\varepsilon^2 \sigma(s)^2} [dX_s - f(\vartheta, s) m(\vartheta, s) ds] = 0.
\end{aligned}$$

Thus

$$\begin{aligned}
& \mathbf{E}_\vartheta \left[\frac{\partial \ln L(\vartheta, \vartheta_0, X^t)}{\partial \vartheta} \right]^2 \\
& = \mathbf{E}_\vartheta \left(\int_0^t \frac{\dot{f}(\vartheta, s) m(\vartheta, s) + f(\vartheta, s) \dot{m}(\vartheta, s)}{\varepsilon \sigma(s)} d\bar{W}_s \right)^2 \\
& = \mathbf{E}_\vartheta \int_0^t \left(\frac{\dot{f}(\vartheta, s) m(\vartheta, s) + f(\vartheta, s) \dot{m}(\vartheta, s)}{\varepsilon \sigma(s)} \right)^2 ds \\
& = \frac{1}{\varepsilon^2} \int_0^t \left(\frac{\dot{f}(\vartheta, s) y(\vartheta, s) + f(\vartheta, s) \dot{y}(\vartheta, s)}{\sigma(s)} \right)^2 ds (1 + O(\varepsilon)) \\
& = \frac{1}{\varepsilon^2} \int_0^t \left(\frac{\dot{M}(\vartheta, s)}{\sigma(s)} \right)^2 ds (1 + O(\varepsilon)) = \frac{1}{\varepsilon^2} \mathbf{I}^t(\vartheta) (1 + O(\varepsilon)). \quad (5.4)
\end{aligned}$$

Using (5.2)-(5.4) we obtain the relation

$$\begin{aligned}
& \left(\int_{\vartheta_0-\nu}^{\vartheta_0+\nu} \mathbf{E}_\vartheta \dot{m}(\vartheta, s) p(\vartheta) \, d\vartheta \right)^2 \\
& \leq \mathbb{E} |\bar{m}_\varepsilon(t) - m(\vartheta, t)|^2 \left(\int_{\vartheta_0-\nu}^{\vartheta_0+\nu} \frac{\mathbf{I}^t(\vartheta)}{\varepsilon^2} (1 + O(\varepsilon)) p(\vartheta) d\vartheta + \mathbf{I}_p \right),
\end{aligned}$$

which can be written as follows

$$\begin{aligned} \varepsilon^{-2} \mathbb{E} |\hat{m}_\varepsilon(t) - m(\vartheta, t)|^2 &\geq \frac{\left(\int_{\vartheta_0 - \nu}^{\vartheta_0 + \nu} \dot{y}(\vartheta, s) p(\vartheta) d\vartheta \right)^2 (1 + O(\varepsilon))}{\int_{\vartheta_0 - \nu}^{\vartheta_0 + \nu} \Gamma^t(\vartheta) p(\vartheta) d\vartheta (1 + O(\varepsilon)) + \varepsilon^2 \mathbf{I}_p} \\ &\xrightarrow{\varepsilon \rightarrow 0} \frac{\left(\int_{\vartheta_0 - \nu}^{\vartheta_0 + \nu} \dot{y}(\vartheta, s) p(\vartheta) d\vartheta \right)^2}{\int_{\vartheta_0 - \nu}^{\vartheta_0 + \nu} \Gamma^t(\vartheta) p(\vartheta) d\vartheta} \xrightarrow{\nu \rightarrow 0} \frac{\dot{y}(\vartheta_0, t)^2}{\Gamma^t(\vartheta_0)}. \end{aligned}$$

Therefore we obtained the van Trees inequality (5.1).

The asymptotically efficient estimator of the conditional expectation $m(\vartheta_0, t)$ is defined as estimator $\hat{m}_\varepsilon(t)$ which satisfies

$$\lim_{\nu \rightarrow 0} \lim_{\varepsilon \rightarrow 0} \sup_{|\vartheta - \vartheta_0| < \nu} \varepsilon^{-2} \mathbf{E}_\vartheta |\hat{m}_\varepsilon(t) - m(\vartheta, t)|^2 = \frac{\dot{y}(\vartheta_0, t)^2}{\Gamma^t(\vartheta_0)} \quad (5.5)$$

for all $\vartheta_0 \in \Theta$.

To construct an asymptotically efficient estimator we have to modify slightly the estimator $m_\varepsilon(t)$. The solutions of the equations (1.3) and (4.4) can be written as follows

$$\begin{aligned} m(\vartheta_0, t) &= y_0 N(\vartheta_0, t) + N(\vartheta_0, t) \int_{\tau_\varepsilon}^t \frac{Q(\vartheta_0, s)}{N(\vartheta_0, s)} dX_s, \quad (5.6) \\ m_\varepsilon(t) &= y_0 N_\varepsilon(\vartheta^*, t) + N_\varepsilon(\vartheta^*, t) \int_{\tau_\varepsilon}^t \frac{Q_\varepsilon(\vartheta^*, s)}{N_\varepsilon(\vartheta^*, s)} dX_s, \end{aligned}$$

where

$$\begin{aligned} N(\vartheta_0, t) &= \exp \left\{ \int_{\tau_\varepsilon}^t \left[a(\vartheta_0, s) - \gamma_*(\vartheta_0, s) f(\vartheta_0, s)^2 \sigma(s)^{-2} \right] ds \right\}, \\ N_\varepsilon(\vartheta^*, t) &= \exp \left\{ \int_{\tau_\varepsilon}^t \left[a(\vartheta_{s,\varepsilon}^*, s) - \gamma_{*,\varepsilon}(s) f(\vartheta_{s,\varepsilon}^*, s)^2 \sigma(s)^{-2} \right] ds \right\}, \\ Q(\vartheta_0, t) &= \gamma_*(\vartheta_0, t) f(\vartheta_0, t) \sigma(t)^{-2}, \quad Q_\varepsilon(\vartheta^*, t) = \gamma_{*,\varepsilon}(t) f(\vartheta_{t,\varepsilon}^*, t) \sigma(t)^{-2}. \end{aligned}$$

We can not put $\vartheta_{t,\varepsilon}^*$ in $m(\vartheta_0, t)$ because the stochastic integral

$$\int_{\tau_\varepsilon}^t F(\vartheta_{t,\varepsilon}^*, s) dX_s$$

is not well defined for $F(\vartheta, t) = Q(\vartheta, s) N(\vartheta, s)^{-1}$. We will use the so-called robust version of the integral. It is given by the formula

$$\int_{\tau_\varepsilon}^t F(\vartheta, s) dX_s = F(\vartheta, t) X_t - F(\vartheta, \tau_\varepsilon) X_{\tau_\varepsilon} - \int_{\tau_\varepsilon}^t F'(\vartheta, s) X_s ds, \quad (5.7)$$

where

$$F'(\vartheta, s) = \frac{Q'(\vartheta, s) N(\vartheta, s) - Q(\vartheta, s) N'(\vartheta, s)}{N(\vartheta, s)^2}, \quad (5.8)$$

$$N'(\vartheta, s) = N(\vartheta, s) \left[a(\vartheta, s) - \gamma_*(\vartheta_0, s) f(\vartheta_0, s)^2 \sigma(s)^{-2} \right].$$

Let us denote the right side of equality (5.7) as $G(\vartheta, X^t, t)$ and introduce the estimator

$$m_\varepsilon^*(t) = y_0 N(\vartheta_{t,\varepsilon}^*, t) + N(\vartheta_{t,\varepsilon}^*, t) G(\vartheta_{t,\varepsilon}^*, X^t), \quad (5.9)$$

which we compare with

$$m(\vartheta_0, t) = y_0 N(\vartheta_0, t) + N(\vartheta_0, t) G(\vartheta_0, X^t). \quad (5.10)$$

Remind that we comprehend $F'(\vartheta_{t,\varepsilon}^*, s)$ as that is written in (5.8), therefore the equality (5.7) will no more be valid if ϑ is replaced by $\vartheta_{t,\varepsilon}^*$. This means that the estimator in (5.9) is not exactly $m(\vartheta_{t,\varepsilon}^*, t)$.

Recall that

$$m(\vartheta_0, t)|_{\varepsilon=0} = y_0 N(\vartheta_0, t) + N(\vartheta_0, t) G(\vartheta_0, x^t) = y_t(\vartheta_0).$$

It is interesting to see if the proposed estimator $m_\varepsilon^*(t)$ is optimal in some sense.

Proposition 5.1. *Let the conditions $\mathcal{R}_1, \tilde{\mathcal{R}}_2, \mathcal{R}'_3, \mathcal{R}_4$ be fulfilled, then the estimator $m_\varepsilon^*(t)$ is asymptotically efficient, i.e., for any $t \in (0, T]$ and any $\vartheta_0 \in \Theta$ we have*

$$\lim_{\nu \rightarrow 0} \lim_{\varepsilon \rightarrow 0} \sup_{|\vartheta - \vartheta_0| < \nu} \varepsilon^{-2} \mathbf{E}_\vartheta |m_\varepsilon^*(t) - m(\vartheta, t)|^2 = \frac{\dot{y}(\vartheta_0, t)^2}{\Gamma^t(\vartheta_0)}.$$

Proof. Let us consider the difference between (5.9) and (5.10). Applying Taylor's expansions,

$$\begin{aligned} m_\varepsilon^*(t) - m(\vartheta, t) &= y_0 [N(\vartheta_{t,\varepsilon}^*, t) - N(\vartheta, t)] \\ &\quad + [N(\vartheta_{t,\varepsilon}^*, t) - N(\vartheta, t)] G(\vartheta_{t,\varepsilon}^*, X^t) \\ &\quad + N(\vartheta, t) [G(\vartheta_{t,\varepsilon}^*, X^t) - G(\vartheta, X^t)] \\ &= (\vartheta_{t,\varepsilon}^* - \vartheta) [y_0 + G(\vartheta_{t,\varepsilon}^*, X^t)] \dot{N}(\tilde{\vartheta}_{t,\varepsilon}^*, t) \\ &\quad + (\vartheta_{t,\varepsilon}^* - \vartheta) N(\vartheta, t) \dot{G}(\tilde{\vartheta}_{t,\varepsilon}^*, X^t) \\ &= (\vartheta_{t,\varepsilon}^* - \vartheta) \left[(y_0 + G(\vartheta, X^t)) \dot{N}(\vartheta, t) + N(\vartheta, t) \dot{G}(\vartheta, X^t) \right] (1 + r_\varepsilon) \\ &= (\vartheta_{t,\varepsilon}^* - \vartheta) \left[(y_0 + G(\vartheta, x^t)) \dot{N}(\vartheta, t) + N(\vartheta, t) \dot{G}(\vartheta, x^t) \right] (1 + \tilde{R}_\varepsilon) \\ &= (\vartheta_{t,\varepsilon}^* - \vartheta) \dot{y}(\vartheta, t) (1 + \tilde{R}_\varepsilon). \end{aligned}$$

The conditions $\tilde{\mathcal{R}}_2$ allow us to verify that the derivatives $\dot{N}(\vartheta, t), \ddot{N}(\vartheta, t)$ are bounded. The derivatives of $G(\vartheta, X^t)$ are

$$\dot{G}(\vartheta, X^t) = X_t \dot{F}(\vartheta, t) - X_{\tau_\varepsilon} \dot{F}(\vartheta, \tau_\varepsilon) - \int_{\tau_\varepsilon}^t \dot{F}'(\vartheta, s) X_s ds,$$

$$\ddot{G}(\vartheta, X^t) = X_t \ddot{F}(\vartheta, t) - X_{\tau_\varepsilon} \ddot{F}(\vartheta, \tau_\varepsilon) - \int_{\tau_\varepsilon}^t \ddot{F}'(\vartheta, s) X_s ds.$$

Therefore we have for any $p > 0$

$$\sup_{|\vartheta - \vartheta_0| < \nu} \mathbf{E}_\vartheta \left| \tilde{R}_\varepsilon \right|^p \xrightarrow{\varepsilon \rightarrow 0} 0.$$

We can write formally

$$\begin{aligned} m_\varepsilon^*(t) - m(\vartheta, t) &= m(\vartheta_{t,\varepsilon}^*, t) - m(\vartheta, t) + o(1) = (\vartheta_{t,\varepsilon}^* - \vartheta) \dot{m}(\vartheta, t) + o(1) \\ &= (\vartheta_{t,\varepsilon}^* - \vartheta) \dot{y}(\vartheta, t) + o(1) \end{aligned}$$

and

$$\begin{aligned} \sup_{|\vartheta - \vartheta_0| < \nu} \varepsilon^{-2} \mathbf{E}_\vartheta |m_\varepsilon^*(t) - m(\vartheta, t)|^2 &= \sup_{|\vartheta - \vartheta_0| < \nu} \frac{\dot{y}(\vartheta, t)^2}{\Gamma^t(\vartheta)} + o(1), \\ \sup_{|\vartheta - \vartheta_0| < \nu} \frac{\dot{y}(\vartheta, t)^2}{\Gamma^t(\vartheta)} &\xrightarrow{\nu \rightarrow 0} \frac{\dot{y}(\vartheta_0, t)^2}{\Gamma^t(\vartheta_0)}. \end{aligned}$$

6. Two examples

We show the construction of preliminary estimators, which are simpler and could be easily realized by numerical method.

Example 1. In this example we have the following system:

$$\begin{aligned} dX_t &= \vartheta f_t Y_t dt + \varepsilon \sigma_t dW_t, \quad X_0 = 0, \quad 0 \leq t \leq T, \\ dY_t &= a_t Y_t dt + \varepsilon b_t dV_t, \quad Y_0 = y_0 \neq 0. \end{aligned}$$

For simplicity of exposition, we suppose that functions $f_t, a_t, \sigma_t, b_t, t \in [0, T]$ are positive and bounded. The function $f_t \in \mathcal{C}_t^{(1)}$. The observations are $X^T = (X_t, 0 \leq t \leq T)$ and the process $\{Y_t, 0 \leq t \leq T\}$ is *hidden*. The unknown parameter $\vartheta \in \Theta = (\alpha, \beta)$, $\alpha > 0$.

The limit ($\varepsilon = 0$) system is

$$x_\tau(\vartheta) = \vartheta \int_0^\tau f_s y_s ds.$$

We put $\tau_\varepsilon = \varepsilon^\delta$ and define the preliminary estimator by the relation

$$\bar{\vartheta}_{\tau_\varepsilon} = \frac{X_{\tau_\varepsilon}}{\int_0^{\tau_\varepsilon} f_s y_s ds} = \frac{x_{\tau_\varepsilon}(\vartheta_0) + \varepsilon \eta_{\tau_\varepsilon}}{\int_0^{\tau_\varepsilon} f_s y_s ds} = \vartheta_0 + \frac{\varepsilon \eta_{\tau_\varepsilon}}{\int_0^{\tau_\varepsilon} f_s y_s ds}.$$

Hence

$$\mathbf{E}_{\vartheta_0} (\bar{\vartheta}_{\tau_\varepsilon} - \vartheta_0)^2 = \varepsilon^2 \left(\int_0^{\tau_\varepsilon} f_s y_s ds \right)^{-2} \mathbf{E}_{\vartheta_0} \eta_{\tau_\varepsilon}^2 \leq C \frac{\varepsilon^2}{\tau_\varepsilon}.$$

Example 2. In the second example the partially observed system is

$$\begin{aligned} dX_t &= f_t Y_t dt + \varepsilon \sigma_t dW_t, & X_0 &= 0, & 0 \leq t \leq T, \\ dY_t &= \vartheta a_t Y_t dt + \varepsilon b_t dV_t, & y_0 &\neq 0. \end{aligned}$$

We have $Y_t = y_t(\vartheta_0) + \varepsilon \xi_t$ and the limit equation is

$$y_t(\vartheta_0) = y_0 + \vartheta_0 \int_0^t a_s y_s(\vartheta_0) ds.$$

Therefore we can write

$$\begin{aligned} X_t &= \int_0^t f_s Y_s ds + \varepsilon \int_0^t \sigma_s dW_s \\ &= \int_0^t f_s y_s(\vartheta_0) ds + \varepsilon \int_0^t f_s \xi_s ds + \varepsilon \int_0^t \sigma_s dW_s \\ &= y_0 \int_0^t f_s ds + \vartheta_0 \int_0^t f_s \int_0^s a_r y_r(\vartheta_0) dr ds + \varepsilon \int_0^t f_s \xi_s ds + \varepsilon \int_0^t \sigma_s dW_s. \end{aligned}$$

For small values of τ we have

$$X_\tau - y_0 \int_0^\tau f_s ds = \vartheta_0 f_0 a_0 y_0 \frac{\tau^2}{2} + O(\tau^3) + \varepsilon \int_0^\tau f_s \xi_s ds + \varepsilon \int_0^\tau \sigma_s dW_s.$$

The preliminary estimator can be defined as follows

$$\bar{\vartheta}_{\tau_\varepsilon} = \frac{2}{f_0 a_0 y_0 \tau_\varepsilon^2} \left(X_{\tau_\varepsilon} - y_0 \int_0^{\tau_\varepsilon} f_s ds \right). \quad (6.1)$$

The error of estimation is

$$\mathbf{E}_{\vartheta_0} (\bar{\vartheta}_{\tau_\varepsilon} - \vartheta_0)^2 \leq C \tau_\varepsilon^2 + C \frac{\varepsilon^2}{\tau_\varepsilon^3} = C (\varepsilon^{2\delta} + \varepsilon^{2-3\delta}).$$

Hence if $\delta \in (0, 2/3)$, then this estimator is consistent. The optimal choice, which minimizes the right hand side of this inequality is $\delta = 2/5$.

In the simulation, we take firstly constant coefficient functions. We put $f_t = f$, $a_t = a$, $\sigma_t = \sigma$, $b_t = b$ for constants f , a , σ and b . Thus we have

$$y_t(\vartheta) = y_0 e^{a\vartheta t} \quad \text{and} \quad x_t(\vartheta) = x_0 + \frac{f y_0}{a\vartheta} e^{a\vartheta t}.$$

With fixed constants and ε , true value $\vartheta_0 = 1$, we simulate processes X_t and Y_t , and then calculate the preliminary estimator $\bar{\vartheta}_{\tau_\varepsilon}$ according to (6.1), with $\tau_\varepsilon = \varepsilon^{1/6}$. Thus we have the estimator-process

$$\vartheta_{t,\varepsilon}^* = \bar{\vartheta}_{\tau_\varepsilon} + I_{\tau_\varepsilon}^t (\bar{\vartheta}_{\tau_\varepsilon})^{-1} \int_{\tau_\varepsilon}^t \frac{\dot{M}(\bar{\vartheta}_{\tau_\varepsilon}, s)}{\sigma(s)^2} [dX_s - f(\bar{\vartheta}_{\tau_\varepsilon}, s)m(\bar{\vartheta}_{\tau_\varepsilon}, s)ds].$$

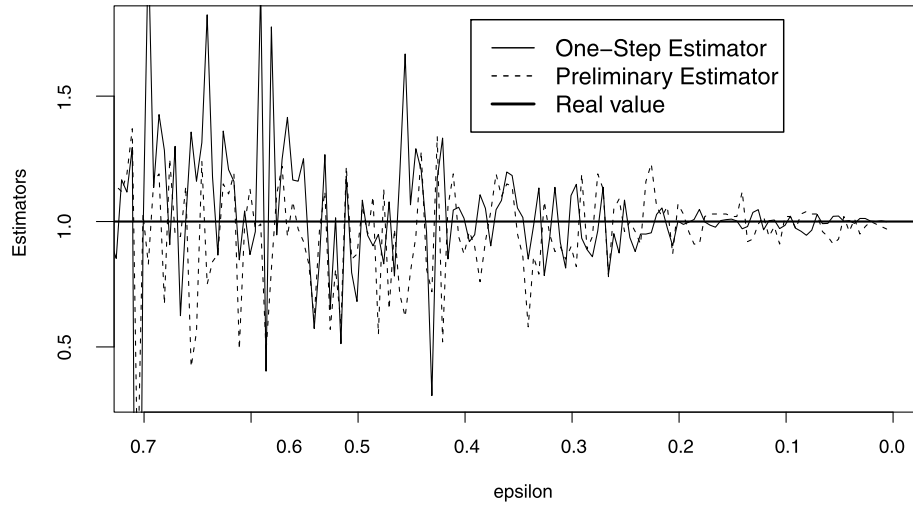


FIG 1. Consistency w.r.t. ε for constant coefficients, $\vartheta_0 = 1$.

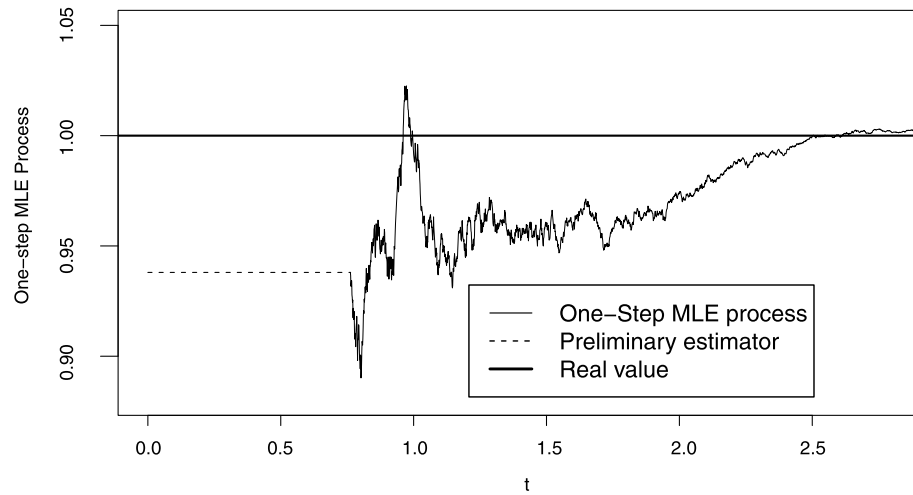


FIG 2. Consistency w.r.t. t for constant coefficients, $\vartheta_0 = 1$, $\varepsilon = 0.2$.

We show in the first graphic Figure 1, the consistency of the preliminary estimator $\vartheta_{T,\varepsilon}$ and the One-step estimator $\vartheta_{T,\varepsilon}^*$ w.r.t. ε . Note that the estimators converge to the true value when ε goes to zero. In the second graphic Figure 2, we show the consistency of the One-step MLE process $\vartheta_{t,\varepsilon}^*$ w.r.t. t , for fixed $\varepsilon = 0.2$. The process begins with a preliminary estimator and converges to the true value after some fluctuation. The fluctuation after the preliminary estimator is caused by the small value of the initial Fisher information. In the third graphic Figure 3, we present the normal approximation of $\varepsilon^{-1}(\vartheta_{t,\varepsilon}^* - \vartheta_0)$ for fixed ε .

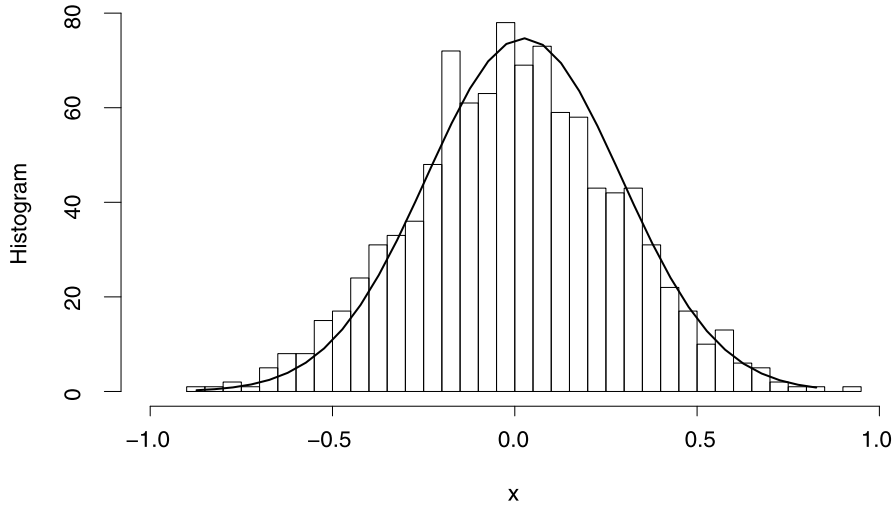


FIG 3. Asymptotical normality for constant coefficients.

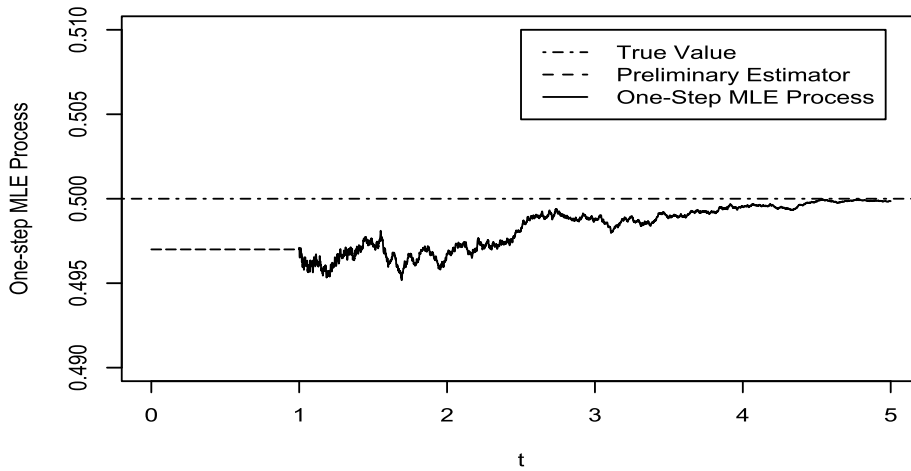


FIG 4. Consistency w.r.t. t for linear coefficients, $\vartheta_0 = 0.5$, $\varepsilon = 0.05$.

Similar work was done for linear coefficients. The convergence of the estimator is shown in Figure 4 and 5 w.r.t. t and ε respectively.

7. Discussion

1. The main conditions in the construction of One-step MLE ϑ_ε^* and One-step MLE-process $\vartheta_{t,\varepsilon}^*, \tau_\varepsilon < t \leq T$ are $|\dot{f}(\vartheta, 0)| > 0$ or $|\dot{a}(\vartheta, 0)| > 0$. Consider the situation where both conditions are not fulfilled. Suppose that there exists

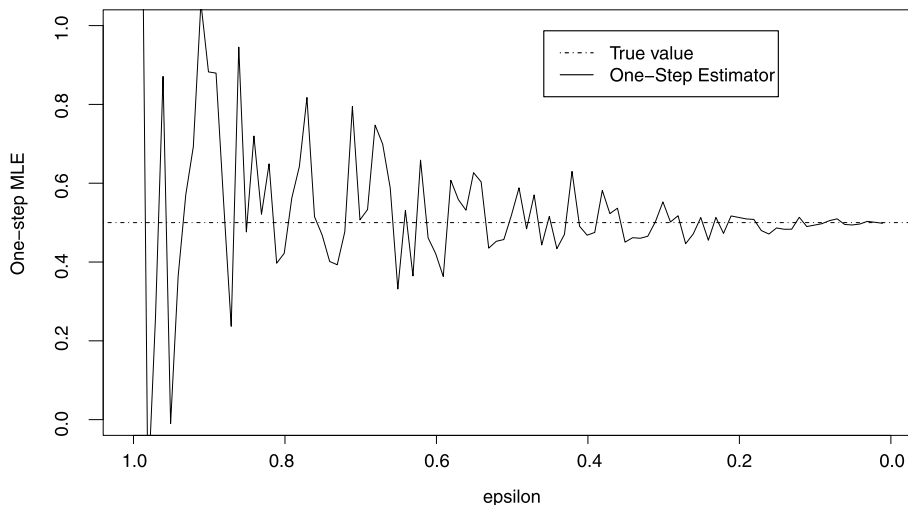


FIG 5. Consistency w.r.t. ε for linear coefficients, $\vartheta_0 = 0.5$.

$\tau_0 \in (0, T)$ such that

$$f(\vartheta, t) = f(t) \mathbb{1}_{\{0 \leq t < \tau_0\}} + \vartheta g(t) \mathbb{1}_{\{\tau_0 \leq t \leq T\}},$$

and $a(\vartheta, t) = a(t)$, $0 \leq t \leq T$. Then the construction of these estimators can be done as follows. First we solve the filtration equations on the interval $[0, \tau_0]$

$$\begin{aligned} dm(t) &= a(t)m(t)dt + \frac{\gamma_*(t)f(t)}{\sigma(t)^2} [dX_t - f(t)m(t)dt], \quad m(0) = y_0, \\ \frac{\partial \gamma_*(t)}{\partial t} &= 2a(t)\gamma_*(t) - \frac{\gamma_*(t)^2 f(t)^2}{\sigma(t)^2} + b(t)^2, \quad \gamma_*(0) = 0. \end{aligned}$$

Then we take $m(\tau_0), \gamma_*(\tau_0)$ as initial values for the filtration equations on the interval $[\tau_0, T]$ and the preliminary estimator $\bar{\vartheta}_{\tau_0 + \tau_\varepsilon}$ can be constructed by observations $X_{\tau_0}^{\tau_\varepsilon} = (X_t, \tau_0 \leq t \leq \tau_0 + \tau_\varepsilon)$.

2. Another question concerns the length of the “learning interval” $[0, \tau_\varepsilon]$, where $\tau_\varepsilon = \varepsilon^\delta$ and $\delta \in (0, 1)$. If we need the approximation of $m(\vartheta_0, t)$ for the values t smaller than the given τ_ε , then we can use the construction of Two-step MLE-process described in the work [18].

3. In the case of multidimensional $\vartheta \in \Theta \subset \mathcal{R}^d$ the main problem is in the construction of the preliminary estimators. One way is to suppose that the observed process is as well multidimensional $X^T = (X_1^T, \dots, X_k^T)$, where $X_j^T = (X_{j,t}, 0 \leq t \leq T)$ and $k \geq d$. For example, suppose that

$$\begin{aligned} dX_{j,t} &= f_j(\vartheta, t)Y_t dt + \varepsilon\sigma_j(t) dW_{j,t}, \quad X_{j,0} = 0, \quad j = 1, \dots, k, \\ dY_t &= a(\vartheta, t)Y_t dt + \varepsilon b(t) dV_t, \quad Y_0 = y_0 \neq 0. \end{aligned}$$

Here $W_{1,t}, \dots, W_{k,t}, V_t, 0 \leq t \leq T$ are independent Wiener processes. Then the limit deterministic system is

$$\frac{dx_{j,t}(\vartheta)}{dt} = f_j(\vartheta, t) y_0 \exp \left\{ \int_0^t a(\vartheta, s) ds \right\}, \quad x_{j,0}(\vartheta) = 0, \quad j = 1, \dots, k.$$

The version of the condition \mathcal{R}_3 in this case is based on the asymptotic (small t) representation of $x_{j,t}(\vartheta)$. The purpose of this condition is to provide the consistency of preliminary estimator. The choice of τ_ε has to yield the good rate of convergence of this preliminary estimator.

Another possibility is to fix the learning interval $[0, \tau]$. Then even for the model of observations (1.1)-(1.2) with $\vartheta \in \Theta \subset \mathcal{R}^d$ we can use the *minimum distance estimator* (MDE) $\bar{\vartheta}_\varepsilon$ defined by the equation

$$\bar{\vartheta}_\varepsilon = \arg \inf_{\vartheta \in \Theta} \int_0^\tau [X_t - x_t(\vartheta)]^2 dt.$$

In this case the condition \mathcal{R}_3 is replaced by identifiability condition: for any $\vartheta_0 \in \Theta$ and any $\nu > 0$ we have

$$\inf_{\|\vartheta - \vartheta_0\| > \nu} \int_0^\tau [x_t(\vartheta) - x_t(\vartheta_0)]^2 dt > 0.$$

It can be shown that the estimator $\bar{\vartheta}_\varepsilon$ is consistent and $\varepsilon^{-1}(\bar{\vartheta}_\varepsilon - \vartheta_0)$ is asymptotically normal (see details in [16]). Then One-step MLE-process can be introduced and studied following the same steps as it was done above.

Acknowledgements

The authors are grateful to the reviewer for useful comments. The first author was supported by RSF project no 20-61-47043. The second author gratefully acknowledge financial support from the National Natural Science Foundation of China (NSFC) under grant 11701329.

References

- [1] Arato, M. (1983) *Linear Stochastic Systems with Constant Coefficients A Statistical Approach*. Lecture Notes in Control and Inform. Sci., 45, New York: Springer-Verlag. [MR0791212](#)
- [2] Cappé, O., Moulines, E. and Rydén, T. (2005) *Inference in Hidden Markov Models*. Springer, N.Y. [MR2159833](#)
- [3] Chigansky, P. (2009) Maximum likelihood estimation for hidden Markov models in continuous time. *Statist. Inference Stoch. Processes*, 12, 2, 139-163. [MR2511676](#)
- [4] Elliott, R.J., Aggoun, L. and Moor, J.B. (1995) *Hidden Markov Models*. Springer, N.Y. [MR1323178](#)

- [5] Ephraim, Y., Mehrav, N. (2002) Hidden Markov processes. *IEEE Trans. Inform. Theory*, 48, 6, 1518-1569. [MR1909472](#)
- [6] Fisher, R.A. (1925) *Theory of Statistical Estimation*. Proc. Camb. Phil. Soc., 22, 700-725.
- [7] Gill, R.D. and Levit, B.Ya. (1995) Application of the van Trees inequality: a Bayesian Cramer-Rao bound. *Bernoulli*, 1, 59-79. [MR1354456](#)
- [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] Gustafsson, F. (2000) *Adaptive Filtering and Change Detection*. J. Wiley&Sons, N.Y.
- [10] Hide, C., Moore, T. and Smith, M. (2003) Adaptive Kalman filtering for low cost ING/GPS. *The Journal of Navigation*, 56, 1, 143-152.
- [11] Hu, C., Chen, W., Chen, Y. and Liu, D. (2003) Adaptive Kalman filtering for vehicule navigation. *J. Global Positioning systems*, 2, 1, 42-47.
- [12] 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](#)
- [13] Khasminskii, R. (2005) Nonlinear filtering of smooth signals. *Stochastics and Dynamics*, 5, 1, 27-35. [MR2118752](#)
- [14] Khasminskii, R.Z. and Kutoyants, Yu.A. (2018) On parameter estimation of hidden telegraph process. *Bernoulli*, 24, 3, 2064-2090. [MR3757523](#)
- [15] Kutoyants, Yu.A. (1984) *Parameter Estimation for Stochastic Processes*. Heldermann, Berlin. [MR0777685](#)
- [16] Kutoyants, Yu.A. (1994) *Identification of Dynamical Systems with Small Noise*. Kluwer Academic Publisher, Dordrecht. [MR1332492](#)
- [17] Kutoyants, Yu.A. (2004) *Statistical Inference for Ergodic Diffusion Processes*. Springer, London. [MR2144185](#)
- [18] Kutoyants, Yu.A. (2017) On the multi-step MLE-process for ergodic diffusion. *Stochastic Process. Appl.*, 127, 2243-2261. [MR3652412](#)
- [19] Kutoyants, Yu.A. (2019) On parameter estimation of the hidden Ornstein-Uhlenbeck process. *J. Multivar. Analysis*, 169, 1, 248-269. [MR3875598](#)
- [20] Kutoyants, Yu.A. (2019) On parameter estimation of the hidden ergodic Ornstein-Uhlenbeck process. *Electronic J. of Statistics*, 13, 4508-4526. [MR4029160](#)
- [21] Kutoyants, Yu.A. and Motrunich, A. (2016) On multi-step MLE-process for Markov sequences. *Metrika*, 79, 705-724. [MR3518583](#)
- [22] Kutoyants, Yu.A. and Zhou, L. (2014) On approximation of the backward stochastic differential equation. *J. Stat. Plann. Infer.* 150, 111-123. [MR3206723](#)
- [23] Liptser, R.S. and Shirayayev, A.N. (2001) *Statistics of Random Processes, I. General Theory*. 2nd Ed., Springer, N.Y. [MR1800857](#)
- [24] Rutan, S.C. (1991) Adaptive Kalman filtering. *Anal. Chem.*, 63 (22), 1103A-1109A.
- [25] Zhou, L. (2014) On asymptotically parameter free test for ergodic diffusion process. *Revue des Annales de l'ISUP*, 58-1-2, 37-56. [MR3236780](#)