

# On approximation of BSDE and multi-step MLE-processes



Yu A. Kutoyants

Received: 2 January 2016 / Accepted: 15 June 2016 / Published online: 16 August 2016

© The Author(s). 2016 **Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

**Abstract** We consider the problem of approximation of the solution of the backward stochastic differential equations in Markovian case. We suppose that the forward equation depends on some unknown finite-dimensional parameter. This approximation is based on the solution of the partial differential equations and multi-step estimator-processes of the unknown parameter. As the model of observations of the forward equation we take a diffusion process with small volatility. First we establish a lower bound on the errors of all approximations and then we propose an approximation which is asymptotically efficient in the sense of this bound. The obtained results are illustrated on the example of the Black and Scholes model.

**Keywords** Backward stochastic differential equation · Parameter estimation · Multi-step MLE-process · Small noise · Black and Scholes model

**MSC 2000 Classification** 62M05

## Introduction

We consider the problem of approximation of the solution of the backward stochastic differential equation (BSDE) in the so-called Markovian case. Let us recall some basics of BSDEs. We are given a stochastic differential equation (called *forward*)

$$dX_t = S(t, X_t) dt + \sigma(t, X_t) dW_t, \quad X_0 = x_0, \quad 0 \leq t \leq T,$$

---

Y.A. Kutoyants (✉)

Laboratoire Manceau des Mathématiques, Université du Maine Le Mans, France and National Research University “MPEI”, Moscow, Russia  
e-mail: kutoyants@univ-lemans.fr

where  $S(t, x)$  is the drift coefficient,  $\sigma(t, x)^2$  is the diffusion coefficient, and  $W_t, 0 \leq t \leq T$  is a standard Wiener process. In addition, we have two functions  $f(t, x, y, z)$  and  $\Phi(x)$  and we must construct such couple of processes  $(Y_t, Z_t, 0 \leq t \leq T)$  that the solution of the equation

$$dY_t = -f(t, X_t, Y_t, Z_t) dt + Z_t dW_t, \quad Y_0, \quad 0 \leq t \leq T,$$

(called *backward*) has the final value  $Y_T = \Phi(X_T)$ . Such BSDEs were first introduced by Bismut in 1973 (Bismut 1973) in the linear case and the general theory was developed by Pardoux and Peng (1990) (Pardoux and Peng 1990). The *Markovian case* considered in this work was studied by Pardoux and Peng (Pardoux and Peng 1992), see Section 4 in El Karoui et al. (1997) as well. This model is also called *forward-backward stochastic differential equation* (FBSDE) (El Karoui et al. 1997).

The construction of the backward equation is realized as follows. Suppose that  $u(t, x)$  satisfies the parabolic partial differential equation

$$\frac{\partial u}{\partial t} + S(t, x) \frac{\partial u}{\partial x} + \frac{1}{2} \sigma(t, x)^2 \frac{\partial^2 u}{\partial x^2} = -f\left(t, x, u, a(t, x) \frac{\partial u}{\partial x}\right),$$

with the final condition  $u(T, x) = \Phi(x)$ . Let us let  $Y_t = u(t, X_t)$ , and  $Z_t = \sigma(t, X_t) u'_x(t, X_t)$ . Then, by Itô's formula

$$\begin{aligned} dY_t &= \left[ \frac{\partial u}{\partial t}(t, X_t) + S(t, X_t) \frac{\partial u}{\partial x}(t, X_t) + \frac{1}{2} \sigma(t, x)^2 \frac{\partial^2 u}{\partial x^2}(t, X_t) \right] dt \\ &\quad + \sigma(t, X_t) \frac{\partial u}{\partial x}(t, X_t) dW_t \\ &= -f(t, X_t, Y_t, Z_t) dt + Z_t dW_t, \quad Y_0 = u(0, X_0), \quad 0 \leq t \leq T. \end{aligned}$$

The final value  $Y_T = u(T, X_T) = \Phi(X_T)$ . Therefore, if we have the solution  $u(t, x)$ , then we immediately obtain the BSDE.

We are interested by the problem of approximation of  $(Y_t, Z_t, 0 \leq t \leq T)$  in the situation, where the forward equation contains some unknown finite-dimensional parameter  $\vartheta$ :

$$dX_t = S(\vartheta, t, X_t) dt + \sigma(\vartheta, t, X_t) dW_t, \quad X_0 = x_0, \quad 0 \leq t \leq T.$$

Then the solution of the PDE  $u = u(t, x, \vartheta)$ . We cannot simply let  $Y_t = u(t, X_t, \vartheta)$  because we do not know  $\vartheta$ . Of course, the *natural* way to approximate  $Y_t$  and  $Z_t$  is to estimate first the unknown parameter  $\vartheta$  with the help of some estimator  $\bar{\vartheta}$  and then to put, say,  $\bar{Y}_t = u(t, X_t, \bar{\vartheta})$ . We can guess that if  $\bar{\vartheta}$  is a good estimator of  $\vartheta$ , then  $\bar{Y}_t$  will be a good estimator of  $Y_t$ . There are several problems, that are interesting to study in this framework. We must understand what the conditions imposed on the estimator  $\bar{\vartheta}$  that allow us to say that it is good. We consider that a good estimator has the following properties.

1. *To estimate  $Y_t$  we need an estimator, which is constructed by the first observations of the solution of the forward equation up to time  $t$ , i.e.,  $\bar{\vartheta}_t = \bar{\vartheta}_t(X_s, 0 \leq s \leq t), 0 < t \leq T$ .*
2. *As we need such estimator for all  $t \in (0, T]$  we suppose that its calculation must be relatively simple.*

3. The error of estimation, say,  $E_{\vartheta_0} (\bar{\vartheta}_t - \vartheta_0)^2$  must be as minimal as possible.

Therefore  $\bar{\vartheta}$  is an estimator-process  $\bar{\vartheta} = (\bar{\vartheta}_t, 0 < t \leq T)$ . Of course, the construction of such estimator-process is an intermediate problem. The main problem is to obtain a good approximations of  $Y_t$  and  $Z_t$ . In particular, we must show that the approximations

$$\bar{Y}_t = u(t, X_t, \bar{\vartheta}_t), \quad \bar{Z}_t = u'_x(t, X_t, \bar{\vartheta}_t) \sigma(\bar{\vartheta}_t, t, X_t)$$

are in some sense asymptotically optimal, i.e., it is impossible to have approximations of these processes with asymptotic errors smaller than that of  $\bar{Y}_t$  and  $\bar{Z}_t$ .

The goal of the study initiated in Kutoyants and Zhou (2014) is to realize such a program for three models of observations of the forward equation. As is usual in statistics, we consider situations where it is possible to have a consistent estimation of the unknown parameters and processes. Therefore, we are interested by the following well-known models of observations.

- Diffusion process with an unknown parameter in the drift coefficient and *small noise or small volatility*

$$dX_t = S(\vartheta, t, X_t) dt + \varepsilon \sigma(t, X_t) dW_t, \quad x_0, \quad 0 \leq t \leq T. \tag{1}$$

Here the time  $T$  of observations  $X^T = (X_t, 0 \leq t \leq T)$  is fixed and the limit corresponds to  $\varepsilon \rightarrow 0$ .

- Diffusion process

$$dX_t = S(t, X_t) dt + \sigma(\vartheta, t, X_t) dW_t, \quad X_0, \quad 0 \leq t \leq T, \tag{2}$$

observed in the discrete times  $X^n = (X_{t_0}, X_{t_1}, \dots, X_{t_n})$ ,  $t_i = i \frac{T}{n}$ . Here the unknown parameter is in the volatility coefficient and the limit corresponds to  $n \rightarrow \infty$  (high frequency model of observations). The time  $T$  of observations is fixed.

- Ergodic diffusion process

$$dX_t = S(\vartheta, X_t) dt + \sigma(X_t) dW_t, \quad X_0, \quad 0 \leq t \leq T. \tag{3}$$

Here the unknown parameter  $\vartheta$  is in the drift coefficient, we have continuous time observations  $X^T = (X_t, 0 \leq t \leq T)$  and the limit is  $T \rightarrow \infty$ .

Of course there are other possible statements. For example, it can be considered the *mixture* of discrete time and ergodic diffusion. This corresponds to the equation

$$dX_t = S(\vartheta, X_t) dt + \sigma(\vartheta, X_t) dW_t, \quad X_0, \quad 0 \leq t \leq T_n$$

and observations  $X^n = (X_{t_0}, X_{t_1}, \dots, X_{t_n})$ . Here  $\max_i |t_i - t_{i-1}| \rightarrow 0$  and  $T_n \rightarrow \infty$ . Such a model of parameter estimation was studied, e.g., in Kamatani and Uchida (2015) and Uchida and Yoshida (2014). It is possible to consider the *mixture* of discrete-time and small noise models, to consider the model with  $X_t \rightarrow \pm\infty$  or the models with null recurrent forward equation *etc*. It will be interesting to see the statements of the statistical problems in non-Markovian cases for more general models.

Let us describe the general framework of the statistical study of the above mentioned three models (1)-(3). For each model we propose an estimator-processes

$\vartheta_t^*$ ,  $0 < t \leq T$  such that  $Y_t^* = u(t, X_t, \vartheta_t^*) \rightarrow Y_t$  and the error of approximation  $\mathbf{E}_{\vartheta} (Y_t^* - Y_t)^2$  is asymptotically minimal. In the earlier works Kutoyants and Zhou (2014) and Gasparyan and Kutoyants (2015); (Abakirova, A and Kutoyants, YA: On approximation of the BSDE. Large samples approach. *In preparation*) (see the review of these works in Kutoyants (2014)) we considered the approximation of the solution of BSDEs with a learning interval of fixed length.

The optimality of estimators of  $Y_t$  and  $Z_t$  is understood as follows. We define for each model a normalization function  $\varphi \rightarrow 0$ , i.e.,  $\varphi_{\varepsilon} \rightarrow 0$  as  $\varepsilon \rightarrow 0$ ,  $\varphi_n \rightarrow 0$  as  $n \rightarrow \infty$ , and  $\varphi_T \rightarrow 0$  as  $T \rightarrow \infty$ .

We propose the lower bounds on the risks of all estimators

$$\lim_{\delta \rightarrow 0} \lim_{\varepsilon, n, T} \sup_{|\vartheta - \vartheta_0| < \delta} \mathbf{E}_{\vartheta} \left| \frac{\bar{Y}_t - Y_t}{\varphi} \right|^2 \geq D(\vartheta_0)^2,$$

which allow us to define the *asymptotically efficient estimators*  $Y_t^*$  of  $Y_t$  as follows

$$\lim_{\delta \rightarrow 0} \lim_{\varepsilon, n, T} \sup_{|\vartheta - \vartheta_0| < \delta} \mathbf{E}_{\vartheta} \left| \frac{Y_t^* - Y_t}{\varphi} \right|^2 = D(\vartheta_0)^2.$$

We suppose that the last equality takes place for all  $\vartheta_0 \in \Theta$  and all  $t \in (0, T]$ . We also have a similar bound in the problem of estimation of  $Z_t$ . For models (1) and (3) these bounds are slight modifications of the Hajek-Le Cam lower bound (Ibragimov and Has'minskii 1981) and for model (2) the lower bound is similar to Jeganathan's lower bound (Jeganathan 1983).

We take the quadratic loss function just for simplicity of exposition. For all mentioned models, the similar lower bounds and corresponding estimator processes can be proved for more general loss functions.

The approximation of the solution of BSDEs in the Markovian case were initiated in the work Kutoyants and Zhou (2014), where the model of small volatility was considered. The parameter  $\vartheta$  was supposed one-dimensional and the approximation-process  $Y_{t,\varepsilon}^*$  was defined for  $t \in [\tau, T]$ , where  $\tau > 0$  is a fixed value.

In the work Gasparyan and Kutoyants (2015), we considered the model of discrete-time observations (2) and the one-step MLE-process which allowed us to construct an estimator-process  $Y_{t_k,n}^*$  for the values  $t_{k,n} \in [\tau, T]$ , where  $\tau > 0$  is fixed.

The case of ergodic diffusion process is considered in the work (Abakirova, A and Kutoyants, YA: On approximation of the BSDE. Large samples approach. *In preparation*), which is still in progress.

The main contribution of the present work is due to a new class of estimator-processes called multi-step MLE-process introduced in Kutoyants (2015). These estimator-processes allow us to construct the approximations of the solutions of BSDEs for three above mentioned models with vanishing learning intervals (models (1) and (2)) or negligible with respect to the whole volume of observations learning interval (model (3)). Here we consider the model (1) only. The models (2) and (3) we leave to study later.

In the present work, we consider the small volatility model where we suppose that the unknown parameter is multi-dimensional and the approximation process  $Y_{t,\varepsilon}^*$  we define for  $t \in [\tau_{\varepsilon}, T]$ , where  $\tau_{\varepsilon} \rightarrow 0$ . This approximation allows us to consider the

case  $\tau_\varepsilon = \varepsilon^\delta \rightarrow 0$  and, moreover, to choose  $\delta$  close to 2. The relations between the choice of  $\delta$  and the multi-step MLE-processes are the following. If  $\delta \in (0, 1)$ , then we use the one-step MLE-process  $\vartheta_{t,\varepsilon}^*$ ; if  $\delta \in [1, \frac{4}{3})$ , then we use the two-step MLE-process  $\vartheta_{t,\varepsilon}^{**}$ ; if  $\delta \in [\frac{4}{3}, \frac{3}{2})$ , then we use the three-step MLE-process  $\vartheta_{t,\varepsilon}^{***}$ .

In the work Kutoyants (2015) we already studied the multi-step MLE-process for ergodic diffusion process, and the structure of estimator-process proposed in the present work is quite similar.

Note that the multi-step, like the well-known one-step ML-estimators, are based on the so-called *Fisher-score* device proposed by Fisher in 1925 (Fisher 1925) and studied by Le Cam in 1956 (Le Cam 1956). Let us recall this construction. Suppose that we have  $n$  i.i.d. r.v.'s  $X^n = (X_1, \dots, X_n)$  with smooth density function  $f(\vartheta, x)$  and denote  $\ell(\vartheta, x) = \ln f(\vartheta, x)$ . The maximum likelihood equation is

$$\sum_{j=1}^n \dot{\ell}(\hat{\vartheta}_n, X_j) = 0.$$

Here and in the rest of the paper dot means derivation w.r.t.  $\vartheta$ . If we expand it at the vicinity of the true value  $\vartheta_0$ , we obtain

$$\sum_{j=1}^n \dot{\ell}(\vartheta_0, X_j) + (\hat{\vartheta}_n - \vartheta_0) \sum_{j=1}^n \ddot{\ell}(\tilde{\vartheta}_n, X_j) = 0.$$

Therefore

$$\hat{\vartheta}_n = \vartheta_0 - \frac{\sum_{j=1}^n \dot{\ell}(\vartheta_0, X_j)}{\sum_{j=1}^n \ddot{\ell}(\tilde{\vartheta}_n, X_j)} = \vartheta_0 + \frac{1}{\sqrt{n}} \frac{\frac{1}{\sqrt{n}} \sum_{j=1}^n \dot{\ell}(\vartheta_0, X_j)}{-\frac{1}{n} \sum_{j=1}^n \ddot{\ell}(\tilde{\vartheta}_n, X_j)}. \tag{4}$$

Note that

$$-\frac{1}{n} \sum_{j=1}^n \ddot{\ell}(\vartheta_0, X_j) \rightarrow \mathbb{I}(\vartheta_0) = \int \dot{\ell}(\vartheta_0, x)^2 f(\vartheta_0, x) dx, \tag{5}$$

where  $\mathbb{I}(\vartheta_0)$  is the Fisher information.

Suppose that we have a preliminary estimator  $\bar{\vartheta}_n$  such that

$$\sqrt{n}(\bar{\vartheta}_n - \vartheta_0) \Rightarrow \mathcal{N}(0, D(\vartheta_0)), \quad D(\vartheta_0) > \mathbb{I}(\vartheta_0)^{-1}.$$

Keeping in mind the relations (4)-(5), the one-step MLE  $\vartheta_n^*$  is defined as follows

$$\vartheta_n^* = \bar{\vartheta}_n + \frac{1}{\sqrt{n}} \frac{\Delta_n(\bar{\vartheta}_n, X^n)}{\mathbb{I}(\bar{\vartheta}_n)}, \quad \Delta_n(\vartheta, X^n) = \frac{1}{\sqrt{n}} \sum_{j=1}^n \dot{\ell}(\vartheta, X_j).$$

This estimator is already asymptotically efficient because its limit variance is  $\mathbb{I}(\vartheta_0)^{-1}$ :

$$\sqrt{n}(\vartheta_n^* - \vartheta_0) \Rightarrow \mathcal{N}(0, \mathbb{I}(\vartheta_0)^{-1}).$$

Therefore, this Fisher-score device allows us to improve the preliminary estimator up to asymptotically efficient (see details, e.g., in Lehmann and Romano (2005)).

Moreover, this device can be applied even in the case of preliminary estimator with the rate of convergence worse than  $\sqrt{n}$  (see, e.g., Robinson (1988) and Kamatani and Uchida (2015)). For continuous-time stochastic processes such a construction was used, for example, in Skorohod and Khasminskii (1996).

The one-step MLE-process, introduced in Kutoyants (2015), for this model of observations can be written as follows. Let us denote  $\bar{\vartheta}_N$  the preliminary estimator constructed by the first  $N = [n^\delta]$  observations  $X^N = (X_1, \dots, X_N)$  with  $\delta \in (\frac{1}{2}, 1)$ . Then the one-step MLE-process  $\vartheta_n^* = (\vartheta_{k,n}^*, N + 1 \leq k \leq n)$  is defined by the equality

$$\vartheta_{k,n}^* = \bar{\vartheta}_N + \mathbb{I}(\bar{\vartheta}_N)^{-1} \frac{1}{k} \sum_{j=N+1}^k \dot{\ell}(\bar{\vartheta}_N, X_j) \tag{6}$$

and for  $k = [sn], s \in (0, 1]$  we have the convergence

$$\sqrt{k} (\vartheta_{k,n}^* - \vartheta_0) \implies \mathcal{N}(0, \mathbb{I}(\vartheta_0)^{-1}).$$

Here  $s$  is fixed and  $n \rightarrow \infty$ . Therefore  $\vartheta_n^*$  is a *good estimator*, i.e.,  $\vartheta_{k,n}^*$  depends on  $X^k = (X_1, \dots, X_k)$ , is easy to calculate and is asymptotically efficient because it is asymptotically equivalent to the MLE. For the details see Kutoyants and Motrunich (2016).

The one-step MLE-process in the case of ergodic diffusion forward Eq. 3 can be illustrated as follows. Suppose that we have a preliminary estimator  $\bar{\vartheta}_{T^\delta}$  constructed by the observations  $X^{T^\delta} = (X_t, 0 \leq t \leq T^\delta)$  with  $\delta \in (\frac{1}{2}, 1)$ . Then the one-step MLE-process  $\vartheta_{t,T}^*, T^\delta < t \leq T$  based on the Fisher-score device (4), (6) has the following form

$$\vartheta_{t,T}^* = \bar{\vartheta}_{T^\delta} + \mathbb{I}(\bar{\vartheta}_{T^\delta})^{-1} \int_{T^\delta}^t \frac{\dot{S}(\bar{\vartheta}_{T^\delta}, X_s)}{t \sigma(X_s)^2} [dX_s - S(\bar{\vartheta}_{T^\delta}, X_s) ds]. \tag{7}$$

This estimator-process is asymptotically efficient ( $t = rT; r \in (0, 1)$ )

$$\sqrt{t} (\vartheta_{t,T}^* - \vartheta_0) \implies \mathcal{N}(0, \mathbb{I}(\vartheta_0)^{-1})$$

(see Kutoyants (2015)) and provides asymptotically efficient estimator-processes

$$Y_{t,T}^* = u(t, X_t, \vartheta_{t,T}^*), \quad Z_{t,T}^* = u'_x(t, X_t, \vartheta_{t,T}^*) \sigma(X_t) \tag{8}$$

of the solution  $(Y_t, Z_t)$  of the BSDE.

### Forward equation with small volatility

We are given the function  $f(t, x, y, z)$  defined on  $[0, T] \times \mathcal{R}^k \times \mathcal{R} \times \mathcal{R}^k$ , function  $\Phi(x), x \in \mathcal{R}^k$  and  $k$ -dimensional diffusion process (forward)

$$dX_t = S(\vartheta, t, X_t) dt + \varepsilon \sigma(t, X_t) dW_t, \quad X_0, 0 \leq t \leq T. \tag{9}$$

Here  $\vartheta \in \Theta \subset \mathcal{R}^d$ ,  $\Theta$  is an open bounded set and  $W_t = (W_t^1, \dots, W_t^k), 0 \leq t \leq T$  is a standard  $k$ -dimensional Wiener process.

Introduce the condition  $\mathcal{L}$ .

The functions  $f(t, x, y, z)$ ,  $\Phi(x)$ , vector  $S(\vartheta, t, x) = (S_l(\vartheta, t, x), l = 1, \dots, k)$  and  $k \times k$  matrix  $\sigma(t, x) = (\sigma_{lm}(t, x))$  are smooth

$$|S(\vartheta, t, x) - S(\vartheta, t, y)| + |\sigma(t, x) - \sigma(t, y)| \leq L|x - y|,$$

$$|f(t, x, y_1, z_1) - f(t, x, y_2, z_2)| \leq C[|y_1 - y_2| + |z_1 - z_2|]$$

and satisfy ( $p > 0$ )

$$|S(\vartheta, t, x)| + |\sigma(t, x)| \leq C(1 + |x|),$$

$$|f(t, x, y, z)| + |\Phi(x)| \leq C(1 + |x|^p).$$

We must find a couple of stochastic processes  $(X_{t,\varepsilon}^*, Z_{t,\varepsilon}^*, 0 \leq t \leq T)$  which approximate well the solution of the BSDE

$$dY_t = -f(t, X_t, Y_t, Z_t) dt + Z_t dW_t, \quad Y_0, \quad 0 \leq t \leq T \tag{10}$$

satisfying the condition  $Y_T = \Phi(X_T)$ .

Let us denote  $x_t(\vartheta) = (x_t^{(1)}(\vartheta), \dots, x_t^{(k)}(\vartheta))$ ,  $0 \leq t \leq T$  the solution of the system of ordinary differential equations

$$\frac{dx_t(\vartheta)}{dt} = S(\vartheta, t, x_t(\vartheta)), \quad x_0, \quad 0 \leq t \leq T.$$

The true value is  $\vartheta_0$  and we let  $x_t = x_t(\vartheta_0)$ . We have the estimates: with probability 1

$$\sup_{0 \leq t \leq T} |X_t - x_t| \leq C\varepsilon \sup_{0 \leq t \leq T} |W_t| \tag{11}$$

and for any  $p > 0$

$$\sup_{0 \leq t \leq T} \mathbf{E}_{\vartheta_0} |X_t - x_t|^p \leq C\varepsilon^p. \tag{12}$$

For the proof see, e.g., Kutoyants (1994).

We have a family of problems of parameter estimation by observations  $X^t = (X_s, 0 \leq s \leq t)$ , where  $t \in (0, T]$  and therefore we need a family of estimators  $\vartheta_{t,\varepsilon}$ ,  $0 < t \leq T$ . Let  $(\mathcal{C}^k([0, t]), \mathfrak{B}_t)$  be a measurable space of continuous vector-functions on  $[0, t]$  with Borelian  $\sigma$ -algebra  $\mathfrak{B}_t$ . Denote by  $\{\mathbf{P}_{\vartheta}^{(\varepsilon, t)}, \vartheta \in \Theta\}$  the family of measures induced in this space by the solutions of (9) with different  $\vartheta \in \Theta$ .

Note that these measures are equivalent (see Liptser and Shiryaev (2001)) and the likelihood ratio function is

$$L(\vartheta, X^t) = \frac{d\mathbf{P}_\vartheta^{(\varepsilon,t)}}{d\mathbf{P}_0^{(\varepsilon,t)}}(X^t) = \exp \left\{ \frac{1}{\varepsilon^2} \int_0^t S(\vartheta, s, X_s) \mathbb{A}(s, X_s)^{-1} dX_t - \frac{1}{2\varepsilon^2} \int_0^t S(\vartheta, s, X_s) \mathbb{A}(s, X_s)^{-1} S(\vartheta, s, X_s) ds \right\}, \quad \vartheta \in \Theta.$$

Here  $\mathbf{P}_0^{(\varepsilon,t)}$  is the measure, which corresponds to the observations (9) with  $S(\vartheta, t, X_t) \equiv 0$ . The matrix  $\mathbb{A}(s, x)$  is

$$\mathbb{A}_{lm}(s, x) = [\sigma(s, x)^* \sigma(s, x)]_{lm}, \quad l, m = 1, \dots, k.$$

Recall that the MLE  $\hat{\vartheta}_{\varepsilon,t}$  is defined by the equation

$$L(\hat{\vartheta}_{\varepsilon,t}, X^t) = \sup_{\vartheta \in \Theta} L(\vartheta, X^t). \tag{13}$$

Introduce the *Regularity conditions*  $\mathfrak{R}$ .

1. The function  $S(\vartheta, t, x)$  is two-times continuously differentiable w.r.t.  $\vartheta$  and the derivatives are Lipschitz in  $x$ .
2. We suppose that there exists a positive constant  $m$  such that for any real  $\lambda \in \mathcal{R}^k$  we have

$$m^{-1} \|\lambda\|^2 \leq \lambda^* \mathbb{A}(s, x) \lambda \leq m \|\lambda\|^2. \tag{14}$$

3. The Fisher information matrix

$$\mathbb{I}_t(\vartheta) = \int_0^t \dot{S}(\vartheta, s, x_s(\vartheta))^* \mathbb{A}(s, x_s(\vartheta))^{-1} \dot{S}(\vartheta, s, x_s(\vartheta)) ds,$$

is uniformly nondegenerate:

$$\inf_{\vartheta \in \Theta} \inf_{|\lambda|=1} \lambda^* \mathbb{I}_t(\vartheta) \lambda > 0$$

Here  $\lambda \in \mathcal{R}^d$  dot means derivation w.r.t.  $\vartheta$  and  $\dot{S}(\vartheta, s, x)$  is  $k \times d$ -matrix.

4. Identifiability condition : for any  $\nu > 0$  and any  $t \in (0, T]$  the estimate

$$\inf_{\vartheta_0 \in \Theta} \inf_{|\vartheta - \vartheta_0| > \nu} \int_0^t \delta(s, x_s, \vartheta, \vartheta_0)^* \mathbb{A}(s, x_s)^{-1} \delta(s, x_s, \vartheta, \vartheta_0) ds > 0$$

holds. Here  $\delta(s, x_s, \vartheta, \vartheta_0) = S(\vartheta, s, x_s) - S(\vartheta_0, s, x_s)$ .

The Regularity conditions allow us to prove the following properties of the MLE  $\hat{\vartheta}_{\varepsilon,t}, t \in (0, T]$ .

1. It is uniformly consistent: for any  $\nu > 0$  and any compact  $\mathbf{K} \subset \Theta$

$$\lim_{\varepsilon \rightarrow 0} \sup_{\vartheta_0 \in \mathbf{K}} \mathbf{P}_{\vartheta_0}^{(\varepsilon,t)} \left( \left| \hat{\vartheta}_{\varepsilon,t} - \vartheta_0 \right| > \nu \right) = 0.$$

2. Uniformly on compacts  $\mathbf{K} \subset \Theta$  asymptotically normal

$$\varepsilon^{-1} \left( \hat{\vartheta}_{\varepsilon,t} - \vartheta_0 \right) \implies \mathcal{N} \left( 0, \mathbb{I}_t(\vartheta_0)^{-1} \right).$$



3. *The polynomial moments converge and it is asymptotically efficient.*

These properties were established in Kutoyants (1994) in the case of the one-dimensional diffusion processes (9). There is no essential difficulties to apply the same proof in our case. The presented below multi-step MLE-processes have exactly the same asymptotic properties, but can be calculated more easily.

Introduce the family of functions

$$\mathcal{U} = \left\{ \left( u(t, x, \vartheta), t \in [0, T], x \in \mathcal{R}^k \right), \vartheta \in \Theta \right\}$$

such that for all  $\vartheta \in \Theta$  the function  $u(t, x, \vartheta)$  satisfies the PDE

$$\begin{aligned} \frac{\partial u}{\partial t} + \sum_{l=1}^k S_l(\vartheta, t, x) \frac{\partial u}{\partial x_l} + \frac{\varepsilon^2}{2} \sum_{l,m=1}^k \mathbb{A}_{l,m}(t, x) \frac{\partial^2 u}{\partial x_l \partial x_m} \\ = -f \left( t, x, u, \varepsilon \sum_{l=1}^k \sigma_{lm}(t, x) \frac{\partial u}{\partial x_m} \right) \end{aligned}$$

and the condition  $u(T, x, \vartheta) = \Phi(x), x \in \mathcal{R}^k$ .

The limit of the function  $u(t, x, \vartheta)$  as  $\varepsilon \rightarrow 0$  we denote as  $u_o(t, x, \vartheta)$ . The function  $u_o(t, x, \vartheta)$  satisfies the equation

$$\frac{\partial u_o}{\partial t} + \sum_{l=1}^k S_l(\vartheta, t, x) \frac{\partial u_o}{\partial x_l} = -f \left( t, x, u_o, \varepsilon \sum_{l=1}^k \sigma_{lm}(t, x) \frac{\partial u_o}{\partial x_m} \right)$$

with the final value  $u_o(T, x, \vartheta) = \Phi(x)$ . Below  $\dot{u}_o(t, x, \vartheta)$  and  $\ddot{u}_o(t, x, \vartheta)$  means the derivative of this function w.r.t.  $\vartheta$ .

Introduce the condition  $\mathcal{U}$

1. *The function  $u(t, x, \vartheta)$  is two-times continuously differentiable w.r.t.  $\vartheta$  and the derivatives  $\dot{u}(t, x, \vartheta)$  and  $\ddot{u}(t, x, \vartheta)$  are Lipschitz w.r.t.  $x$  uniformly in  $\vartheta \in \Theta$ .*
2. *The function  $u(t, x, \vartheta)$  and its derivatives  $\dot{u}(t, x, \vartheta)$  and  $\dot{u}'_x(t, x, \vartheta)$  converge uniformly in  $t \in [0, T]$  to  $u_o(t, x, \vartheta), \dot{u}_o(t, x, \vartheta), \dot{u}'_{o,x}(t, x, \vartheta)$  respectively.*

The sufficient conditions providing these properties of  $u(t, x, \vartheta)$  can be found in Freidlin and Wentzell (1998), Theorem 2.3.1. Note that the derivatives  $\dot{u}(t, x, \vartheta)$  and  $\ddot{u}(t, x, \vartheta)$  satisfy the linear PDE of the same type.

If we let  $Y_t = u(t, X_t, \vartheta)$ , then by Itô's formula we obtain BSDE (10) with

$$Z_t = \left( Z_t^1, \dots, Z_t^k \right), \quad Z_t^m = \varepsilon \sum_{l=1}^k \sigma_{ml}(t, X_t) u'_{x_l}(t, X_t, \vartheta).$$

Recall that our goal is to construct an asymptotically efficient approximation of the couple  $(Y_t, Z_t)$ . To compare all possible estimators we introduce the lower bounds on the mean-square risks. This is a version of the well-known Hajek-Le Cam minimax risk bound (see, e.g., Ibragimov and Has'minskii (1981), Theorem 2.12.1).

**Theorem 1** Suppose that the conditions  $\mathfrak{L}, \mathfrak{R}$  and  $\mathfrak{U}$  are fulfilled. Then for all estimators  $\bar{Y}_t$  and  $\bar{Z}_t$  and all  $t \in (0, T]$  we have the relations

$$\lim_{\nu \rightarrow 0} \lim_{\varepsilon \rightarrow 0} \sup_{|\vartheta - \vartheta_0| \leq \nu} \varepsilon^{-2} \mathbf{E}_{\vartheta} |\bar{Y}_t - Y_t|^2 \geq \dot{u}_o(t, x_t, \vartheta_0)^* \mathbb{I}_t(\vartheta_0)^{-1} \dot{u}_o(t, x_t, \vartheta_0), \quad (15)$$

$$\lim_{\nu \rightarrow 0} \lim_{\varepsilon \rightarrow 0} \sup_{|\vartheta - \vartheta_0| \leq \nu} \varepsilon^{-4} \mathbf{E}_{\vartheta} |\bar{Z}_t - Z_t|^2 \geq \left| (\dot{u}_o)'_x(t, x_t, \vartheta_0)^* \mathbb{I}_t(\vartheta_0)^{-\frac{1}{2}} \sigma(t, x_t) \right|^2. \quad (16)$$

*Proof* We first verify that the family of measures is locally asymptotically normal (LAN) and then we apply the proof of the Hajek-Le Cam lower bound (Ibragimov and Has'minskii 1981), which provides us (15), (16). We present here the necessary modification of the proof given in Ibragimov and Has'minskii (1981). Usually this inequality is considered for the risk like  $\mathbf{E}_{\vartheta} |\bar{\vartheta}_{\varepsilon} - \vartheta|^2$  and we are interested by the risk  $\mathbf{E}_{\vartheta} |\bar{Y}_{t,\varepsilon} - Y_t|^2$ , where  $Y_t$  is a random process. Another point, the random vector  $\Delta_t$  (see below), in general, is asymptotically normal and, in our case, it has Gaussian distribution, that is why the proof is slightly simplified.  $\square$

Let us denote  $\varphi_{\varepsilon} = \varepsilon \mathbb{I}_t^{-\frac{1}{2}}$  where  $\mathbb{I}_t = \mathbb{I}_t(\vartheta_0)$  and introduce the normalized likelihood ratio

$$Z_{t,\varepsilon}(v) = \frac{L(\vartheta_0 + \varphi_{\varepsilon} v, X^t)}{L(\vartheta_0, X^t)}, \quad v \in V_{\varepsilon} = \{v : \vartheta_0 + \varphi_{\varepsilon} v \in \Theta\}.$$

We can write

$$\begin{aligned} \ln Z_{t,\varepsilon}(v) &= \frac{1}{\varepsilon} \int_0^t [S(\vartheta_0 + \varphi_{\varepsilon} v, s, X_s) - S(\vartheta_0, s, X_s)] \sigma(s, X_s)^{-1} dW_s \\ &\quad - \frac{1}{2\varepsilon^2} \int_0^t \left| [S(\vartheta_0 + \varphi_{\varepsilon} v, s, X_s) - S(\vartheta_0, s, X_s)]^* \sigma(s, X_s)^{-1} \right|^2 ds \\ &= v^* \Delta_t - \frac{1}{2} |v|^2 + r_{\varepsilon}, \end{aligned}$$

where  $r_{\varepsilon} \rightarrow 0$  and the vector

$$\Delta_t = \mathbb{I}_t(\vartheta_0)^{-\frac{1}{2}} \int_0^t \dot{S}(\vartheta_0, s, x_s) \sigma(s, x_s)^{-1} dW_s \sim \mathcal{N}(0, \mathbb{J}).$$

Here  $\mathbb{J}$  is a unit  $d \times d$  matrix.

Hence, the family of measures  $\{\mathbf{P}_{\vartheta}^{(\varepsilon,t)}, \vartheta \in \Theta\}$  is LAN in  $\Theta$  (Ibragimov and Has'minskii 1981, Kutoyants 1994).

Below,  $M > 0$ ,  $\mathcal{K}_M$  is a cube in  $\mathcal{R}^d$  whose vertices have coordinates  $\pm M$ , so that its volume is  $(2M)^d$  and  $\vartheta_v = \vartheta_0 + \varphi_\varepsilon v$ .

$$\begin{aligned} \sup_{|\vartheta - \vartheta_0| \leq v} \mathbf{E}_\vartheta \left| \bar{Y}_{t,\varepsilon} - Y_t \right|^2 &= \sup_{|\vartheta - \vartheta_0| \leq v} \mathbf{E}_\vartheta \left| \bar{Y}_t - u(t, X_t, \vartheta) \right|^2 \\ &= \sup_{|\varphi_\varepsilon v| \leq v} \mathbf{E}_{\vartheta_v} \left| \bar{Y}_t - u(t, X_t, \vartheta_0 + \varphi_\varepsilon v) \right|^2 \\ &\geq \sup_{v \in \mathcal{K}_M} \mathbf{E}_{\vartheta_v} \left| \bar{Y}_t - u(t, X_t, \vartheta_0 + \varphi_\varepsilon v) \right|^2 \\ &\geq \frac{1}{(2M)^d} \int_{\mathcal{K}_M} \mathbf{E}_{\vartheta_v} \left| \bar{Y}_t - u(t, X_t, \vartheta_0 + \varphi_\varepsilon v) \right|^2 dv \\ &= \frac{1}{(2M)^d} \int_{\mathcal{K}_M} \mathbf{E}_{\vartheta_0 Z_{t,\varepsilon}(v)} \left| \bar{Y}_t - u(t, X_t, \vartheta_0 + \varphi_\varepsilon v) \right|^2 dv. \end{aligned}$$

We have

$$\begin{aligned} u(t, X_t, \vartheta_0 + \varphi_\varepsilon v) &= u(t, X_t, \vartheta_0) + \dot{u}(t, X_t, \vartheta_0)^* \varphi_\varepsilon v \\ &\quad + \varepsilon \int_0^1 [\dot{u}(t, X_t, \vartheta_0 + r\varphi_\varepsilon v) - \dot{u}(t, X_t, \vartheta_0)]^* \mathbb{I}_t^{-\frac{1}{2}} v dr \\ &= u(t, X_t, \vartheta_0) + \dot{u}(t, X_t, \vartheta_0)^* \varphi_\varepsilon v + \varphi_\varepsilon h_\varepsilon, \end{aligned}$$

where  $|h_\varepsilon| \leq C\varepsilon$ . Hence, if we denote

$$\bar{b}_\varepsilon = \varepsilon^{-1} (\bar{Y}_t - u(t, X_t, \vartheta_0)), \quad \dot{u} = \dot{u}(t, X_t, \vartheta_0)^* \mathbb{I}_t^{-\frac{1}{2}}$$

and introduce such vector  $\bar{v}_\varepsilon$  that  $\bar{b}_\varepsilon = \dot{u}(t, X_t, \vartheta_0)^* \mathbb{I}_t^{-\frac{1}{2}} \bar{v}_\varepsilon$ , then we can write

$$\begin{aligned} \varepsilon^{-2} \mathbf{E}_{\vartheta_0 Z_{t,\varepsilon}(v)} \left| \bar{Y}_t - u(t, X_t, \vartheta_0 + \varphi_\varepsilon v) \right|^2 &= \mathbf{E}_{\vartheta_0 Z_{t,\varepsilon}(v)} \left| \bar{b}_\varepsilon - \dot{u}(t, X_t, \vartheta_0)^* \mathbb{I}_t^{-\frac{1}{2}} v \right|^2 (1 + O(\varepsilon)) \\ &= \mathbf{E}_{\vartheta_0 Z_{t,\varepsilon}(v)} \left| \dot{u}^* (\bar{v}_\varepsilon - v) \right|^2 (1 + O(\varepsilon)). \end{aligned}$$

Further, we use the following result known as Scheffé’s lemma

**Lemma 1** *Let the random variables  $Z_\varepsilon \geq 0$ ,  $\varepsilon \in (0, 1]$  converge in probability to the random variable  $Z \geq 0$  as  $\varepsilon \rightarrow 0$  and  $\mathbf{E}Z_\varepsilon = \mathbf{E}Z = 1$ , then*

$$\lim_{\varepsilon \rightarrow 0} \mathbf{E} |Z_\varepsilon - Z| = 0.$$

For the proof see, e.g., Theorem A.4 in Ibragimov and Has’minskii (1981).

Recall that  $\mathbf{E}_{\vartheta_0} Z_{t,\varepsilon}(v) = \mathbf{E}_{\vartheta_0} Z_t(v) = 1$ , where  $\ln Z_t(v) = v^* \Delta_t - \frac{1}{2} |v|^2$ . Hence for any  $K > 0$

$$\mathbf{E}_{\vartheta_0} Z_{t,\varepsilon}(v) \left| \dot{u}^* (\bar{v}_\varepsilon - v) \right|_K^2 = \mathbf{E}_{\vartheta_0} Z_t(v) \left| \dot{u}^* (\bar{v}_\varepsilon - v) \right|_K^2 (1 + o(1)).$$

Here we denoted  $|D|_K^2 = |D|^2 \wedge K$ . These allow us to write

$$\begin{aligned} & \frac{1}{(2M)^d} \int_{\mathcal{K}_M} \mathbf{E}_{\vartheta_0} Z_{t,\varepsilon}(v) |\bar{Y}_t - u(t, X_t, \vartheta_0 + \varepsilon v)|^2 dv \\ & \geq \frac{1}{(2M)^d} \int_{\mathcal{K}_M} \mathbf{E}_{\vartheta_0} Z_{t,\varepsilon}(v) |\bar{Y}_t - u(t, X_t, \vartheta_0 + \varepsilon v)|_K^2 dv \\ & = \frac{1}{(2M)^d} \int_{\mathcal{K}_M} \mathbf{E}_{\vartheta_0} Z_t(v) |\dot{u}^*(\bar{v}_\varepsilon - v)|_K^2 dv (1 + o(1)). \end{aligned}$$

Then

$$Z_t(v) = \exp \left\{ v^* \Delta_t - \frac{1}{2} |v|^2 \right\} = \exp \left\{ -\frac{1}{2} |v - \Delta_t|^2 \right\} \exp \left\{ \frac{1}{2} |\Delta_t|^2 \right\}$$

and

$$\begin{aligned} \mathbf{E}_{\vartheta_0} Z_t(v) |\dot{u}^*(\bar{v}_\varepsilon - v)|_K^2 &= \mathbf{E}_{\vartheta_0} e^{-\frac{1}{2}|v-\Delta_t|^2} |\dot{u}^*(v_\varepsilon - v)|_K^2 e^{\frac{1}{2}|\Delta_t|^2} \\ &= \mathbf{E}_{\vartheta_0} e^{-\frac{1}{2}|w|^2} |\dot{u}^*(\bar{v}_\varepsilon - \Delta_t - w)|_K^2 e^{\frac{1}{2}|\bar{\Delta}_t|^2} \\ &= \mathbf{E}_{\vartheta_0} e^{-\frac{1}{2}|w|^2} |\dot{u}^*(\tilde{w}_\varepsilon - w)|_K^2 e^{\frac{1}{2}|\bar{\Delta}_t|^2} \end{aligned}$$

where  $w = v - \Delta_t$  and  $\tilde{w}_\varepsilon = \bar{v}_\varepsilon - \Delta_t$ . Introduce the set  $\mathcal{C}_M$  such that each coordinate of  $\Delta_t = (\Delta_t^{(1)}, \dots, \Delta_t^{(d)})$  is less than  $M - \sqrt{M}$ , i.e.,  $|\Delta_t^{(l)}| \leq M - \sqrt{M}$ . Then

$$\begin{aligned} & \frac{1}{(2M)^d} \mathbf{E}_{\vartheta_0} e^{\frac{1}{2}|\Delta_t|^2} \int_{w+\Delta_t \in \mathcal{K}_M} |\dot{u}^*(\tilde{w}_\varepsilon - w)|_K^2 e^{-\frac{1}{2}|w|^2} dv \\ & \geq \frac{1}{(2M)^d} \mathbf{E}_{\vartheta_0} e^{\frac{1}{2}|\Delta_t|^2} \mathbb{1}_{\{\Delta_t \in \mathcal{C}_M\}} \int_{\mathcal{C}_M} |\dot{u}^*(\tilde{w}_\varepsilon - w)|_K^2 e^{-\frac{1}{2}|w|^2} dv \\ & \geq \frac{1}{(2M)^d} \mathbf{E}_{\vartheta_0} e^{\frac{1}{2}|\Delta_t|^2} \mathbb{1}_{\{\Delta_t \in \mathcal{C}_M\}} \int_{\mathcal{K}_{\sqrt{M}}} |\dot{u}^*(\tilde{w}_\varepsilon - w)|_K^2 e^{-\frac{1}{2}|w|^2} dv \end{aligned}$$

because  $\mathcal{K}_{\sqrt{M}} \subset \mathcal{C}_M$ . By Andersen’s Lemma (see, e.g., Ibragimov and Has’minskii (1981), Lemma 2.10.2)

$$\int_{\mathcal{K}_{\sqrt{M}}} |\dot{u}^*(\tilde{w}_\varepsilon - w)|_K^2 e^{-\frac{1}{2}|w|^2} dv \geq \int_{\mathcal{K}_{\sqrt{M}}} |\dot{u}^* w|_K^2 e^{-\frac{1}{2}|w|^2} dv.$$

Note that as  $M \rightarrow \infty$  we obtain the limits

$$\frac{1}{(2M)^d} \mathbf{E}_{\vartheta_0} e^{\frac{1}{2}|\Delta_t|^2} \mathbb{1}_{\{\Delta_t \in \mathcal{C}_M\}} \longrightarrow \frac{1}{(2\pi)^{\frac{d}{2}}}$$

and

$$\frac{1}{(2\pi)^{\frac{d}{2}}} \int_{\mathcal{K}_{\sqrt{M}}} |\dot{u}^* w|_K^2 e^{-\frac{1}{2}|w|^2} dv \longrightarrow \mathbf{E}_{\vartheta_0} |\dot{u}^* \Delta_t|_K^2.$$

The last steps are  $\varepsilon \rightarrow 0$  and  $K \rightarrow \infty$

$$\mathbf{E}_{\vartheta_0} |\dot{u}^* \Delta_t|_K^2 \longrightarrow \dot{u}_\circ(t, x_t, \vartheta_0)^* \mathbb{I}_t(\vartheta_0)^{-1} \dot{u}_\circ(t, x_t, \vartheta_0).$$

The detailed proof can be found in Ibragimov and Has’minskii (1981), Theorem 2.12.1.

Therefore the bound (15) is verified. The bound (16) is proved in a similar way. Note that  $Z_t = \varepsilon u'_x(t, X_t, \vartheta) \sigma(t, X_t)$ . An arbitrary estimator  $\bar{Z}_t$  of  $Z_t$  we write as  $\bar{Z}_t = \varepsilon \tilde{Z}_t$ . Then, for  $\varepsilon^{-1}(\bar{Z}_t - Z_t)$  we follow the proof given above.

**Definition** *Suppose that the conditions  $\mathfrak{L}, \mathfrak{R}, \mathfrak{U}$  are fulfilled. Then we call the estimator-processes  $Y_t^*, Z_t^*, 0 < t \leq T$  asymptotically efficient if for all  $\vartheta_0 \in \Theta$  and all  $t \in (0, T]$  we have the equalities*

$$\lim_{\nu \rightarrow 0} \lim_{\varepsilon \rightarrow 0} \sup_{|\vartheta - \vartheta_0| \leq \nu} \varepsilon^{-2} \mathbf{E}_{\vartheta} |Y_t^* - Y_t|^2 = \dot{u}_o(t, x_t, \vartheta_0)^* \mathbb{I}_t(\vartheta_0)^{-1} \dot{u}_o(t, x_t, \vartheta_0), \tag{17}$$

$$\lim_{\nu \rightarrow 0} \lim_{\varepsilon \rightarrow 0} \sup_{|\vartheta - \vartheta_0| \leq \nu} \varepsilon^{-4} \mathbf{E}_{\vartheta} |Z_t^* - Z_t|^2 = \left| (\dot{u}_o)'_x(t, x_t, \vartheta_0)^* \mathbb{I}_t(\vartheta_0)^{-\frac{1}{2}} \sigma(t, x_t) \right|^2. \tag{18}$$

As we do not know the value  $\vartheta$  we propose first to estimate it using some estimator-process  $\vartheta_{\varepsilon,t}^*, 0 < t \leq T$  and then to put

$$Y_t^* = u(t, X_t, \vartheta_{\varepsilon,t}^*), \quad Z_t^* = \varepsilon \sum_{l=1}^k u'_{x_l}(t, X_t, \vartheta_{\varepsilon,t}^*) \sigma_l(t, X_t).$$

Recall that formally the MLE-process  $\hat{\vartheta}_{\varepsilon,t}, 0 < t \leq T$  “solves” the problem and it can be shown that under the supposed regularity conditions the estimator-processes  $\hat{Y}_{t,\varepsilon} = u(t, X_t, \hat{\vartheta}_{\varepsilon,t})$  and  $\hat{Z}_{t,\varepsilon} = u'_x(t, X_t, \hat{\vartheta}_{\varepsilon,t}) \sigma(t, X_t)$  are asymptotically efficient in the sense of the relations (17) and (18), respectively, but this solution can not be called acceptable because the calculation of  $\hat{\vartheta}_{\varepsilon,t}$  for all  $t \in (0, T]$ , in the general case, is a computationally difficult problem. That is why we propose to use the so-called multi-step MLE-process (Kutoyants 2015), which is introduced as follows. First we construct a preliminary estimator  $\bar{\vartheta}_{\tau_{\varepsilon}}$  by the observations  $X^{\tau_{\varepsilon}} = (X_s, 0 \leq s \leq \tau_{\varepsilon})$  on some learning interval  $[0, \tau_{\varepsilon}]$ , where  $\tau_{\varepsilon} = \varepsilon^{\delta}$  with  $0 < \delta < 1$  and then we propose an estimator-process  $\vartheta_{t,\varepsilon}^*, \tau_{\varepsilon} \leq t \leq T$  based on this preliminary estimator. Finally we show that the corresponding estimators, say,  $Y_{t,\varepsilon}^* = u(t, X_t, \vartheta_{t,\varepsilon}^*), \tau_{\varepsilon} \leq t \leq T$  are asymptotically efficient.

As a preliminary we propose the minimum distance estimator (MDE)  $\bar{\vartheta}_{\tau_{\varepsilon}}$  defined by the relation

$$\left\| X - \hat{X}(\bar{\vartheta}_{\tau_{\varepsilon}}) \right\|_{\tau_{\varepsilon}}^2 = \inf_{\vartheta \in \Theta} \left\| X - \hat{X}(\vartheta) \right\|_{\tau_{\varepsilon}}^2 = \inf_{\vartheta \in \Theta} \int_0^{\tau_{\varepsilon}} \left[ X_t - \hat{X}_t(\vartheta) \right]^2 dt.$$

Here the family of random processes  $\left\{ \left( \hat{X}_t(\vartheta), 0 \leq t \leq \tau_{\varepsilon} \right), \vartheta \in \Theta \right\}$  is defined as follows

$$\hat{X}_t(\vartheta) = x_0 + \int_0^t S(\vartheta, s, X_s) ds, \quad 0 \leq t \leq \tau_{\varepsilon}, \quad \vartheta \in \Theta.$$

These estimators were studied in Kutoyants (1994) in the case of fixed  $\tau_{\varepsilon} = \tau$  and are called the *trajectory fitting estimators* as well, because we choose an estimator  $\bar{\vartheta}_{\tau_{\varepsilon}}$ , which provides a trajectory  $\hat{X}_t(\bar{\vartheta}_{\tau_{\varepsilon}}), 0 \leq t \leq \tau_{\varepsilon}$  closest to the observations

$X_t, 0 \leq t \leq \tau_\varepsilon$ . It was shown that if the conditions of regularity and the condition of identifiability: for any  $\nu > 0$

$$\inf_{\vartheta_0 \in \Theta} \inf_{|\vartheta - \vartheta_0| > \nu} \int_0^\tau \left| \int_0^t [S(\vartheta, s, x_s) - S(\vartheta_0, s, x_s)] ds \right|^2 dt > 0 \tag{19}$$

hold and the matrix

$$\mathbb{J}_\tau(\vartheta_0) = \int_0^\tau \int_0^t \dot{S}(\vartheta_0, s, x_s)^* ds \int_0^t \dot{S}(\vartheta_0, s, x_s) ds dt$$

is uniformly nondegenerate (below  $\lambda \in \mathcal{R}^d$ )

$$\inf_{\vartheta_0 \in \Theta} \inf_{|\lambda|=1} \lambda^* \mathbb{J}_\tau(\vartheta_0) \lambda > 0, \tag{20}$$

then the MDE is asymptotically normal

$$\varepsilon^{-1}(\vartheta_\tau - \vartheta_0) \implies \mathbb{J}_\tau(\vartheta_0)^{-1} \int_0^\tau \int_0^t \dot{S}(\vartheta_0, s, x_s)^* ds \int_0^t \sigma(s, x_s)^* dW_s dt.$$

Note that if we have the Regularity condition 3 (identifiability) with  $T = \tau$ , then the identifiability condition (19) is also fulfilled. Indeed, suppose that there exists  $\vartheta_1 \neq \vartheta_0$  such that

$$\int_0^\tau \left| \int_0^t [S(\vartheta_1, s, x_s) - S(\vartheta_0, s, x_s)] ds \right|^2 dt = 0.$$

Then for all  $t \in [0, \tau]$

$$\int_0^t S(\vartheta_1, s, x_s) ds = \int_0^t S(\vartheta_0, s, x_s) ds,$$

which implies

$$S(\vartheta_1, s, x_s) = S(\vartheta_0, s, x_s), \quad 0 \leq s \leq \tau.$$

The last equality, of course, contradicts Regularity condition 3.

Now suppose that  $\tau_\varepsilon = \varepsilon^\delta$  with  $\delta < 1$  and the matrix

$$\mathbb{C}(\vartheta_0) = \dot{S}(\vartheta_0, 0, x_0)^* \dot{S}(\vartheta_0, 0, x_0)$$

is uniformly nondegenerate in  $\vartheta_0 \in \Theta$  (below  $\lambda \in \mathcal{R}^d$ )

$$\inf_{\vartheta_0 \in \Theta} \inf_{|\lambda|=1} \lambda^* \mathbb{C}(\vartheta_0) \lambda > 0 \tag{21}$$

Then, we can obtain the asymptotics

$$\varepsilon^{-1}(\vartheta_\tau - \vartheta_0) = \frac{3}{2\sqrt{\tau_\varepsilon}} \mathbb{C}(\vartheta_0)^{-1} \sigma(0, x_0) \int_0^1 [1 - r^2] dW(r) (1 + o(1)).$$

Note that

$$\begin{aligned} \mathbb{J}_{\tau_\varepsilon}(\vartheta_0) &= \int_0^{\tau_\varepsilon} t^2 \dot{S}(\vartheta_0, 0, x_0)^* \dot{S}(\vartheta_0, 0, x_0) dt (1 + o(1)) \\ &= \frac{\tau_\varepsilon^3}{3} \dot{S}(\vartheta_0, 0, x_0)^* \dot{S}(\vartheta_0, 0, x_0) (1 + o(1)) \end{aligned}$$

and

$$\begin{aligned} & \int_0^{\tau_\varepsilon} \int_0^t \dot{S}(\vartheta_0, s, x_s)^* ds \int_0^t \sigma(s, x_s)^* dW_s dt \\ &= \dot{S}(\vartheta_0, 0, x_0)^* \sigma(0, x_0) \int_0^{\tau_\varepsilon} t W_t dt (1 + o(1)) \\ &= \frac{\tau_\varepsilon^{\frac{5}{2}}}{2} \dot{S}(\vartheta_0, 0, x_0)^* \sigma(0, x_0) \int_0^1 (1 - r^2) dW(r) (1 + o(1)). \end{aligned}$$

Therefore, the family of random vectors  $\varepsilon^{-1+\frac{\delta}{2}} (\bar{\vartheta}_{\tau_\varepsilon} - \vartheta_0)$  is asymptotically normal. Moreover, following Kutoyants (1994) it can be shown that the moments are bounded, i.e.,

$$\sup_{\vartheta_0 \in \mathbf{K}} \mathbf{E}_{\vartheta_0} \left| \varepsilon^{-1+\frac{\delta}{2}} (\bar{\vartheta}_{\tau_\varepsilon} - \vartheta_0) \right|^p < C, \tag{22}$$

where the constant  $C = C(p) > 0$  does not depend on  $\varepsilon$  for all  $p > 0$ .

Let us introduce the one-step MLE-process  $\vartheta_{t,\varepsilon}^*, \tau_\varepsilon \leq t \leq T$

$$\begin{aligned} \vartheta_{t,\varepsilon}^* &= \bar{\vartheta}_{\tau_\varepsilon} \\ &+ \mathbb{I}_t (\bar{\vartheta}_{\tau_\varepsilon})^{-1} \int_{\tau_\varepsilon}^t \dot{S}(\bar{\vartheta}_{\tau_\varepsilon}, s, X_s)^* \mathbb{A}(s, X_s)^{-1} [dX_s - S(\bar{\vartheta}_{\tau_\varepsilon}, s, X_s) ds]. \end{aligned} \tag{23}$$

Its properties are described in the following proposition.

**Proposition 1** *Let the conditions  $\mathcal{L}, \mathfrak{R}$  be fulfilled and  $\delta \in (0, 1)$ , then for all  $t \in (0, T]$*

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

and this estimator-process is asymptotically efficient. Moreover, we have the uniform consistency, i.e., for any  $\nu > 0$

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

*Proof* Note that the estimator  $\vartheta_{t,\varepsilon}^*$  is defined for  $t \in [\tau_\varepsilon, T]$ , but as  $\tau_\varepsilon \rightarrow 0$  we obtain for any positive  $t$  the relation  $t > \tau_\varepsilon$ . □

The substitution of the observations (9) provides us the equality

$$\begin{aligned} \vartheta_{t,\varepsilon}^* - \vartheta_0 &= \bar{\vartheta}_{\tau_\varepsilon} - \vartheta_0 + \varepsilon \mathbb{I}_t (\bar{\vartheta}_{\tau_\varepsilon})^{-1} \int_{\tau_\varepsilon}^t \dot{S}(\bar{\vartheta}_{\tau_\varepsilon}, s, X_s)^* \sigma(s, X_s)^{-1} dW_s \\ &+ \mathbb{I}_t (\bar{\vartheta}_{\tau_\varepsilon})^{-1} \int_{\tau_\varepsilon}^t \dot{S}(\bar{\vartheta}_{\tau_\varepsilon}, s, X_s)^* \mathbb{A}(s, X_s)^{-1} [S(\vartheta_0, s, X_s) - S(\bar{\vartheta}_{\tau_\varepsilon}, s, X_s)] ds. \end{aligned}$$

Recall that the vector-process  $(X_s, 0 \leq s \leq T)$  converges uniformly in  $s$  to the deterministic vector-function  $(x_s, 0 \leq s \leq T)$  and the estimator  $\bar{\vartheta}_{\tau_\varepsilon}$  is consistent. Therefore, we have the convergence in probability

$$\begin{aligned} & \mathbb{I}_t(\bar{\vartheta}_{\tau_\varepsilon})^{-1} \int_{\tau_\varepsilon}^t \dot{\mathbb{S}}(\bar{\vartheta}_{\tau_\varepsilon}, s, X_s)^* \sigma(s, X_s)^{-1} dW_s \\ & \longrightarrow \mathbb{I}_t(\vartheta_0)^{-1} \int_0^t \dot{\mathbb{S}}(\vartheta_0, s, x_s)^* \sigma(s, x_s)^{-1} dW_s \sim \mathcal{N}\left(0, \mathbb{I}_t(\vartheta_0)^{-1}\right). \end{aligned}$$

For the other terms, we first write the Taylor expansion

$$S(\vartheta_0, s, X_s) - S(\bar{\vartheta}_{\tau_\varepsilon}, s, X_s) = \dot{S}(\vartheta_0, s, X_s)^* (\vartheta_0 - \bar{\vartheta}_{\tau_\varepsilon}) + O(\varepsilon^{2-\delta})$$

because  $\vartheta_0 - \bar{\vartheta}_{\tau_\varepsilon} = O(\varepsilon^{1-\frac{\delta}{2}})$ . Then, we denote

$$\mathbb{D}_\varepsilon = \mathbb{I}_t(\bar{\vartheta}_{\tau_\varepsilon}) - \int_{\tau_\varepsilon}^t \dot{\mathbb{S}}(\bar{\vartheta}_{\tau_\varepsilon}, s, X_s)^* \mathbb{A}(s, X_s)^{-1} \dot{\mathbb{S}}(\vartheta_0, s, X_s)^* ds$$

and write

$$\begin{aligned} & \bar{\vartheta}_{\tau_\varepsilon} - \vartheta_0 + \mathbb{I}_t(\bar{\vartheta}_{\tau_\varepsilon})^{-1} \int_{\tau_\varepsilon}^t \dot{\mathbb{S}}(\bar{\vartheta}_{\tau_\varepsilon}, s, X_s)^* \mathbb{A}(s, X_s)^{-1} [S(\vartheta_0, s, X_s) - S(\bar{\vartheta}_{\tau_\varepsilon}, s, X_s)] ds \\ & = (\bar{\vartheta}_{\tau_\varepsilon} - \vartheta_0) \mathbb{I}_t(\bar{\vartheta}_{\tau_\varepsilon})^{-1} \mathbb{D}_\varepsilon + O(\varepsilon^{2-\delta}) \end{aligned}$$

The following estimate can be easily verified

$$\mathbb{D}_\varepsilon = O(\varepsilon^\delta) + O(\varepsilon^{1-\frac{\delta}{2}}) + O(\varepsilon)$$

because  $X_s - x_s = O(\varepsilon)$ ,  $\bar{\vartheta}_{\tau_\varepsilon} - \vartheta_0 = O(\varepsilon^{1-\frac{\delta}{2}})$  and

$$\mathbb{I}_t(\vartheta_0) - \int_{\tau_\varepsilon}^t \dot{\mathbb{S}}(\vartheta_0, s, x_s)^* \mathbb{A}(s, x_s)^{-1} \dot{\mathbb{S}}(\vartheta_0, s, x_s) ds = O(\varepsilon^\delta).$$

Hence

$$\begin{aligned} & \varepsilon^{-1} (\vartheta_{t,\varepsilon}^* - \vartheta_0) - \mathbb{I}_t(\vartheta_0)^{-1} \int_0^t \dot{\mathbb{S}}(\vartheta_0, s, x_s)^* \sigma(s, x_s)^{-1} dW_s \\ & = \varepsilon^{-1} (\bar{\vartheta}_{\tau_\varepsilon} - \vartheta_0) O(\varepsilon^{1-\frac{\delta}{2}}) = O(\varepsilon^{1-\delta}) \longrightarrow 0. \end{aligned} \tag{24}$$

The uniform consistency can be shown following the proof of such uniform consistency presented in Kutoyants (2015), Theorem 1.

Let us define the estimator-processes  $Y_\varepsilon^* = (Y_{t,\varepsilon}^*, \tau_\varepsilon \leq t \leq T)$  and  $Z_\varepsilon^* = (Z_{t,\varepsilon}^*, \tau_\varepsilon \leq t \leq T)$  as follows

$$Y_{t,\varepsilon}^* = u(t, X_t, \vartheta_{t,\varepsilon}^*), \quad Z_{t,\varepsilon}^* = \varepsilon u'_x(t, X_t, \vartheta_{t,\varepsilon}^*) \sigma(t, X_t).$$

**Theorem 2** *Suppose the conditions  $\mathfrak{L}, \mathfrak{R}, \mathfrak{U}$  and (21) hold, then the estimator-processes  $Y_\varepsilon^*, Z_\varepsilon^*$  admit the representations*

$$Y_{t,\varepsilon}^* = Y_t + \varepsilon \dot{u}_o(t, x_t, \vartheta_0)^* \xi_t(\vartheta_0) (1 + o(1)), \tag{25}$$

$$Z_{t,\varepsilon}^* = Z_t + \varepsilon^2 \dot{u}'_{o,x}(t, x_t, \vartheta_0)^* \xi_t(\vartheta_0) \sigma(t, x_t) (1 + o(1)), \tag{26}$$



where the Gaussian process

$$\xi_t(\vartheta_0) = \mathbb{I}_t(\vartheta_0)^{-1} \int_0^t \dot{\mathbb{S}}(\vartheta_0, s, x_s)^* \sigma(s, x_s)^{-1} dW_s, \quad \tau_\varepsilon \leq t \leq T.$$

The random processes

$$\begin{aligned} \eta_{t,\varepsilon} &= \varepsilon^{-1} (Y_{t,\varepsilon}^* - Y_t), \quad \tau \leq t \leq T, \\ \zeta_{t,\varepsilon} &= \varepsilon^{-2} (Z_{t,\varepsilon}^* - Z_t), \quad \tau \leq t \leq T \end{aligned}$$

for any  $\tau \in (0, T]$  converge in probability to the processes

$$\begin{aligned} \eta_t &= \dot{u}_o(t, x_t, \vartheta_0)^* \xi_t(\vartheta_0), \quad \tau \leq t \leq T, \\ \zeta_t &= \dot{u}'_{o,x}(t, x_t, \vartheta_0)^* \xi_t(\vartheta_0) \sigma(t, x_t), \quad \tau \leq t \leq T, \end{aligned}$$

respectively, uniformly in  $t \in [\tau, T]$ . Moreover, these approximations are asymptotically efficient in the sense of (17), (18).

*Proof* By the condition  $\mathcal{U}$ , we obtain the representation

$$\begin{aligned} Y_{t,\varepsilon}^* - Y_t &= u(t, X_t, \vartheta_{t,\varepsilon}^*) - u(t, X_t, \vartheta_0) = \dot{u}(t, X_t, \vartheta_0)^* (\vartheta_{t,\varepsilon}^* - \vartheta_0) (1 + o(1)), \\ Z_{t,\varepsilon}^* - Z_t &= \varepsilon [u'_x(t, X_t, \vartheta_{t,\varepsilon}^*) - u'_x(t, X_t, \vartheta_0)] \sigma(t, X_t) \\ &= \varepsilon \dot{u}'_x(t, X_t, \vartheta_0)^* (\vartheta_{t,\varepsilon}^* - \vartheta_0) \sigma(t, X_t) (1 + o(1)), \end{aligned}$$

and for any  $\tau \in (0, T]$  we have the convergence in probability

$$\begin{aligned} \sup_{\tau \leq t \leq T} |\dot{u}(t, X_t, \vartheta_0) - \dot{u}_o(t, x_t, \vartheta_0)| &\leq \sup_{\tau \leq t \leq T} |\dot{u}(t, X_t, \vartheta_0) - \dot{u}(t, x_t, \vartheta_0)| \\ &+ \sup_{\tau \leq t \leq T} |\dot{u}(t, x_t, \vartheta_0) - \dot{u}_o(t, x_t, \vartheta_0)| \longrightarrow 0, \end{aligned} \tag{27}$$

$$\begin{aligned} \sup_{\tau \leq t \leq T} |\dot{u}'_x(t, X_t, \vartheta_0) - \dot{u}'_x(t, x_t, \vartheta_0)| &\leq \sup_{\tau \leq t \leq T} |\dot{u}'_x(t, X_t, \vartheta_0) - \dot{u}'_x(t, x_t, \vartheta_0)| \\ &+ \sup_{\tau \leq t \leq T} |\dot{u}'_x(t, x_t, \vartheta_0) - \dot{u}'_{o,x}(t, x_t, \vartheta_0)| \longrightarrow 0. \end{aligned} \tag{28}$$

□

Therefore, the representations (25),(26) follow now from (24).

More detailed analysis shows that the convergences  $O(1)$  in (24),(25) are uniform in  $t \in [\tau, T]$  due to (11). Moreover, we have the convergence of moments uniform on compacts  $\vartheta_0 \in \mathbf{K}$  as well, because we have (12) and the moments of the preliminary estimator are bounded (22). Therefore, the estimates used above can be also written for the moments. This convergence of moments provides the asymptotic efficiency of the estimators  $Y_\varepsilon^*, Z_\varepsilon^*$ .

The estimators  $Y_{t,\varepsilon}^*, Z_{t,\varepsilon}^*, \tau_\varepsilon \leq t \leq T$  are given for the values  $t > \tau_\varepsilon = \varepsilon^\delta$  with  $\delta \in (0, 1)$ . It is interesting to have a shorter learning interval and, therefore, longer estimation period for  $Y_t, Z_t$ . That is why we propose the two-step MLE-process

which uses the preliminary estimator with the worse rate of convergence. Let us take  $\delta \in [1, \frac{4}{3})$ , introduce the second preliminary estimator-process

$$\bar{\vartheta}_{t,\varepsilon} = \bar{\vartheta}_{\tau_\varepsilon} + \mathbb{I}_t (\bar{\vartheta}_{\tau_\varepsilon})^{-1} \int_{\tau_\varepsilon}^t \dot{S}(\bar{\vartheta}_{\tau_\varepsilon}, s, X_s)^* \mathbb{A}(s, X_s)^{-1} [dX_s - S(\bar{\vartheta}_{\tau_\varepsilon}, s, X_s) ds]$$

and the two-step MLE-process  $\vartheta_{t,\varepsilon}^{**}, \tau_\varepsilon \leq t \leq T$

$$\vartheta_{t,\varepsilon}^{**} = \bar{\vartheta}_{t,\varepsilon} + \mathbb{I}_t (\bar{\vartheta}_{t,\varepsilon})^{-1} \int_{\tau_\varepsilon}^t \dot{S}(\bar{\vartheta}_{t,\varepsilon}, s, X_s)^* \mathbb{A}(s, X_s)^{-1} [dX_s - S(\bar{\vartheta}_{t,\varepsilon}, s, X_s) ds].$$

For the preliminary estimator we obtain the same estimate (22), but with different  $\tau_\varepsilon$ . Further, for the first preliminary estimator similar calculations as above provide us the estimates

$$\varepsilon^{-\gamma} (\bar{\vartheta}_{t,\varepsilon} - \vartheta_0) = \varepsilon^{-\gamma} |\bar{\vartheta}_{\tau_\varepsilon} - \vartheta_0|^2 O(1) + o(1) = \varepsilon^{-\gamma+2-\delta} O(1) + o(1).$$

For the two-step MLE-process we have

$$\begin{aligned} \varepsilon^{-1} (\vartheta_{t,\varepsilon}^{**} - \vartheta_0) &= \varepsilon^{\gamma-\frac{\delta}{2}} (\varepsilon^{-\gamma} |\bar{\vartheta}_{t,\varepsilon} - \vartheta_0|) (\varepsilon^{-1+\frac{\delta}{2}} |\bar{\vartheta}_{\tau_\varepsilon} - \vartheta_0|) O(1) \\ &\quad + \mathbb{I}_t (\vartheta_0)^{-1} \int_{\tau_\varepsilon}^t \dot{S}(\vartheta_0, s, x_s)^* \mathbb{A}(s, x_s)^{-1} dW_s + o(1). \end{aligned}$$

Therefore if we take  $\gamma$  such that  $\gamma + \delta < 2$  and  $\gamma - \frac{\delta}{2} > 0$ , say,  $\gamma < \frac{2}{3}$ , then we obtain

$$\varepsilon^{-1} (\vartheta_{t,\varepsilon}^{**} - \vartheta_0) \implies \mathcal{N}(0, \mathbb{I}_t (\vartheta_0)^{-1}).$$

Now the estimator-processes  $Y_\varepsilon^{**}, Z_\varepsilon^{**}$  defined with the help of two-step MLE-process

$$\begin{aligned} Y_{t,\varepsilon}^{**} &= u(t, X_t, \vartheta_{t,\varepsilon}^{**}), \quad \tau_\varepsilon \leq t \leq T, \\ Z_{t,\varepsilon}^{**} &= u'_x(t, X_t, \vartheta_{t,\varepsilon}^{**}) \sigma(t, X_t), \quad \tau_\varepsilon \leq t \leq T \end{aligned}$$

are known for the larger time interval  $[\tau_\varepsilon, T]$ .

Of course, we can continue this process and to reduce the learning interval once more by introducing the three-step MLE-process  $\vartheta_{t,\varepsilon}^{***}$  as follows. The learning interval is  $[0, \tau_\varepsilon]$ ,  $\tau_\varepsilon = \varepsilon^\delta$ , where  $\delta \in [\frac{4}{3}, \frac{3}{2})$ . The first preliminary estimator-process is

$$\bar{\vartheta}_{t,\varepsilon} = \bar{\vartheta}_{\tau_\varepsilon} + \mathbb{I}_t (\bar{\vartheta}_{\tau_\varepsilon})^{-1} \int_{\tau_\varepsilon}^t \dot{S}(\bar{\vartheta}_{\tau_\varepsilon}, s, X_s)^* \mathbb{A}(s, X_s)^{-1} [dX_s - S(\bar{\vartheta}_{\tau_\varepsilon}, s, X_s) ds],$$

the second is

$$\bar{\bar{\vartheta}}_{t,\varepsilon} = \bar{\vartheta}_{t,\varepsilon} + \mathbb{I}_t (\bar{\vartheta}_{t,\varepsilon})^{-1} \int_{\tau_\varepsilon}^t \dot{S}(\bar{\vartheta}_{t,\varepsilon}, s, X_s)^* \mathbb{A}(s, X_s)^{-1} [dX_s - S(\bar{\vartheta}_{t,\varepsilon}, s, X_s) ds],$$

and the three-step MLE-process

$$\vartheta_{t,\varepsilon}^{***} = \bar{\bar{\vartheta}}_{t,\varepsilon} + \mathbb{I}_t (\bar{\bar{\vartheta}}_{t,\varepsilon})^{-1} \int_{\tau_\varepsilon}^t \dot{S}(\bar{\vartheta}_{t,\varepsilon}, s, X_s)^* \mathbb{A}(s, X_s)^{-1} [dX_s - S(\bar{\bar{\vartheta}}_{t,\varepsilon}, s, X_s) ds].$$

The similar calculations will provide us the relations :  $\bar{\vartheta}_{\tau_\varepsilon} - \vartheta_0 = \varepsilon^{1-\frac{\delta}{2}} O(1)$ ,

$$\varepsilon^{-\gamma_1} (\bar{\vartheta}_{t,\varepsilon} - \vartheta_0) = \varepsilon^{2-\gamma_1-\delta} O(1), \quad \varepsilon^{-\gamma_2} (\bar{\bar{\vartheta}}_{t,\varepsilon} - \vartheta_0) = \varepsilon^{-\gamma_2+\gamma_1+1-\frac{\delta}{2}} O(1),$$

and

$$\begin{aligned} \varepsilon^{-1} (\vartheta_{t,\varepsilon}^{***} - \vartheta_0) &= \varepsilon^{\gamma_2 - \frac{\delta}{2}} \left( \varepsilon^{-\gamma_2} \left| \bar{\vartheta}_{t,\varepsilon} - \vartheta_0 \right| \right) (\varepsilon^{-1 + \frac{\delta}{2}} |\bar{\vartheta}_{\tau_\varepsilon} - \vartheta_0|) O(1) \\ &\quad + \mathbb{I}_t(\vartheta_0)^{-1} \int_{\tau_\varepsilon}^t \dot{\mathbb{S}}(\vartheta_0, s, x_s)^* \mathbb{A}(s, x_s)^{-1} dW_s + o(1). \end{aligned}$$

Hence if we chose  $\delta, \gamma_1$  and  $\gamma_2$  such that

$$\delta < 2, \quad \gamma_1 + \delta < 2, \quad \gamma_1 - \gamma_2 + 1 - \frac{\delta}{2} > 0, \quad \gamma_2 > \frac{\delta}{2},$$

then once more we obtain asymptotically efficient MLE-process

$$\varepsilon^{-1} (\vartheta_{t,\varepsilon}^{***} - \vartheta_0) \implies \mathcal{N} \left( 0, \mathbb{I}_t(\vartheta_0)^{-1} \right).$$

Therefore, we obtain the corresponding approximations  $Y_{t,\varepsilon}^{***}, Z_{t,\varepsilon}^{***}$  for the values  $t \in [\tau_\varepsilon, T]$  with essentially smaller  $\tau_\varepsilon$  than in the case of one-step MLE-process.

### Example

*Black and Scholes model.* Suppose that the forward equation is

$$dX_t = \vartheta X_t dt + \varepsilon \sigma X_t dW_t, \quad X_0 = x_0 > 0, \quad 0 \leq t \leq T,$$

and the functions  $f(x, y, z) = -\beta y - \gamma xz$  and  $\Phi(x)$  are given. The function  $\Phi(x)$  is continuous and satisfies the condition  $|\Phi(x)| \leq C(1 + |x|^p)$  with some constants  $C > 0$  and  $p > 0$ . We have to find  $(Y_t, Z_t)$  such that

$$dY_t = [\beta Y_t + \gamma X_t Z_t] dt + Z_t dW_t, \quad 0 \leq t \leq T,$$

and  $Y_T = \Phi(X_T)$ .

The corresponding PDE is

$$\frac{\partial u}{\partial t} + \frac{\varepsilon^2 \sigma^2 x^2}{2} \frac{\partial^2 u}{\partial x^2} + (\vartheta - \varepsilon \sigma \gamma) x \frac{\partial u}{\partial x} - \beta u = 0, \quad u(T, x, \vartheta) = \Phi(x).$$

To write its solution we change the variables  $s = T - t, \bar{x} = \ln x$  and let  $u(t, \bar{x}, \vartheta) = e^{\mu(\vartheta)\bar{x} + \lambda(\vartheta)s} v(s, \bar{x}, \vartheta)$ , where

$$\mu(\vartheta) = \frac{2\varepsilon\sigma\gamma + \varepsilon\sigma^2 - 2\vartheta}{2\varepsilon^2\sigma^2}, \quad \lambda(\vartheta) = \beta + \frac{(2\varepsilon\sigma\gamma + \varepsilon\sigma^2 - 2\vartheta)^2}{8\varepsilon^2\sigma^2}.$$

Then, we obtain the reduced equation

$$\frac{\partial v}{\partial s} = \frac{\varepsilon^2 \sigma^2}{2} \frac{\partial^2 v}{\partial \bar{x}^2}, \quad 0 \leq s \leq T, \quad v(0, \bar{x}, \vartheta) = e^{-\mu(\vartheta)\bar{x}} \Phi(e^{\bar{x}}),$$

whose solution is well-known

$$v(s, \bar{x}, \vartheta) = \frac{1}{\sqrt{2\pi s \varepsilon^2 \sigma^2}} \int_{-\infty}^{\infty} \exp \left\{ -\frac{(\bar{x} - z)^2}{2\varepsilon^2 \sigma^2 s} \right\} e^{-\mu(\vartheta)z} \Phi(e^z) dz.$$

Let us fix  $\tau_\varepsilon = \varepsilon^{\frac{3}{4}}$  and introduce the preliminary TFE is

$$\bar{\vartheta}_{\tau_\varepsilon} = \frac{\int_0^{\tau_\varepsilon} (X_s - x_0) H_t dt}{\int_0^{\tau_\varepsilon} H_t^2 dt}, \quad H_t = \int_0^t X_s ds.$$

Of course, we also can write the MLE  $\hat{\vartheta}_{\tau_\varepsilon}$

$$\hat{\vartheta}_{\tau_\varepsilon} = \frac{1}{\tau_\varepsilon} \int_0^{\tau_\varepsilon} \frac{dX_s}{X_s},$$

but as in our work we used the TFE, we show how to calculate  $\bar{\vartheta}_{\tau_\varepsilon}$ . The Fisher information is  $\mathbb{I}_t(\vartheta) = t\sigma^{-2}$ . The one-step MLE-process is

$$\vartheta_{t,\varepsilon}^* = \bar{\vartheta}_{\tau_\varepsilon} + \frac{1}{t} \int_{\tau_\varepsilon}^t \frac{1}{X_s} [dX_s - \bar{\vartheta}_{\tau_\varepsilon} X_s ds].$$

Moreover, it is easy to see that

$$\begin{aligned} \vartheta_{t,\varepsilon}^* &= \bar{\vartheta}_{\tau_\varepsilon} + \frac{1}{t} \int_{\tau_\varepsilon}^t \frac{dX_s}{X_s} - \bar{\vartheta}_{\tau_\varepsilon} \frac{t - \tau_\varepsilon}{t} = \frac{1}{t} \int_{\tau_\varepsilon}^t \frac{dX_s}{X_s} + \bar{\vartheta}_{\tau_\varepsilon} \frac{\tau_\varepsilon}{t} \\ &= \hat{\vartheta}_{t,\varepsilon} + \bar{\vartheta}_{\tau_\varepsilon} \frac{\tau_\varepsilon}{t} - \frac{1}{t} \int_0^{\tau_\varepsilon} \frac{dX_s}{X_s} = \hat{\vartheta}_{t,\varepsilon} + o(\varepsilon). \end{aligned}$$

Hence, the estimators  $\vartheta_{t,\varepsilon}^*$  and  $\hat{\vartheta}_{t,\varepsilon}$  have the same limit distributions. Therefore the estimator-process

$$Y_{t,\varepsilon}^* = X_t^{\mu(\bar{\vartheta}_{\tau_\varepsilon})} \frac{e^{\lambda(\bar{\vartheta}_{\tau_\varepsilon})(T-t)}}{\sqrt{2\pi(T-t)\varepsilon^2\sigma^2}} \int_{-\infty}^{\infty} e^{-\frac{(\ln X_t - z)^2}{2\varepsilon^2\sigma^2(T-t)} - \mu(\bar{\vartheta}_{\tau_\varepsilon})z} \Phi(e^z) dz.$$

It is easy to see that  $Y_{t,\varepsilon}^* \rightarrow \Phi(X_T)$  as  $t \rightarrow T$ . The expression for  $Z_{t,\varepsilon}^*$  can be written as well.

### Discussions

Note that we approximate the solution of the BSDE and not the equation itself. Of course, it is also possible to write the stochastic differential for  $Y_{t,\varepsilon}^*$ . For simplicity of notation we consider the case  $k = 1, d = 1$ . Indeed, the process  $Y_{t,\varepsilon}^* = u(t, X_t, \vartheta_{t,\varepsilon}^*)$ , where  $X_t$  has stochastic differential (9) and  $\vartheta_{t,\varepsilon}^*$  given by (23) can be written as follows

$$\begin{aligned} \vartheta_{t,\varepsilon}^* &= \bar{\vartheta}_{\tau_\varepsilon} + \varepsilon \mathbb{I}_t(\bar{\vartheta}_{\tau_\varepsilon})^{-1} \int_{\tau_\varepsilon}^t \dot{S}(\bar{\vartheta}_{\tau_\varepsilon}, s, X_s) \sigma(s, X_s)^{-1} dW_s \\ &\quad + \mathbb{I}_t(\bar{\vartheta}_{\tau_\varepsilon})^{-1} \int_{\tau_\varepsilon}^t \dot{S}(\bar{\vartheta}_{\tau_\varepsilon}, s, X_s) \mathbb{A}(s, X_s)^{-1} [S(\vartheta_0, s, X_s) - S(\bar{\vartheta}_{\tau_\varepsilon}, s, X_s)] ds \\ &= \bar{\vartheta}_{\tau_\varepsilon} + \varepsilon I_t^{-1} \int_{\tau_\varepsilon}^t \alpha_{s,\varepsilon} dW_s + I_t^{-1} \int_{\tau_\varepsilon}^t \beta_{s,\varepsilon} ds \end{aligned} \tag{29}$$

with obvious notations. Therefore, the stochastic differential for  $Y_{t,\varepsilon}^*$  can be written (Itô formula).

It was shown that the right-hand side of (29) tends to a constant  $\vartheta_0$  as  $\varepsilon \rightarrow 0$  and we can verify that  $d\vartheta_{t,\varepsilon}^* \rightarrow 0$ .

More detailed analysis shows that

$$dY_{t,\varepsilon}^* = -f(t, X_t, Y_t, Z_t) dt + Z_t dW_t + \varepsilon d\eta_t + o(\varepsilon), \quad \tau_\varepsilon \leq t \leq T,$$

where the Gaussian process  $\eta_t$  is defined in Theorem 2. We used the relation  $Y_{t,\varepsilon}^* = Y_t + \varepsilon \eta_t + o(\varepsilon)$ .

The multi-step MLE-processes used in this work can be useful in similar problems of BSDE approximations for discrete-time observations and ergodic diffusion models mentioned in the introduction (see (Abakirova, A and Kutoyants, YA: On approximation of the BSDE. Large samples approach. *In preparation*) and Gasparyan and Kutoyants (2015)).

**Acknowledgements** This work was done with partial financial support of the RSF grant number 14-49-10079.

## Competing interests

I declare that there is no competing interests.

## Authors' contributions

I read and approved the final manuscript.

## References

- Bismut, JM: Conjugate convex functions in optimal stochastic control. *J. Math. Anal. Appl.* **44**, 384–404 (1973)
- El Karoui, N, Peng, S, Quenez, M: Backward stochastic differential equations in finance. *Math. Fin.* **7**, 1–71 (1997)
- Fisher, RA: Theory of statistical estimation. *Proc. Cambridge Philosophical Society* **22**, 700–725 (1925)
- Freidlin, MI, Wentzell, AD *Random Perturbations of Dynamical Systems*, 2nd Ed. Springer, NY (1998)
- Gasparyan, S, Kutoyants, YA: On approximation of the BSDE with unknown volatility in forward equation. *Armenian J. Math.* **7**(1), 59–79 (2015)
- Ibragimov, IA, Has'minskii, RZ: *Statistical Estimation - Asymptotic Theory*. Springer, New York (1981)
- Jeganathan, P: Some asymptotic properties of risk functions when the limit of the experiment is mixed normal. *Sankhya: The Indian Journal of Statistics* **45**(Series A, Pt.1), 66–87 (1983)
- Kamatani, K, Uchida, M: Hybrid multi-step estimators for stochastic differential equations based on sampled data. *Statist. Inference Stoch. Processes* **18**(2), 177–204 (2015)
- Kutoyants, YA: *Identification of Dynamical Systems with Small Noise*. Kluwer Academic Publisher, Dordrecht (1994)
- Kutoyants, YA: On approximation of the backward stochastic differential equation. Small noise, large samples and high frequency cases. *Proceed. Steklov Inst. Mathematics* **287**, 133–154 (2014)
- Kutoyants, YA: On Multi-Step MLE-Process for Ergodic Diffusion. arXiv:1504.01869 [math.ST] (2015)
- Kutoyants, YA, Motrunich, A: On multi-step MLE-process for Markov sequences. *Metrika* **79**(6), 705–724 (2016)
- Kutoyants, YA, Zhou, L: On approximation of the backward stochastic differential equation. (arXiv:1305.3728). *J. Stat. Plann. Infer* **150**, 111–123 (2014)
- Le Cam, L: On the asymptotic theory of estimation and testing hypotheses. *Proc. 3rd Berkeley Symposium*, vol. 1, pp. 129–156 (1956)

- Lehmann, EL, Romano, JP Testing Statistical Hypotheses, 3rd ed. Springer, NY (2005)
- Liptser, R, Shiryaev, AN Statistics of Random Processes, v.'s 1 and 2, 2-nd ed. Springer, NY (2001)
- Pardoux, E, Peng, S: Adapted solution of a backward stochastic differential equation. System Control Letter **14**, 55–61 (1990)
- Pardoux, E, Peng, S: Backward stochastic differential equations and quasilinear parabolic partial differential equations. *Stochastic Partial Differential Equations and their Applications*, pp. 200–217. Springer, Berlin (1992). (Lect. Notes Control Inf. Sci. 176)
- Robinson, PM: The stochastic difference between econometric statistics. *Econometrica* **56**(3), 531–548 (1988)
- Skorohod, AV, Khasminskii, RZ: On parameter estimation by indirect observations. Prob. Inform. Transm **32**, 58–68 (1996)
- Uchida, M, Yoshida, N: Adaptive Bayes type estimators of ergodic diffusion processes from discrete observations. *Statist. Inference Stoch. Processes* **17**(2), 181–219 (2014)