Armenian Journal of Mathematics
Volume 7, Number 1, 2015, 59–79

On Approximation of the BSDE with Unknown Volatility in Forward Equation

Samvel B. Gasparyan and Yury A. Kutoyants
Université du Maine, Yerevan State University and Université du Maine,
National Research University "MPEI"

Abstract. We consider the problem of the construction of the backward stochastic differential equation in the Markovian case. We suppose that the forward equation has a diffusion coefficient depending on some unknown parameter. We propose an estimator of this parameter constructed by the discrete time observations of the forward equation and then we use this estimator for approximation of the solution of the backward equation. The question of asymptotic optimality of this approximation is also discussed.

Key Words: Backward SDE, approximation of the solution, volatility estimation.
Mathematics Subject Classification 2000: 62M05, 62F12

1 Introduction

We consider the following problem. Suppose that we have a stochastic differential equation (called *forward*)

$$\mathrm{d}X_t = S\left(t, X_t\right)\mathrm{d}t + \sigma\left(\vartheta, t, X_t\right)\mathrm{d}W_t, \quad X_0, 0 \le t \le T$$

and we are given two functions $f\left(t, x, y, z\right)$ and $\Phi\left(x\right)$. We have to find a couple of stochastic processes $\left(Y_t, Z_t\right)$ such that it satisfies the stochastic differential equation (called *backward*)

$$\mathrm{d}Y_t = -f\left(t, X_t, Y_t, Z_t\right)\mathrm{d}t + Z_t\mathrm{d}W_t, \quad 0 \le t \le T$$

with the final value $Y_T = \Phi\left(X_T\right)$.

The solution of this problem is well-known. We have to solve a special partial differential equation, to find its solution $u\left(t, x, \vartheta\right)$ and to put $Y_t = u\left(t, X_t, \vartheta\right)$ and $Z_t = \sigma\left(\vartheta, t, X_t\right)u'_x\left(t, X_t, \vartheta\right)$.

We are interested by the problem of approximation of the solution (Y_t, Z_t) in the situation where the parameter ϑ is unknown. Therefore we first estimate this parameter with the help of some good estimator $\vartheta^\star_{t,n}, 0 < t \leq T$ and then we propose the approximations $Y_t^\star = u\left(t, X_t, \vartheta^\star_{t,n}\right)$ and $Z_t^\star = \sigma\left(\vartheta^\star_{t,n}, t, X_t\right) u'_x\left(t, X_t, \vartheta^\star_{t,n}\right)$. Moreover we show that the proposed approximations are in some sense asymptotically optimal.

The main difficulty in the construction of this approximation is to find an estimator-process $\vartheta^\star_{t,n}, 0 < t \leq T$ which can be easily calculated for all $t \in (0,T]$ and at the same time has asymptotically optimal properties. Unfortunately we can not use the well-studied pseudo-MLE based on the pseudo-maximum likelihood function because its calculation is related to the solution of nonlinear equations and numerically is sufficiently difficult problem.

We propose here a one-step MLE-process, which was recently introduced in the case of ergodic diffusion [10] and in the case of diffusion process with small noise [11], [12]. The review of statistical problems for the BSDE model of observations can be found in [9].

Note that the problem of volatility parameter estimation by discrete time observations is actually a well developed branch of statistics (see [19] and references therein). The particularity of our approach is due to the need of updated *on-line* estimator $\vartheta^\star_{t,n}$ which depends on the first observations up to time t.

Recall that the BSDE was first introduced in the linear case by Bismuth [2] and in general case this equation was studied by Pardoux and Peng [16]. Since that time the BSDE attracts attention of probabilists working in financial mathematics and obtained an intensive development (see, e.g. El Karoui *et al.* [7], Ma and Yong [15] and the references therein). The detailed exposition of the current state of this theory can be found in Pardoux and Rascanu [18].

Note that the approach developed in the present paper is carried out for the following three models of observations (forward equations).

Diffusion process with *small noise* ($\varepsilon \to 0$)

$$\mathrm{d}X_t = S\left(\vartheta, t, X_t\right)\mathrm{d}t + \varepsilon\sigma\left(t, X_t\right)\,\mathrm{d}W_t, \quad x_0,\ 0 \leq t \leq T,$$

(see [11], [12]).

Ergodic diffusion process (*large samples,* $T \to \infty$)

$$\mathrm{d}X_t = S\left(\vartheta, X_t\right)\mathrm{d}t + \sigma\left(X_t\right)\,\mathrm{d}W_t, \quad X_0,\ 0 \leq t \leq T,$$

(see [1]).

High frequency observations $X^n = (X_{t_0}, X_{t_1}, \ldots X_{t_n}),\ t_i = i\frac{T}{n}, (n \to \infty)$ of diffusion process with volatility depending on the unknown parameter

$$\mathrm{d}X_t = S\left(t, X_t\right)\mathrm{d}t + \sigma\left(\vartheta, t, X_t\right)\,\mathrm{d}W_t, \quad X_0,\ 0 \leq t \leq T,$$

see present work.

We propose for these three models of observations three MLE-processes $\vartheta_t^* = \vartheta_t^*(X_s, 0 \leq s \leq t)$ such that the corresponding approximations $\hat{Y}_t = u(t, X_t, \vartheta_t^*) \to Y_t$ and the error of approximation $\mathbf{E}_\vartheta \left(\hat{Y}_t - Y_t\right)^2$ for each model of observations is asymptotically ($\varepsilon \to 0, T \to \infty, n \to \infty$) minimal.

2 Auxiliary results

Let us recall what is the BSDE in the Markovian case. Suppose that we are given a filtered probability space $\left(\Omega, (\mathcal{F}_t)_{t\in[0,T]}, \mathbb{P}\right)$ with the filtration $(\mathcal{F}_t)_{t\in[0,T]}$ satisfying the *usual conditions*. Define the stochastic differential equation (called *forward*)

$$\mathrm{d}X_t = S(t, X_t)\,\mathrm{d}t + \sigma(t, X_t)\,\mathrm{d}W_t, \quad X_0,\ 0 \leq t \leq T,$$

where $(W_t, \mathcal{F}_t, 0 \leq t \leq T)$ is the standard Wiener process and X_0 is $\mathcal{F}_0$ measurable initial value. The trend coefficient $S(t,x)$ and diffusion coefficient $\sigma(t,x)^2$ satisfy the Lipschitz and linear growth conditions

$$|S(t,x) - S(t,y)| + |\sigma(t,x) - \sigma(t,y)| \leq L\,|x-y|, \tag{1}$$
$$|S(t,x)| + |\sigma(t,x)| \leq C\,(1+|x|), \tag{2}$$

for all $x, y \in \mathbb{R}$ and for all $t \in [0,T]$. Here $L > 0$ and $C > 0$ are some constants. By these conditions the stochastic differential equation has a unique strong solution (see Liptser and Shiryaev [14]).

Further, we are given two functions $f(t,x,y,z)$ and $\Phi(x)$ and we have to construct a couple of processes (Y_t, Z_t) such that the solution of the stochastic differential equation

$$\mathrm{d}Y_t = -f(t, X_t, Y_t, Z_t)\,\mathrm{d}t + Z_t\,\mathrm{d}W_t, \quad 0 \leq t \leq T,$$

(called *backward*) has the terminal value $Y_T = \Phi(X_T)$.

This equation is often written as follows

$$Y_t = \Phi(X_T) + \int_t^T f(s, X_s, Y_s, Z_s)\,\mathrm{d}s - \int_t^T Z_s\,\mathrm{d}W_s, \quad 0 \leq t \leq T.$$

We suppose that the functions $f(t,x,y,z)$ and $\Phi(x)$ satisfy the conditions

$$|f(t,x,y_1,z_1) - f(t,x,y_2,z_2)| \leq L\,(|y_1-y_2| + |z_1-z_2|),$$
$$|f(t,x,y,z)| + |\Phi(x)| \leq C\,(1+|x|^p),$$

for all $x, y, z, y_i, z_i \in \mathbb{R}$, $i = 1,2$ and for all $t \in [0,T]$. Here $p \geq 1/2$.

This is the so-called *Markovian case*. For the existence and uniqueness of the solution see Pardoux and Peng [17].

The solution (Y_t, Z_t) can be constructed as follows. Suppose that $u(t,x)$ satisfies the equation

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

with the terminal condition $u(T,x)=\Phi(x)$.

Let us put $Y_t=u(t,X_t)$, then we obtain by Itô's formula

$$\mathrm{d}Y_t=\left[\frac{\partial u}{\partial t}+S(t,X_t)\frac{\partial u}{\partial x}+\frac{\sigma(t,X_t)^2}{2}\frac{\partial^2 u}{\partial x^2}\right]\mathrm{d}t+\sigma(t,X_t)\frac{\partial u}{\partial x}\,\mathrm{d}W_t.$$

Hence if we denote $Z_t=\sigma(t,X_t)\,u'_x(t,X_t)$ then this equation become

$$\mathrm{d}Y_t=-f(t,X_t,Y_t,Z_t)\,\mathrm{d}t+Z_t\,\mathrm{d}W_t,\qquad Y_0=u(0,X_0).$$

The terminal value $Y_T=u(T,X_T)=\Phi(X_T)$.

We consider the problem of the approximation of the solution (Y_t,Z_t) of BSDE in the situations, where the forward equation contains an unknown finite-dimensional parameter ϑ:

$$\mathrm{d}X_t=S(t,X_t)\,\mathrm{d}t+\sigma(\vartheta,t,X_t)\,\mathrm{d}W_t,\quad X_0,\ 0\le t\le T. \tag{3}$$

Then the solution u of the corresponding partial differential equation

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

depends on ϑ, i.e., $u=u(t,x,\vartheta)$. The backward equation

$$\mathrm{d}Y_t=-f(t,X_t,Y_t,Z_t)\,\mathrm{d}t+Z_t\,\mathrm{d}W_t,\quad 0\le t\le T,$$

we obtain if we put $Y_t=u(t,X_t,\vartheta)$ and $Z_t=u'_x(t,X_t,\vartheta)\,\sigma(\vartheta,t,X_t)$. But as ϑ is unknown we propose the "natural" approximations

$$\hat{Y}_t=u(t,X_t,\vartheta_t^*),\qquad \hat{Z}_t=u'_x(t,X_t,\vartheta_t^*)\,\sigma(\vartheta_t^*,t,X_t).$$

Here $\vartheta_t^*, 0\le t\le T$ is some *good estimator-process* of ϑ with *small error*. In this problem the good estimator means the following

- $\vartheta_t^\star=\vartheta_t^\star(X^t)$, i.e., *it depends on* observations $X^t=(X_s, 0\le s\le t)$ up to time t.

- *Easy to calculate for each* $t\in(0,T]$.

- *Provides an asymptotically efficient estimator of Y_t, i.e., we have in some sense*

$$\mathbf{E}_\vartheta \left(\hat Y_t - Y_t\right)^2 \to \min_{\bar Y_t} \mathbf{E}_\vartheta \left(\bar Y_t - Y_t\right)^2.$$

As we have already fixed the approximation Y_t as $\hat Y_t = u(t, X_t, \vartheta_t^*)$ the main problem is *how to find a good estimator-process* $\vartheta_t^*, 0 \leq t \leq T$?

Observe that the problem of estimation of ϑ is singular, i.e., the parameter ϑ can be estimated by continuous time observations without error.

3 Continuous time observations

Let us remind the situation which we have in the case of continuous time observations of the solution of the stochastic differential equation (3) (see, e.g., [20]).

By Itô's formula

$$X_t^2 = X_0^2 + 2\int_0^t X_s \,\mathrm{d}X_s + \int_0^t \sigma\left(\vartheta_0, s, X_s\right)^2 \mathrm{d}s,$$

where ϑ_0 is the true value.

The trajectory fitting estimator (TFE) ϑ^* of the parameter ϑ can be defined as follows

$$\inf_{\vartheta\in\Theta} \int_0^T \left[X_t^2 - X_0^2 - 2\int_0^t X_s \,\mathrm{d}X_s - \int_0^t \sigma\left(\vartheta, s, X_s\right)^2 \mathrm{d}s\right]^2 \mathrm{d}t$$
$$= \int_0^T \left[X_t^2 - X_0^2 - 2\int_0^t X_s \,\mathrm{d}X_s - \int_0^t \sigma\left(\vartheta^*, s, X_s\right)^2 \mathrm{d}s\right]^2 \mathrm{d}t$$

Of course, $\vartheta^* = \vartheta_0$ under the following mild identifiability condition: for any $\nu > 0$ with probability 1 we have

$$\inf_{|\vartheta-\vartheta_0|>\nu} \int_0^T \left[\int_0^t \sigma\left(\vartheta, s, X_s\right)^2 \mathrm{d}s - \int_0^t \sigma\left(\vartheta_0, s, X_s\right)^2 \mathrm{d}s\right]^2 \mathrm{d}t > 0.$$

If this condition is not fulfilled, then on an event of positive probability, for some $\vartheta_1 \neq \vartheta_0$ we have

$$\int_0^t \sigma\left(\vartheta_1, s, X_s\right)^2 \mathrm{d}s = \int_0^t \sigma\left(\vartheta_0, s, X_s\right)^2 \mathrm{d}s, \ \forall t \in [0,T],$$

which implies that for all $t \in [0,T]$

$$\sigma\left(\vartheta_1, t, X_t\right)^2 = \sigma\left(\vartheta_0, t, X_t\right)^2.$$

In such a situation no estimation method can provide us a consistent estimator.

Let us illustrate this situation by several examples.

Example 1. Suppose that $\sigma_t(\vartheta, x) = \sqrt{\vartheta} h_t(x), \vartheta \in (\alpha, \beta), \alpha > 0$, and the observed process is

$$\mathrm{d}X_t = S_t(X)\,\mathrm{d}t + \sqrt{\vartheta} h_t(X)\,\mathrm{d}W_t, \quad X_0, \quad 0 \leq t \leq T,$$

where $S_t(X)$ and $h_t(X)$ are some functionals of the past, say,

$$S_t(X) = M(t, X_t) + \int_0^t N(s, X_s)\,\mathrm{d}s, \quad h_t(X) = P(t, X_t) + \int_0^t q(s, X_s)\,\mathrm{d}s,$$

where $M(\cdot), N(\cdot), P(\cdot), q(\cdot)$ are smooth functions. This is an example of so-called *diffusion type* process [14].

To estimate ϑ without error we use two approaches. The first one is the TFE

$$\begin{aligned}\vartheta^* &= \arg\inf_{\vartheta\in\Theta} \int_0^T \left[X_t^2 - X_0^2 - 2\int_0^t X_s \mathrm{d}X_s - \vartheta \int_0^t h_s(X)^2\,\mathrm{d}s\right]^2 \mathrm{d}t \\ &= D_T(h)^{-1} \int_0^T \left[X_t^2 - X_0^2 - 2\int_0^t X_s \mathrm{d}X_s\right] \int_0^t h_s(X)^2\,\mathrm{d}s\,\mathrm{d}t = \vartheta_0,\end{aligned}$$

where

$$D_T(h) = \int_0^T \left(\int_0^t h_s(X)^2\,\mathrm{d}s\right)^2 \mathrm{d}t$$

The second possibility is the following. Let $G(x)$ be a two-times continuously differentiable function. By the Itô's formula for $G(X_t)$ we can write

$$G(X_t) = G(X_0) + \int_0^t G'(X_s)\,\mathrm{d}X_s + \frac{\vartheta_0}{2}\int_0^t G''(X_s)\,h_s(X)^2\,\mathrm{d}s.$$

We solve this equation w.r.t. ϑ_0 and obtain for all $t \in (0, T]$ with probability 1 the equality

$$\bar{\vartheta}_t = \frac{2G(X_t) - 2G(X_0) - 2\int_0^t G'(X_s)\,\mathrm{d}X_s}{\int_0^t G''(X_s)\,h_s(X)^2\,\mathrm{d}s} = \vartheta_0.$$

Therefore we have for all $t \in (0, T]$ the estimator $\bar{\vartheta}_t = \vartheta_0$. Note that we need not know $S(\cdot)$ and the only condition we use is that for all $t \in (0, T]$

$$\int_0^t G''(X_s)\,h_s(X)^2\,\mathrm{d}s \neq 0.$$

Therefore we obtain an "estimator" of unknown the parameter without error.
Example 2. Suppose that the unknown parameter is $\vartheta = (\vartheta_1, \ldots, \vartheta_d) \in \mathbb{R}_+^d$ and the diffusion coefficient

$$\sigma(\vartheta, t, X_t)^2 = \lambda + \sum_{l=1}^{d} \vartheta_l h_l(t, X_t),$$

where $\lambda > 0$ and the functions $h_l(\cdot) > 0$ are known and $\vartheta_l > 0, l = 1, \ldots, d$. The observed diffusion process is

$$\mathrm{d}X_t = S(t, X_t)\,\mathrm{d}t + \sigma(\vartheta, t, X_t)\,\mathrm{d}W_t, \quad X_0, \quad 0 \leq t \leq T.$$

If we denote the vector $H_t = (H_{l,t}, \ldots, H_{l,t})$

$$H_{l,t} = \int_0^t h_l(s, X_s)\,\mathrm{d}s, \qquad l = 1, \ldots, d,$$

and introduce the $d \times d$ matrix $\mathbb{H}_t$ and vector $\tilde{X}_t$ by the relations

$$\mathbb{H}_t = \int_0^t H_t H_t^{\mathbb{T}}\,\mathrm{d}s, \qquad \tilde{X}_t = \left[X_t^2 - X_0^2 - 2\int_0^t X_s \mathrm{d}X_s - \lambda t\right] H_t,$$

then the TFE is

$$\vartheta_t^* = \mathbb{H}_t^{-1} \int_0^t \left[X_s^2 - X_0^2 - 2\int_0^s X_v \mathrm{d}X_v - \lambda s\right] H_s\,\mathrm{d}s = \vartheta_0.$$

Here we suppose that the matrix $\mathbb{H}_t$ for some values of t is non degenerate and we calculate the estimator for these values of t. We see that once more we estimate the unknown parameter without error.

Therefore in the case of continuous time observations the approximations $\hat{Y}_t = u\left(t, X_t, \bar{\vartheta}_t\right)$ and $\hat{Z}_t = u'_x\left(t, X_t, \bar{\vartheta}_t\right)\sigma\left(\bar{\vartheta}_t, t, X_t\right)$ or $\hat{Y}_t = u(t, X_t, \vartheta_t^*)$ and $\hat{Z}_t = u'_x(t, X_t, \vartheta_t^*)\,\sigma(\vartheta_t^*, t, X_t)$ are without errors: $\hat{Y}_t = Y_t,\ \hat{Z}_t = Z_t$.

4 High frequency asymptotics

The problem becomes more interesting if we consider the discrete time observations. Suppose that the solution of the equation (4) is observed at discrete times $t_i = i\frac{T}{n}$ and we have to study the approximations

$$\hat{Y}_{t_k} = u(t_k, X_{t_k}, \hat{\vartheta}_{t_k}), \quad \hat{Z}_{t_k} = \sigma\left(\hat{\vartheta}_{t_k}, t_k, X_{t_k}\right) u'_x(t_k, X_{t_k}, \hat{\vartheta}_{t_k}), k = 1, \ldots, n,$$

of the solution Y_t, Z_t of BSDE (5). Here k satisfies the conditions $t_k \leq t < t_{k+1}$ and the estimator $\hat{\vartheta}_{t_k}$ can be constructed by the observations $X^k =$

$(X_0, X_{t_1}, \ldots, X_{t_k})$ up to time t_k. We study the properties of estimators in the so-called *higher frequency* asymptotics: $n \to \infty$. Observe that the problem of estimation of the parameter ϑ in the case of discrete-time observations of the processes like (4) was extensively studied last years (see, e.g., [19] and the references therein).

Suppose that the forward equation is

$$\mathrm{d}X_t = S\left(t, X_t\right) \mathrm{d}t + \sigma\left(\vartheta, t, X_t\right) \,\mathrm{d}W_t, \quad X_0,\ 0 \le t \le T \tag{4}$$

where $\vartheta \in \Theta$, the set Θ is an open, convex, bounded subset of $\mathcal{R}^d$. The BSDE is

$$\mathrm{d}Y_t = -f\left(t, X_t, Y_t, Z_t\right) \mathrm{d}t + Z_t \mathrm{d}W_t, \quad 0 \le t \le T \tag{5}$$

with $Y_T = \Phi\left(X_T\right)$. It can be obtained with the help of the of functions

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

satisfying for all $\vartheta \in \Theta$ the partial differential equation

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

and the terminal condition $u(T,x,\vartheta) = \Phi\left(x\right)$. The equation (5) we obtain by the Itô's formula for the function $Y_t = u\left(t, X_t, \vartheta\right)$.

As before our goal is to find an estimator $\hat\vartheta_t$ such that the approximation $\hat Y_t = u(t, X_t, \hat\vartheta_t)$ has good properties.

Recall that in the case of continuous time observations there is no statistical problem of estimation of ϑ because the measures $\left\{\mathbf{P}_\vartheta^{(T)}, \vartheta \in \Theta\right\}$ corresponding to different values of ϑ are singular.

Then in Example 1 with $h_t\left(X\right) = h\left(t, X_t\right)$ and $G\left(x\right) = x^2$ we obtain the well-known estimator

$$\bar\vartheta_{k,n} = \frac{X_{t_k}^2 - X_0^2 - 2\sum_{j=1}^k X_{t_{j-1}}\left(X_{t_j} - X_{t_{j-1}}\right)}{\sum_{j=1}^k h\left(t_{j-1}, X_{t_{j-1}}\right)^2 \delta}, \quad \delta = \frac{T}{n}.$$

It can be easily shown that if $n \to \infty$ then for a fixed t and corresponding $k = \left[\frac{nt}{T}\right]$ we have these convergences in probability

$$\sum_{j=1}^k X_{t_{j-1}}\left(X_{t_j} - X_{t_{j-1}}\right) \longrightarrow \int_0^t X_s \mathrm{d}X_s,$$

$$\sum_{j=1}^k h\left(t_{j-1}, X_{t_{j-1}}\right)^2 \delta \rightarrow \int_0^t h\left(s, X_s\right)^2 \mathrm{d}s$$

and therefore, in probability,

$$\bar{\vartheta}_{k,n} \longrightarrow \frac{X_t^2 - X_0^2 - 2\int_0^t X_s \mathrm{d}X_s}{\int_0^t h\left(s, X_s\right)^2 \mathrm{d}s} = \vartheta_0.$$

Of course we can use it in the approximation of Y_t and Z_t as follows $\hat{Y}_{t_k,n} = u(t_k, X_{t_k}, \bar{\vartheta}_{k,n})$ and $\hat{Z}_{t_k,n} = \bar{\vartheta}_{k,n} u'(t_k, X_{t_k}, \bar{\vartheta}_{k,n}) h\left(t_k, X_{t_k}\right)$. Note that this type of approximation is not asymptotically efficient and we will seek other estimators of ϑ which can provide smaller error of estimation.

Below we study the distribution of the error $\sqrt{n}\left(\hat{Y}_{t_k,n} - Y_{t_k}\right)$ in a more general situation and with the help of a different estimator of ϑ.

We proceed as follows. The observed forward equation (4) can be written as

$$X_{t_j} - X_{t_{j-1}} = \int_{t_{j-1}}^{t_j} S\left(s, X_s\right) \mathrm{d}s + \int_{t_{j-1}}^{t_j} \sigma\left(\vartheta, s, X_s\right) \mathrm{d}W_s. \tag{6}$$

but we consider a (wrong) model which we obtain if we replace the functions $S\left(s, X_s\right)$ and $\sigma\left(\vartheta, s, X_s\right)$ in these integrals by the piecewise constant function with values $S\left(t_{j-1}, X_{t_{j-1}}\right)$ and $\sigma\left(\vartheta, t_{j-1}, X_{t_{j-1}}\right)$ respectively on the interval $[t_{j-1}, t_j]$. Then we obtain

$$X_{t_j} - X_{t_{j-1}} = S\left(t_{j-1}, X_{t_{j-1}}\right)\delta + \sigma\left(\vartheta, t_{j-1}, X_{t_{j-1}}\right)\left(W_{t_j} - W_{t_{j-1}}\right). \tag{7}$$

Note that if (7) is true then the random variables

$$\frac{X_{t_j} - X_{t_{j-1}} - S\left(t_{j-1}, X_{t_{j-1}}\right)\delta}{\sigma\left(\vartheta, t_{j-1}, X_{j-1}\right)\sqrt{\delta}} \qquad j = 1, \ldots, n$$

are i.i.d. with the standard Gaussian distribution $\mathcal{N}\left(0,1\right)$.

Introduce the log pseudo-likelihood for the model (7)

$$\begin{aligned} L_{t,k}\left(\vartheta, X^k\right) = &-\frac{1}{2}\sum_{j=0}^{k} \ln\left[2\pi\sigma\left(\vartheta, t_{j-1}, X_{t_{j-1}}\right)^2\delta\right] \\ &- \sum_{j=1}^{k} \frac{\left[X_{t_j} - X_{t_{j-1}} - S\left(t_{j-1}, X_{t_{j-1}}\right)\delta\right]^2}{2\sigma\left(\vartheta, t_{j-1}, X_{t_{j-1}}\right)^2\delta} \end{aligned}$$

and define the pseudo-maximum likelihood estimator (PMLE) $\hat{\vartheta}_{t,n}$ by the equation

$$L_{t,k}(\hat{\vartheta}_{t,n}, X^k) = \sup_{\theta\in\Theta} L_{t,k}\left(\theta, X^k\right).$$

As it was already explained such estimator can not be used for the construction of the approximations of BSDE due to the complexity of the calculations of the solution of this equation for all k in nonlinear case. Below we will use this estimator as a preliminary one for the construction of a one-step MLE-process.

Regularity conditions. (R)

R$_1$ *The functions $S(\cdot)$ and $\sigma(\cdot)$ satisfy the conditions of Lipschiz and of linear growth.*

R$_2$ *The function $\sigma(\vartheta,t,x)$ is strictly positive and has two continuous derivatives w.r.t. ϑ.*

$$\left|\dot\sigma(\vartheta,t,x)\right|+\left\|\ddot\sigma(\vartheta,t,x)\right\|\leq C\left(1+\left|x\right|^p\right) \tag{8}$$

R$_3$ *With probability one, the information matrix*

$$\mathbb{I}_t(\vartheta)=2\int_0^t\frac{\dot\sigma(\vartheta,s,X_s)\,\dot\sigma(\vartheta,s,X_s)^{\mathbb{T}}}{\sigma(\vartheta,s,X_s)^2}\,\mathrm{d}s$$

is strictly positive defined for all $t\in(0,T]$.

R$_4$ *The function $u(t,x,\vartheta)$ is continuously differentiable w.r.t. ϑ and the derivative* satisfies the condition

$$|\dot u(t,x,\vartheta)|\leq \mathrm{C}(1+|x|^p).$$

It will be convenient to replace the likelihood ratio function by the contrast function

$$\begin{aligned}U_{t,k}\left(\vartheta,X^k\right)&=\sum_{j=1}^k\delta\ln a\left(\vartheta,t_{j-1},X_{t_{j-1}}\right)\\&\qquad+\sum_{j=1}^k\frac{\left(X_{t_j}-X_{t_{j-1}}-S\left(t_{j-1},X_{t_{j-1}}\right)\delta\right)^2}{a\left(\vartheta,t_{j-1},X_{t_{j-1}}\right)},\end{aligned}$$

where $a(\vartheta,t,x)=\sigma(\vartheta,t,x)^2$. The estimator $\hat\vartheta_{t,n}$ satisfies the equation

$$U_{t,k}\left(\hat\vartheta_{t,n},X^k\right)=\inf_{\vartheta\in\Theta}U_{t,k}\left(\vartheta,X^k\right). \tag{9}$$

The contrast function converges to the following limit

$$\begin{aligned}U_{t,k}\left(\vartheta,X^k\right)\longrightarrow U_t\left(\vartheta,X^t\right)&=\int_0^t\left[\frac{a(\vartheta_0,s,X_s)}{a(\vartheta,s,X_s)}-\ln\frac{a(\vartheta_0,s,X_s)}{a(\vartheta,s,X_s)}\right]\mathrm{d}s\\&\qquad+\int_0^t\ln a(\vartheta_0,s,X_s)\,\mathrm{d}s.\end{aligned}$$

Identifiability condition.

I_1 With probability one the random function $U_t\left(\vartheta, X^t\right), \vartheta \in \Theta, 0 < t \leq T$ has a unique minimum at the point $\vartheta = \vartheta_0$

$$\inf_{\vartheta \in \Theta} U_t\left(\vartheta, X^t\right) = U_t\left(\vartheta_0, X^t\right), \quad 0 < t \leq T.$$

In realty this condition requires only the uniqueness of the solution since the function $x \mapsto x - \ln x$ attains its minimum at $x = 1$ implies that $U_t(\vartheta, X^t)$ is always larger than $U_t(\vartheta_0, X^t)$. Hence ϑ_0 is always a minimum point of the mapping $\vartheta \mapsto U_t(\vartheta, X^t)$.

Introduce the vector-process

$$\xi_t\left(\vartheta_0\right) = \mathbb{I}_t\left(\vartheta_0\right)^{-1} \sqrt{2} \int_0^t \frac{\dot{\sigma}\left(\vartheta_0, s, X_s\right)}{\sigma\left(\vartheta_0, s, X_s\right)} \,\mathrm{d}w\left(s\right), \quad 0 < t \leq T.$$

Note that the Wiener process $w\left(s\right), 0 \leq s \leq T$ here is independent of the diffusion process $X_s, 0 \leq s \leq T$.

For given $t \in (0, T]$ the value t_k in the estimator $\hat{\vartheta}_{t_k,n}$ satisfies the condition $t_k \leq t < t_{k+1}$.

Theorem 1 *Suppose that the Regularity and Identifiability conditions are fulfilled. Then for all $t \in (0, T]$ the estimator $\hat{\vartheta}_{t_k,n}$ is consistent and asymptotically conditionally normal (stable convergence)*

$$\sqrt{\frac{n}{T}}\left(\hat{\vartheta}_{t,n} - \vartheta_0\right) \Longrightarrow \xi_t\left(\vartheta_0\right). \tag{10}$$

Moreover this estimator is asymptotically efficient.

The proofs of this theorem can be found in [3] (lower bound, $d = 1$) and in [4] (properties of estimator, $d \geq 1$).

Let us give here some lines of the proof. Suppose that the consistency of the estimator $\hat{\vartheta}_{t_k,n}$ defined by the equation (9) is already proved.

Introduce the independent random variables

$$w_j = \left(2\delta\right)^{-1/2}\left[\left(W_{t_j} - W_{t_{j-1}}\right)^2 - \delta\right], \quad \mathbf{E}w_j = 0, \quad \mathbf{E}w_j^2 = \delta, \quad \mathbf{E}w_j w_i = 0$$

for $j \neq i$ and note that the empirical Fisher information matrix

$$\mathbb{I}_{t,n}\left(\vartheta_0\right) = 2\sum_{j=1}^{k} \frac{\dot{\sigma}(\vartheta_0, t_{j-1}, X_{t_{j-1}})\dot{\sigma}(\vartheta_0, t_{j-1}, X_{t_{j-1}})^{\mathbb{T}}}{\sigma(\vartheta, t_{j-1}, X_{t_{j-1}})^2}\,\delta \longrightarrow \mathbb{I}_t\left(\vartheta_0\right) \tag{11}$$

as $n \to \infty$. Then by Taylor expansion of the solution $\hat{\vartheta}_{t,n}$ of the system of d-equations

$$\frac{\partial U_{t,k}\left(\vartheta, X^k\right)}{\partial \vartheta} = 0$$

we can write the representation of the MCE

$$\delta^{-1/2}\left(\hat\vartheta_{t,n}-\vartheta_0\right)$$
$$=\mathbb{I}_{t,n}\left(\vartheta_0\right)^{-1}\sqrt{2}\sum_{j=1}^{k}\frac{\dot\sigma(\vartheta_0,t_{j-1},X_{t_{j-1}})}{\sigma(\vartheta_0,t_{j-1},X_{t_{j-1}})}\,w_j\left(1+o\left(1\right)\right).$$

Throughout this paper the symbols o, O are understood in the sense of convergence in probability. Now the convergence (10) follows from (11) and (stable convergence)

$$\sum_{j=1}^{k}\frac{\dot\sigma(\vartheta_0,t_{j-1},X_{t_{j-1}})}{\sigma(\vartheta_0,t_{j-1},X_{t_{j-1}})}\,w_j\Longrightarrow\int_0^t\frac{\dot\sigma(\vartheta_0,s,X_s)}{\sigma(\vartheta,s,X_s)}\,{\rm d}w\left(s\right).$$

Note that the approximation $\hat Y_{t,n}=u(t,X_{t_k},\hat\vartheta_{t,n})$ is computationally difficult to realize because solving equation (9) for all $t_k, k=1,\ldots,n$ especially in non linear case is almost impossible. That is why we propose the one-step MLE-process as follows.

Let us fix some (small) $\tau>0$ and denote by $\hat\vartheta_{\tau,n}$ the MCE constructed by the observations $X^{\tau,n}=\left(X_0,X_{t_{1,n}},\ldots,X_{t_{N,n}}\right)$, where $t_{N,n}\leq\tau<t_{N+1,n}$.

By Theorem 1, this estimator is consistent and asymptotically conditionally normal

$$\sqrt{\frac{n}{T}}\left(\hat\vartheta_{\tau,n}-\vartheta_0\right)=\mathbb{I}_{\tau,n}\left(\vartheta_0\right)^{-1}\sqrt{2}\sum_{j=1}^{N}\frac{\dot\sigma(\vartheta_0,t_{j-1},X_{t_{j-1}})}{\sigma(\vartheta_0,t_{j-1},X_{t_{j-1}})}\,w_j+o\left(1\right)$$
$$\Longrightarrow\xi_\tau\left(\vartheta_0\right)=\mathbb{I}_\tau\left(\vartheta_0\right)^{-1}\sqrt{2}\int_0^\tau\frac{\dot\sigma\left(\vartheta_0,s,X_s\right)}{\sigma\left(\vartheta_0,s,X_s\right)}{\rm d}w\left(s\right).$$

Here the random Fisher information matrix is

$$\mathbb{I}_\tau\left(\vartheta_0\right)=2\int_0^\tau\frac{\dot\sigma\left(\vartheta_0,s,X_s\right)\dot\sigma\left(\vartheta_0,s,X_s\right)^{\mathbb{T}}}{\sigma\left(\vartheta_0,s,X_s\right)^2}\,{\rm d}s.$$

Introduce the pseudo score-function $\left(A_{j-1}\left(\vartheta\right)=\sigma\left(\vartheta,t_{j-1},X_{t_{j-1}}\right)^2\right)$

$$\Delta_{k,n}\left(\vartheta,X^k\right)=\sum_{j=1}^{k}\dot\ell\left(\vartheta,X_{t_{j-1}},X_{t_j}\right)$$
$$=\sum_{j=1}^{k}\frac{\left[\left(X_{t_j}-X_{t_{j-1}}-S_{j-1}\,\delta\right)^2-A_{j-1}\left(\theta\right)\delta\right]\dot A_{j-1}\left(\vartheta\right)}{2A_{j-1}\left(\vartheta\right)^2\sqrt{\delta}}.$$

For any $t\in[\tau,T]$ define k by the condition $t_k\leq t<t_{k+1}$ and the one-step PMLE-process by the relation

$$\vartheta^\star_{k,n}=\hat\vartheta_{\tau,n}+\sqrt{\delta}\,\mathbb{I}_{k,n}(\hat\vartheta_{\tau,n})^{-1}\Delta_{k,n}(\hat\vartheta_{\tau,n},X^k),\qquad k=N+1,\ldots,n.\qquad(12)$$

Our goal is to show that the corresponding approximation

$$Y^{\star}_{t_k,n}=u\left(t_k,X_{t_k},\vartheta^{\star}_{k,n}\right),k=N+1,\ldots,n$$

is asymptotically efficient. To do this we need to present the lower bound on the risks of all estimators and then to show that for the proposed approximation this lower bound is achieved.

First we recall some known results on asymptotically efficient estimation of the parameter ϑ. The family of measures $\left\{\mathbf{P}^{(k,n)}_{\vartheta},\vartheta\in\Theta\right\}$ induced in R^k by the observations X^k are *locally asymptotically mixed normal* (LAMN), i.e., the likelihood ratio

$$Z_{k,n}\left(v\right)=\frac{\mathrm{d}\mathbf{P}^{(k,n)}_{\vartheta_0+\frac{v}{\sqrt{n}}}}{\mathrm{d}\mathbf{P}^{(k,n)}_{\vartheta_0}},\qquad v\in V_n=\left\{v:\ \vartheta_0+\frac{v}{\sqrt{n}}\in\Theta\right\},$$

admits the representation

$$Z_{k,n}\left(v\right)=\exp\left\{\langle v,\Delta_{k,n}\left(\vartheta_0,X^k\right)\rangle-\frac{1}{2}v\mathbb{I}_{k,n}\left(\vartheta_0\right)v^{\mathbb{T}}+r_n\right\},$$

where $r_n=r_n(v,\vartheta_0)\rightarrow 0$ in probability for fixed $\vartheta_0\in\Theta$ and fixed $v\in R$. The proof can be found in [3] ($d=1$) and in [5] ($d\geq 1$).

In statistical problems with such property of families of measures we have, so-called, Jeganathan-type lower bound on the risks of all estimators $\bar\vartheta_{k,n}$:

$$\varliminf_{\varepsilon\rightarrow 0}\varliminf_{n\rightarrow\infty}\sup_{\left|\vartheta-\vartheta_0\right|\leq\varepsilon}\mathbf{E}_{\vartheta}\ell\left(\delta^{-1/2}\left(\bar\vartheta_{k,n}-\vartheta\right)\right)\geq\mathbf{E}_{\vartheta_0}\ell\left(\xi_t\left(\vartheta_0\right)\right).$$

Here $\ell\left(v\right),v\in R^d$ is some symmetric, non decreasing loss function (see the conditions in [6]).

Therefore we can call an estimator $\vartheta^{*}_{k,n}$ asymptotically efficient if for some function $\ell\left(\cdot\right)$ and all $\vartheta_0\in\Theta$ we have the equality

$$\lim_{\varepsilon\rightarrow 0}\lim_{n\rightarrow\infty}\sup_{\left|\vartheta-\vartheta_0\right|\leq\varepsilon}\mathbf{E}_{\vartheta}\ell\left(\delta^{-1/2}\left(\vartheta^{*}_{k,n}-\vartheta\right)\right)=\mathbf{E}_{\vartheta_0}\ell\left(\xi_t\left(\vartheta_0\right)\right).$$

We say that the estimator-process $\vartheta^{*}_{k,n},k=N+1,\ldots,n$ is asymptotically efficient for the values $t\in[\tau_*,T]$, if we have this equality for all $t\in[\tau_*,T]$. Here $0<\tau<\tau_*<T$.

Theorem 2 *The one-step MLE-process $\vartheta^{\star}_{k,n},k=N+1,\ldots,n$ is consistent, asymptotically conditionally normal (stable convergence)*

$$\delta^{-1/2}\left(\vartheta^{\star}_{k,n}-\vartheta_0\right)\Longrightarrow\xi_t\left(\vartheta_0\right)\tag{13}$$

and is asymptotically efficient for $t\in[\tau_*,T]$ where $\tau<\tau_*<T$ and the loss functions is bounded.

Proof. The proof follows the main steps of the similar proof given in [10].

We have for any $\nu > 0$ the estimates

$$\mathbf{P}_{\vartheta_0}^{(k,n)}\left\{\left|\vartheta^\star_{k,n}-\vartheta_0\right|>\nu\right\}\leq \mathbf{P}_{\vartheta_0}^{(k,n)}\left\{\left|\hat\vartheta_{\tau,n}-\vartheta_0\right|>\frac{\nu}{2}\right\}$$
$$+\,\mathbf{P}_{\vartheta_0}^{(k,n)}\left\{\left|\sqrt{\delta}\,\mathbb{I}_{k,n}(\hat\vartheta_{\tau,n})^{-1}\Delta_{k,n}(\hat\vartheta_{\tau,n},X^k)\right|>\frac{\nu}{2}\right\}.$$

We can write

$$\left|\mathbb{I}_{k,n}(\hat\vartheta_{\tau,n})-\mathbb{I}_{k,n}(\vartheta_0)\right|\leq C\left|\hat\vartheta_{\tau,n}-\vartheta_0\right|\longrightarrow 0$$

and

$$\sqrt{\delta}\left|\Delta_{k,n}(\hat\vartheta_{\tau,n},X^k)-\Delta_{k,n}(\vartheta_0,X^k)\right|\longrightarrow 0.$$

Further, it can be shown that

$$\mathbf{P}_{\vartheta_0}^{(k,n)}\left\{\left|\sqrt{\delta}\,\mathbb{I}_{k,n}(\vartheta_0)^{-1}\Delta_{k,n}(\vartheta_0,X^k)\right|>\frac{\nu}{2}\right\}.$$

Moreover, more detailed analysis allows to verify the uniform consistency as well:

$$\mathbf{P}_{\vartheta_0}^{(k,n)}\left\{\max_{N+1\leq k\leq n}\left|\vartheta^\star_{k,n}-\vartheta_0\right|>\nu\right\}\longrightarrow 0.$$

See the similar problem in [10], Theorem 1. The asymptotic conditional normality as well follows from the similar steps. We have

$$\begin{aligned}\delta^{-1/2}\left(\vartheta^\star_{k,n}-\vartheta_0\right)&=\delta^{-1/2}\left(\hat\vartheta_{\tau,n}-\vartheta_0\right)+\mathbb{I}_{k,n}(\hat\vartheta_{\tau,n})^{-1}\Delta_{k,n}(\hat\vartheta_{\tau,n},X^k)\\&=\delta^{-1/2}\left(\hat\vartheta_{\tau,n}-\vartheta_0\right)+\mathbb{I}_{k,n}(\hat\vartheta_{\tau,n})^{-1}\Delta_{k,n}(\vartheta_0,X^k)\\&\quad+\mathbb{I}_{k,n}(\hat\vartheta_{\tau,n})^{-1}\left[\Delta_{k,n}(\hat\vartheta_{\tau,n},X^k)-\Delta_{k,n}(\vartheta_0,X^k)\right].\end{aligned}$$

The central statistics

$$\mathbb{I}_{k,n}(\hat\vartheta_{\tau,n})^{-1}\Delta_{k,n}(\vartheta_0,X^k)\Longrightarrow\xi_t\left(\vartheta_0\right).$$

We have to show that

$$b_n=\delta^{-1/2}\left(\hat\vartheta_{\tau,n}-\vartheta_0\right)+\mathbb{I}_{k,n}(\hat\vartheta_{\tau,n})^{-1}\left[\Delta_{k,n}(\hat\vartheta_{\tau,n},X^k)-\Delta_{k,n}(\vartheta_0,X^k)\right]\longrightarrow 0.$$

The representation

$$\begin{aligned}&\Delta_{k,n}(\hat\vartheta_{\tau,n},X^k)-\Delta_{k,n}(\vartheta_0,X^k)\\&\qquad=\sum_{j=1}^{k}\int_0^1\dot\ell\left(\vartheta_0+v\left(\hat\vartheta_{\tau,n}-\vartheta_0\right),X_{t_{j-1}},X_{t_j}\right)\left(\hat\vartheta_{\tau,n}-\vartheta_0\right){\rm d}v\end{aligned}$$

allows us to write

$$\mathbb{I}_{k,n}(\hat\vartheta_{\tau,n})b_n$$
$$=\left[\mathbb{I}_{k,n}(\hat\vartheta_{\tau,n})+\sum_{j=1}^{k}\int_0^1\ddot\ell\left(\vartheta\left(v\right),X_{t_{j-1}},X_{t_j}\right){\rm d}v\,\sqrt{\delta}\right]\frac{\left(\hat\vartheta_{\tau,n}-\vartheta_0\right)}{\sqrt{\delta}},$$

where $\vartheta\left(v\right)=\vartheta_0+v\left(\hat\vartheta_{\tau,n}-\vartheta_0\right)$. Further

$$\sum_{j=1}^{k}\ddot\ell\left(\vartheta\left(v\right),X_{t_{j-1}},X_{t_j}\right)=\sum_{j=1}^{k}\ddot\ell\left(\vartheta_0,X_{t_{j-1}},X_{t_j}\right)+O\left(\hat\vartheta_{\tau,n}-\vartheta_0\right)$$
$$=-\sum_{j=1}^{k}\frac{\dot A_{j-1}\left(\vartheta_0\right)\dot A_{j-1}\left(\vartheta_0\right)^{\mathbb{T}}}{2A_{j-1}\left(\vartheta_0\right)^2}\sqrt{\delta}+o\left(1\right)$$

because in two other terms after the differentiation

$$\dot\ell\left(\vartheta,X_{t_{j-1}},X_{t_j}\right)=\frac{\left[\left(X_{t_j}-X_{t_{j-1}}-S_{j-1}\,\delta\right)^2-A_{j-1}\left(\theta\right)\delta\right]\dot A_{j-1}\left(\vartheta\right)}{2A_{j-1}\left(\vartheta\right)^2\sqrt{\delta}}$$

contains the quantity

$$\left[X_{t_j}-X_{t_{j-1}}-S_{j-1}\delta\right]^2-A_{j-1}\left(\vartheta_0\right)\delta=-\sigma\left(t_{j-1},X_{t_{j-1}},\vartheta_0\right)^2\delta$$
$$+\left(\int_{t_{j-1}}^{t_j}\left[S\left(s,X_s\right)-S\left(t_{j-1},X_{t_{j-1}}\right)\right]{\rm d}s+\int_{t_{j-1}}^{t_j}\sigma\left(s,X_s,\vartheta_0\right){\rm d}W_s\right)^2$$
$$=2\int_{t_{j-1}}^{t_j}\tilde X_s\left[S\left(s,X_s\right)-S\left(t_{j-1},X_{t_{j-1}}\right)\right]{\rm d}s+2\int_{t_{j-1}}^{t_j}\tilde X_s\sigma\left(s,X_s,\vartheta_0\right){\rm d}W_s$$
$$+\int_{t_{j-1}}^{t_j}\left[\sigma\left(s,X_s,\vartheta_0\right)^2-\sigma\left(t_{j-1},X_{t_{j-1}},\vartheta_0\right)^2\right]{\rm d}s=O\left(\delta^2\right)+O\left(\delta\right).$$

Here $\tilde X_s=X_s-X_{t_{j-1}}-S_{j-1}s$. Hence

$$\tilde X_s=\int_{t_{j-1}}^{s}\left[S\left(r,X_r\right)-S\left(t_{j-1},X_{t_{j-1}}\right)\right]{\rm d}r+\int_{t_{j-1}}^{s}\sigma\left(r,X_r,\vartheta_0\right){\rm d}W_r$$
$$=O\left(\delta^{3/2}\right)+\sigma\left(t_{j-1},X_{t_{j-1}},\vartheta_0\right)\left[W_s-W_{t_{j-1}}\right]+O\left(\delta\right).$$

Note that for the stochastic integral as $n\to\infty$ we have

$$\int_{t_{j-1}}^{t_j}\tilde X_s\sigma\left(s,X_s,\vartheta_0\right){\rm d}W_s$$
$$=\sigma\left(t_{j-1},X_{t_{j-1}},\vartheta_0\right)^2\int_{t_{j-1}}^{t_j}\left[W_s-W_{t_{j-1}}\right]{\rm d}W_s\left(1+o\left(1\right)\right)$$
$$=\sigma\left(t_{j-1},X_{t_{j-1}},\vartheta_0\right)^2\left[\frac{\left(W_{t_j}-W_{t_{j-1}}\right)^2-\delta}{2}\right].$$

Therefore

$$\begin{aligned}
&\mathbb{I}_{k,n}(\hat\vartheta_{\tau,n})+\sum_{j=1}^{k}\int_0^1\ddot\ell\left(\vartheta\left(v\right),X_{t_{j-1}},X_{t_j}\right){\rm d}v\,\sqrt{\delta}\\
&\quad=\mathbb{I}_{k,n}(\hat\vartheta_{\tau,n})-\sum_{j=1}^{k}\frac{\dot A_{j-1}\left(\vartheta_0\right)\dot A_{j-1}\left(\vartheta_0\right)^{\mathbb{T}}}{2A_{j-1}\left(\vartheta_0\right)^2}\delta+o\left(1\right)\\
&\quad=\mathbb{I}_{k,n}(\hat\vartheta_{\tau,n})-\mathbb{I}_{k,n}(\vartheta_0)+o\left(1\right)\longrightarrow 0.
\end{aligned}$$

The obtained relations prove the convergence (13). More detailed analysis shows that this convergence is locally uniform. Hence the one-step MLE-process is asymptotically efficient for the bounded loss functions.

Introduce for the values $t_k\in[\tau,T]$ the estimators

$$Y^\star_{t_k,n}=u(t_k,X_{t_k},\vartheta^\star_{k,n}),\qquad Z^\star_{t_k,n}=u'_x(t_k,X_{t_k},\vartheta^\star_{k,n})\sigma(t_k,X_{t_k},\vartheta^\star_{k,n}).$$

Theorem 3 *Suppose that the conditions of regularity hold, then the estimators* $\left(Y^\star_{t,n},t\in[\tau,T]\right)$ *and* $\left(Z^\star_{t,n},t\in[\tau,T]\right)$ *are consistent*

$$Y^\star_{t_k,n}\longrightarrow Y_t,\qquad Z^\star_{t_k,n}\longrightarrow Z_t,$$

and are asymptotically conditionally normal (stable convergence)

$$\begin{aligned}
&\delta^{-1/2}\left(Y^\star_{t_k,n}-Y_{t_k}\right)\Longrightarrow\langle\dot u\left(t,X_t,\vartheta_0\right),\ \xi_t\left(\vartheta_0\right)\rangle,\\
&\delta^{-1/2}\left(Z^\star_{t_k,n}-Z_{t_k}\right)\Longrightarrow\sigma\left(t,X_t,\vartheta_0\right)\ \langle\dot u'_x\left(t,X_t,\vartheta_0\right),\ \xi_t\left(\vartheta_0\right)\rangle\\
&\qquad\qquad\qquad\qquad+u'_x\left(t,X_t,\vartheta_0\right)\langle\dot\sigma\left(t,X_t,\vartheta_0\right),\ \xi_t\left(\vartheta_0\right)\rangle.
\end{aligned}$$

Proof. Let us denote $v^\star_{k,n}=\delta^{-1/2}\left(\vartheta^\star_{k,n}-\vartheta_0\right)$ and write the Taylor expansion

$$\begin{aligned}
Y^\star_{t_k,n}&=u(t_k,X_{t_k},\vartheta_0+\delta^{1/2}v^\star_{k,n})=u(t_k,X_{t_k},\vartheta_0)\\
&\quad+\delta^{1/2}\langle v^\star_{k,n},\dot u(t_k,X_{t_k},\vartheta_0)\rangle+o\left(\delta^{-1/2}\right),\\
Z^\star_{t_k,n}&=u'_x(t_k,X_{t_k},\vartheta_0+\delta^{1/2}v^\star_{k,n})\sigma(t_k,X_{t_k},\vartheta_0+\delta^{1/2}v^\star_{k,n})\\
&=\sigma\left(t_k,X_{t_k},\vartheta_0\right)\ \langle\dot u'_x\left(t_k,X_{t_k},\vartheta_0\right),\ \xi_{t_k}\left(\vartheta_0\right)\rangle\\
&\quad+u'_x\left(t_k,X_{t_k},\vartheta_0\right)\langle\dot\sigma\left(t_k,X_{t_k},\vartheta_0\right),\ \xi_{t_k}\left(\vartheta_0\right)\rangle+o\left(\delta^{-1/2}\right).
\end{aligned}$$

Now the proof follows from the Theorem 2 and the regularity of the functions $u\left(\cdot\right)$, $u'_x\left(\cdot\right)$ and $\sigma\left(\cdot\right)$.

Remark. Note that we do not evaluate the difference $\delta^{-1/2}(Y^\star_{t,n}-Y_t)$ for $t\in[t_k,t_{k+1})$ because in the representation

$$\delta^{-1/2}(Y^\star_{t,n}-Y_t)=\delta^{-1/2}(Y^\star_{t,n}-Y_{t_k})+\delta^{-1/2}(Y_{t_k}-Y_t)$$

for the second term we have the relation

$$
\begin{aligned}
\delta^{-1/2}(Y_t - Y_{t_k}) &= \delta^{-1/2}u'_x(t,\tilde X_{t_k},\vartheta_0)\,(X_t - X_{t_k})\\
&= u'_x(t,X_{t_k},\vartheta_0)\sigma\left(\vartheta_0,t_k,X_{t_k}\right)\frac{(W_t - W_{t_k})}{\sqrt{t_{k+1}-t_k}}\left(1+o\left(1\right)\right)\\
&= u'_x(t,X_{t_k},\vartheta_0)\sigma\left(\vartheta_0,t_k,X_{t_k}\right)\zeta_t\,\sqrt{v_n\left(t\right)}\left(1+o\left(1\right)\right).
\end{aligned}
$$

Here $\zeta_t\sim\mathcal{N}\left(0,1\right)$ and $v_n\left(t\right)=\left(t_{k+1}-t_k\right)^{-1}\left(t-t_k\right)$. The study of the limit of $v_n\left(t\right)$ for all $t\in[0,T]$ is a special interesting problem.

To prove the optimality of the presented approximations $Y^\star_{t_k,n}$ and $Z^\star_{t_k,n}$ we need the lower bound of Jeganathan type given in the following proposition. Below

$$
\eta\left(t,X_t,\vartheta_0\right)=\left\langle \dot u'_x\left(t,X_t,\vartheta_0\right),\ \xi_t\left(\vartheta_0\right)\right\rangle + u'_x\left(t,X_t,\vartheta_0\right)\left\langle\dot\sigma\left(t,X_t,\vartheta_0\right),\ \xi_t\left(\vartheta_0\right)\right\rangle
$$

Proposition 1 *Let the conditions of regularity be fulfilled. Then for all estimators $\bar Y_{t_k,n}$ and $\bar Z_{t_k,n}$ we have*

$$
\begin{aligned}
&\varliminf_{\varepsilon\to0}\varliminf_{n\to\infty}\sup_{|\vartheta-\vartheta_0|\le\varepsilon}\mathbf{E}_\vartheta\ell\left(\delta^{-1/2}\left(\bar Y_{t_k,n}-Y_{t_k}\right)\right)\geq \mathbf{E}_{\vartheta_0}\ell\left(\left\langle\dot u\left(t,X_t,\vartheta_0\right),\ \xi_t\left(\vartheta_0\right)\right\rangle\right),\\
&\varliminf_{\varepsilon\to0}\varliminf_{n\to\infty}\sup_{|\vartheta-\vartheta_0|\le\varepsilon}\mathbf{E}_\vartheta\ell\left(\delta^{-1/2}\left(\bar Z_{t_k,n}-Z_{t_k}\right)\right)\geq \mathbf{E}_{\vartheta_0}\ell\left(\eta\left(t,X_t,\vartheta_0\right)\right).
\end{aligned}
$$

The proof of this proposition is similar to the proof of the lower bound in the problem of approximation of the solution of BSDE in the asymptotics of small noise given in the works [11] and [12].

Example. Black-Scholes model. We are given the forward equation

$$
\mathrm{d}X_t=\alpha X_t\mathrm{d}t+\vartheta X_t\,\mathrm{d}W_t,\quad X_0=x_0,\quad 0\le t\le T
$$

and two functions $f\left(x,y,z\right)=\beta y+\gamma xz$ and $\Phi\left(x\right)$. We have to approximate the solution of the backward equation

$$
\mathrm{d}Y_t=-\beta Y_t\mathrm{d}t-\gamma X_tZ_t\mathrm{d}t+Z_t\mathrm{d}W_t,\quad Y_T=\Phi\left(X_T\right)
$$

in the situation where $\vartheta\in\left(a,b\right), a>0$ and is unknown.

The corresponding partial differential equation is

$$
\frac{\partial u}{\partial t}+\left(\alpha+\vartheta\gamma\right)x\frac{\partial u}{\partial x}+\frac{\vartheta^2x^2}{2}\frac{\partial^2u}{\partial x^2}+\beta u=0,\qquad u\left(T,x,\vartheta\right)=\Phi\left(x\right).
$$

The solution of this equation is the function

$$
u\left(t,x,\vartheta\right)=\frac{e^{\beta(T-t)}}{\sqrt{2\pi\vartheta^2\left(T-t\right)}}\int_{-\infty}^{\infty}e^{-\frac{z^2}{2\vartheta^2(T-t)}}\,\Phi\left(xe^{\left(\alpha+\vartheta\gamma-\frac{\vartheta^2}{2}\right)(T-t)-z}\right)\mathrm{d}z.
$$

The discrete-time observations are $X^n = (X_0, X_{t_1} \ldots, X_{t_n})$. We can calculate the pseudo MLE-process

$$\hat\vartheta_{t_k,n} = \left(\frac{1}{t_k}\sum_{j=0}^{k-1}\frac{(X_{t_{j+1}} - X_{t_j} - \alpha X_{t_j}\delta)^2}{X_{t_j}^2}\right)^{\frac{1}{2}}.$$

The estimator of $Y_t = u\,(t, X_t, \vartheta_0)$ is

$$\hat Y_{t_k} = \int_{-\infty}^{\infty}\frac{e^{-\frac{z^2}{2\hat\vartheta_{t_k,n}^2(T-t_k)}+\beta(T-t_k)}}{\sqrt{2\pi\hat\vartheta_{t_k,n}^2\,(T-t_k)}}\Phi\left(X_{t_k}e^{(\alpha+\hat\vartheta_{t_k,n}\gamma-\frac{\hat\vartheta_{t_k,n}^2}{2})(T-t_k)-z}\right)\mathrm{d}z,$$

where $k = \left[\frac{t}{T}n\right]$.

Approximation of $\hat Z_t = \vartheta_0 X_t u'\,(t, X_t, \vartheta_0)$ can be written explicitly as well.

The one-step MLE-process is constructed as follows. Let us fix a learning interval $[0,\tau]\,, \tau < T$ and introduce the estimator $\hat\vartheta_{\tau,n}$ constructed by the observations $X^N = (X_0, X_{t_1} \ldots, X_{t_N})$, where $N = \left[\frac{\tau}{T}n\right]$ is as preliminary. Then we have

$$\vartheta_{k,n}^\star = \hat\vartheta_{\tau,n} + \frac{1}{2\hat\vartheta_{\tau,n}}\sum_{j=1}^{k}\left[\left(X_j - X_{t_{j-1}} - \alpha X_{t_{j-1}}\delta\right)^2 - \hat\vartheta_{\tau,n}^2 X_{t_{j-1}}^2\delta\right].$$

The corresponding approximations are

$$\hat Y_{t_k}^\star = u\left(t_k, X_{t_k}, \vartheta_{k,n}^\star\right), \qquad Z_t^\star = \vartheta_{k,n}^\star X_t u'\left(t, X_t, \vartheta_{k,n}^\star\right), \qquad N+1 \le k \le n$$

and by Theorem 3 and by Proposition 1 these approximations are asymptotically conditionally normal and asymptotically efficient.

5 Discussions

The approximation of the solution of the BSDE is done in several steps. First we estimate the unknown parameter on the learning interval $[0;\tau]$ and then using this estimator we constructed the one-step MLE process $\vartheta_{k,n}^\star, \tau \le t_k \le T$. Then we take the solution of partial differential equation $u\,(t,x,\vartheta)$ and put $Y_{t_k}^\star = u\left(t_k, X_{t_k}, \vartheta_{k,n}^\star\right), \tau \le t_k \le T$ and $Z_t^\star = \vartheta_{k,n}^\star X_t u'\left(t, X_t, \vartheta_{k,n}^\star\right), \tau \le t_k \le T$. The obtained approximation of (Y_t, Z_t) is valid for the values $t \in [\tau, T]$. This restriction is due to the condition that the preliminary estimator has to be consistent with good rate of convergence. Note that it is possible to obtain such approximations on the interval $[0, \tau_n]$, where $\tau_n \to 0$ with special rate. The preliminary estimator in such situation is $\sqrt{\frac{n}{N_n}}$-rate (which tends to ∞) consistent, but the construction of one-step (and more generally multi-step) MLE-processes allow nevertheless to construct the asymptotically efficient estimator-processes with the good rate. Such estimators were already studied in the works [10] (ergodic diffusion process) and [12] (dynamical systems with small noise).

Acknowledgment

This work was done under partial financial support of the grant of RSF number 14-49-00079.

References

[1] Abakirova, A. and Kutoyants, Yu.A. (2015) On approximation of the backward stochastic differential equation. Large samples approach. Working paper.

[2] Bismut, J.M. (1973) Conjugate convex functions in optimal stochastic control. *J. Math. Anal. Appl.*, 44, 384-404.

[3] Dohnal, G. (1987) On estimating the diffusion coefficient. *J. Appl. Probab.*, 24, 1, 105-114.

[4] Genon-Catalot, V. and Jacod, J. (1993) On the estimation of diffusion coefficient for multi-dimensional diffusion. *Ann. IHP*, Sec. B, 29, 1, 119-151.

[5] Gobet, E. (2001) Local asymptotic mixed normality property for elliptic diffusion: a Malliavin calculus approach. Bernoulli 7(6), 899-912.

[6] Jeganathan, P. (1983) 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.

[7] El Karoui N., Peng S. and Quenez M. (1997) Backward stochastic differential equations in finance. *Math. Fin.*, 7, 1-71.

[8] Kamatani, K. and Uchida, M. (2014) Hybrid multi-step estimators for stochastic differential equations based on sampled data. To appear in *Statist. Inference Stoch. Processes*

[9] Kutoyants, Yu.A. (2014) Approximation of the backward stochastic differential equation. Small noise, large samples and high frequency cases. *Proceedings of the Steklov Institute of Mathematics.*, 287, 133-154.

[10] Kutoyants, Yu.A. (2015) On multi-step MLE-processes for ergodic diffusion. submitted.

[11] Kutoyants, Yu.A. and Zhou, L. (2014) On approximation of the backward stochastic differential equation. *J. Stat. Plann. Infer.* 150, 111-123.

[12] Kutoyants, Yu.A. and Zhou, L. (2014) On estimation of the solution of BSDE in the case of small volatility. Submitted.

[13] Levanony, D., Shwartz, A. and Zeitouni, O. (1994) Recursive identification in continuous-time stochastic process. *Stochastic Process. Appl.*, 49, 245-275.

[14] Liptser, R. and Shiryaev, A.N. (2005) *Statistics of Random Processes.* v. 2, 2-nd ed. Springer, N.Y.

[15] Ma, J. and Yong, J. (1999) *Forward-Backward Stochastic Differential Equations and their Applications.* Lecture Notes in Mathematics. Springer, Berlin.

[16] Pardoux, E. and Peng, S. (1990) Adapted solution of a backward stochastic differential equation. *System Control Letter*, 14, 55-61.

[17] Pardoux, E. and Peng, S. (1992) Backward stochastic differential equation and quasilinear parabolic differential equations. In *Stochastic Partial Differential Equations and Their Applications,* Lecture Notes in Control and Information Sciences, 176, 200-217.

[18] Pardoux, E. and Răscanu, S. (2014) *Stochastic Differential Equations, Backward SDEs, Partial Differential Equations.* Springer, N.Y.

[19] Sørensen, M. (2009) Estimating functions for diffusion-type processes. in *Statistical Methods for Stochastic Differential Equations*, Kessler, Lindner and Sørensen (Ed's), Chapman & Hall/CRC, Boca Raton, 1-107.

[20] Wong, E. (1971) Representation of martingales, quadratic variation and applications. *SIAM Journal on Control*, 9, 4, 621-633.

Samvel B. Gasparyan
Université du Maine
Avenue Olivier Messiaen - 72085 Le Mans Cedex 9, France

Yerevan State University
Alex Manoogian, 0025 Yerevan, Armenia
samvel.gasparyan.etu@univ-lemans.fr

Yury A. Kutoyants
Université du Maine
Avenue Olivier Messiaen - 72085 Le Mans Cedex 9, France

National Research University "MPEI"
14 Krasnokazarmennaya, 111250 Moscow, Russia kutoyants@univ-lemans.fr

Please, cite to this paper as published in
Armen. J. Math., V. **7**, N. 1(2015), pp. 59–79