ARTICLE IN PRESS
JID: ECOSTA [m3Gsc;January 27, 2023;14:45]

Econometrics and Statistics xxx (xxxx) xxx

Contents lists available at ScienceDirect

Econometrics and Statistics

journal homepage: www.elsevier.com/locate/ecosta

# Approximation of BSDE with hidden forward equation and unknown volatility

Oleg V. Chernoyarov[a,c], Yury A. Kutoyants[b,c,*]

[a] National Research University "MPEI", Krasnokazarmennaya 14, Moscow, 111250, Russia
[b] Le Mans University, av. O. Messiaen, Le Mans, 72085, France
[c] Tomsk State University, av. Lenin 36, Tomsk, 634050, Russia

ARTICLE INFO

*Article history:*
Received 2 June 2021
Revised 8 January 2023
Accepted 11 January 2023
Available online xxx

*MSC:*
62P05
62M05

*Keywords:*
BSDE
solution approximation
perturbed dynamical systems
Kalman filtration
volatility estimation

ABSTRACT

The focus is on the approximation of the solution of BSDE in the case where the solution of forward equation is observed in the presence of small Gaussian noise. The volatility of the forward equation is considered to depend on some unknown parameter. This approximation is made in several steps. First a preliminary estimator of the unknown volatility is obtained, then using Kalman-Bucy filtration equations and Fisher-score device one-step MLE-process of this parameter is constructed. The solution of BSDE is approximated by means of the solution of PDE and the One-step MLE-process. The error of approximation is described in different metrics.



## 1. Introduction

The backward stochastic differential equation (BSDE) was first introduced in the linear case by Bismuth Bismut (1973). The general (nonlinear) case was initiated by Pardoux and Peng Pardoux and Peng (1990). The theory of BSDE has been extensively developed due to its importance in financial mathematics (see, e.g. El Karoui et al. El Karoui et al. (1997), Ma and Yong Ma and Yong (1999) and the references therein).

For example, BSDE can be used to describe risk-based optimal investment problem of insurer Elliott and Siu (2011), where a simplified continuous-time economy with two investment vehicles, namely, a fixed interest security and a share, is considered. The BSDE approach is used to solve the game problem of two players: the insurer and the market. In the work Elliott and Siu (2012) a BSDE approach is used to evaluate convex risk measures for unhedged positions of derivative securities in a continuous-time economy. The convex risk measure is represented as the solution of a BSDE. BSDE is extensively used in the problems of continuous time economies Eberlein et al. (2014). All these are just several examples.

Despite the detailed and well developed probabilistic description of BSDE and its wide applications, the statistical study of BSDE is relatively limited. This work is the continuation of the study of statistical approximation of a solution of BSDE initiated in the work Kutoyants and Zhou (2014) and then developed in Gasparyan and Kutoyants (2015) and Kutoyants (2014).

* Corresponding author.
*E-mail address:* kutoyants@univ-lemans.fr (Y.A. Kutoyants).

https://doi.org/10.1016/j.ecosta.2023.01.002

In all these works the forward equation was supposed to depend on some unknown finite-dimensional parameter and could be directly observed. In the present work we propose a new statistical problem by considering a partially observable linear model, where the forward equation is observed in small white Gaussian noise and its volatility depends on an unknown one-dimensional parameter. We propose an estimator-process for this parameter, which is then used for the construction of the approximation of the solution of BSDE. The properties of the estimator and approximation are described in the asymptotic of small noise. Another nonparametric approach was considered in the work Su and Lin (2009). We refer the interested reader to Javaheri (2015) and Wells (2010), where the problem of volatility estimation with the help of Kalman filter and other applications of Kalman filtering in financial mathematics are considered respectively.

Let us recall the construction of the BSDE in the Markovian case with observable forward equation following El Karoui et al. (1997). For simplicity of exposition we consider only the one-dimensional processes. 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*)

$$\mathrm{d}X_t = S(t, X_t)\,\mathrm{d}t + \sigma(t, X_t)\,\mathrm{d}W_t, \quad X_0,\ \ 0 \le t \le T, \tag{1}$$

where $W_t, \mathcal{F}_t, 0 \le t \le T$ is standard Wiener process, $X_0$ 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: for all $t \in [0,T]$

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

$$|S(t,x)| + |\sigma(t,x)| \le C(1 + |x|), \tag{3}$$

where $L > 0$ and $C > 0$ are some constants. By these conditions the stochastic differential equation has a unique strong solution (see Liptser and Shiryaev Liptser and Shiryaev (2005)).

The main problem is the following: *given two functions* $\Phi(x)$ *and* $F(t,x,z,\sigma)$, *we have to construct two processes* $Z_t, \Sigma_t, \mathcal{F}_t, 0 \le t \le T$ *such that the solution of the stochastic differential equation*

$$\mathrm{d}Z_t = -F(t, X_t, Z_t, \Sigma_t)\,\mathrm{d}t + \Sigma_t\,\mathrm{d}W_t, \quad 0 \le t \le T, \tag{4}$$

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

This equation is often written in integral form as follows

$$Z_t = \Phi(X_T) + \int_t^T F(s, X_s, Z_s, \Sigma_s)\,\mathrm{d}s - \int_t^T \Sigma_s\,\mathrm{d}W_s, \quad 0 \le t \le T.$$

We suppose that the functions $F(t,x,y,z)$ and $\Phi(x)$ for all $t \in [0,T]$ satisfy the conditions

$$|F(t,x,z_1,\sigma_1) - F(t,x,z_2,\sigma_2)| \le L(|z_1 - z_2| + |\sigma_1 - \sigma_2|), \tag{5}$$

$$|F(t,x,z,\sigma)| + |\Phi(x)| \le C\big(1 + |x|^p\big), \tag{6}$$

where $p \ge 1/2$.

This is the so-called *Markovian case*. For the existence and uniqueness of the solution see Pardoux and Peng Pardoux and Peng (1992).

The processes $Z_t, \Sigma_t, \mathcal{F}_t, 0 \le t \le T$ can be constructed as follows. Suppose that $u(t,x)$ is the solution of the partial differential 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 set $Z_t = u(t, X_t)$, then we obtain by Itô's formula

$$\mathrm{d}Z_t = \left[\frac{\partial u(t,X_t)}{\partial t} + S(t,X_t)\frac{\partial u(t,X_t)}{\partial x} + \frac{\sigma(t,X_t)^2}{2}\frac{\partial^2 u(t,X_t)}{\partial x^2}\right]\mathrm{d}t$$
$$+\,\sigma(t,X_t)\frac{\partial u(t,X_t)}{\partial x}\,\mathrm{d}W_t, \qquad Z_0 = u(0, X_0).$$

In the sequel, we use the notation

$$\frac{\partial u(t,X_t)}{\partial x} = u'_x(t,X_t) = \left.\frac{\partial u(t,x)}{\partial x}\right|_{x=X_t}.$$

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

$$\mathrm{d}Z_t = -F(t, X_t, Z_t, \Sigma_t)\mathrm{d}t + \Sigma_t\,\mathrm{d}W_t, \qquad Z_0 = u(0, X_0).$$

The terminal value $Z_T = u(T, X_T) = \Phi(X_T)$. Therefore the problem is solved and the Equation (4) with the given terminal value is obtained.

We are interested in the problem of the approximation of the solution $Z_t, \Sigma_t, \mathcal{F}_t, 0 \le t \le T$ of BSDE in the cases where the forward Equation (1) contains a finite-dimensional parameter $\vartheta$:

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

Then the solution $u$ of the corresponding partial differential equation depends on $\vartheta$. The "natural" approximations $\hat Z_t, \hat \Sigma_t, \mathcal{F}_t, 0 \le t \le T$ can be constructed as follows. Suppose that $u(t, x, \vartheta)$ is a solution of the partial differential equation

$$\frac{\partial u}{\partial t} + S(\vartheta, t, x)\frac{\partial u}{\partial x} + \frac{\sigma(\vartheta, t, x)^2}{2}\frac{\partial^2 u}{\partial x^2} = -F\left(t, x, u, \sigma(t, x)\frac{\partial u}{\partial x}\right), \tag{7}$$

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

We are interested in the case where the true value of this parameter denoted $\vartheta = \vartheta_0$ is unknown. It is worth mentionning that we cannot set $Z_t = u(t, X_t, \vartheta_0)$ since $\vartheta_0$ is unknown. One way to obtain approximation $\hat Z_t, \hat \Sigma_t$ of $Z_t, \Sigma_t$ is to find first an estimator-process $\vartheta_t^*, 0 < t \le T$ of $\vartheta_0$ and then to set

$$\hat Z_t = u(t, X_t, \vartheta_t^*), \qquad \hat \Sigma_t = u'_x(t, X_t, \vartheta_t^*)\, \sigma(\vartheta_t^*, t, X_t).$$

If the estimator has good properties, say, $\vartheta_t^* - \vartheta$ is small in some sense, then the error

$$\hat Z_t - Z_t = u(t, X_t, \vartheta_t^*) - u(t, X_t, \vartheta_0) \approx \frac{\partial u(t, X_t, \vartheta_0)}{\partial \vartheta_0}(\vartheta_t^* - \vartheta_0)$$

will be small as well.

Here $\vartheta_t^*, 0 \le t \le T$ is some *good estimator-process* of $\vartheta_0$ in the sense that

- *The estimator $\vartheta_t^*$ depends on $X^t = (X_s, 0 \le s \le t)$.*
- *It can be easily calculated for each $t \in (0, T]$.*
- *Provides the following asymptotically efficient estimation of $Z_t$*

$$\mathbf{E}_{\vartheta_0}\left(\hat Z_t - Z_t\right)^2 \to \min.$$

Therefore the main problem is *how to find a good estimator-process*. Such problems were studied in the works Gasparyan and Kutoyants (2015); Kutoyants (2014); Kutoyants and Zhou (2014).

The solutions of the different problems there were studied following the same general program, which is illustrated as follows. Consider the forward equation with small volatility: $\sigma(t, X_t, \vartheta) = \varepsilon\sigma(t, X_t)$, where $\sigma(t, x)$ satisfies the conditions (2), (3), $\varepsilon \in (0, 1]$ is a small parameter and we consider asymptotics $\varepsilon \to 0$. Introduce a learning interval $[0, \tau_\varepsilon]$, where $\tau_\varepsilon \to 0$ but *slowly*. A preliminary consistent estimator $\vartheta^*_{\tau_\varepsilon}$ of $\vartheta$ is constructed by means of the observations $X^{\tau_\varepsilon} = (X_t, 0 \le t \le \tau_\varepsilon)$. Then with the help of slightly modified Fisher-score device this estimator is improved up to the asymptotically ($\varepsilon \to 0$) efficient One-step MLE-process $\vartheta^\star_{t,\varepsilon}, \tau_\varepsilon, < t \le T$. Now the approximation of $(Z_t, \Sigma_t)$ is given, by the relations

$$\hat Z_t = u\left(t, X_t, \vartheta^\star_{t,\varepsilon}\right), \qquad \hat\Sigma_t = \varepsilon\sigma(t, X_t)u'_x\left(t, X_t, \vartheta^\star_{t,\varepsilon}\right)$$

It is shown that these approximations are asymptotically efficient. For the details see Kutoyants (2014); Kutoyants and Zhou (2014). In the work Gasparyan and Kutoyants (2015) it is supposed that the volatility $\sigma(t, X_t, \vartheta)$ depends on $\vartheta$ and the forward Equation (1) is observed in discrete times. Then a similar program of approximation is followed.

In particular, the proposed here One-step MLE-process for volatility estimation is similar to the adaptive recursive quasi-maximum likelihood estimators of volatility estimation constructed following stochastic approximation procedure and studied in Werge and Wintenberger (2022). The difference is in the structure of estimators. In our case the right hand part of (25) depends on just one value of unknown parameter given by the preliminary estimator. Therefore the calculation of One-step MLE-process requires the solution of two equations of filtration. The application of stochastic approximation algorithms in this problem of approximation will require solutions of such equations for each $t \in [0, T]$ which makes the numerical realization of such estimators difficult.

Note that a different approach of identification of BSDE equation was considered in Su and Lin (2009). There both processes $X_t, Y_t$ (similar to (9),(10) below) are supposed to be observable in discrete times and these observations are used for constructing a nonparametric estimator of the random process $\Sigma_t^2$.

The main goal of this work is to construct the approximation of the process $Z_t$ in the case where the forward equation is linear but cannot be directly observed. The volatility of the forward equation depends on unknown parameter $\vartheta_0$. The observations $X^T = (X_t, 0 \le t \le T)$ satisfy a linear differential equation where the trend coefficient depending on the solution $Y_t, 0 \le t \le T$ of forward equation. The approximation is realized in several steps. First we note that the optimal estimator of the solution $Y_t$ of forward equation is conditional expectation $m(\vartheta_0, t) = \mathbf{E}_{\vartheta_0}(Y_t | X_s, 0 < s \le t)$ satisfying Kalman-Bucy filtration equations. Then we replace the unobserved forward equation for $Y_t, 0 \le t \le T$ by the equation for $m(\vartheta_0, t), 0 \le t \le T$ and consider the problem of approximating the corresponding BSDE

$$dZ_t = -F(t, m(\vartheta_0, t), Z_t, s(t))dt + s(t)d\bar W_t, \quad Z_T = \Phi(m(\vartheta_0, T)). \tag{8}$$

To make this approximation we consider the PDE which corresponds to the construction of such BSDE. Since all mentioned equations depend on unknown parameter $\vartheta_0$, we propose the following One-step MLE-process estimation procedure. Introduce a learning interval $[0,\tau]$ and take as preliminary stimator the MLE $\hat\vartheta_{\tau,\varepsilon}$ or substitution estimator constructed by observations $X^\tau=(X_t,0\le t\le\tau)$ (Section 2.2). Then this estimator is used for construction of the One-step MLE-process $\vartheta^\star_{t,\varepsilon},\tau<t\le T$ (Section 2.3). The next step is the approximation $\hat m_t,\tau<t\le T$ of the process $m(\vartheta_0,t),\tau<t\le T$ by the solution of the Equation (34). The last step (Section 2.4) is the approximation

$$\hat Z_t=u\left(t,\hat m_t,\vartheta^\star_{t,\varepsilon},\varepsilon\right)$$

of the solution $Z_t$ of the BSDE (8). The error of this approximation is described in Theorem 1.

Our approach to hidden Markov models can be called *adaptive maximum likelihood*. Bayesian approach to such models was discussed in Liu and Song (2021).

## 2. Main result

### 2.1. Model of observations and auxiliary results

Suppose that the forward equation is

$$\mathrm{d}Y_t=-a(t)Y_t\mathrm{d}t+b(\vartheta,t)\mathrm{d}V_t,\quad Y_0=y_0,\quad 0\le t\le T, \tag{9}$$

whereas the solution $Y^T=(Y_t,0\le t\le T)$ of this equation can not be directly observed. The observations are given by the equation

$$\mathrm{d}X_t=f(t)Y_t\mathrm{d}t+\varepsilon\sigma(t)\mathrm{d}W_t,\qquad X_0=0,\quad 0\le t\le T \tag{10}$$

only. Here $V_t,0\le t\le T$ and $W_t,0\le t\le T$ are independent Wiener processes. The parameter $\vartheta\in\Theta=(\alpha,\beta)$, where $|\alpha|+|\beta|<\infty$. Here $a(\cdot),b(\cdot),\ f(\cdot)$ and $\sigma(\cdot)$ are known functions and $\varepsilon\in(0,1]$. These functions satisfy the following regularity conditions.

*Conditions $\mathcal{A}$.*

$\mathcal{A}_1$. *The functions $a(t),b(\vartheta,t),f(t)$ and $\sigma(t)$ have continuous derivatives w.r.t. $t\in[0,T]$.*

$\mathcal{A}_2$. *The functions $b(\vartheta,t),f(t)$ and $\sigma(t)$ are separated from zero by a constant, which does not depend neither on $\vartheta$ nor on $t$.*

Given two functions $F(t,y,u,s)$, $\Phi(y)$ and observations $X^T=(X_t,0\le t\le T)$, we have to construct the corresponding BSDE. It is worth noticing that the construction of the following BSDE is not feasible

$$\mathrm{d}\tilde Z_t=-F\left(t,Y_t,\tilde Z_t,s_t\right)\mathrm{d}t+s_t\mathrm{d}V_t,\quad \tilde Z_T=\Phi(Y_T),\quad 0\le t\le T \tag{11}$$

for two reasons: first we have no access to the process $Y^T$ (no Wiener process $V_t$) and even if $Y^T$ is available the solution $U=U(t,y,\vartheta)$ of the corresponding PDE

$$\frac{\partial U}{\partial t}-a(t)y\frac{\partial U}{\partial y}+\frac{b(\vartheta,t)^2}{2}\frac{\partial^2U}{\partial y^2}=-F\left(t,y,U,b(\vartheta,t)\frac{\partial U}{\partial y}\right),$$
$$U(T,y,\vartheta)=\Phi(y) \tag{12}$$

depends on the unknown parameter $\vartheta$. Therefore we can not set $\tilde Z_t=U(t,Y_t,\vartheta)$ since neither $Y_t$ nor $\vartheta$ are known.

As we have no solution $Y^T$ of the forward equation we change the statement of the problem and propose BSDE based on the best in the mean squared estimator of these process. Introduce the conditional expectation $\hat Y^T=(\hat Y_t,0\le t\le T)$, where $\hat Y_t=\mathbf{E}_\vartheta(Y_t|X_s,0\le s\le t)$. Now the corresponding BSDE can be written as follows

$$\mathrm{d}Z_t=-F(t,\hat Y_t,Z_t,s(t))\mathrm{d}t+s(t)\mathrm{d}\bar W_t,\qquad Z_T=\Phi(\hat Y_T),\quad 0\le t\le T. \tag{13}$$

The Wiener process $\bar W_t,0\le t\le T$ is described below. To construct the Equation (13) we need the equations of Kalman-Bucy filtration for $\hat Y_t$, which we remind here. In order to show the dependence on $\vartheta$, we will denote $\hat Y_t=m(\vartheta,t)$. The equation for $m(\vartheta,t)$ is (see Liptser and Shiryaev (2005))

$$\mathrm{d}m(\vartheta,t)=-\left[a(t)+\frac{\gamma(\vartheta,t)f(t)^2}{\varepsilon^2\sigma(t)^2}\right]m(\vartheta,t)\mathrm{d}t+\frac{\gamma(\vartheta,t)f(t)}{\varepsilon^2\sigma(t)^2}\mathrm{d}X_t. \tag{14}$$

Here $m(\vartheta,0)=0$ and $\gamma(\vartheta,t)=\mathbf{E}_\vartheta(Y_t-m(\vartheta,t))^2$ is the solution of Riccati equation

$$\frac{\partial\gamma(\vartheta,t)}{\partial t}=-2a(t)\gamma(\vartheta,t)-\frac{\gamma(\vartheta,t)^2f(t)^2}{\varepsilon^2\sigma(t)^2}+b(\vartheta,t)^2,\quad \gamma(\vartheta,0)=0. \tag{15}$$

We further introduce

$$\gamma_*(\vartheta,t)=\frac{\gamma(\vartheta,t)}{\varepsilon},\quad \gamma_0(\vartheta,t)=\frac{b(\vartheta,t)\sigma(t)}{f(t)},\quad A_\varepsilon(\vartheta,t)=\frac{\gamma_*(\vartheta,t)f(t)}{\sigma(t)^2},$$

$$A_0(\vartheta,t)=\frac{b(\vartheta,t)}{\sigma(t)},\qquad q_\varepsilon(\vartheta,t)=a(t)+\frac{A_\varepsilon(\vartheta,t)f(t)}{\varepsilon}.$$

The true value is denoted by $\vartheta_0$. The Equation (14) for $m(\vartheta_0,t)$ and Riccati Equation (15) can be re-written as follows

$$\mathrm{d}m(\vartheta_0,t)=-a(t)m(\vartheta_0,t)\mathrm{d}t+A_\varepsilon(\vartheta_0,t)\sigma(t)\mathrm{d}\bar W_t,$$
$$\frac{\partial\gamma_*(\vartheta_0,t)}{\partial\tau}=-2a(t)\gamma_*(\vartheta_0,t)-\frac{A_\varepsilon(\vartheta_0,t)^2\sigma(t)^2}{\varepsilon}+\frac{b(\vartheta_0,t)^2}{\varepsilon},$$

with initial values $m(\vartheta_0,0)=0$ and $\gamma_*(\vartheta_0,0)=0$ respectively. Here $\bar W_t,\mathcal{F}_t,\ 0\le t\le T$ is the *innovation* Wiener process defined by the relation

$$\mathrm{d}X_t=f(t)m(\vartheta_0,t)\mathrm{d}t+\varepsilon\sigma(t)\mathrm{d}\bar W_t,\qquad X_0=0$$

(see Liptser and Shiryaev (2005), Theorem 7.12).

**Lemma 1.** *Let the conditions $\mathcal{A}$ be fulfilled. Then for any $t_0\in(0,T]$ we have the convergence*

$$\sup_{t_0\le t\le T}|\gamma_*(\vartheta,t)-\gamma_0(\vartheta,t)|\longrightarrow 0,\qquad \sup_{t_0\le t\le T}|A_\varepsilon(\vartheta,t)-A_0(\vartheta,t)|\longrightarrow 0 \tag{16}$$

*as $\varepsilon\to 0$.*

**Proof.** See Lemma 2 in Kutoyants (2019). □

This lemma allows us to verify the following obvious result

**Lemma 2.** *Let the conditions $\mathcal{A}$ be fulfilled. Then for any $t_0\in(0,T]$ we have the convergence*

$$\sup_{t_0\le t\le T}\mathbf{E}_{\vartheta_0}|m(\vartheta_0,t)-Y_t|^2\le C\varepsilon\to 0 \tag{17}$$

*as $\varepsilon\to 0$.*

**Proof.** For the difference $\delta_t=m(\vartheta_0,t)-Y_t$ we have the equation

$$\begin{aligned}\mathrm{d}\delta_t&=-a(t)\delta_t\mathrm{d}t-b(\vartheta_0,t)\mathrm{d}V_t+A_\varepsilon(\vartheta_0,t)\sigma(t)\mathrm{d}\bar W_t\\&=-q_\varepsilon(\vartheta_0,t)\delta_t\mathrm{d}t-b(\vartheta_0,t)\mathrm{d}V_t+A_\varepsilon(\vartheta_0,t)\sigma(t)\mathrm{d}W_t\end{aligned}$$

where we denote

$$q_\varepsilon(\vartheta_0,t)=a(t)+\varepsilon^{-1}A_\varepsilon(\vartheta_0,t)f(t).$$

Hence

$$\begin{aligned}\delta_t&=-\int_0^t e^{-\int_s^t q_\varepsilon(\vartheta_0,v)\mathrm{d}v}b(\vartheta_0,s)\mathrm{d}V_s\\&\quad+\int_0^t e^{-\int_s^t q_\varepsilon(\vartheta_0,v)\mathrm{d}v}A_\varepsilon(\vartheta_0,s)\sigma(s)\mathrm{d}W_s\end{aligned} \tag{18}$$

and

$$\begin{aligned}\mathbf{E}_{\vartheta_0}|m(\vartheta_0,t)-Y_t|^2&=\int_0^t e^{-2\int_s^t q_\varepsilon(\vartheta_0,v)\mathrm{d}v}\left[b(\vartheta_0,s)^2+A_\varepsilon(\vartheta_0,s)^2\sigma(s)^2\right]\mathrm{d}s\\&\le C\int_0^t e^{-\frac{2}{\varepsilon}\int_s^t A_\varepsilon(\vartheta_0,v)f(0,v)\mathrm{d}v}\mathrm{d}s\le C\int_0^t e^{-\frac{c(t-s)}{\varepsilon}}\mathrm{d}s\le C\varepsilon\left[1-e^{-\frac{ct}{\varepsilon}}\right].\end{aligned}$$

Here the condition $\mathcal{A}_2$ and the boundedness of all functions were used. □

Therefore for small $\varepsilon$, the random process $m(\vartheta_0,t)$ is a good approximation of the solution $Y_t$ of the forward equation.

Note that if $\vartheta_0$ is known, then in order to construct (13) with innovation Wiener process the solution of partial differential equation is necessary

$$\begin{aligned}\frac{\partial u}{\partial t}-a(t)y\frac{\partial u}{\partial y}+\frac{B_\varepsilon(\vartheta_0,t)^2}{2}\frac{\partial^2 u}{\partial y^2}&=-F\left(t,y,u,B_\varepsilon(\vartheta_0,t)\frac{\partial u}{\partial y}\right),\\u(T,y,\vartheta_0,\varepsilon)&=\Phi(y),\end{aligned} \tag{19}$$

where $B_\varepsilon(\vartheta_0,t)=A_\varepsilon(\vartheta_0,t)\sigma(t)$. If this equation is solved, then

$$Z_t=u(t,m(\vartheta_0,t),\vartheta_0,\varepsilon),\ \ s(t)=A_\varepsilon(\vartheta_0,t)\sigma(t)\frac{\partial u}{\partial y}(t,m(\vartheta_0,t),\vartheta_0,\varepsilon)$$

form the Equation (13).

In the sequel we use the notation

$$\frac{\partial}{\partial t}u(t,y,\vartheta,\varepsilon)=u'_t(t,y,\vartheta,\varepsilon),\qquad \frac{\partial}{\partial y}u(t,y,\vartheta,\varepsilon)=u'_y(t,y,\vartheta,\varepsilon),$$
$$\frac{\partial}{\partial \vartheta}u(t,y,\vartheta,\varepsilon)=\dot u(t,y,\vartheta,\varepsilon),\qquad \frac{\partial}{\partial \varepsilon}u(t,y,\vartheta,\varepsilon)=u'_\varepsilon(t,y,\vartheta,\varepsilon).$$

Suppose that we have some estimator-process $\vartheta^*_{t,\varepsilon}$ which is consistent: for any $t\in(0,T]$ the estimator $\vartheta^*_{t,\varepsilon}\to\vartheta_0$. Let us set

$$\hat Z_t=u\left(t,\hat m_t,\vartheta^*_{t,\varepsilon},\varepsilon\right),\qquad \hat s(t)=A_\varepsilon\left(\vartheta^*_{t,\varepsilon},t\right)\sigma(t)u'\left(t,\hat m_t,\vartheta^*_{t,\varepsilon},\varepsilon\right),$$

where $\hat m_t$ is an approximation of $m(\vartheta_0,t)$. Then the relation between the solution $Z_t$ of the Equations (11), the solution $Z_t$ of the Equations (13) and the approximation $\hat Z_t$ is investigated.

As we have the convergences

$$\gamma_*(\vartheta,t)\longrightarrow\gamma_0(\vartheta,t),\qquad A_\varepsilon(\vartheta,t)\longrightarrow A_0(\vartheta,t),\qquad m(\vartheta_0,t)\longrightarrow Y_t,$$

the coefficient $B_\varepsilon(\vartheta,t)^2$ in the Equation (19) converges to $b(\vartheta,t)^2$ in the Equation (12). Hence under regularity conditions (5), (6) the solution $u(\cdot,\cdot,\cdot)$ of (19) converges to the solution $u(\cdot,\cdot,\cdot)$ of (12).

Note that no BSDE is available for the approximation process

$$\mathrm{d}\hat Z_t=-F(t,\hat Z_t,\hat m_t,\hat s(t))\mathrm{d}t+\hat s(t)\mathrm{d}\bar W_t,\qquad \hat Z_T=\Phi(\hat m_T).$$

The stochastic differential for the random process $\hat Z_t=u\left(t,\hat m_t,\vartheta^*_{t,\varepsilon},\varepsilon\right)$ can be written (it is different of given above), but it is quite cumbersome and is not used in the proofs. Our goal is to propose an approximation of the solution $Z_t$ of the Equation (11) and study the error of approximation, say, $\mathbf{E}_{\vartheta_0}(Z_t-\hat Z_t)^2$. Moreover, the optimality of such approximation is discussed.

### *2.2. Preliminary estimators*

Remind that to approximate the solution of BSDE, we need to define a *good estimator-process* $\vartheta_{t,\varepsilon},\tau\le t\le T$ and therefore to construct a preliminary estimator $\bar\vartheta_{\tau,\varepsilon}$ based on the first observations $X^\tau=(X_t,0\le t\le\tau)$ on the (small) time interval $[0,\tau]$ where $\tau\in(0,T]$. In this section we propose two such estimators. One is the MLE $\hat\vartheta_{\tau,\varepsilon}$ and the other is the estimator of substitution which uses the estimator of the quadratic variation of the derivative of limit of the observed process.

The likelihood ratio function is (see Liptser and Shiryaev (2005))

$$L(\vartheta,X^\tau)=\exp\left\{\int_0^\tau\frac{f(t)m(\vartheta,t)}{\varepsilon^2\sigma(t)^2}\mathrm{d}X_t-\int_0^\tau\frac{f(t)^2m(\vartheta,t)^2}{2\varepsilon^2\sigma(t)^2}\mathrm{d}t\right\},\ \vartheta\in\Theta,$$

and the corresponding maximum likelihood estimator (MLE) $\hat\vartheta_{\tau,\varepsilon}$ is defined by the relation

$$L(\hat\vartheta_{\tau,\varepsilon},X^\tau)=\sup_{\vartheta\in\Theta}L(\vartheta,X^\tau).\tag{20}$$

In the following we introduce the notation

$$\mathrm{I}^\tau(\vartheta)=\int_0^\tau\frac{f(t)\dot b(\vartheta,t)^2}{2b(\vartheta,t)\sigma(t)}\,\mathrm{d}t,\quad G_\tau(\vartheta,\vartheta_0)=\int_0^\tau\frac{f(t)[b(\vartheta,t)-b(\vartheta_0,t)]^2}{2b(\vartheta,t)\sigma(t)}\,\mathrm{d}t.$$

*Conditions $\mathcal{B}$.*

$\mathcal{B}_1$. *The function $b(\vartheta,t)$ has three continuous derivatives w.r.t. $\vartheta\in\Theta$.*

$\mathcal{B}_2$. *Identifiability condition: For any $\tau\in(0,T]$ and $\nu>0$*

$$\inf_{\vartheta_0\in\Theta}\inf_{|\vartheta-\vartheta_0|>\nu}G_\tau(\vartheta,\vartheta_0)>0.$$

$\mathcal{B}_3$. *Non degeneracy of Fisher information: For any $\tau\in(0,T]$*

$$\inf_{\vartheta\in\Theta}\mathrm{I}^\tau(\vartheta)>0.$$

Note that if $f(0)>0$ and $\inf_{\vartheta\in\Theta}\left|\dot b(\vartheta,0)b(\vartheta,0)^{-1}\right|>0$, then the condition $\mathcal{B}_3$ is fulfilled.

**Proposition 1.** *The MLE $\hat\vartheta_{\tau,\varepsilon}$ under regularity conditions $\mathcal{A}$, $\mathcal{B}$ is consistent, asymptotically normal*

$$\sqrt{\frac{\mathrm{I}^\tau(\vartheta_0)}{\varepsilon}}\left(\hat\vartheta_{\tau,\varepsilon}-\vartheta_0\right)\Longrightarrow\zeta\sim\mathcal{N}(0,1),\tag{21}$$

*asymptotically efficient and the moments converge: for any* $p > 0$

$$\left|\frac{\mathrm{I}^{\tau}(\vartheta_0)}{\varepsilon}\right|^{p/2}\mathbf{E}_{\vartheta_0}\left|\hat\vartheta_{\tau,\varepsilon}-\vartheta_0\right|^p\longrightarrow\mathbf{E}|\zeta|^p, \tag{22}$$

For the proof see Kutoyants (2019), Theorem 1.

Note that we can not use the MLE-process $\hat\vartheta_{t,\varepsilon}, 0 < t \le T$ as a good estimator-process since to solve Equation (20) for all $t \in (0,T]$ we need the solutions $m(\vartheta,s), 0 \le s \le t$ of Equations (14) for all $\vartheta \in \Theta$ and all $t \in (0,T]$. Since this problem is difficult to be solved numerically, a good estimator-process needs to be"easy to calculate". Then we can consider $\hat\vartheta_{\tau,\varepsilon}$ as preliminary estimator whose calculation is simpler since the Equation (20) needs to be solved just once.

Note that even the calculation of the preliminary MLE $\hat\vartheta_{\tau,\varepsilon}$ by (20) requires the solutions of the filtration equations for many values of $\vartheta$. Below we propose another estimator which requires much simpler calculations.

Note that the observed process $X^{\tau}$ converges with probability 1 to $x^{\tau}$:

$$\sup_{0\le t\le\tau}|X_t-x_t|\longrightarrow 0,$$

where $x^{\tau}=(x_t, 0\le t\le\tau)$ satisfies the limit $(\varepsilon=0)$ relation

$$x_t=\int_0^t f(s)Y_s\,\mathrm{d}s,\qquad 0\le t\le\tau.$$

Here $Y_t$ is the solution of the forward Equation (9). Let us set

$$\frac{\mathrm{d}x_t}{\mathrm{d}t}=N_t=f(t)Y_t.$$

Then for the quadratic variation of the process $N_\tau, 0\le\tau\le T$ we obtain

$$\lim_{\varphi\to0}\sum_{i=1}^{[\frac{\tau}{\varphi}]}\left[N_{i\varphi}-N_{(i-1)\varphi}\right]^2=\int_0^\tau f(t)^2b(\vartheta_0,t)^2\mathrm{d}t=\Psi_\tau(\vartheta_0),$$

where the last equality defines the function $\Psi_\tau(\vartheta), \vartheta\in\Theta$.

Therefore the substitution estimator $\check\vartheta_{\tau,\varepsilon}$ can be constructed as follows. Suppose that an estimator of the derivative $x'_t, 0\le t\le\tau$ of the limit of observed process $X_t, 0\le t\le\tau$ and the estimator $\Psi_{\tau,\varepsilon}$ of the quadratic variation of the estimator of the derivative are obtained. The substitution estimator $\check\vartheta_{\tau,\varepsilon}$ is defined as the solution of the equation

$$\Psi_{\tau,\varepsilon}=\Psi_\tau\left(\check\vartheta_{\tau,\varepsilon}\right).$$

This construction is realized via the statistic

$$\Psi_{\tau,\varepsilon}=\sum_{i=0}^{K_{\tau,\varepsilon}-1}\left(\frac{X_{t_{i+1}+\delta_\varepsilon}-X_{t_{i+1}}}{\delta_\varepsilon}-\frac{X_{t_i+\delta_\varepsilon}-X_{t_i}}{\delta_\varepsilon}\right)^2,\qquad 0<\tau\le T.$$

Here $t_i=i\varphi_\varepsilon, K_{\tau,\varepsilon}=\left[\frac{\tau}{\varphi_\varepsilon}\right]$, the rates $\varphi_\varepsilon\to0, \delta_\varepsilon\to0$ will be defined as follows. Note that as the first step is the derivation, the rate $\delta_\varepsilon\to0$ is expected to be faster than the step of discretization $\varphi_\varepsilon\to0$. It is shown that the choice $\delta_\varepsilon=\varepsilon$ and $t_{i+1}-t_i=\varepsilon^{1/3}$ provides the desired properties of the statistic $\Psi_{\tau,\varepsilon}$ (see Kutoyants (2022)). Further, suppose that the function $\Psi(\vartheta), \vartheta\in\Theta$ is monotone increasing and denote

$$\psi_m=\inf_{\vartheta\in\Theta}\Psi(\vartheta),\quad \psi_M=\sup_{\vartheta\in\Theta}\Psi(\vartheta),\qquad \psi_m=\Psi(\alpha),\quad \psi_M=\Psi(\beta),$$

$$G(\psi)=\Psi^{-1}(\psi),\quad \psi_m<\psi<\psi_M,\qquad \alpha<G(\psi)<\beta,\qquad \eta_\varepsilon=G(\hat\Psi_\varepsilon),$$

$$\mathbb{B}_m=\left\{\omega:\quad \hat\Psi_\varepsilon\le\psi_m\right\},\qquad \mathbb{B}_M=\left\{\omega:\quad \hat\Psi_\varepsilon\ge\psi_M\right\},$$

$$\mathbb{B}=\left\{\omega:\quad \psi_m<\hat\Psi_\varepsilon<\psi_M\right\},\qquad g(\nu)=\inf_{\vartheta_0\in\Theta}\inf_{|\vartheta-\vartheta_0|>\nu}|\Psi(\vartheta)-\Psi(\vartheta_0)|.$$

The substitution estimator (SE) is introduced as follows

$$\check\vartheta_{\tau,\varepsilon}=\alpha\mathbb{1}_{\{\mathbb{B}_m\}}+\eta_\varepsilon\mathbb{1}_{\{\mathbb{B}\}}+\beta\mathbb{1}_{\{\mathbb{B}_M\}}. \tag{23}$$

It has the following properties.

**Proposition 2.** *Suppose that the conditions* $\mathcal{A}_1,\mathcal{B}_1$ *are fulfilled,* $\inf_{\vartheta\in\Theta}\dot\Psi(\vartheta)>0$ *and for any (small)* $\nu>0$, $g(\nu)>0$. *Then the SE* $\check\vartheta_{\tau,\varepsilon}$ *is uniformly consistent and for any* $p>0$ *there exists a constant* $C=C(p)>0$ *such that*

$$\sup_{\vartheta_0\in\Theta}\varepsilon^{-p/2}\mathbf{E}_{\vartheta_0}\left|\check\vartheta_{\tau,\varepsilon}-\vartheta_0\right|^p\le C. \tag{24}$$

For the proof and more general results see Kutoyants (2022).

### 2.3. One-step MLE-process

Below we consider the MLE $\hat\vartheta_{\tau,\varepsilon}$ as the preliminary estimator. Following the same steps it can be shown that the SE $\check\vartheta_{\tau,\varepsilon}$ can also be used as the preliminary one. Recall that this estimator can be calculated more easily and the property (24) is sufficient for the proof of the Proposition 3 below.

Let us introduce the statistic

$$\vartheta^\star_{t,\varepsilon}=\hat\vartheta_{\tau,\varepsilon}+\frac{1}{\mathrm{I}_\tau^t(\hat\vartheta_{\tau,\varepsilon})}\int_\tau^t\frac{f(s)\dot m(\hat\vartheta_{\tau,\varepsilon},s)}{\varepsilon\sigma(s)^2}\Big[\mathrm{d}X_s-f(s)m(\hat\vartheta_{\tau,\varepsilon},s)\mathrm{d}s\Big], \tag{25}$$

where $\tau<t\leq T$ and $\mathrm{I}_\tau^t(\vartheta)$ is the Fisher information

$$\mathrm{I}_\tau^t(\vartheta)=\int_\tau^t\frac{f(s)\dot b(\vartheta,s)^2}{2b(\vartheta,s)\sigma(s)}\,\mathrm{d}s.$$

Concrning the calculations of the quantities $\dot m(\hat\vartheta_{\tau,\varepsilon},s)$ and $m(\hat\vartheta_{\tau,\varepsilon},s)$ since according to (14) we have

$$\begin{aligned}m(\vartheta,t)&=\int_0^t e^{-\int_s^t q_\varepsilon(\vartheta,v)\mathrm{d}v}\frac{\gamma(\vartheta,s)f(s)}{\varepsilon^2\sigma(s)^2}\mathrm{d}X_s\\&=e^{-\int_0^t q_\varepsilon(\vartheta,v)\mathrm{d}v}\int_0^t e^{\int_0^s q_\varepsilon(\vartheta,v)\mathrm{d}v}\frac{\gamma(\vartheta,s)f(s)}{\varepsilon^2\sigma(s)^2}\mathrm{d}X_s\\&=h(\vartheta,t)\int_0^t H(\vartheta,s)\mathrm{d}X_s\end{aligned}$$

with obvious notation. In such situations the stochastic integral is replaced by an ordinary one as follows. We have

$$\int_0^t H(\vartheta,s)\mathrm{d}X_s=H(\vartheta,t)X_t-\int_0^t X_sH_s'(\vartheta,s)\mathrm{d}s.$$

Let us denote $N\big(\vartheta,t,X^t\big)$ the right hand side of this equality. Then we can set

$$m(\hat\vartheta_{\tau,\varepsilon},s)=h(\hat\vartheta_{\tau,\varepsilon},s)N(\hat\vartheta_{\tau,\varepsilon},s,X^s).$$

The similar relation can also be written for $\dot m(\hat\vartheta_{\tau,\varepsilon},s)$.

Further, introduce the random processes

$$\eta_{t,\varepsilon}=\frac{\vartheta^\star_{t,\varepsilon}-\vartheta_0}{\sqrt{\varepsilon}},\qquad \tau\leq t\leq T,$$

$$\eta_t=\frac{1}{\mathrm{I}_\tau^t(\vartheta_0)}\int_\tau^t\frac{\dot b(\vartheta_0,s)\sqrt{f(s)}}{\sqrt{2b(\vartheta_0,s)\sigma(s)}}\,\mathrm{d}w(s),\qquad \tau\leq t\leq T,$$

where $w(s), 0\leq s\leq T$ is some standard Wiener process.

We add one more condition.

$\mathcal{B}_4$.*Non degeneracy of Fisher information: For any* $t_0\in(\tau,T]$

$$\inf_{\vartheta\in\Theta}\mathrm{I}_\tau^{t_0}(\vartheta)>0.$$

**Proposition 3.** *Let the conditions $\mathcal{A}$, $\mathcal{B}$ be fulfilled. Then the One-step MLE-process $\vartheta^\star_{t,\varepsilon}$, $\tau<t\leq T$ is uniformly consistent: for any $\nu>0$ and any $t_0\in(\tau,T)$*

$$\mathbf{P}_{\vartheta_0}\left(\sup_{t_0\leq t\leq T}\left|\vartheta^\star_{t,\varepsilon}-\vartheta_0\right|>\nu\right)\longrightarrow 0, \tag{26}$$

*the stochastic process $\eta_{t,\varepsilon}, t_0\leq t\leq T$ converges in distribution in the measurable space $(\mathcal{C}[t_0,T],\mathscr{B})$ to the random process $\eta_t, t_0\leq t\leq T$*

$$\eta_{\cdot,\varepsilon}\Longrightarrow\eta_\cdot,\qquad \eta_t\sim\mathcal{N}\Big(0,\mathrm{I}_\tau^t(\vartheta_0)^{-1}\Big). \tag{27}$$

**Proof.** Consider the normalized difference

$$\begin{aligned}\frac{\vartheta^\star_{t,\varepsilon}-\vartheta_0}{\sqrt{\varepsilon}}&=\frac{\hat\vartheta_{\tau,\varepsilon}-\vartheta_0}{\sqrt{\varepsilon}}+\frac{1}{\mathrm{I}_\tau^t(\hat\vartheta_{\tau,\varepsilon})}\int_\tau^t\frac{f(s)\dot m(\hat\vartheta_{\tau,\varepsilon},s)}{\sqrt{\varepsilon}\sigma(s)}\mathrm{d}\bar W_s\\&\quad+\frac{1}{\mathrm{I}_\tau^t(\hat\vartheta_{\tau,\varepsilon})}\int_\tau^t\frac{f(s)^2\dot m(\hat\vartheta_{\tau,\varepsilon},s)}{\varepsilon^{3/2}\sigma(s)^2}\Big[m(\vartheta_0,s)-m(\hat\vartheta_{\tau,\varepsilon},s)\Big]\mathrm{d}s.\end{aligned}$$

We have the relations (Lemma 6 in Kutoyants (2019))

$$\begin{aligned}
&m(\hat\vartheta_{\tau,\varepsilon},s)-m(\vartheta_0,s)=(\hat\vartheta_{\tau,\varepsilon}-\vartheta_0)\dot m(\tilde\vartheta_{\tau,\varepsilon},s),\\
&m(\hat\vartheta_{\tau,\varepsilon},s)-m(\vartheta_0,s)=(\hat\vartheta_{\tau,\varepsilon}-\vartheta_0)\dot m(\vartheta_0,s)+\frac12(\hat\vartheta_{\tau,\varepsilon}-\vartheta_0)^2\ddot m(\tilde\vartheta_{\tau,\varepsilon},s),\\
&\dot m(\vartheta_0,s)=\sqrt{\frac{\varepsilon\sigma(s)}{2b(\vartheta_0,s)f(s)}}\,\dot b(\vartheta_0,s)\xi_{s,\varepsilon}+\varepsilon\, R_{t,\varepsilon},\\
&\mathrm{I}_\tau^t(\hat\vartheta_{\tau,\varepsilon})^{-1}=\mathrm{I}_\tau^t(\vartheta_0)^{-1}+(\hat\vartheta_{\tau,\varepsilon}-\vartheta_0)\,Q_{t,\varepsilon}.
\end{aligned}$$

Here $\xi_{s,\varepsilon},s\in[\tau,T]$ are Gaussian, asymptotically independent random variables, i.e., $\xi_{s,\varepsilon}\Longrightarrow\xi_s\sim\mathcal N(0,1)$, where $\xi_s,s\in[\tau,T]$ are mutually independent.

For any $\nu>0$ we can write

$$\begin{aligned}
\mathbf P_{\vartheta_0}\left(\sup_{t_0\le t\le T}\left|\vartheta^\star_{t,\varepsilon}-\vartheta_0\right|>\nu\right)&\le\mathbf P_{\vartheta_0}\left(\left|\hat\vartheta_{\tau,\varepsilon}-\vartheta_0\right|\ge\frac\nu3\right)\\
&+\mathbf P_{\vartheta_0}\left(\mathrm{I}_\tau^{t_0}(\hat\vartheta_{\tau,\varepsilon})^{-1}\left|\int_\tau^T\frac{f(s)\dot m(\hat\vartheta_{\tau,\varepsilon},s)}{\sigma(s)}\mathrm d\bar W_s\right|\ge\frac\nu3\right)\\
&+\mathbf P_{\vartheta_0}\left(\frac{\left|\hat\vartheta_{\tau,\varepsilon}-\vartheta_0\right|}{\mathrm{I}_\tau^{t_0}(\hat\vartheta_{\tau,\varepsilon})}\int_\tau^T\frac{f(s)^2\left|\dot m(\hat\vartheta_{\tau,\varepsilon},s)\ddot m(\tilde\vartheta_{\tau,\varepsilon},s)\right|}{\sqrt\varepsilon\sigma(s)^2}\mathrm ds\ge\frac\nu3\right).
\end{aligned}$$

Now the convergence (26) follows from the consistency of $\hat\vartheta_{\tau,\varepsilon}$ and the following estimate of the moments of $\dot m(\cdot,\cdot)$: for any $p>0$

$$\sup_{\vartheta_0\in\Theta}\sup_{\tau\le t\le T}\mathbf E_{\vartheta_0}|\dot m(\vartheta_0,t)|^p\le C\varepsilon^{p/2}. \tag{28}$$

The proof of this estimate follows from the proof of Lemma 6 in Kutoyants (2019).

Moreover, we have the convergences

$$\begin{aligned}
&\int_\tau^t\frac{f(s)^2\dot m(\hat\vartheta_{\tau,\varepsilon},s)^2}{\varepsilon\sigma(s)^2}\mathrm ds=\int_\tau^t\frac{f(s)\dot b(\vartheta_0,s)^2\xi_{s,\varepsilon}^2}{2b(\vartheta_0,s)\sigma(s)}\,\mathrm ds(1+o(1))\longrightarrow\mathrm{I}_\tau^t(\vartheta_0),\\
&\int_\tau^t\frac{f(s)\dot m(\hat\vartheta_{\tau,\varepsilon},s)}{\sqrt\varepsilon\sigma(s)}\mathrm d\bar W_s=\int_\tau^t\frac{\sqrt{f(s)}\,\dot b(\vartheta_0,s)\,\xi_{s,\varepsilon}}{\sqrt{2b(\vartheta_0,s)\sigma(s)}}\mathrm d\bar W_s(1+o(1))\\
&\qquad\qquad\Longrightarrow\mathcal N\big(0,\mathrm{I}_\tau^t(\vartheta_0)\big).
\end{aligned}$$

The random processes $R_{t,\varepsilon},Q_{t,\varepsilon}$ have bounded polynomial moments.

Hence we can write the representation

$$\begin{aligned}
\frac{\vartheta^\star_{t,\varepsilon}-\vartheta_0}{\sqrt\varepsilon}&=\frac{\hat\vartheta_{\tau,\varepsilon}-\vartheta_0}{\sqrt\varepsilon}+\frac{1}{\mathrm{I}_\tau^t(\vartheta_0)}\int_\tau^t\frac{\dot b(\vartheta_0,s)\sqrt{f(s)}\,\xi_{s,\varepsilon}}{\sqrt{2b(\vartheta_0,s)\sigma(s)}}\mathrm d\bar W_s+o(1)\\
&\quad-\frac{(\hat\vartheta_{\tau,\varepsilon}-\vartheta_0)}{\sqrt\varepsilon}\frac{1}{\mathrm{I}_\tau^t(\vartheta_0)}\int_\tau^t\frac{\dot b(\vartheta_0,s)^2f(s)\,\xi_{s,\varepsilon}^2}{2b(\vartheta_0,s)\sigma(s)}\,\mathrm ds+\frac{(\hat\vartheta_{\tau,\varepsilon}-\vartheta_0)^2}{\sqrt\varepsilon}P_{t,\varepsilon}\\
&=\frac{1}{\mathrm{I}_\tau^t(\vartheta_0)}\int_\tau^t\frac{\dot b(\vartheta_0,s)\sqrt{f(s)}\xi_{s,\varepsilon}}{\sqrt{2b(\vartheta_0,s)\sigma(s)}}\,\mathrm d\bar W_s+o(1)+\left(\frac{\hat\vartheta_{\tau,\varepsilon}-\vartheta_0}{\sqrt\varepsilon}\right)^2P_{t,\varepsilon}\sqrt\varepsilon.
\end{aligned}\tag{29}$$

The random processes $P_{t,\varepsilon}$ has bounded polynomial moments. From this representation it follows that the One-step MLE-process is asymptotically normal: for all $t\in(\tau,T]$

$$\frac{\vartheta^\star_{t,\varepsilon}-\vartheta_0}{\sqrt\varepsilon}\Longrightarrow\eta_t\sim\mathcal N\big(0,\mathrm{I}_\tau^t(\vartheta_0)^{-1}\big).$$

The representation (29) allows us to obtain the convergence of the finite dimensional distributions as well

$$\left(\eta_{t_1,\varepsilon},\ldots,\eta_{t_k,\varepsilon}\right)\Longrightarrow\left(\eta_{t_1},\ldots,\eta_{t_k}\right) \tag{30}$$

for any $k\ge2$ and any $t_0\le t_1<\ldots<t_k\le T$.

Let us verify the condition

$$\mathbf E_{\vartheta_0}|\eta_{t_1,\varepsilon}-\eta_{t_2,\varepsilon}|^4\le C|t_1-t_2|^2 \tag{31}$$

which together with the convergence (30) provides the weak convergence (27) of the random process $\eta_{t,\varepsilon}, t_0 \leq t \leq T$. We denote

$$J_1(t) = \int_\tau^t \frac{f(s)\dot m(\hat\vartheta_{\tau,\varepsilon},s)}{\sqrt{\varepsilon}\sigma(s)}\mathrm{d}\bar W_s, \qquad K(t) = \mathrm{I}_\tau^t(\hat\vartheta_{\tau,\varepsilon})^{-1},$$

$$J_2(t) = \int_\tau^t \frac{f(s)^2\dot m(\vartheta_{\tau,\varepsilon},s)[m(\vartheta_0,s)-m(\hat\vartheta_{\tau,\varepsilon},s)]}{\varepsilon\sigma(s)^2}\mathrm{d}s$$

Then we obtain

$$\begin{aligned}\mathbf{E}_{\vartheta_0}|\eta_{t_1,\varepsilon}-\eta_{t_2,\varepsilon}|^4 &\leq C\mathbf{E}_{\vartheta_0}|K(t_1)J_1(t_1)-K(t_2)J_1(t_2)|^4\\ &\quad + C\mathbf{E}_{\vartheta_0}|K(t_1)J_2(t_1)-K(t_2)J_2(t_2)|^4\\ &\leq C\mathbf{E}_{\vartheta_0}|(K(t_1)-K(t_2))J_1(t_1)|^4\\ &\quad + C\mathbf{E}_{\vartheta_0}|(J_1(t_1)-J_1(t_2))K(t_2)|^4\\ &\quad + C\mathbf{E}_{\vartheta_0}|(K(t_1)-K(t_2))J_2(t_1)|^4\\ &\quad + C\mathbf{E}_{\vartheta_0}|(J_2(t_1)-J_2(t_2))K(t_2)|^4.\end{aligned}$$

Using once more the estimates (28) we obtain

$$\begin{aligned}&\mathbf{E}_{\vartheta_0}|K(t_1)-K(t_2)|^8 \leq C|t_2-t_1|^8, \qquad \mathbf{E}_{\vartheta_0}|J_1(t_2)|^8 \leq C,\\ &\mathbf{E}_{\vartheta_0}|(J_1(t_1)-J_1(t_2))|^8 \leq C|t_2-t_1|^4, \qquad \mathbf{E}_{\vartheta_0}|K(t_2)|^8 \leq C,\\ &\mathbf{E}_{\vartheta_0}|(J_2(t_1)-J_2(t_2))|^8 \leq C|t_2-t_1|^8.\end{aligned}$$

These estimates and Cauchy-Schwartz inequality allow us verify (31) and therefore obtain (27). □

Let us notice that the similar convergence is discussed in Kutoyants (2022) in a more general case; however, in the proof given there some calculations were omited.

### *2.4. Approximation*

Consider the family of solutions $u(t,y,\vartheta,\varepsilon), \vartheta \in \Theta, \varepsilon \in (0,1]$ of the equations

$$\begin{aligned}u_t' - a(t)y\,u_y' + \frac{1}{2}B_\varepsilon(\vartheta,t)^2u_{yy}'' &= -F\left(t,y,u,B_\varepsilon(\vartheta,t)u_y'\right),\\ u(T,y,\vartheta,\varepsilon) &= \Phi(y)\end{aligned} \tag{32}$$

and equation

$$\begin{aligned}U_t' - a(t)yU_y' + \frac{1}{2}b(\vartheta,t)^2U_{yy}'' &= -F\left(t,y,U,b(\vartheta,t)U_y'\right),\\ U(T,y,\vartheta) &= \Phi(y).\end{aligned} \tag{33}$$

By continuity we suppose that $U(t,y,\vartheta) = u(t,y,\vartheta,0)$. Recall that $B_\varepsilon(\vartheta,t) \to b(\vartheta,t)$ as $\varepsilon \to 0$.

*Conditions $\mathcal{C}$*

$\mathcal{C}_1$. *The functions $F(t,y,u,s)$ and $\Phi(y)$ satisfy conditions (5), (6).*

$\mathcal{C}_2$. *The function $u(t,y,\vartheta,\varepsilon), t \in (0,T], y \in \mathcal{R}, \vartheta \in \Theta, \varepsilon \in [0,1]$ has continuous derivatives $u_y'(\cdot), \dot u(\cdot), u_\varepsilon'(\cdot)$.*

Remark, that $Z_t = u(t,m(\vartheta_0,t),\vartheta_0,\varepsilon)$ is solution of the BSDE

$$\mathrm{d}Z_t = -F(t,m(\vartheta_0,t),Z_t,s(t))\mathrm{d}t + s(t)\mathrm{d}\bar W_t, \qquad Z_T = \Phi(m(\vartheta_0,T)),$$

where $s(t) = B_\varepsilon(\vartheta_0,t)u'(t,m(\vartheta_0,t),\vartheta_0,\varepsilon)$. As $u(\cdot) \to U(\cdot)$ the corresponding limit BSDE is

$$\mathrm{d}Z_t = -F(t,Y_t,Z_t,s_t)\mathrm{d}t + s_t\mathrm{d}V_t, \qquad Z_T = \Phi(Y_T),$$

where $Z_t = U(t,Y_t,\vartheta_9), s_t = b(\vartheta_0,t)U_y'(t,Y_t,\vartheta_0)$.

We do not set $m\left(\vartheta_{t,\varepsilon}^\star,t\right)$ and $\hat Z_t = u\left(t,m\left(\vartheta_{t,\varepsilon}^\star,t\right),\vartheta_{t,\varepsilon}^\star,\varepsilon\right)$ since in this case we need solutions of the Equations (14) for many values of $\vartheta$ and computationally this can be a difficult problem from a numerical point of view. Let us now introduce the recurrent equation

$$\mathrm{d}\hat m_t = -q_\varepsilon\left(\vartheta_{t,\varepsilon}^\star,t\right)\hat m_t\,\mathrm{d}t + \varepsilon^{-1}A_\varepsilon\left(\vartheta_{t,\varepsilon}^\star,t\right)\mathrm{d}X_t, \quad \tau < t \leq T \tag{34}$$

where the initial value is $\hat m_\tau = m(\hat\vartheta_{\tau,\varepsilon},\tau)$.

Let us set

$$\hat Z_t = u\left(t,\hat m_t,\vartheta_{t,\varepsilon}^\star,\varepsilon\right), \qquad \hat s_t = B_\varepsilon\left(\vartheta_{t,\varepsilon}^\star,t\right)u_y'\left(t,\hat m_t,\vartheta_{t,\varepsilon}^\star,\varepsilon\right)$$

The main result of this work is the following theorem.

**Theorem 1.** *Let the conditions $\mathcal{A},\mathcal{B},\mathcal{C}$ be fulfilled. Then for any $t\in(\tau,T]$ we have the convergence*

$$\frac{\hat Z_t-Z_t}{\sqrt{\varepsilon}}\Longrightarrow U_y'(t,Y_t,\vartheta_0)\sqrt{\frac{b(\vartheta_0,t)\sigma(t)}{2f(t)}}\left[\hat\zeta_t-\hat\xi_t\right]+\frac{\dot U(t,Y_t,\vartheta_0)}{\mathrm{I}_\tau^t(\vartheta_0)}\int_\tau^t\sqrt{\frac{\dot b(\vartheta_0,s)^2f(s)}{2b(\vartheta_0,s)\sigma(s)}}\,\mathrm{d}w(s). \tag{35}$$

*Here $\hat\zeta_t\sim\mathcal{N}(0,1),\hat\xi_t\sim\mathcal{N}(0,1)$ and $w(s),\tau\le s\le T$ are mutually independent random variables and Wiener process, respectively.*

**Proof.** First, we have the relation

$$\begin{aligned}\hat Z_t-Z_t&=u\big(t,\hat m_t,\vartheta_{t,\varepsilon}^\star,\varepsilon\big)-U(t,Y_t,\vartheta_0)\\&=u\big(t,\hat m_t,\vartheta_{t,\varepsilon}^\star,\varepsilon\big)-u\big(t,m(\vartheta_0,t),\vartheta_{t,\varepsilon}^\star,\varepsilon\big)\\&\quad+u\big(t,m(\vartheta_0,t),\vartheta_{t,\varepsilon}^\star,\varepsilon\big)-u(t,m(\vartheta_0,t),\vartheta_0,\varepsilon)\\&\quad+u(t,m(\vartheta_0,t),\vartheta_0,\varepsilon)-u(t,Y_t,\vartheta_0,\varepsilon)\\&\quad+u(t,Y_t,\vartheta_0,\varepsilon)-u(t,Y_t,\vartheta_0,0)\\&=u_y'\big(t,\tilde m_t,\vartheta_{t,\varepsilon}^\star,\varepsilon\big)\big(\hat m_t-m(\vartheta_0,t)\big)\\&\quad+u_y'\big(t,\bar m_t,\vartheta_{t,\varepsilon}^\star,\varepsilon\big)(m(\vartheta_0,t)-Y_t)\\&\quad+\dot u(t,m(\vartheta_0,t),\tilde\vartheta,\varepsilon)\big(\vartheta_{t,\varepsilon}^\star-\vartheta_0\big)+u_\varepsilon'(t,Y_t,\vartheta,\tilde\varepsilon)\varepsilon.\end{aligned}\tag{36}$$

Here $\tilde m_t,\bar m_t,\tilde\vartheta$ are some intermediate points in the corresponding expansions. Second, we have to study the quantities $\delta_t=\hat m_t-m(\vartheta_0,t)$ and $m(\vartheta_0,t)-Y_t$. Recall the equations for $m(\vartheta_0,t)$ and $\hat m_t$

$$\mathrm{d}m(\vartheta_0,t)=-a(t)m(\vartheta_0,t)\mathrm{d}t+\frac{\gamma_*(\vartheta_0,t)f(t)}{\sigma(t)}\mathrm{d}\bar W_t,\quad m(\vartheta_0,0)=0,$$

$$\mathrm{d}\hat m_t=-a(t)\hat m_t\mathrm{d}t-\frac{\gamma_*\big(\vartheta_{t,\varepsilon}^\star,t\big)f(t)^2}{\varepsilon\sigma(t)^2}\,\delta_t\,\mathrm{d}t+\frac{\gamma_*\big(\vartheta_{t,\varepsilon}^\star,t\big)f(t)}{\sigma(t)}\mathrm{d}\bar W_t,$$

where $\hat m_\tau=m\big(\vartheta_{t,\varepsilon}^\star,\tau\big)$. Therefore for $\delta_t$ we obtain the equation

$$\mathrm{d}\delta_t=-q_\varepsilon\big(\vartheta_{t,\varepsilon}^\star,t\big)\delta_t\mathrm{d}t+\frac{\big[\gamma_*\big(\vartheta_{t,\varepsilon}^\star,t\big)-\gamma_*(\vartheta_0,t)\big]f(t)^2}{\sigma(t)^2}\mathrm{d}\bar W_t,\qquad\tau<t\le T,$$

where $\delta_\tau=m(\hat\vartheta_{\tau,\varepsilon}^\star,\tau)-m(\vartheta_0,\tau)$ and

$$q_\varepsilon\big(\vartheta_{t,\varepsilon}^\star,t\big)=a(t)+\frac{\gamma_*\big(\vartheta_{t,\varepsilon}^\star,t\big)f(t)^2}{\varepsilon\sigma(t)^2}.$$

The solution of this equation on the time interval $[\tau,T]$ is

$$\begin{aligned}\delta_t&=\delta_\tau e^{-\int_\tau^t q_\varepsilon\left(\vartheta_{v,\varepsilon}^\star,v\right)\mathrm{d}v}\\&\quad+e^{-\int_\tau^t q_\varepsilon\left(\vartheta_{v,\varepsilon}^\star,v\right)\mathrm{d}v}\int_\tau^t e^{\int_\tau^s q_\varepsilon\left(\vartheta_{v,\varepsilon}^\star,v\right)\mathrm{d}v}\frac{\big[\gamma_*\big(\vartheta_{s,\varepsilon}^\star,s\big)-\gamma_*(\vartheta_0,s)\big]f(s)^2}{\sigma(s)^2}\mathrm{d}\bar W_s.\end{aligned}$$

Note that at the vicinity of the point $t$ we have the expansion

$$q_\varepsilon\big(\vartheta_{s,\varepsilon}^\star,s\big)=\frac{1}{\varepsilon}\frac{\gamma_*(\vartheta_0,t)f(t)^2}{\sigma(t)^2}\big(1+O(\varepsilon)+O(s-t)+O\big(\sqrt{\varepsilon}\big)\big),$$

where we used the relation $\vartheta_{s,\varepsilon}^\star-\vartheta_0=O\big(\sqrt{\varepsilon}\big)$. Let us denote $K(\vartheta,t)=\gamma_*(\vartheta_0,t)f(t)^2\sigma(t)^{-2}$ and note that

$$q_\varepsilon\big(\vartheta_{v,\varepsilon}^\star,v\big)=q_\varepsilon(\vartheta_0,v)\big(1+O\big(\sqrt{\varepsilon}\big)\big).$$

Using the same arguments as in the proof of Lemma 2 in Kutoyants (2019) for the stochastic integral we obtain

$$\begin{aligned}&e^{-\int_\tau^t q_\varepsilon\left(\vartheta_{v,\varepsilon}^\star,v\right)\mathrm{d}v}\int_\tau^t e^{\int_\tau^s q_\varepsilon\left(\vartheta_{v,\varepsilon}^\star,v\right)\mathrm{d}v}\frac{\big[\gamma_*\big(\vartheta_{s,\varepsilon}^\star,s\big)-\gamma_*(\vartheta_0,s)\big]f(s)^2}{\sigma(s)^2}\mathrm{d}\bar W_s\\&\quad=\int_\tau^t e^{-\int_s^t q_\varepsilon(\vartheta_0,v)\mathrm{d}v}\frac{\dot\gamma_*(\vartheta_0,s)\big[\vartheta_{s,\varepsilon}^\star-\vartheta_0\big]f(s)^2}{\sigma(s)^2}\mathrm{d}\bar W_s(1+o(1))\end{aligned}$$

$$
\begin{aligned}
&= \int_\tau^t e^{-\frac{1}{\varepsilon}K(\vartheta_0,t)(t-s)} \frac{\dot\gamma_*(\vartheta_0,s)\ \eta^\star_{s,\varepsilon}\ f(s)^2}{\sigma(s)^2} \mathrm{d}\bar W_s \sqrt{\varepsilon}(1+o(1))\\
&= \frac{\dot\gamma_*(\vartheta_0,t)\ \eta^\star_{t,\varepsilon}\ f(t)^2}{\sigma(t)^2\sqrt{2K(\vartheta_0,t)}}\ \zeta_{t,\varepsilon}\ \varepsilon(1+o(1))\\
&= \frac{\dot\gamma_*(\vartheta_0,t)\ \eta^\star_{t,\varepsilon}\ f(t)}{\sigma(t)\sqrt{2\gamma_*(\vartheta_0,t)}}\ \zeta_{t,\varepsilon}\ \varepsilon(1+o(1)).
\end{aligned}
$$

Here $\zeta_{t,\varepsilon} \sim \mathcal{N}(0,1), t \in (\tau, T]$ are independent random variables. By Lemma 1 we have

$$
\gamma_*(\vartheta_0,t) \longrightarrow \frac{b(\vartheta_0,t)\sigma(t)}{f(t)}, \qquad \dot\gamma_*(\vartheta_0,t) \longrightarrow \frac{\dot b(\vartheta_0,t)\sigma(t)}{f(t)}.
$$

The second limit here can be obtained like the first one in Lemma 2Kutoyants (2019).

For the initial value we have

$$
\begin{aligned}
\delta_\tau &= \dot m(\tilde\vartheta,\tau)(\hat\vartheta_{\tau,\varepsilon}-\vartheta_0)e^{-\int_\tau^t q_\varepsilon(\vartheta^\star_{v,\varepsilon},v)\mathrm{d}v}\\
&= \dot m(\vartheta_0,\tau)(\hat\vartheta_{\tau,\varepsilon}-\vartheta_0)\ e^{-\int_\tau^t q_\varepsilon(\vartheta_0,v)\mathrm{d}v}(1+o(1))\\
&= \dot m(\vartheta_0,\tau)\hat\eta_{\tau,\varepsilon}\ e^{-\frac{1}{\varepsilon}\int_\tau^t K(\vartheta_0,v)\mathrm{d}v}\ \sqrt{\varepsilon}\ (1+o(1)) = O\Big(e^{-\frac{c_*}{\varepsilon}(t-\tau)}\Big),
\end{aligned}
$$

where we denoted $c_* = \inf_{\tau<v\le\tau} K(\vartheta_0,v)$.

Finally, we obtain the representation

$$
\hat m_t - m(\vartheta_0,t) = \sqrt{\frac{\dot b(\vartheta_0,t)^2 f(t)}{2b(\vartheta_0,t)\sigma(t)}}\ \eta^\star_{t,\varepsilon}\ \zeta_{t,\varepsilon}\ \varepsilon(1+o(1)). \tag{37}
$$

For the difference $m(\vartheta_0,t)-Y_t$ we have the representation (18)

$$
\begin{aligned}
m(\vartheta_0,t)-Y_t &= \int_0^t e^{-\frac{1}{\varepsilon}K(\vartheta_0,t)(t-s)} \frac{\gamma_*(\vartheta_0,s)f(s)}{\sigma(s)}\mathrm{d}W_s(1+o(1))\\
&\qquad - \int_0^t e^{-\frac{1}{\varepsilon}K(\vartheta_0,t)(t-s)} b(\vartheta_0,s)\mathrm{d}V_s(1+o(1))\\
&= \frac{\gamma_*(\vartheta_0,t)f(t)}{\sigma(t)}\int_0^t e^{-\frac{1}{\varepsilon}K(\vartheta_0,t)(t-s)}\mathrm{d}W_s(1+o(1))\\
&\qquad - b(\vartheta_0,t)\int_0^t e^{-\frac{1}{\varepsilon}K(\vartheta_0,t)(t-s)}\mathrm{d}V_s(1+o(1))\\
&= \frac{\gamma_*(\vartheta_0,t)f(t)}{\sigma(t)\sqrt{2K(\vartheta_0,t)}}\ \hat\zeta_{t,\varepsilon}\ \sqrt{\varepsilon}\ (1+o(1))\\
&\qquad - \frac{b(\vartheta_0,t)}{\sqrt{2K(\vartheta_0,t)}}\ \hat\xi_{t,\varepsilon}\ \sqrt{\varepsilon}\ (1+o(1))\\
&= \sqrt{\frac{b(\vartheta_0,t)\sigma(t)}{2f(t)}}\Big[\hat\zeta_{t,\varepsilon}-\hat\xi_{t,\varepsilon}\Big]\sqrt{\varepsilon}\ (1+o(1)).
\end{aligned}
$$

From the convergences $m(\vartheta_0,t)\to Y_t$, $\vartheta^\star_{t,\varepsilon}\to\vartheta_0$ as $\varepsilon\to 0$ and continuity of derivatives we obtain the representation

$$
\begin{aligned}
\frac{\hat Z_t - Z_t}{\sqrt{\varepsilon}} &= u'_y(t,Y_t,\vartheta_0,0)\sqrt{\frac{b(\vartheta_0,t)\sigma(t)}{2f(t)}}\Big[\hat\zeta_{t,\varepsilon}-\hat\xi_{t,\varepsilon}\Big](1+o(1))\\
&\qquad + \frac{\dot u(t,Y_t,\vartheta_0,0)}{\mathrm{I}_\tau^t(\vartheta_0)}\int_\tau^t\sqrt{\frac{\dot b(\vartheta_0,s)^2 f(s)}{2b(\vartheta_0,s)\sigma(s)}}\ \xi_{s,\varepsilon}\mathrm{d}\bar W_s\ (1+o(1)).
\end{aligned} \tag{38}
$$

Therefore

$$
\begin{aligned}
\frac{\hat Z_t - Z_t}{\sqrt{\varepsilon}} &\Longrightarrow u'_y(t,Y_t,\vartheta_0,0)\sqrt{\frac{b(\vartheta_0,t)\sigma(t)}{2f(t)}}\Big[\hat\zeta_t-\hat\xi_t\Big]\\
&\qquad + \frac{\dot u(t,Y_t,\vartheta_0,0)}{\mathrm{I}_\tau^t(\vartheta_0)}\int_\tau^t\sqrt{\frac{\dot b(\vartheta_0,s)^2 f(s)}{2b(\vartheta_0,s)\sigma(s)}}\ \mathrm{d}w(s).
\end{aligned}
$$

Here $\hat\zeta_t\sim\mathcal{N}(0,1), \hat\xi_t\sim\mathcal{N}(0,1)$ and $w(s), 0\le s\le T$ are mutually independent random variables and Wiener process. □

Let us introduce the random process

$$z(t,\vartheta_0,Y_t)=\frac{\dot u(t,Y_t,\vartheta_0,0)}{\mathrm{I}_\tau^t(\vartheta_0)}\int_\tau^t\sqrt{\frac{\dot b(\vartheta_0,s)^2f(s)}{2b(\vartheta_0,s)\sigma(s)}}\,\mathrm{d}w(s)$$

and note that the Gaussian process $Y_t, 0\le t\le T$ and Wiener process $w(t), 0\le t\le T$ are independent.

At this point, let us introduce an additional condition.

*$\mathcal{D}$. The derivatives $u_y'(\cdot),\dot u(\cdot),u_\varepsilon'(\cdot)$ have polynomial majorants in y.*

**Corollary 1.** *Let the conditions $\mathcal{A},\mathcal{B},\mathcal{C},\mathcal{D}$ be fulfilled, then for any continuous functions $h(\cdot)$ we have the relation*

$$\varepsilon^{-1/2}\int_\tau^T h(t)\left[\hat Z_t-Z_t\right]\mathrm{d}t\Longrightarrow\int_\tau^T h(t)z(s,\vartheta_0,Y_s)\mathrm{d}s. \tag{39}$$

**Proof.** The proof follows from the limits

$$\int_\tau^T r(Y_s,s)\hat\zeta_{s,\varepsilon}\mathrm{d}s\longrightarrow 0,\qquad\qquad \int_\tau^T g(Y_s,s)\hat\xi_{s,\varepsilon}\mathrm{d}s\longrightarrow 0$$

for any continuous functions $r(\cdot),g(\cdot)$ with finite moments (see Kutoyants (2019)).

Concerning the process

$$\hat\zeta_{t,\varepsilon}=\sqrt{\frac{2K(\vartheta_0,t)}{\varepsilon}}\int_0^t e^{-\frac{1}{\varepsilon}K(\vartheta_0,t)(t-s)}\mathrm{d}W_s$$

we have

$$\mathbf{E}_{\vartheta_0}\hat\zeta_{t_1,\varepsilon}\hat\zeta_{t_2,\varepsilon}=\sqrt{\frac{K_1K_2}{K_1+K_2}}\left[e^{-\frac{1}{\varepsilon}K|t_1-t_2|}-e^{-\frac{1}{\varepsilon}[K_1t_1+K_2t_2]}\right]\longrightarrow 0$$

where $K_i=K(\vartheta_0,t_i), i=1,2$ and $\quad K=K_1\ 1_{\{t_1>t_2\}}+K_2\ 1_{\{t_1\le t_2\}}\quad.$ □

**Corollary 2.** *Let the conditions $\mathcal{A},\mathcal{B},\mathcal{C},\mathcal{D}$ be fulfilled. Then for any $t\in(\tau,T]$*

$$\varepsilon^{-1}\mathbf{E}_{\vartheta_0}\left(\hat Z_t-Z_t\right)^2\longrightarrow\frac{b(\vartheta_0,t)\sigma(t)}{f(t)}\mathbf{E}_{\vartheta_0}U_y'(t,Y_t,\vartheta_0)^2+\frac{\mathbf{E}_{\vartheta_0}\dot U(t,Y_t,\vartheta_0)^2}{\mathrm{I}_\tau^t(\vartheta_0)}.$$

**Proof.** For the proof we use the same arguments as in Corollary 1. □

## 3. Conclusions

It is shown that from four components of the error of approximation (35) only two of them have main contribution (Theorem 1). Moreover, if the integrated error is considered, then just one term remains(Corollary 1).

The contribution of the approximation of the conditional expectation $m(\vartheta_0,t)$ by the values $\hat m_t$ of solution of recurrent equation (adaptive filtration) is negligible. This means that for the model of observations (9), (10) with unknown parameter of volatility function, Equation (34) proposes the approximation $\hat m_t$ of $m(\vartheta_0,t)$ with error of order $\varepsilon$. This result can be applied in filtration theory.

Note that there are several possible generalizations which can be straightforward. For example, if we suppose as in Kutoyants (2019) that the function $f(t)=f(\vartheta,t)$, then the construction of approximation of $Z_t$ is similar to this one presented here.

## Declaration of Competing Interest

Dear Editor, by this letter we confirm that there is no conflict of interests related with this submission.

## Acknowledgment

We would like to thank AE, reviewers and I. Votsi for useful comments. This research was financially supported by the Ministry of Education and Science of the Russian Federation (project no. FSWF-2023-0012).

## References

Bismut, J.M., 1973. Conjugate convex functions in optimal stochastic control. J. Math. Anal. Appl. 44, 384–404.
Eberlein, E., Madan, D., Pistorius, M., Schoutens, W., Yor, M., 2014. Two price economies in continuous time. Ann. Finance 10, 71–100.
El Karoui, N., Peng, S., Quenez, M., 1997. Backward stochastic differential equations in finance. Math. Fin. 7, 1–71.
Elliott, R.J., Siu, T.K., 2011. A BSDE approach to a risk-based optimal investment of an insurer. Automatica 47 (2), 253–261.
Elliott, R.J., Siu, T.K., 2012. A BSDE approach to convex measures for derivative securities. Stochastic Analysis and Applications. 30 (6), 1083–1101.
Gasparyan, S., Kutoyants, Y.A., 2015. On approximation of the BSDE with unknown volatility in forward equation. Armenian J. of Mathematics 7 (1), 59–79.
Javaheri, A., 2015. Inside Volatility Filtering. Secrets of the Skew. 2nd Ed. Wiley, New Jersey.
Kutoyants, Y.A., 2014. Approximation of the solution of the backward stochastic differential equation. small noise, large samples and high frequency cases. In: Proceedings of the steklov institute of mathematics, moscow, vol. 287, pp. 133–154.
Kutoyants, Y.A., 2019. On parameter estimation of hidden ornstein-uhlenbeck process. Journal of Multivariate Analysis 169 (1), 248–263.
Kutoyants, Y.A., 2022. Volatility estimation of hidden markov processes and adaptive filtration. submitted. arXiv:2010.07603.
Kutoyants, Y.A., Zhou, L., 2014. On approximation of the backward stochastic differential equation. J. Stat. Plann. Infer. 150, 111–123.
Liptser, R., Shiryaev, A.N., 2005. Statistics of Random Processes. v. 2, 2-nd ed. Springer, N.Y..
Liu, H., Song, X., 2021. Bayesian analysis of hidden markov structural equation models with an unknown number of hidden states. Econometrics and Statistics 18, 29–43.
Ma, J., Yong, J., 1999. Forward-backward stochastic differential equations and their applications. In: Lecture notes in mathematics. Springer, Berlin.
Pardoux, E., Peng, S., 1990. Adapted solution of a backward stochastic differential equation. Systems and Control Letters 14 (1), 55–61.
Pardoux, E., Peng, S., 1992. Backward stochastic differential equation and quasilinear parabolic differential equations. In: Stochastic partial differential equations and their applications, vol. 176. Lecture Notes in Control and Information Sciences, pp. 200–217.
Su, Y., Lin, L., 2009. Semi-parametric estimation for forward-backward stochastic differential equations. Communications in Statistics - Theory and Methods, 38 (11), 1759–1775.
Wells, C., 2010. The kalman filter in finance. In: Advanced studies in theoretical and applied econometrics. Springer, N.Y..
Werge, N., Wintenberger, O., 2022. Adavol: An adaptive recursive volatility prediction method. Econometrics and Statistics 19–35.