Теорія Ймовір. та Матем. Статист.
Вип. 96, 2017

Theor. Probability and Math. Statist.
No. 96, 2018, Pages 1–13
https://doi.org/10.1090/tpms/1030
Article electronically published on October 5, 2018

STOCHASTIC DIFFERENTIAL EQUATIONS WITH GENERALIZED STOCHASTIC VOLATILITY AND STATISTICAL ESTIMATORS

UDC 519.21

M. BEL HADJ KHLIFA, YU. MISHURA, K. RALCHENKO, G. SHEVCHENKO, AND M. ZILI

Abstract. We study a stochastic differential equation, the diffusion coefficient of which is a function of some adapted stochastic process. The various conditions for the existence and uniqueness of weak and strong solutions are presented. The drift parameter estimation in this model is investigated, and the strong consistency of the least squares and maximum likelihood estimators is proved. As an example, the Ornstein–Uhlenbeck model with stochastic volatility is considered.

1. Introduction

In this article we investigate the stochastic differential equation of the form

$$X_t = X_0 + \theta a(t, X_t)\, dt + \sigma(t, X_t, Y_t)\, dW_t, \qquad t \in [0, T],$$

where W is a Wiener process, Y is some additional stochastic process, and θ is an unknown drift parameter. The models of such type have been known in mathematical finance since the late eighties; see [8]. Later the various models with stochastic volatility were proposed and studied by Stein and Stein [15], Heston [7], and Fouque et al. [5, 6] among others. For the recent results on this topic we refer to [9, 10], and the references cited therein. The problem of the parameter estimation in stochastic volatility models was considered in [1].

The case when the coefficient σ is a product of the form $\sigma_1(t, X_t)\sigma_2(t, Y_t)$ was studied in detail in [3], where the existence-uniqueness theorems for weak and strong solutions under various assumptions were proved, and the maximum likelihood estimator (MLE) was constructed and investigated. Here we obtain similar results for the case of a general diffusion coefficient $\sigma(t, X_t, Y_t)$. Moreover, we also propose the least squares estimator (LSE) for θ. Unlike the MLE, this estimator does not depend on the process Y. This is its crucial advantage, since in the financial applications the volatility process usually is not observed. As an example, we study the Ornstein–Uhlenbeck process with stochastic volatility and establish the strong consistency of both estimators for it.

The paper is organized as follows. In Section 2 we discuss the existence and uniqueness of weak and strong solutions. The drift parameter estimation is studied in Section 3. Section 4 is devoted to numerics. Some auxiliary results are proved in Section 5.

2010 *Mathematics Subject Classification.* Primary 60H10, 62F10, 62F12.

Key words and phrases. Stochastic differential equation, weak and strong solutions, stochastic volatility, drift parameter estimation, maximum likelihood estimator, strong consistency.

2. Existence and uniqueness results

Let $(\Omega, \mathfrak{F}, \{\mathfrak{F}_t\}_{t\geq 0}, \mathsf{P})$ be a complete probability space with filtration satisfying the standard assumptions. Let us consider the stochastic differential equation

$$X_t = X_0 + \int_0^t a(s, X_s)\, ds + \int_0^t \sigma(s, X_s, Y_s)\, dW_s, \qquad t \in [0, T], \tag{1}$$

where $X_0 \in \mathbb{R}$ is a constant, $a\colon [0,T] \times \mathbb{R} \to \mathbb{R}$ and $\sigma\colon [0,T] \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ are non-random functions, $W = \{W_t, \mathfrak{F}_t, t \in [0,T]\}$ is a standard Wiener process, and $Y = \{Y_t, \mathfrak{F}_t, t \in [0,T]\}$ is some stochastic process.

In this section we consider the existence and uniqueness of weak and strong solutions for the equation (1), adapting the approaches of Skorokhod [14], Stroock and Varadhan [16, 17], Yamada and Watanabe [18], and the standard Lipschitz conditions. Most of the results of this section can be proved similarly to the corresponding theorems of [3], so we omit their proofs.

2.1. Existence of weak solutions in terms of the Skorokhod conditions.

The proof of the following result follows the scheme from [14, Ch. 3, §3] and is similar to [3, Th. 1].

Theorem 2.1. *Let $Y = \{Y_t, \mathfrak{F}_t, t \in [0,T]\}$ be a stochastically continuous stochastic process, i.e.,*

$$\lim_{h\to 0} \sup_{|t_1 - t_2| \leq h} \mathsf{P}\left(|Y_{t_1} - Y_{t_2}| > \varepsilon\right) = 0.$$

Assume that the coefficients $a(t,x)$ and $\sigma(t,x,y)$ satisfy the following assumptions:

(i) *$a(t,x)$ and $\sigma(t,x,y)$ are jointly continuous with respect to $t \in [0,T]$ and $x, y \in \mathbb{R}$,*
(ii) *there exists a constant $K > 0$ such that*

$$a(t,x)^2 + \sigma(t,x,y)^2 \leq K\left(1 + x^2\right),$$

for all $x, y \in \mathbb{R}$.

Then the equation (1) has a weak solution.

2.2. Existence and uniqueness of weak solution in terms of Stroock–Varadhan conditions.

In this approach we assume additionally that the process Y is also a solution of some diffusion stochastic differential equation. Let W^1 and W^2 be two Wiener processes, possibly correlated, so that $dW_t^1 W_t^2 = \rho\, dt$ for some $|\rho| \leq 1$. In this case we can present $W_t^2 = \rho W_t^1 + \sqrt{1-\rho^2}\, W_t^3$, where W^3 is a Wiener process independent of W^1.

Theorem 2.2. *Consider the system of stochastic differential equations*

$$\begin{cases} dX_t = a(t, X_t)\, dt + \sigma(t, X_t, Y_t)\, dW_t^1, \\ dY_t = \alpha(t, Y_t)\, dt + \beta(t, Y_t)\, dW_t^2, \end{cases} \tag{2}$$

where all coefficients a, σ, α, and β are non-random measurable and bounded functions, and σ and β are continuous in all arguments. Let $|\rho| < 1$, $\beta(t,y) > 0$, $\sigma(t,x,y) > 0$ for all t, x, y. Then the weak existence and uniqueness in law hold for system (2), and, in particular, the weak existence and uniqueness in law hold for the first equation of (2) with Y being a weak solution of the second equation of (2).

Proof. Equations in (2) are equivalent to the two-dimensional stochastic differential equation

$$dZ(t) = A(t, Z_t)\, dt + B(t, Z_t)\, dW(t),$$

where $Z(t) = \binom{X(t)}{Y(t)}$, $W(t) = \binom{W^1(t)}{W^3(t)}$ is a two-dimensional Wiener process,

$$A(t,x,y) = \begin{pmatrix} a(t,x) \\ \alpha(t,y) \end{pmatrix}, \qquad B(t,x,y) = \begin{pmatrix} \sigma(t,x,y) & 0 \\ \rho\beta(t,y) & \sqrt{1-\rho^2}\beta(t,y) \end{pmatrix}.$$

It follows from measurability and boundedness of a and α and continuity and boundedness of σ and β that coefficients of matrices A and B are non-random, measurable, and bounded, and additionally coefficients of B are continuous in all arguments. Then we can apply [16, Ths. 4.2 and 5.6] (see also [4, Prop. 1.14]) and deduce that we have to prove the following relation: for any $(t,x,y) \in \mathbb{R}^+ \times \mathbb{R}^2$ there exists $\varepsilon(t,x,y) > 0$ such that for all $\lambda \in \mathbb{R}^2$,

$$\|B(t,x,y)\lambda\| \ge \varepsilon(t,x,y) \|\lambda\| . \tag{3}$$

Relation (3) is equivalent to the following (we omit arguments):

$$\sigma^2\lambda_1^2 + \beta^2 \left(\rho\lambda_1 + \sqrt{1-\rho^2}\lambda_2\right)^2 \ge \varepsilon^2 \left(\lambda_1^2 + \lambda_2^2\right)$$

or

$$\left(\sigma^2 + \beta^2\rho^2\right)\lambda_1^2 + \beta^2\left(1-\rho^2\right)\lambda_2^2 + 2\rho\sqrt{1-\rho^2}\beta^2\lambda_1\lambda_2 \ge \varepsilon^2\left(\lambda_1^2+\lambda_2^2\right). \tag{4}$$

The quadratic form

$$Q(\lambda_1,\lambda_2) = \left(\sigma^2 + \beta^2\rho^2\right)\lambda_1^2 + \beta^2\left(1-\rho^2\right)\lambda_2^2 + 2\rho\sqrt{1-\rho^2}\beta^2\lambda_1\lambda_2$$

in the left-hand side of (4) is positive definite, since its discriminant

$$D = \rho^2\left(1-\rho^2\right)\beta^4 - \beta^2\left(1-\rho^2\right)\left(\sigma^2+\beta^2\rho^2\right) = -\beta^2\left(1-\rho^2\right)\sigma^2 < 0.$$

The continuity of $Q(\lambda_1,\lambda_2)$ implies the existence of $\min_{\lambda_1^2+\lambda_2^2=1} Q(\lambda_1,\lambda_2) > 0$. Then putting $\varepsilon = \min_{\lambda_1^2+\lambda_2^2=1} Q(\lambda_1,\lambda_2)$ and using homogeneity, we get (4). □

2.3. **Existence and uniqueness of strong solution in terms of Yamada–Watanabe conditions.** Now we consider strong existence-uniqueness conditions for equation (1), adapting the Yamada–Watanabe conditions for inhomogeneous coefficients from [2].

Theorem 2.3. *Let a and σ be non-random measurable and bounded functions such that*

(i) *there exist a positive increasing function $\rho(u)$, $u \in (0,\infty)$, satisfying $\rho(0) = 0$ and a positive measurable bounded function ψ such that*

$$|\sigma(t,x_1,y) - \sigma(t,x_2,y)| \le \psi(y)\rho(|x_1-x_2|),$$

for all $t \ge 0$, $x_1,x_2,y \in \mathbb{R}$, and $\int_0^\infty \rho^{-2}(u)\,du = +\infty$;

(ii) *there exists a positive increasing concave function $k(u)$, $u \in (0,\infty)$, satisfying $k(0) = 0$ such that*

$$|a(t,x) - a(t,y)| \le k(|x-y|),$$

for all $t \ge 0$, $x,y \in \mathbb{R}$, and $\int_0^\infty k^{-1}(u)\,du = +\infty$.

Also, let Y be an adapted continuous stochastic process. Then the pathwise uniqueness of a solution holds for the equation (1), and hence it has the unique strong solution.

2.4. **Existence and uniqueness of strong solution in terms of Lipschitz conditions.**

Theorem 2.4. *Let a and σ be non-random measurable functions and let Y be an adapted continuous stochastic process. Consider the following assumptions:*

(i) *there exists $K > 0$ such that for all $t \ge 0$, $x \in \mathbb{R}$, $y \in \mathbb{R}$,*

$$|\sigma(t,x,y)|^2 + |a(t,x)|^2 \le K^2 \left(1 + |x|^2\right);$$

(ii) *for any $N \in \mathbb{N}$ there exist $K_N > 0$ and $C_N > 0$ such that for all $t \ge 0$ and for all (x_1, x_2, y) satisfying $|x_1| \le N$, $|x_2| \le N$, and $|y| \le N$,*

$$|a(t,x_1) - a(t,x_2)| \le K_N |x_1 - x_2|$$

and

$$|\sigma(t,x_1,y) - \sigma(t,x_2,y)| \le K_N \varphi(t,y) |x_1 - x_2|,$$

where φ is a positive and measurable function such that

$$\sup_{s\ge 0} \sup_{|x|\le N} |\varphi(s,x)| \le C_N.$$

Then equation (1) *has a unique strong solution.*

This result can be proved by using the successive approximation method; see, e.g., [13, Th. 1.2].

3. Drift parameter estimation

Let $(\Omega, \mathfrak{F}, \overline{\mathfrak{F}}, \mathsf{P})$ be a complete probability space with filtration $\overline{\mathfrak{F}} = \{\mathfrak{F}_t, t \ge 0\}$ satisfying the standard assumptions. It is assumed that all processes under consideration are adapted to the filtration $\overline{\mathfrak{F}}$. Consider a parametrized version of equation (1),

$$X_t = X_0 + \theta \int_0^t a(s,X_s)\,ds + \int_0^t \sigma(s,X_s,Y_s)\,dW_s, \qquad t \in [0,T], \tag{5}$$

where W is a Wiener process. Assume that equation (1) has a unique strong solution $X = \{X_t, t \in [0,T]\}$. Our main problem is to estimate the unknown parameter θ by the continuous observations of X and Y.

3.1. **Least squares estimation.** Assume that

$$\mathsf{E} \int_0^t a^2(s,X_s)\,ds < \infty, \tag{6}$$

$$\int_0^\infty a^2(s,X_s)\,ds = \infty \quad \text{almost surely}, \tag{7}$$

$$|\sigma(t,X_t,Y_t)| \le C \quad \text{almost surely}, \tag{8}$$

for all $t > 0$ and for some constant $C > 0$. Consider the following least squares estimator:

$$\tilde\theta_T = \frac{\int_0^T a(t,X_t)\,dX_t}{\int_0^T a^2(t,X_t)\,dt}.$$

Theorem 3.1. *Under the assumptions* (6)–(8)*, the estimator $\tilde\theta_T$ is strongly consistent, as $T \to \infty$.*

Proof. Using (5), the estimator $\tilde\theta_T$ can be written as

$$\tilde\theta_T = \theta + \frac{Z_T}{L_T},$$

where

$$Z_T = \int_0^T a(t, X_t)\sigma(t, X_t, Y_t)\, dW_t, \qquad L_T = \int_0^T a^2(t, X_t)\, dt.$$

Under assumptions (6)–(8) the process Z_t is a square-integrable martingale with quadratic variation $\langle Z\rangle_t = \int_0^t a^2(s, X_s)\sigma^2(s, X_s, Y_s)\, ds$, and L_t is an increasing process such that $L_0 = 0$, and $L_\infty = \infty$ almost surely. According to the strong law of large numbers for martingales [12, Ch. 2, §6, Th. 10], in order to prove the almost sure convergence $Z_T/L_T \to 0$, it suffices to verify that $\int_0^\infty (1+L_t)^{-2}\, d\langle Z\rangle_t < \infty$. This condition is satisfied, because

$$\int_0^\infty \frac{d\langle Z\rangle_t}{(1+L_t)^2} = \int_0^\infty \frac{a^2(t, X_t)\sigma^2(t, X_t, Y_t)}{(1+L_t)^2}\, dt \le C^2 \int_0^\infty \frac{dL_t}{(1+L_t)^2} = C^2. \qquad \square$$

3.2. **Maximum likelihood estimation.** Denote

$$f(t, x, y) = \frac{a(t, x)}{\sigma^2(t, x, y)}, \qquad g(t, x, y) = \frac{a(t, x)}{\sigma(t, x, y)}.$$

Assume that for all $t > 0$,

$$\sigma(t, X_t, Y_t) \ne 0 \quad \text{almost surely,} \tag{9}$$

$$\mathsf{E}\int_0^t g^2(s, X_s, Y_s)\, ds < \infty, \tag{10}$$

$$\int_0^\infty g^2(s, X_s, Y_s)\, ds = \infty \quad \text{almost surely.} \tag{11}$$

Then a likelihood function for equation (1) has the form

$$\frac{d\,\mathsf{P}_\theta(T)}{d\,\mathsf{P}_0(T)} = \exp\left\{\theta \int_0^T f(t, X_t, Y_t)\, dX_t - \frac{\theta^2}{2}\int_0^T g^2(t, X_t, Y_t)\, dt\right\};$$

see [11, Ch. 7]. Hence, the maximum likelihood estimator of parameter θ constructed by the observations of X and Y on the interval $[0, T]$ has the form

$$\hat\theta_T = \frac{\int_0^T f(t, X_t, Y_t)\, dX_t}{\int_0^T g^2(t, X_t, Y_t)\, dt} = \theta + \frac{\int_0^T g(t, X_t, Y_t)\, dW_t}{\int_0^T g^2(t, X_t, Y_t)\, dt}. \tag{12}$$

Theorem 3.2. *Under the assumptions* (9)–(11)*, the estimator $\hat\theta_T$ is strongly consistent, as $T \to \infty$.*

Proof. Note that under condition (10) the process $M_t = \int_0^t g(s, X_s, Y_s)\, dW_s$ is a square-integrable martingale with quadratic variation $\langle M\rangle_t = \int_0^t g^2(s, X_s, Y_s)\, ds$. According to [12, Ch. 2, §6, Th. 10, Cor. 1], under condition $\langle M\rangle_T \to \infty$ almost surely, as $T \to \infty$, we have that $M_T/\langle M\rangle_T \to 0$ almost surely, as $T \to \infty$. Therefore, it follows from representation (12) that $\hat\theta_T$ is strongly consistent. $\square$

3.3. **Drift parameter estimation for the Ornstein–Uhlenbeck process with stochastic volatility.** As an example let us consider the following model:

$$X_t = X_0 + \theta \int_0^t X_s\,ds + \int_0^t \sigma(Y_s)\,dW_s, \qquad t \in [0,T], \tag{13}$$

where the process Y is independent of the Wiener process W, and the diffusion coefficient $\sigma(Y)$ satisfies the following condition: for all $t \ge 0$, $y \in \mathbb{R}$,

$$\sigma_1 \le \sigma(Y_s) \le \sigma_2 \tag{14}$$

almost surely for some positive constants σ_1 and σ_2.

By Theorem 2.4, the equation (13) has a unique strong solution. It is not hard to see that this solution is given by

$$X_t = X_0 e^{\theta t} + \int_0^t \sigma(Y_s) e^{\theta(t-s)}\,dW_s, \qquad t \in [0,T].$$

Note that when σ is a constant, we obtain the well-known Ornstein–Uhlenbeck model. Therefore, we will call the process X *the Ornstein–Uhlenbeck process with stochastic volatility.*

The LSE and MLE for θ are equal to

$$\tilde\theta_T = \frac{\int_0^T X_t\,dX_t}{\int_0^T X_t^2\,dt}, \qquad \hat\theta_T = \frac{\int_0^T f(X_t,Y_t)\,dX_t}{\int_0^T g^2(X_t,Y_t)\,dt},$$

where $f(x,y) = x/\sigma^2(y)$, $g(x,y) = x/\sigma(y)$.

Theorem 3.3. *In the model* (13)*, under the assumption* (14)*, both estimators $\tilde\theta_T$ and $\hat\theta_T$ are strongly consistent, as $T \to \infty$.*

Proof. Since Y is independent of W, we can assume that $\mathsf{P} = \mathsf{P}_W \times \mathsf{P}_Y$, $\Omega = \Omega_W \times \Omega_Y$, $\omega = (\omega_W, \omega_Y)$, $W_t(\omega) = W_t(\omega_W)$, $Y_t(\omega) = Y_t(\omega_Y)$. Thus it is sufficient to show the strong consistency with respect to P_W for a. a. $\omega_Y \in \Omega_Y$. In other words, we can assume that $\sigma(Y_t) = \sigma(t)$ is deterministic. More precisely, let

$$X_t = X_0 e^{\theta t} + \int_0^t \sigma(s) e^{\theta(t-s)}\,dW_s, \qquad t \in [0,T]. \tag{15}$$

Note that under the assumption (14), the conditions (8) and (9) are satisfied. Furthermore, the conditions (6)–(7) and (10)–(11) are equivalent to

$$\mathsf{E}\int_0^t X_s^2\,ds < \infty, \tag{16}$$

$$\int_0^\infty X_s^2\,ds = \infty \quad \text{almost surely.} \tag{17}$$

Clearly, the assumption (16) is satisfied, because

$$\begin{aligned}
\mathsf{E}\int_0^t X_s^2\,ds &\le 2\left(X_0\int_0^t e^{\theta s}\,ds\right)^2 + 2\,\mathsf{E}\left(\int_0^t\int_0^s \sigma(u)e^{\theta(s-u)}\,dW_u\,ds\right)^2 \\
&= \left(X_0\int_0^t e^{\theta s}\,ds\right)^2 + 2\,\mathsf{E}\left(\int_0^t\int_u^t \sigma(u)e^{\theta(s-u)}\,ds\,dW_u\right)^2 \\
&\le \left(X_0\int_0^t e^{\theta s}\,ds\right)^2 + 2\sigma_2^2\int_0^t\left(\int_u^t e^{\theta(s-u)}\,ds\right)^2 du < \infty.
\end{aligned}$$

It remains to verify the assumption (17). Let us consider two cases.

Case $\theta \geq 0$. It suffices to prove that for $\lambda > 0$ the Laplace transform

$$\Psi_t(\lambda) := \mathsf{E} \exp\left\{-\lambda \int_0^t X_s^2\, ds\right\}$$

converges to zero, as $t \to \infty$. Since

$$\int_0^t X_s^2 \geq \int_{t-1}^t X_s^2\, ds \geq \left(\int_{t-1}^t X_s\, ds\right)^2,$$

we have

$$\Psi_t(\lambda) \leq \mathsf{E} \exp\left\{-\lambda \left(\int_{t-1}^t X_s\, ds\right)^2\right\}.$$

Note that $\int_{t-1}^t X_s\, ds$ is Gaussian. For a Gaussian random variable $\xi \sim \mathcal{N}(\mu, s^2)$,

$$\mathsf{E} \exp\left\{-\lambda \xi^2\right\} = \left(2\lambda s^2 + 1\right)^{-1/2} \exp\left\{-\frac{\lambda \mu^2}{2\lambda s^2 + 1}\right\} \leq \left(2\lambda s^2 + 1\right)^{-1/2}.$$

Therefore,

$$\Psi_t(\lambda) \leq \left(2\lambda V(t) + 1\right)^{-1/2},$$

where

$$V(t) = \operatorname{Var}\left[\int_{t-1}^t X_s\, ds\right].$$

However, by Lemma 5.1, $V(t) \to \infty$ as $t \to \infty$, whence the proof follows.

Case $\theta < 0$. We will prove a stronger property than (17), namely,

$$\mathsf{P}\left(\limsup_{t\to\infty} \int_t^{t+1} X_s^2\, ds = \infty\right) = 1.$$

Evidently, it suffices to prove that for all $C > 0$,

$$\mathsf{P}\left(\limsup_{t\to\infty} \int_t^{t+1} X_s^2\, ds \geq C\right) = 1$$

or

$$\mathsf{P}\left(\liminf_{t\to\infty} \int_t^{t+1} X_s^2\, ds \leq C\right) = 0.$$

By the Cauchy–Schwarz inequality,

$$\left|\int_t^{t+1} X_s\, ds\right|^2 \leq \int_t^{t+1} X_s^2\, ds.$$

Therefore,

$$\mathsf{P}\left(\liminf_{t\to\infty} \int_t^{t+1} X_s^2\, ds \leq C\right) \leq \mathsf{P}\left(\liminf_{t\to\infty} \left|\int_t^{t+1} X_s\, ds\right|^2 \leq C\right)$$
$$\leq \mathsf{P}\left(\bigcup_{N\in\mathbb{N}} \bigcap_{t\geq N} A_t\right) \leq \sum_{N\in\mathbb{N}} \mathsf{P}\left(\bigcap_{t\geq N} A_t\right),$$

where $A_t = \left\{\left|\int_t^{t+1} X_s\, ds\right|^2 \leq C + 1\right\}$. Now it suffices to show that for all N,

$$\mathsf{P}\left(\textstyle\bigcap_{t\geq N} A_t\right) = 0. \tag{18}$$

For any $k \geq 1$ and $N < N_1 < N_2 < \cdots < N_k$,

$$\mathsf{P}\left(\bigcap_{t\geq N} A_t\right) \leq \mathsf{P}(A_N)\,\mathsf{P}\left(A_{N_1} \mid A_N\right)\mathsf{P}\left(A_{N_2} \mid A_{N_1} \cap A_N\right)\dots \\ \times \mathsf{P}\left(A_{N_k} \mid A_{N_1} \cap \dots \cap A_{N_{k-1}} \cap A_N\right).$$

By Lemma 5.2, $\mathsf{P}(A_N) \leq \delta < 1$, where a constant $\delta = \delta(\theta, C)$ does not depend on N. Since Z is a Gaussian process, the conditional distribution of $\zeta_{N_1} = \int_{N_1}^{N_1+1} X_s\,ds$ given $\sigma(X_s, s \leq N)$ is Gaussian. Moreover, in view of (15) we can decompose $\zeta_{N_1} = \zeta'_{N_1} + \zeta''_{N_1}$, where

$$\zeta'_{N_1} = \int_{N_1}^{N_1+1} \int_0^N \sigma(s) e^{\theta(t-s)}\,dW_s\,dt$$

is $\sigma(X_s, s \leq N)$-measurable, and

$$\zeta''_{N_1} = \int_{N_1}^{N_1+1} \left(X_0 e^{\theta t} + \int_N^t \sigma(s) e^{\theta(t-s)}\,dW_s \right) dt$$

is independent from $\sigma(X_s, s \leq N)$. Then $\zeta'_{N_1} \to 0$ in probability, as $N_1 \to \infty$, since

$$\begin{aligned} \mathsf{E}\left(\zeta'_{N_1}\right)^2 &= \left(\int_{N_1}^{N_1+1} e^{\theta t}\,dt\right)^2 \int_0^N \sigma^2(s) e^{-2\theta s}\,ds \\ &\leq e^{2\theta N_1} \frac{\left(e^{\theta t} - 1\right)^2}{\theta^2} \sigma_2^2 \int_0^N e^{-2\theta s}\,ds \to 0, \end{aligned}$$

as $N_1 \to \infty$. Therefore, for any $\varepsilon > 0$,

$$\begin{aligned} \limsup_{N_1\to\infty} \mathsf{P}\left(A_{N_1} \mid A_N\right) &= \limsup_{N_1\to\infty} \frac{\mathsf{P}\left(\zeta_{N_1}^2 \leq C+1, \zeta_N^2 \leq C+1\right)}{\mathsf{P}\left(\zeta_N^2 \leq C+1\right)} \\ &\leq \limsup_{N_1\to\infty} \frac{\mathsf{P}\left(\left|\zeta'_{N_1}\right| \geq \varepsilon\right) + \mathsf{P}\left(\left|\zeta''_{N_1}\right| \leq \sqrt{C+1} + \varepsilon, \zeta_N^2 \leq C+1\right)}{\mathsf{P}\left(\zeta_N^2 \leq C+1\right)} \\ &= \limsup_{N_1\to\infty} \mathsf{P}\left(\left|\zeta''_{N_1}\right| \leq \sqrt{C+1} + \varepsilon\right). \end{aligned}$$

Letting $\varepsilon \to 0$, we get

$$\limsup_{N_1\to\infty} \mathsf{P}\left(A_{N_1} \mid A_N\right) \leq \limsup_{N_1\to\infty} \mathsf{P}\left(\left|\zeta''_{N_1}\right|^2 \leq C+1\right) < \delta,$$

by Lemma 5.2, since $\zeta''_{N_1} = \int_{N_1}^{N_1+1} X_t^{(N)} dt$ in terms of the notation (19). Hence there exists $N_1 > N$ such that

$$\mathsf{P}\left(A_{N_1} \mid A_N\right) < \frac{1+\delta}{2}.$$

Similarly, there exists $N_2 > N_1$ such that

$$\mathsf{P}\left(A_{N_2} \mid A_{N_1} \cap A_N\right) < \frac{1+\delta}{2},$$

and so on. Then

$$\mathsf{P}\left(\bigcap_{t\geq N} A_t\right) \leq \left(\frac{1+\delta}{2}\right)^k.$$

Letting $k \to \infty$, we get (18). This completes the proof. $\square$

4. Simulations

In this section we illustrate the quality of the estimators by simulations. Assume that the process X is described by the model (5), where Y is a unique strong solution of the homogeneous stochastic differential equation

$$Y_t = Y_0 + \int_0^t \alpha(Y_s)\,ds + \int_0^t \beta(Y_s)\,d\widetilde{W}_s, \qquad t \in [0,T],$$

$\widetilde{W} = \{\widetilde{W}_t, \mathfrak{F}_t, t \in [0,T]\}$ is a Wiener process, independent of W. More precisely, we consider the following four examples of Y:

1. constant coefficients: $\alpha(y) = \alpha$, $\beta(y) = \beta$ (we choose $\alpha = 1$, $\beta = 2$);
2. geometric Brownian motion: $\alpha(y) = \alpha y$, $\beta(y) = \beta y$ (we choose $\alpha = 2$, $\beta = 1$);
3. Ornstein–Uhlenbeck model: $\alpha(y) = -\alpha y$, $\beta(y) = \beta$ (we choose $\alpha = \beta = 1$);
4. Cox–Ingersoll–Ross model: $\alpha(y) = \alpha_1(\alpha_2 - y)$, $\beta(y) = \beta\sqrt{y}$ (we choose $\alpha_1 = 1$, $\alpha_2 = 2$, $\beta = 1$).

We simulate 100 sample paths of X for each set of parameters and compute means and standard deviations of LSE and MLE. Since the influence of the initial values X_0 and Y_0 on the behavior of the estimators is quite small, we choose $X_0 = Y_0 = 1$ everywhere.

Table 1. $dX_t = \theta(2 + \sin X_t)\,dt + \big(2 + \cos(X_t + Y_t)\big)\,dW_t$, $\theta = 2$

$\alpha(y)$	$\beta(y)$	Est.	Mean / Std.dev.	T = 10	50	100	200
1	2	$\tilde\theta$	Mean	1.9870	1.9965	1.9899	1.9887
			Std.dev.	0.2839	0.1447	0.1077	0.0813
		$\hat\theta$	Mean	1.9935	1.9919	1.9937	1.9940
			Std.dev.	0.2538	0.1163	0.0862	0.0629
$2y$	y	$\tilde\theta$	Mean	2.0262	2.0048	1.9975	1.9908
			Std.dev.	0.2885	0.1344	0.1015	0.0673
		$\hat\theta$	Mean	2.0141	1.9935	1.9874	1.9893
			Std.dev.	0.2194	0.1006	0.0861	0.0562
$-y$	1	$\tilde\theta$	Mean	2.0164	1.9885	1.9990	2.0058
			Std.dev.	0.3293	0.1482	0.1113	0.0836
		$\hat\theta$	Mean	2.0305	1.9951	2.0072	2.0081
			Std.dev.	0.2649	0.1139	0.0825	0.0606
$2-y$	$\sqrt{y}$	$\tilde\theta$	Mean	2.0283	2.0143	2.0094	2.0017
			Std.dev.	0.3177	0.1427	0.0964	0.0642
		$\hat\theta$	Mean	2.0167	2.0122	2.0079	2.0042
			Std.dev.	0.2403	0.1080	0.0771	0.0527

At first, let the coefficients $a(t,x)$ and $\sigma(t,x,y)$ be bounded away from zero and infinity: $a(t,x) = 2 + \sin x$, $\sigma(t,x,y) = 2 + \cos(x+y)$. Evidently, in this case all assumptions of Theorems 3.1 and 3.2 are satisfied. The results of simulations for $\theta = 2$ are reported in Table 1. We see that both estimators converge to the true value of θ and demonstrate quite similar asymptotic behavior. Therefore, we can conclude that LSE is preferable, since it has simpler form and does not depend on the process Y, which can be unobservable.

Now let us take the unbounded diffusion coefficient $\sigma(t,x,y) = 1 + y^2$ (as before, $a(t,x) = 2 + \sin x$). We see from Table 2 that MLE converges to θ for all four examples of Y. In the case of constants α and β, as well as for the geometric Brownian motion, the LSE does not work. This means that the assumption (8) in Theorem 3.1 is substantial. However, the LSE converges to the true value of the parameter for two other examples of Y. Moreover, in the Ornstein–Uhlenbeck model the behavior of two estimators is similar, while in the Cox–Ingersoll–Ross model the MLE clearly outperforms the LSE, since it has smaller standard deviation.

TABLE 2. $dX_t = \theta(2 + \sin X_t)\,dt + \left(1 + Y_t^2\right) dW_t$, $\theta = 2$

$\alpha(y)$	$\beta(y)$	Est.	Mean / Std.dev.	T = 10	50	100	200
1	2	$\tilde\theta$	Mean	1.5750	−8.5535	3.6113	78.1776
			Std.dev.	15.3463	84.8756	241.035	623.109
		$\hat\theta$	Mean	2.0408	2.0402	2.0407	2.0408
			Std.dev.	0.9771	0.9020	0.9004	0.9000
$2y$	y	$\tilde\theta$	Mean	$2.1 \cdot 10^{18}$	$4.2 \cdot 10^{76}$	$7.8 \cdot 10^{153}$	$8.9 \cdot 10^{281}$
			Std.dev.	$1.6 \cdot 10^{19}$	$4.1 \cdot 10^{77}$	$7.8 \cdot 10^{154}$	$8.9 \cdot 10^{282}$
		$\hat\theta$	Mean	2.2443	2.2443	2.2443	2.2443
			Std.dev.	1.9967	1.9967	1.9967	1.9967
$-y$	1	$\tilde\theta$	Mean	2.0189	2.0000	1.9978	1.9978
			Std.dev.	0.2712	0.1371	0.0984	0.0627
		$\hat\theta$	Mean	1.9954	1.9979	1.9988	1.9962
			Std.dev.	0.2112	0.1010	0.0686	0.0449
$2-y$	$\sqrt{y}$	$\tilde\theta$	Mean	2.1090	1.9942	1.9632	1.9641
			Std.dev.	1.1786	0.5412	0.4200	0.2720
		$\hat\theta$	Mean	1.9883	2.0080	1.9897	2.0024
			Std.dev.	0.4935	0.2092	0.1669	0.0976

Finally, we consider the Ornstein–Uhlenbeck model (13) with the stochastic volatility $\sigma(Y_t) = 2 + \cos Y_t$. The results for $\theta = -2$ and $\theta = 2$ are reported in Tables 3 and 4 respectively. We see that in both cases the simulation studies confirm the theoretical results on strong consistency for both estimators. However, the rate of convergence for the positive value of θ is much higher.

TABLE 3. $dX_t = \theta X_t\,dt + (2 + \cos Y_t)\,dW_t$, $\theta = -2$

$\alpha(y)$	$\beta(y)$	Est.	Mean / Std.dev.	T = 10	50	100	200
1	2	$\tilde\theta$	Mean	−2.3413	−2.03574	−1.9980	−2.0093
			Std.dev.	0.8153	0.3134	0.2120	0.1628
		$\hat\theta$	Mean	−2.2603	−2.0242	−2.0046	−2.0150
			Std.dev.	0.6732	0.2534	0.1811	0.1361
$2y$	y	$\tilde\theta$	Mean	−2.2009	−2.0411	−2.0234	−2.0113
			Std.dev.	0.6545	0.2865	0.2114	0.1537
		$\hat\theta$	Mean	−2.1521	−2.0368	−2.0310	−2.0162
			Std.dev.	0.4669	0.2087	0.1459	0.1039
$-y$	1	$\tilde\theta$	Mean	−2.1340	−2.0895	−2.0495	−2.0406
			Std.dev.	0.6116	0.3010	0.2006	0.1479
		$\hat\theta$	Mean	−2.1329	−2.0883	−2.0471	−2.0419
			Std.dev.	0.5863	0.3058	0.2039	0.1473
$2-y$	$\sqrt{y}$	$\tilde\theta$	Mean	−2.2316	−2.0792	−2.0266	−2.0266
			Std.dev.	0.6980	0.3406	0.2196	0.1546
		$\hat\theta$	Mean	−2.2041	−2.0647	−2.0256	−2.0211
			Std.dev.	0.6180	0.2629	0.1870	0.1342

TABLE 4. $dX_t = \theta X_t\,dt + (2+\cos Y_t)\,dW_t$, $\theta = 2$

$\alpha(y)$	$\beta(y)$	Est.	Mean / Std.dev.	T = 10	50	100	200
1	2	$\tilde\theta$	Mean	2.000	2.000	2.000	2.000
			Std.dev.	$4.3\cdot10^{-8}$	$7.7\cdot10^{-15}$	$8.9\cdot10^{-15}$	$7.0\cdot10^{-15}$
		$\hat\theta$	Mean	2.000	2.000	2.000	2.000
			Std.dev.	$3.0\cdot10^{-8}$	$8.4\cdot10^{-15}$	$8.9\cdot10^{-15}$	$8.0\cdot10^{-15}$
$2y$	y	$\tilde\theta$	Mean	2.000	2.000	2.000	2.000
			Std.dev.	$2.6\cdot10^{-8}$	$8.3\cdot10^{-15}$	$7.2\cdot10^{-15}$	$7.0\cdot10^{-15}$
3-8		$\hat\theta$	Mean	2.000	2.000	2.000	2.000
			Std.dev.	$2.8\cdot10^{-8}$	$1.1\cdot10^{-14}$	$9.0\cdot10^{-15}$	$8.0\cdot10^{-15}$
$-y$	1	$\tilde\theta$	Mean	2.000	2.000	2.000	2.000
			Std.dev.	$4.4\cdot10^{-8}$	$7.6\cdot10^{-15}$	$8.6\cdot10^{-15}$	$7.0\cdot10^{-15}$
		$\hat\theta$	Mean	2.000	2.000	2.000	2.000
			Std.dev.	$4.3\cdot10^{-8}$	$7.0\cdot10^{-15}$	$7.3\cdot10^{-15}$	$7.0\cdot10^{-15}$
$2-y$	$\sqrt{y}$	$\tilde\theta$	Mean	2.000	2.000	2.000	2.000
			Std.dev.	$3.6\cdot10^{-6}$	$8.4\cdot10^{-15}$	$7.8\cdot10^{-15}$	$7.0\cdot10^{-15}$
		$\hat\theta$	Mean	2.000	2.000	2.000	2.000
			Std.dev.	$1.7\cdot10^{-6}$	$8.5\cdot10^{-15}$	$8.8\cdot10^{-15}$	$7.0\cdot10^{-15}$

5. Appendix

Let X be the Ornstein–Uhlenbeck process with deterministic volatility defined by (15). Consider an auxiliary process

$$X_t^{(t_0)} := X_0 e^{\theta t} + \int_{t_0}^t \sigma(s) e^{\theta(t-s)}\,dW_s, \qquad t \ge t_0 \ge 0. \tag{19}$$

(Note that $X_t = X_t^{(0)}$.)

Lemma 5.1. *For any $\theta \in \mathbb{R}$ there exists a constant $\varepsilon = \varepsilon(\theta) > 0$ such that for all $t \ge t_0 \ge 0$,*

$$V(t_0,t) := \operatorname{Var}\left[\int_t^{t+1} X_s^{(t_0)}\,ds\right] \ge \varepsilon. \tag{20}$$

Moreover, if $\theta \ge 0$, then $V(t_0,t) \to \infty$, as $t \to \infty$.

Proof. Denote

$$U_t^{(t_0)} = \int_{t_0}^t \sigma(u) e^{\theta(t-u)}\,dW_u = X_t^{(t_0)} - X_0 e^{\theta t}.$$

Then

$$V(t_0,t) = \mathsf{E}\left(\int_t^{t+1} U_s^{(t_0)}\,ds\right)^2.$$

By Itô's isometry, for $s \ge t_0$, $v \ge t_0$,

$$\mathsf{E}\,U_s^{(t_0)} U_v^{(t_0)} = \int_{t_0}^{\min\{s,v\}} \sigma^2(u) e^{\theta(s-u)} e^{\theta(v-u)}\,du \ge \sigma_1^2 \int_{t_0}^{\min\{s,v\}} e^{\theta(s+v-2u)}\,du.$$

Hence

$$V(t_0,t) = \int_t^{t+1}\int_t^{t+1} \mathsf{E}\,U_s^{(t_0)} U_v^{(t_0)}\,ds\,dv \ge \sigma_1^2 \int_t^{t+1}\int_t^{t+1}\int_{t_0}^{\min\{s,v\}} e^{\theta(s+v-2u)}\,du\,ds\,dv.$$

If $\theta = 0$, then

$$V(t_0,t) \geq \sigma_1^2 \int_t^{t+1}\int_t^{t+1} \left(\min\{s,v\} - t_0\right) ds\, dv = \sigma_1^2 \left(t + \frac{1}{3} - t_0\right) \geq \frac{\sigma_1^2}{3};$$

that is, (20) holds with $\varepsilon = \sigma_1^2/3$, and $V(t_0,t) \to \infty$, as $t \to \infty$.

In what follows we assume that $\theta \neq 0$. We have

$$\begin{aligned} V(t_0,t) &\geq \frac{\sigma_1^2}{2\theta} \int_t^{t+1}\int_t^{t+1} e^{\theta(s+v)} \left(e^{-2\theta t_0} - e^{-2\theta \min\{s,v\}}\right) ds\, dv \\ &= \frac{\sigma_1^2}{2\theta} \int_t^{t+1}\int_t^{t+1} \left(e^{\theta(s+v-2t_0)} - e^{\theta|s-v|}\right) ds\, dv \\ &= \frac{\sigma_1^2}{2\theta} \left(e^{-2\theta t_0} \left(\int_t^{t+1} e^{\theta s}\, ds\right)^2 - 2\int_t^{t+1}\int_t^{v} e^{\theta(v-s)}\, ds\, dv\right) \\ &= \frac{\sigma_1^2}{2\theta^3} \left(e^{2\theta(t-t_0)} \left(e^\theta - 1\right)^2 - 2\left(e^\theta - 1 - \theta\right)\right). \end{aligned} \tag{21}$$

The right-hand side of (21) increases with respect to $t \in [t_0,\infty)$ for $\theta > 0$ as well as for $\theta < 0$. Therefore, it attains its minimum value at the point $t = t_0$. Hence

$$V(t_0,t) \geq \frac{\sigma_1^2}{2\theta^3} \left(\left(e^\theta - 1\right)^2 - 2\left(e^\theta - 1 - \theta\right)\right) =: \frac{\sigma_1^2}{2\theta^3} h(\theta).$$

Note that $h(0) = 0$ and the derivative $h'(\theta) = 2\left(e^\theta - 1\right)^2 > 0$ for $\theta \neq 0$. This implies that $h(\theta) < 0$ for $\theta < 0$, and $h(\theta) > 0$ for $\theta > 0$. Thus, (20) holds with $\varepsilon = \sigma_1^2 h(\theta)/2\theta^3 > 0$ for all $\theta \neq 0$. Moreover, it follows from (21) that for $\theta > 0$, $V(t_0,t) \to \infty$, as $t \to \infty$. This concludes the proof. □

Lemma 5.2. *Let $C > 0$, $\theta \in \mathbb{R}$. Then there exists a constant $\delta = \delta(\theta, C)$ such that for all $t \geq t_0 \geq 0$,*

$$\mathsf{P}\left(\left|\int_t^{t+1} X_s^{(t_0)}\, ds\right|^2 \leq C + 1\right) \leq \delta < 1.$$

Proof. For a Gaussian random variable $\xi_{\mu,s^2} \sim \mathcal{N}(\mu, s^2)$ one has

$$\mathsf{P}\left(\left|\xi_{\mu,s^2}\right| \leq x\right) \leq \mathsf{P}\left(\left|\xi_{0,s^2}\right| \leq x\right) = 2\Phi\left(\frac{x}{s}\right) - 1, \qquad x > 0,$$

where Φ denotes the cdf of the standard normal distribution. Taking into account that the random variable $\int_t^{t+1} X_s^{(t_0)}\, ds$ is Gaussian with variance $V(t_0,t)$ and applying the previous lemma, we get

$$\mathsf{P}\left(\left|\int_t^{t+1} X_s^{(t_0)}\, ds\right|^2 \leq C + 1\right) \leq 2\Phi\left(\frac{\sqrt{C+1}}{\sqrt{V(t_0,t)}}\right) - 1 \leq 2\Phi\left(\frac{\sqrt{C+1}}{\sqrt{\varepsilon}}\right) - 1 =: \delta < 1. \quad □$$

Bibliography

1. Y. Aït-Sahalia and R. Kimmel, *Maximum likelihood estimation of stochastic volatility models*, J. Financ. Econ. **83** (2007), 413–452.
2. S. Altay and U. Schmock, *Lecture notes on the Yamada–Watanabe condition for the pathwise uniqueness of solutions of certain stochastic differential equations*, `http://fam.tuwien.ac.at/~schmock/notes/Yamada-Watanabe.pdf`, 2013.
3. M. Bel Hadj Khlifa, Y. Mishura, K. Ralchenko, and M. Zili, *Drift parameter estimation in stochastic differential equation with multiplicative stochastic volatility*, Mod. Stoch. Theory Appl. **3** (2016), no. 4, 269–285. MR3593112
4. A. S. Cherny and H.-J. Engelbert, *Singular Stochastic Differential Equations*, Lecture Notes in Mathematics, vol. 1858, Springer-Verlag, Berlin, 2005. MR2112227

5. J.-P. Fouque, G. Papanicolaou, and K. R. Sircar, *Derivatives in Financial Markets with Stochastic Volatility*, Cambridge University Press, Cambridge, 2000. MR1768877
6. J.-P. Fouque, G. Papanicolaou, and K. R. Sircar, *Mean-reverting stochastic volatility*, Int. J. Theor. Appl. Finance **3** (2000), no. 1, 101–142.
7. S. Heston, *A closed-form solution of options with stochastic volatility with applications to bond and currency options*, The Review of Financial Studies **6** (1993), no. 2, 327–343.
8. J. Hull and A. White, *The pricing of options on assets with stochastic volatilities*, J. Finance **42** (1987), 281–300.
9. S. Kuchuk-Iatsenko and Y. Mishura, *Option pricing in the model with stochastic volatility driven by Ornstein–Uhlenbeck process. Simulation*, Mod. Stoch. Theory Appl. **2** (2015), no. 4, 355–369. MR3456143
10. S. Kuchuk-Iatsenko and Y. Mishura, *Pricing the European call option in the model with stochastic volatility driven by Ornstein–Uhlenbeck process. Exact formulas*, Mod. Stoch. Theory Appl. **2** (2015), no. 3, 233–249. MR3407504
11. R. S. Liptser and A. N. Shiryaev, *Statistics of Random Processes. I*, Applications of Mathematics, vol. 5, Springer-Verlag, Berlin, 2001. MR1800858
12. R. S. Liptser and A. N. Shiryayev, *Theory of Martingales*, Mathematics and its Applications (Soviet Series), vol. 49, Kluwer Academic Publishers Group, Dordrecht, 1989. MR1022664
13. M. Nisio, *Stochastic Control Theory. Dynamic Programming Principle*, Second edition, Probability Theory and Stochastic Modelling, vol. 72, Springer, Tokyo, 2015. MR3290231
14. A. V. Skorokhod, *Studies in the Theory of Random Processes*, Addison-Wesley Publishing Co., Inc., Reading, Mass., 1965. MR0185620
15. E. M. Stein and J. C. Stein, *Stock price distributions with stochastic volatility: an analytic approach*, Review of Financial Studies **4** (1991), no. 4, 727–752.
16. D. W. Stroock and S. R. S. Varadhan, *Diffusion processes with continuous coefficients. I*, Comm. Pure Appl. Math. **22** (1969), 345–400. MR0253426
17. D. W. Stroock and S. R. S. Varadhan, *Diffusion processes with continuous coefficients. II*, Comm. Pure Appl. Math. **22** (1969), 479–530. MR0254923
18. T. Yamada and S. Watanabe, *On the uniqueness of solutions of stochastic differential equations*, J. Math. Kyoto Univ. **11** (1971), 155–167. MR0278420

Department of Mathematics, Faculty of Sciences of Monastir, University of Monastir, Avenue de l'Environnement, 5000, Monastir, Tunisia
Email address: meriem.bhk@outlook.fr

Department of Probability Theory, Statistics and Actuarial Mathematics, Faculty of Mechanics and Mathematics, Taras Shevchenko National University of Kyiv, Volodymyrska, 64/13, Kyiv, Ukraine, 01601
Email address: myus@univ.kiev.ua

Department of Probability Theory, Statistics and Actuarial Mathematics, Faculty of Mechanics and Mathematics, Taras Shevchenko National University of Kyiv, Volodymyrska, 64/13, Kyiv, Ukraine, 01601
Email address: k.ralchenko@gmail.com

Department of Probability Theory, Statistics and Actuarial Mathematics, Faculty of Mechanics and Mathematics, Taras Shevchenko National University of Kyiv, Volodymyrska, 64/13, Kyiv, Ukraine, 01601
Email address: zhora@univ.kiev.ua

Department of Mathematics, Faculty of Sciences of Monastir, University of Monastir, Avenue de l'Environnement, 5000, Monastir, Tunisia
Email address: Mounir.Zili@fsm.rnu.tn

Received 25/JAN/2017
Originally published in English