Electronic Journal of Statistics
Vol. 4 (2010) 1443–1469
ISSN: 1935-7524
DOI: 10.1214/10-EJS590

Parametric estimation for discretely observed stochastic processes with jumps

Hoang-Long Ngo

Department of Mathematical Sciences, Ritsumeikan University
1-1-1 Noji-Higashi, Kusatsu, Shiga, 525-8577 Japan
e-mail: gr041086@ed.ritsumei.ac.jp

Department of Mathematics and Informatics, Hanoi National University of Education
136 Xuan Thuy, Cau Giay, Ha Noi, Vietnam

Abstract: We consider a two dimensional stochastic process (X, Y), which may have jump components and is not necessarily ergodic. There is an unknown parameter θ within the coefficients of (X, Y). The aim of this paper is to estimate θ from a regularly spaced sample of the process (X, Y). When the dynamic of X is known, an estimator is constructed by using a moment-based method. We show that our estimators will work if the Blumenthal-Getoor index of the jump part of Y is less than 2. What is perhaps the most interesting is the rate at which the estimators converge: it is $1/\sqrt{n}$ (as when the underlying processes are not contaminated by jumps) when that index is not greater than 1. When the dynamic of X is unknown, we introduce a spot volatility estimator-based approach to estimate θ. This approach can work even if the sample is contaminated by microstructure noise.

AMS 2000 subject classifications: Primary 60M09, 60J75; secondary 60M05, 62G05.
Keywords and phrases: Blumenthal-Getoor index, discrete observation, jump process, microstructure noise, non-ergodic process, nonparameric estimation, parametric estimation.

Received May 2010.

1. Introduction

In this work, we consider a process (X, Y) defined by the following stochastic differential equation

$$dY_t = b(\theta, X_t, Y_t)dt + \sigma(\theta, X_t)dW_t + dJ_t,$$

where W denotes a standard Brownian motion, J a Lévy process with no Brownian part. The process (X, Y) depends on an unknown parameter θ. The goal of this note is to estimate this parameter θ from regularly spaced observations of the process (X, Y).

The parametric estimations for discretely observed processes have been intensively studied in the case that the underlying processes (X, Y) possess some ergodic properties (see [12, 33, 34] and the references therein). To the best of our knowledge there are very few results about the non-ergodic situation and most are in the case of continuous diffusion processes (see [11, 31]).

This note is thus the first attempt to construct estimators which work even when the underlying processes X and Y contain jump components, and without any assumption about ergodicity. It should be mentioned here that from a practical point of view, one may think of X for instance as either the (log) price of an asset or the exchange rates process and Y as a state variable such as the (log) price of another asset which is correlated with X (see [4, 9]). Allowing X and Y to have jump components is nowadays of great interest. Some recent researches show that the models where jumps occur are able to fit skews and smiles that can hardly be captured by continuous models (see [5] and the references therein). It is needless to say that many classical estimation schemes for continuous diffusion processes are not suitable for processes with jumps.

In this paper, we present two classes of estimators for the parameter θ. When the dynamic of X is known, an estimator is constructed by using a moment-based method. This estimator is in the spirit of Jacod's recent work [11] for continuous diffusion processes. As far as the author knows, there are basically two ways to overcome the difficulty while working with jump processes. The first approach makes use of a threshold parameter (see [18, 33]). Although this approach can deal with jump processes of infinite activity, its results depend very sensitively on the threshold parameter, which is very difficult to efficiently detect in general. We adopt here the second approach called multipower method, which has been developed recently in [2, 23, 29, 30]. We will show that our estimator will work whenever the Blumenthal-Getoor index α of J is less than 2. In particular, if $\alpha \leq 1$ then the estimator $\hat{\theta}_n$ will converge to the true parameter θ^* at the optimal rate $1/\sqrt{n}$ (as when the underlying processes are not contaminated by jumps, see [11]) in the sense that

$$n^{\delta}(\hat{\theta}_n - \theta) \xrightarrow{\mathbb{P}} 0,$$

for any $\delta < 1/2$. On the other hand, if the index $\alpha \in (1,2)$, the level of activity of the jump process J does effect the behavior of the estimator $\hat{\theta}_n$. More precisely, the rate of convergence of $\hat{\theta}_n$ is $1/n^{1/\alpha - 1/2}$.

When the dynamics of X are unknown, we introduce a new method called spot volatility estimator-based approach to estimate θ. More precisely, under some assumptions, we approximate $\sigma(\theta, X(t_i^n))^2, i = 1, \ldots, n$ by a sequence of statistics $\hat{\sigma}(t_i^n)^2, i = 1, \ldots, n$ which depends only on the observation data of Y. Then the estimator $\hat{\theta}_n$ of θ is selected such that it minimizes

$$\frac{1}{n}\sum_{i=1}^{n}\left(\frac{\sigma(\theta, X(t_i^n))^2 - \hat{\sigma}(t_i^n)^2}{A(X_i^n)^2}\right)^2,$$

with a suitable function A. An interesting feature of this method is that it can work even if the observations of Y are contaminated by microstructure noise. This situation happens, for example, when process Y is observed on a high frequency time scale (e. g. intradaily data) while X is observed on a lower frequency time scale (e. g. daily data). A naïve way to avoid the effect of microstructure noise to the estimators is to sample Y over longer time scale. However, it is not wise to accept that throwing away such a lot of data can be an optimal solution.

Nevertheless, it is visibly clear that the rate of convergence of estimators $\hat{\theta}_n$ depends on the efficiency of the spot volatility estimators and hence, when microstructure noise and jump effects occur, the rates are slower than in the non-noisy case and a $1/\sqrt{n}$-rate can not be attained. Estimators $\hat{\theta}_n$ could also hardly reach the rate which is achieved when the dynamic of X is known. A comprehensive discussion about spot volatility estimation in various situations can be found in [1, 15–17, 19, 23–28] and [21][1].

The present paper is organized in the following way. The moment-based approach and spot volatility estimator-based approach are presented in Sections 2 and 3, respectively. In each section a numerical example is carried out to illustrate the behavior of the estimators. Some spot volatility estimation schemes are provided in Section 4.

2. Moment-based approach

2.1. Preliminary

Throughout this section, we consider the process (X, Y) defined on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ by

$$\begin{cases} dY_t = b(\theta, X_t, Y_t)dt + \sigma(\theta, X_t)dW_t + dJ_t \\ dX_t = a(\theta, X_t)dt + \tilde{\sigma}(\theta, X_t)d\tilde{W}_t + d\tilde{J}_t \end{cases} \quad (0 \leq t \leq T), \tag{2.1}$$

where $W, \tilde{W}$ denote two Brownian motions which can be correlated but the sigma algebra $\sigma\{W_t - W_s, \tilde{W}_t - \tilde{W}_s; t \geq s\}$ is independent of $\mathcal{F}_s$ for all $s \in [0, T)$. Functions $a(\theta, x), \sigma(\theta, x), \tilde{\sigma}(\theta, x)$ are known; function b is unknown. Parameter θ belongs to the set Θ which is a compact subset of $\mathbb{R}$. J and $\tilde{J}$ are Lévy processes with no Brownian component. We also assume that the sigma algebra $\sigma\{J_t - J_s, \tilde{J}_t - \tilde{J}_s; t \geq s\}$ is independent of $\mathcal{F}_s$ for all $s \in [0, T)$. The common assumptions in the literature suppose that the jump processes J and $\tilde{J}$ are either independent of the sigma algebra $\sigma\{W, \tilde{W}\}$ or of finite activity (see [2]). Nevertheless, we remark that in our discussion we do not need these assumptions. The rate of convergence of our estimator depends on the Blumenthal-Getoor index α of J which is defined by

$$\alpha := \inf\left\{p > 0 : \int_{|x|\leq 1} |x|^p \nu(dx) < \infty\right\},$$

where ν is the Lévy measure of J. Necessarily, $\alpha \in [0, 2]$. In this paper, we suppose that process J has finite second moment and Blumenthal-Getoor index $\alpha < 2$. Furthermore, we suppose that the Lévy process $\tilde{J}$ has a characteristic

[1] The author would like to thank the first Referee for pointing him to paper [21].

triplet $(\tilde{\mu}, 0, \tilde{\nu})$ which is known. Here $\tilde{\mu}$ and $\tilde{\nu}$ denote the drift and the Lévy measure of $\tilde{J}$, respectively (see [5]). In order to simplify our argument, $\tilde{J}$ is also supposed to have finite moment of all orders.

We now assume that the coefficients of equations (2.1) satisfy the following conditions.

(A1).

(i) The functions $a, \sigma, \tilde{\sigma}$ are three times differentiable in θ;

(ii) the functions $\frac{\partial^j a}{\partial \theta^j}, \frac{\partial^j \sigma}{\partial \theta^j}, \frac{\partial^j \tilde{\sigma}}{\partial \theta^j}$, for $j = 0, 1, 2, 3$, are three times differentiable in x;

(iii) the functions $\frac{\partial^{j+k} a}{\partial \theta^j \partial x^k}, \frac{\partial^{j+k} \sigma}{\partial \theta^j \partial x^k}, \frac{\partial^{j+k} \tilde{\sigma}}{\partial \theta^j \partial x^k}$, for $j = 0, 1, 2, 3$ and $k = 1, 2, 3$, are bounded by a constant;

(iv) we have

$$\sum_{j=0}^{3} \left| \frac{\partial^j a(\theta, x)}{\partial \theta^j} \right| + \left| \frac{\partial^j \sigma(\theta, x)}{\partial \theta^j} \right| + \left| \frac{\partial^j \tilde{\sigma}(\theta, x)}{\partial \theta^j} \right| \leq A(x), \quad |b(\theta, x, y)| \leq A(x),$$

for some C^∞ function $A : \mathbb{R} \to [1, \infty)$, whose derivatives of any order $m \geq 1$ are bounded and such that $A(x) \leq C(1 + |x|)$, where C is a positive constant.

Let denote θ^* the true value of the parameter θ. Following the paper of Jacod [11], we introduce the following mild assumptions about the identifiablitily of θ^* from the diffusion term $\sigma(\theta, x)$.

(A2). For any $\epsilon > 0$,

$$\inf_{\theta \in \Theta : |\theta - \theta^*| > \epsilon} \int_0^T \frac{(\sigma(\theta, X_s)^2 - \sigma(\theta^*, X_s)^2)^2}{A(X_s)^6} ds > 0 \quad a.s,$$

furthermore,

$$\int_0^T A(X_s)^{-6} \left[\frac{\partial}{\partial \theta} \sigma(\theta^*, X_s)^2 \right]^2 ds > 0 \quad a.s.$$

In the following we denote $\Delta_n = T/n, t_i^n = i\Delta_n, X_i^n = X_{t_{i-1}^n}, \Delta_i^n Z = Z_{t_i^n} - Z_{t_{i-1}^n}$ for any process Z.

Before establishing our estimator for the parameter θ, we give some remarks on the model (2.1). First, in our setting, though process X is observable, the discrete sample of X only may not be sufficient to make inferences about θ. This situation happens when, for example, the dynamic of X does not depend on θ at all (see Section 2.3 for a non-trivial example). Second, the model (2.1) appears in mathematical control and system theory of stochastic systems where processes X and Y are respectively the input and output of a real-time stochastic system (see [35] and the references therein). Third, if we consider a special case where processes X and Y coincide, the model (2.1) becomes

$$dY_t = dX_t = a(\theta, X_t)dt + \tilde{\sigma}(\theta, X_t)d\tilde{W} + d\tilde{J}_t.$$

This jump-diffusion model is widely used not only in finance to model the asset price [6, 20, 32] but also in soil moisture model [22], hydrology [3], population model [8], etc. However, as we already mentioned above, it seems that the estimation for θ in the non-ergodic setting has not been discussed so far in literature.

2.2. Estimator

For each $x \in \mathbb{R}, \theta \in \Theta$, let $X_t^{\theta,x}, 0 \le t \le T$, be the solution of

$$dX_t^{\theta,x} = a(\theta, X_t^{\theta,x})dt + \tilde{\sigma}(\theta, X_t^{\theta,x})d\tilde{W}_t + d\tilde{J}_t, \quad X_0^{\theta,x} = x \ a.s. \tag{2.2}$$

Thank to Assumption (A1), the above equation always has a unique solution. Let

$$\psi_n(\theta, x) = \frac{1}{\Delta_n}\int_0^{\Delta_n} \mathbb{E}(\sigma(\theta, X_s^{\theta,x}))^2 ds. \tag{2.3}$$

In practice, because functions a, $\tilde{\sigma}$ and the triple charateristics of $\tilde{J}$ are known, ψ can be calculated by using for instance Monte Carlo methods. We denote

$$\zeta_i^n(\theta) = \frac{1}{nA(X_i^n)^6}\left(\psi_n(\theta, X_i^n)^2 - \psi_n(\theta, X_i^n)\frac{|\Delta_i^n Y \Delta_{i+1}^n Y|}{\Delta_n}\pi\right), \quad i = 1, \dots, n-1. \tag{2.4}$$

The contrast function is defined by

$$U_n(\theta) = \sum_{i=1}^{n-1} \zeta_i^n(\theta). \tag{2.5}$$

We will show that function U_n is continuous in θ; hence it attains a minimum on the compact set Θ and due to the measurable selection theorem we can find a measurable variable $\hat{\theta}_n$ such that

$$U_n(\hat{\theta}_n) = \min_{\theta\in\Theta} U_n(\theta). \tag{2.6}$$

Denote $\delta_0 = \min\{1/2, 1/\alpha - 1/2\}$. We now state the main result of this paper.

Theorem 2.1. *Assume that θ^* is in the interior of Θ, then*

$$n^\delta(\hat{\theta}_n - \theta^*) \xrightarrow{\mathbb{P}} 0, \tag{2.7}$$

for any δ less than δ_0.

It should be noted that in [33], an estimator for θ^* is proposed, whose rate of convergence is $n^{1/2}$ independently of the jump behavior of J. However, the situation in [33] is quite different from ours because they treat the case that the underlying processes are ergodic and let the terminal time T tend to infinity in the estimator.

2.3. Numerical Example

In this section we consider the following toy model

$$\begin{cases} dY_t = \sqrt{\theta} X_t dW_t + dJ_t \\ dX_t = \theta X_t dt + X_t d\tilde{W}_t, \end{cases}$$

where $Y(0) = X(0) = 1,\ t \in [0,1], \theta \in [\theta_1, \theta_2] \subset (0, \infty)$ and J is a α stable Lévy process with jumps truncated by 1, or in other words, J is a Lévy process with no Brownian component and has a Lévy measure ν defined by

$$\nu(dx) = \left(\frac{A}{x^{\alpha+1}} I_{[0<x\le 1]} + \frac{B}{|x|^{\alpha+1}} I_{[-1\le x<0]} \right) dx,$$

for some positive constants A and B (see [5]).

It is worth to mention here that in this model, though process X does depend on parameter θ, a discrete sample of X, or even a continuous one, is not enough to infer a consistent estimation for θ when the terminal time T is fixed (see [12]). It is easy to verify that this model satisfies conditions (A1) and (A2) with $A(x) = L\sqrt{1+x^2}$ for some positive constant L. The experiment is designed as follows: we fix $\theta = \theta^* = 2$ and simulate the values of processes X and Y by using Euler's method with a very small time-discretization step. We consider the error defined by

$$Error_n = \hat{\theta}_n - \theta^*.$$

We consecutively take the number of observations $n = 10^3, n = 10^4$ and the Blumenthal-Getoor index $\alpha = 0, \alpha = 0.7, \alpha = 1.2$. After iterating the simulation 1000 times for each case, we get the histograms for the distribution of $Error$ in Figures 1 and 2.

We see that the asymptotic behavior of our estimator is better for a small α. The quality becomes slightly worse when the jump part has higher activity, however it remains acceptable.

2.4. Proofs

From now on the symbol C stands for a positive generic constant which can be changed from a line to another but not depend on t, n or θ.

2.4.1. Estimates on moments

We first state a few lemmata which will be used later.

Lemma 2.2. *Let $f(t,\omega)$ be a nonanticipative function with*

$$\mathbb{E} \int_0^T f(t,\omega)^{2m} dt < \infty \quad a.s,$$

for some $m \in \mathbb{Z}^+$. Then, for any $0 \le s \le t \le T$,

$$\mathbb{E}\left(\left(\int_s^t f(u,\omega) dW_u \right)^{2m} |\mathcal{F}_s \right) \le [m(2m-1)]^m (t-s)^{m-1} \mathbb{E}\left(\int_s^t f(u,\omega)^{2m} du |\mathcal{F}_s \right).$$

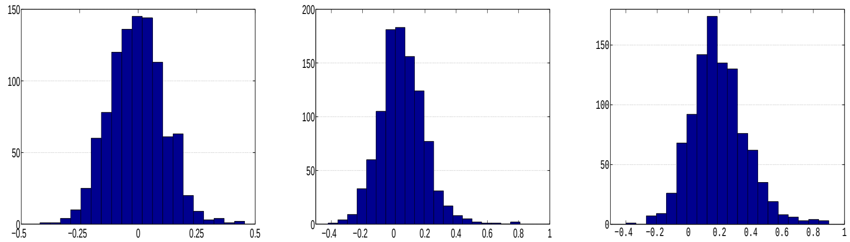

Fig 1. *Histogram of Error with $n = 10^3$, $\alpha = 0$(left), $\alpha = 0.7$(center), $\alpha = 1.2$(right).*

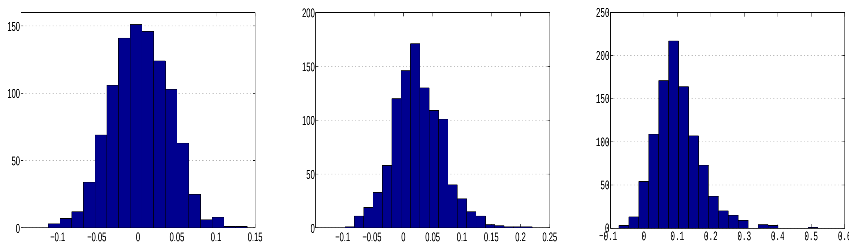

Fig 2. *Histogram of Error with $n = 10^4$, $\alpha = 0$(left), $\alpha = 0.7$(center), $\alpha = 1.2$(right).*

This lemma can be proved by carefully following the argument in [13], Lemma 4.12.

The next lemma gives a bound for conditional moments of process X defined in (2.1).

Lemma 2.3. *Under assumption (A1), for each $p \geq 1$, there exists a constant C_p such that for any $0 \leq s < t \leq T$,*

$$\mathbb{E}\big(|X_t - X_s|^{2p}|\mathcal{F}_s\big) \leq C_p A(X_s)^{2p}(t-s). \tag{2.8}$$

Proof. For any $0 \leq s < t \leq T$ we have

$$\begin{aligned}
&|X_t - X_s|^{2p} \\
&= \bigg|\tilde{J}_t - \tilde{J}_s + a(\theta, X_s)(t-s) + \tilde{\sigma}(\theta, X_s)(\tilde{W}_t - \tilde{W}_s) \\
&\qquad + \int_s^t (a(\theta, X_u) - a(\theta, X_s))du + \int_s^t (\tilde{\sigma}(\theta, X_u) - \tilde{\sigma}(\theta, X_s))dW_u\bigg|^{2p} \\
&\leq 5^{2p-1}\bigg(|\tilde{J}_t - \tilde{J}_s|^{2p} + A(X_s)^{2p}(t-s)^{2p} + A(X_s)^{2p}|\tilde{W}_t - \tilde{W}_s|^{2p} \\
&\qquad + \bigg|\int_s^t (a(\theta, X_u) - a(\theta, X_s))du\bigg|^{2p} + \bigg|\int_s^t (\tilde{\sigma}(\theta, X_u) - \tilde{\sigma}(\theta, X_s))d\tilde{W}_u\bigg|^{2p}\bigg).
\end{aligned}$$

On the other hand, by Theorem 1 in [14], for any $p \geq 1$, there exists a constant $K = K(p, T, \tilde{J})$ such that

$$\mathbb{E}(|\tilde{J}_t - \tilde{J}_s|^{2p}) \leq K(t-s),$$

for any $0 \leq s < t \leq T$. Hence, it follows from Jensen's inequality, assumption (A1) and Lemma 2.2 that

$$\begin{aligned}
&\mathbb{E}\big(|X_t - X_s|^{2p}|\mathcal{F}_s\big) \\
&\leq K5^{2p-1}(t-s) + CA(X_s)^{2p}(t-s)^p + C\int_s^t \mathbb{E}(|X_u - X_s|^{2p}|\mathcal{F}_s)du \\
&\leq CA(X_s)^{2p}(t-s) + C\int_s^t \mathbb{E}(|X_u - X_s|^{2p}|\mathcal{F}_s)du.
\end{aligned}$$

Applying Gronwall's inequality, we get the desired result. □

Now we split the proof of the Theorem 3.1 into several lemmata. First we give some estimates for function $\psi(\theta, x)$.

Lemma 2.4. *For any $x \in \mathbb{R}$, $j = 1, 2, 3$,*

i) $|\psi_n(\theta, x)| \leq CA(x)^2$;
ii) $|\psi_n(\theta, x) - \sigma(\theta, x)^2| \leq CA(x)^2\sqrt{\Delta_n}$;
iii) $\left|\dfrac{\partial^j \psi_n(\theta, x)}{\partial \theta^j}\right| \leq CA(x)^{j+1}$.

Proof. i) Because of condition (A1), there exists a constant C such that

$$|\sigma(\theta, X_s^{\theta,x}) - \sigma(\theta, x)| \leq C|X_s^{\theta,x} - x|,$$

hence, for any $s \leq 1$,

$$\begin{aligned}
\mathbb{E}|\sigma(\theta, X_s^{\theta,x})|^2 &\leq C(\sigma(\theta, x)^2 + \mathbb{E}|X_s^{\theta,x} - x|^2) \\
&\leq C(A(x)^2 + A(x)^2 s) \\
&\leq CA(x)^2.
\end{aligned}$$

ii) Since $\sigma(\theta, x)$ is differentiable in x, there exists u such that

$$\begin{aligned}
|\sigma(\theta, X_s^{\theta,x})^2 - \sigma(\theta, x)^2| &= 2|(X_s^{\theta,x} - x)\sigma(\theta, u)\sigma_x'(\theta, u)| \\
&\leq C|X_s^{\theta,x} - x|(|\sigma(\theta, x)| + |\sigma(\theta, u) - \sigma(\theta, x)|) \\
&\leq C(A(x)|X_s^{\theta,x} - x| + |X_s^{\theta,x} - x|^2).
\end{aligned}$$

Hence

$$\begin{aligned}
|\psi_n(\theta, x) - \sigma(\theta, x)^2| &\leq \frac{1}{\Delta_n}\int_0^{\Delta_n} \mathbb{E}|\sigma(\theta, X_s^{\theta,x})^2 - \sigma(\theta, x)^2|ds \\
&\leq \frac{C}{\Delta_n}\int_0^{\Delta_n} (A(x)\mathbb{E}|X_s^{\theta,x} - x| + \mathbb{E}|X_s^{\theta,x} - x|^2)ds \\
&\leq CA(x)^2\sqrt{\Delta_n}.
\end{aligned}$$

iii) By classical differentiation properties for stochastic differential equations (see for example [31]), we have

$$\frac{\partial X_t^{\theta,x}}{\partial\theta} = \int_0^t \left(a'_\theta(\theta, X_s^{\theta,x}) + a'_x(\theta, X_s^{\theta,x}) \frac{\partial X_s^{\theta,x}}{\partial\theta} \right) ds + \int_0^t \left(\tilde{\sigma}'_\theta(\theta, X_s^{\theta,x}) + \tilde{\sigma}'_x(\theta, X_s^{\theta,x}) \frac{\partial X_s^{\theta,x}}{\partial\theta} \right) d\tilde{W}_s.$$

Using condition (A1), Gronwall's lemma and following a routine argument in SDE theory, we could end up with

$$\mathbb{E}\left|\frac{\partial X_t^{\theta,x}}{\partial\theta}\right|^p \leq C_p A(x)^p t^{p/2}, \quad \forall p \geq 1, t \in (0,1),$$

where C_p is a constant which depends only on p. Next, we have

$$\frac{\partial\psi_n(\theta,x)}{\partial\theta} = \frac{1}{\Delta_n}\mathbb{E}\int_0^{\Delta_n} 2\sigma(\theta, X_s^{\theta,x})\left(\sigma'_\theta(\theta, X_s^{\theta,x}) + \sigma'_x(\theta, X_s^{\theta,x})\frac{\partial X_s^{\theta,x}}{\partial\theta}\right) ds,$$

and hence

$$\begin{aligned}\left|\frac{\partial\psi_n(\theta,x)}{\partial\theta}\right| &\leq \frac{C}{\Delta_n}\int_0^{\Delta_n} \left(\mathbb{E}|\sigma(\theta, X_s^{\theta,x})| + \mathbb{E}\left|\sigma(\theta, X_s^{\theta,x})\frac{\partial X_s^{\theta,x}}{\partial\theta}\right|\right) ds \\ &\leq \frac{C}{\Delta_n}\int_0^{\Delta_n} \left(A(x) + (A(x)^2 A(x)^2 s)^{1/2}\right) ds.\end{aligned}$$

This estimation implies *iii*) for $j = 1$. The demonstration for the case $j = 2, 3$ is similar and will be omitted. □

Lemma 2.5. *Let us define*

$$\varphi_n(\theta,x) = \frac{1}{(\Delta_n)^2}\mathbb{E}\left(\int_0^{\Delta_n} \sigma(\theta, X_s^{\theta,x}) dW_s\right)^4,$$

then there exists a constant C *such that*

$$|\varphi_n(\theta,x)| \leq CA(x)^4, \tag{2.9}$$

and

$$|\varphi_n(\theta,x) - 3\sigma(\theta,x)^4| \leq CA(x)^4\sqrt{\Delta_n}. \tag{2.10}$$

Proof. Since for any $s \in (0,1)$,

$$\begin{aligned}\mathbb{E}\sigma(\theta, X_s^{\theta,x})^4 &\leq C(\sigma(\theta,x)^4 + \mathbb{E}(X_s^{\theta,x} - x)^4) \\ &\leq C(A(x)^4 + A(x)^4 s) \\ &\leq CA(x)^4,\end{aligned}$$

we have

$$\mathbb{E}\bigg(\int_0^{\Delta_n}\sigma(\theta,X_s^{\theta,x})dW_s\bigg)^4 \le 36\Delta_n\int_0^{\Delta_n}\mathbb{E}\sigma(\theta,X_s^{\theta,x})^4ds \le CA(x)^4(\Delta_n)^2,$$

which implies (2.9). Next, we denote

$$\xi_1=\sigma(\theta,x)W_{\Delta_n} \quad \text{and} \quad \xi_2=\int_0^{\Delta_n}(\sigma(\theta,X_s^{\theta,x})-\sigma(\theta,x))dW_s.$$

For each $k=1,2,3$, we have

$$E(|\xi_1|^{2k})\le CA(x)^{2k}(\Delta_n)^k, \qquad \mathbb{E}(|\xi_1|^4)=3(\Delta_n)^2\sigma(\theta,x)^4,$$

and

$$\begin{aligned}\mathbb{E}(|\xi_2|^{2k}) &\le C(\Delta_n)^{k-1}\int_0^{\Delta_n}\mathbb{E}|\sigma(\theta,X_s^{\theta,x})-\sigma(\theta,x)|^{2k}ds\\ &\le C(\Delta_n)^{k-1}\int_0^{\Delta_n}\mathbb{E}|X_s^{\theta,x}-x|^{2k}ds\\ &\le C(\Delta_n)^{k+1}A(x)^{2k}.\end{aligned}$$

Hence it follows from Hölder's inequality that

$$\begin{aligned}|\varphi_n(\theta,x)-3\sigma(\theta,x)^4| &= \left|\frac{1}{(\Delta_n)^2}\mathbb{E}(\xi_1+\xi_2)^4-3\sigma(\theta,x)^4\right|\\ &\le \frac{C}{(\Delta_n)^2}\sum_{j=1}^4\mathbb{E}(|\xi_1|^{4-j}|\xi_2|^j)\\ &\le CA(x)^4\sqrt{\Delta_n}.\end{aligned}$$

□

2.4.2. Contrast function

We denote

$$Z_t=\int_0^t b(\theta,X_s,Y_s)ds+\int_0^t\sigma(\theta,X_s)dW_s, \quad 0\le t\le T.$$

$$\tilde{\zeta}_i^n(\theta)=\frac{1}{nA(X_i^n)^6}\bigg(\psi_n(\theta,X_i^n)^2-\psi_n(\theta,X_i^n)\frac{|\Delta_i^nZ\Delta_{i+1}^nZ|}{\Delta_n}\pi\bigg), \quad i=1,\dots,n-1,$$

$$\tilde{U}_n(\theta)=\sum_{i=1}^{n-1}\tilde{\zeta}_i^n(\theta).$$

Lemma 2.6. *For any $j=0,1,2,3$ and $k=1,2,\dots$, we have*

$$\sup_{i=2,\dots,n}\mathbb{E}\bigg(\bigg|\frac{\partial^j\tilde{\zeta}_i^n(\theta)}{\partial\theta^j}\bigg|^k|\mathcal{F}_{i-1}^n\bigg)\le\frac{C_k}{n^k}, \tag{2.11}$$

where $\mathcal{F}_i^n=\mathcal{F}_{t_i^n}$.

Proof. By Lemma 2.2, it follows that

$$\begin{aligned}
&\mathbb{E}\bigg(\Big(\int_{t_{i-1}^n}^{t_i^n} \sigma(\theta^*, X_s)dW_s\Big)^{2k} |\mathcal{F}_{i-1}^n\bigg) \\
&\leq [k(2k-1)]^k(\Delta_n)^{k-1}\mathbb{E}\bigg(\int_{t_{i-1}^n}^{t_i^n} \sigma(\theta^*, X_s)^{2k}ds|\mathcal{F}_{i-1}^n\bigg) \\
&\leq C(\Delta_n)^{k-1}\int_{t_{i-1}^n}^{t_i^n} \Big(|\sigma(\theta^*, X_i^n)|^{2k} + \mathbb{E}(|X_s - X_i^n|^{2k}|\mathcal{F}_{i-1}^n)\Big)ds \\
&\leq C\bigg(A(X_i^n)^{2k}(\Delta_n)^k + (\Delta_n)^{k-1}\int_{t_{i-1}^n}^{t_i^n} A(X_i^n)^{2k}(s - t_{i-1}^n)ds\bigg) \\
&\leq CA(X_i^n)^{2k}(\Delta_n)^k.
\end{aligned}$$

Furthermore,

$$\begin{aligned}
&\mathbb{E}\bigg(\Big(\int_{t_{i-1}^n}^{t_i^n} b(\theta, X_s, Y_s)ds\Big)^{2k} |\mathcal{F}_{i-1}^n\bigg) \\
&\leq (\Delta_n)^{2k-1}\mathbb{E}\bigg(\int_{t_{i-1}^n}^{t_i^n} A(X_s)^{2k}ds|\mathcal{F}_{i-1}^n\bigg) \\
&\leq C\bigg(A(X_i^n)^{2k}(\Delta_n)^{2k} + (\Delta_n)^{2k-1}\int_{t_{i-1}^n}^{t_i^n} \mathbb{E}(|X_s - X_i^n|^{2k}|\mathcal{F}_{i-1}^n)ds\bigg) \\
&\leq CA(X_i^n)^{2k}(\Delta_n)^{2k}.
\end{aligned}$$

Hence

$$\mathbb{E}(|\Delta_i^n Z|^{2k}|\mathcal{F}_{i-1}^n) \leq CA(X_i^n)^{2k}(\Delta_n)^k. \tag{2.12}$$

Similarly,

$$\mathbb{E}(|\Delta_{i+1}^n Z|^{2k}|\mathcal{F}_{i-1}^n) \leq CA(X_i^n)^{2k}(\Delta_n)^k. \tag{2.13}$$

Now we prove (2.11) for $j = 0$. We have

$$\begin{aligned}
&\mathbb{E}(|\tilde{\zeta}_i^n(\theta)|^k|\mathcal{F}_{i-1}^n) \\
&\leq \frac{2^{k-1}}{n^k A(X_i^n)^{6k}}\bigg(\psi_n(\theta, X_i^n)^{2k} + \psi_n(\theta, X_i^n)^k\mathbb{E}\big((\frac{\pi|\Delta_i^n Z\Delta_{i+1}^n Z|}{\Delta_n})^k|\mathcal{F}_{i-1}^n\big)\bigg),
\end{aligned}$$

by Lemma 2.4, it follows that

$$\begin{aligned}
&\mathbb{E}(|\tilde{\zeta}_i^n(\theta)|^k|\mathcal{F}_{i-1}^n) \\
&\leq \frac{C_k}{n^k A(X_i^n)^{4k}}\Big(A(X_i^n)^{2k} + (\Delta_n)^{-k}\big(\mathbb{E}(|\Delta_i^n Z|^{2k}|\mathcal{F}_{i-1}^n)\mathbb{E}(|\Delta_{i+1}^n Z|^{2k}|\mathcal{F}_{i-1}^n)\big)^{1/2}\Big).
\end{aligned}$$

Taking into account (2.12) and (2.13) we get the desired result. The demonstration for $j = 1, 2, 3$ is similar and will be omitted. □

We recall the following result about moment estimate for Lévy process (see [14, 23]).

Lemma 2.7. *For each $q > \alpha$ and $r \in [0, 2]$, there exists a constant C such that*

$$\mathbb{E}|J(t)|^r \leq Ct^{min\{r,1,r/q\}}, \quad \forall t \in (0, T).$$

The following lemma gives a uniform estimate for $U_n(\theta) - \tilde{U}_n(\theta)$.

Lemma 2.8. *For each $\delta \in (0, \delta_0)$, there exists a constant C_δ such that*

$$\mathbb{E} \sup_{\theta \in \Theta} \left| \frac{\partial^k (U_n(\theta) - \tilde{U}_n(\theta))}{\partial \theta^k} \right| \leq C_\delta (\Delta_n)^\delta, \tag{2.14}$$

with $k = 0, 1, 2$.

Proof. By virtue of Lemma 2.4, we get

$$\begin{aligned} &\left| \frac{\partial^k (U_n(\theta) - \tilde{U}_n(\theta))}{\partial \theta^k} \right| \\ &\leq \pi \sum_{i=1}^{n-1} \frac{1}{A(X_i^n)^6} \left| \frac{\partial^k \psi_n(\theta, X_i^n)}{\partial \theta^k} \big(\Delta_i^n Z \Delta_{i+1}^n Z - \Delta_i^n Y \Delta_{i+1}^n Y\big) \right| \\ &\leq C \sum_{i=1}^{n-1} \left(\left| \frac{\Delta_i^n Z}{A(X_i^n)} \Delta_{i+1}^n J \right| + \left| \frac{\Delta_{i+1}^n Z}{A(X_i^n)} \Delta_i^n J \right| + |\Delta_i^n J \Delta_{i+1}^n J| \right). \end{aligned}$$

We denote $r_1 = (\delta + \frac{1}{2})^{-1} > \max\{1, \ \alpha\}$ and $r_2 = \frac{r_1}{r_1 - 1} > 0$. It follows from Hölder's inequality, (2.12), (2.13) and Lemma 2.7 that

$$\begin{aligned} &\mathbb{E}\left(\left| \frac{\Delta_i^n Z}{A(X_i^n)} \Delta_{i+1}^n J \right| + \left| \frac{\Delta_{i+1}^n Z}{A(X_i^n)} \Delta_i^n J \right| + |\Delta_i^n J \Delta_{i+1}^n J| \right) \\ &\leq \left\| \frac{\Delta_i^n Z}{A(X_i^n)} \right\|_{r_2} \|\Delta_{i+1}^n J\|_{r_1} + \left\| \frac{\Delta_i^n Z}{A(X_{i+1}^n)} \right\|_{r_2} \|\Delta_i^n J\|_{r_1} + \mathbb{E}|\Delta_i^n J| \mathbb{E}|\Delta_{i+1}^n J| \\ &\leq C(\Delta_n)^{1+\delta}, \end{aligned}$$

this estimation implies (2.14). □

We introduce the following auxiliary function

$$F(\theta, x) = \frac{1}{A(x)^6} \big(\sigma(\theta, x)^4 - 2\sigma(\theta, x)^2 \sigma(\theta^*, x)^2\big). \tag{2.15}$$

Lemma 2.9. *$F(\theta, x)$ is three times differentiable in θ, and*

i) $\left| \mathbb{E}(\tilde{\zeta}_i^n(\theta) | \mathcal{F}_{i-1}^n) - \frac{1}{n} F(\theta, X_i^n) \right| \leq C(\Delta_n)^{3/2}$;

ii) $\left| \mathbb{E}\left(\frac{\partial^2 \tilde{\zeta}_i^n(\theta)}{\partial \theta^2} \middle| \mathcal{F}_{i-1}^n \right) - \frac{1}{n} \frac{\partial^2 F(\theta, X_i^n)}{\partial \theta^2} \right| \leq C(\Delta_n)^{3/2}$.

Proof. It follows by Lemma 2.4 that

$$
\begin{aligned}
&\left|\mathbb{E}(\tilde{\zeta}_i^n(\theta)|\mathcal{F}_{i-1}^n) - \frac{1}{n}F(\theta, X_i^n)\right| \\
&\leq \frac{1}{nA(X_i^n)^6}\bigg(|\psi_n(\theta, X_i^n)^2 - \sigma(\theta, X_i^n)^4| + 2\sigma(\theta^*, X_i^n)|\sigma(\theta, X_i^n)^2 - \psi(\theta, X_i^n)| \\
&\qquad + 2\psi_n(\theta, X_i^n)|\sigma(\theta^*, X_i^n)^2 - \frac{\pi}{2\Delta_n}\mathbb{E}(|\Delta_i^n Z \Delta_{i+1}^n Z||\mathcal{F}_{i-1}^n)|\bigg) \\
&\leq C\frac{\sqrt{\Delta_n}}{n} + C\frac{1}{nA(X_i^n)^4}\left|\sigma^2(\theta^*, X_i^n) - \frac{\pi}{2\Delta_n}\mathbb{E}(|\Delta_i^n Z \Delta_{i+1}^n Z||\mathcal{F}_{i-1}^n)\right|.
\end{aligned}
$$

We denote

$$
\xi_j = \int_{t_{i-j}^n}^{t_{i-j+1}^n} b(\theta, X_s, Y_s)dW(s) + \int_{t_{i-j}^n}^{t_{i-j+1}^n} (\sigma(\theta^*, X_s) - \sigma(\theta^*, X_i^n))dW_s, \quad j = 0, 1.
$$

By following the similar argument as in the proof of Lemma 2.6 we can show that

$$
\mathbb{E}(\xi_j^2|\mathcal{F}_{i-1}^n) \leq CA(X_i^n)^2\Delta_n^2, \quad j = 0, 1.
$$

Hence

$$
\begin{aligned}
&|\sigma^2(\theta^*, X_i^n) - \frac{\pi}{2\Delta_n}\mathbb{E}(|\Delta_i^n Z \Delta_{i+1}^n Z||\mathcal{F}_{i-1}^n)| \\
&\leq \left|\sigma^2(\theta^*, X_i^n) - \frac{\pi}{2\Delta_n}\mathbb{E}(\sigma(\theta^*, X_i^n)^2|\Delta_i^n W \Delta_{i+1}^n W||\mathcal{F}_{i-1}^n)\right| \\
&\qquad + \frac{\pi}{2\Delta_n}\mathbb{E}\big(|\xi_0\xi_1 + \sigma(\theta^*, X_i^n)\xi_0\Delta_i^n W + \sigma(\theta^*, X_i^n)\xi_1\Delta_{i+1}^n W||\mathcal{F}_{i-1}^n\big) \\
&\leq CA(X_i^n)(\Delta_n)^{1/2},
\end{aligned}
$$

this estimation implies the first part of this lemma. A proof for the second part can be carried out by a similar argument as above. □

Lemma 2.10. *For each $j = 0, 2$,*

$$
\sup_{\theta\in\Theta}\left|\frac{\partial^j U_n(\theta)}{\partial\theta^j} - \frac{1}{T}\int_0^T \frac{\partial^j F(\theta, X_s)}{\partial\theta^j}ds\right| \xrightarrow{\mathbb{P}} 0.
$$

Proof. Thank to condition (A1), $A(x)^j\frac{\partial^j F(\theta, X_i^n)}{\partial\theta^j}$ are uniformly bounded by a constant. By applying Itô's formula for functions $F(\theta, x)$ and $\frac{\partial^2 F(\theta, x)}{\partial\theta^2}$, we have

$$
\mathbb{E}\left|\frac{1}{n}\sum_{i=1}^n \frac{\partial^j F(\theta, X_i^n)}{\partial\theta^j} - \frac{1}{T}\int_0^T \frac{\partial^j F(\theta, X_s)}{\partial\theta^j}ds\right| \leq C\sqrt{\Delta_n}. \tag{2.16}
$$

This fact, together with Lemma 2.9, leads to

$$
\sum_{i=1}^{n-1}\mathbb{E}\left(\frac{\partial^j\tilde{\zeta}_i^n(\theta)}{\partial\theta^j}\middle|\mathcal{F}_{i-1}^n\right) - \frac{1}{T}\int_0^T \frac{\partial^j F(\theta, X_s)}{\partial\theta^j}ds \xrightarrow{\mathbb{P}} 0. \tag{2.17}
$$

On the other hand, it follows from Lemma 2.6 that

$$\sum_{i=1}^{n-1} \mathbb{E}\left(\left| \frac{\partial^j \tilde{\zeta}_i^n(\theta)}{\partial \theta^j} \right|^2 |\mathcal{F}_{i-1}^n \right) \leq C\Delta_n.$$

This relation and (2.17) yield

$$V_n^j(\theta) := \frac{\partial^j \tilde{U}_n(\theta)}{\partial \theta^j} - \frac{1}{T} \int_0^T \frac{\partial^j F(\theta, X_s)}{\partial \theta^j} ds \xrightarrow{\mathbb{P}} 0. \tag{2.18}$$

Furthermore, by Lemma 2.6, we have

$$\sup_{\theta \in \Theta} \left(\mathbb{E}\left| \frac{\partial^{j+1} \tilde{U}_n(\theta)}{\partial \theta^{j+1}} \right|^2 + \mathbb{E}\left(\frac{1}{T} \int_0^T \frac{\partial^{j+1} F(\theta, X_s)}{\partial \theta^{j+1}} ds \right)^2 \right) \leq C,$$

hence, for any $\theta_1, \theta_2 \in \Theta$,

$$\begin{aligned} \mathbb{E}(V_n^j(\theta_1) - V_n^j(\theta_2))^2 &\leq |\theta_1 - \theta_2| \left| \int_{\theta_1}^{\theta_2} \mathbb{E}\left(\frac{\partial V_n^j(u)}{\partial \theta} \right)^2 du \right| \\ &\leq C(\theta_1 - \theta_2)^2. \end{aligned} \tag{2.19}$$

It also follows from Lemma 2.6 that there exists a constant C such that

$$\sup_{\theta \in \Theta} \mathbb{E}(V_n^j(\theta))^2 < C.$$

This fact, together with (2.18), (2.19) and Theorem 20 in [10], yields

$$\sup_{\theta \in \Theta} \left| \frac{\partial^j \tilde{U}_n(\theta)}{\partial \theta^j} - \frac{1}{T} \int_0^T \frac{\partial^j F(\theta, X_s)}{\partial \theta^j} ds \right| \xrightarrow{\mathbb{P}} 0.$$

By taking into account Lemma 2.8 we get the desired result. □

Lemma 2.11. *For each $\delta \in (0, \delta_0)$, there exists a constant C such that*

$$\mathbb{E}\left| \frac{\partial U_n(\theta^*)}{\partial \theta} \right| \leq C(\Delta_n)^\delta.$$

Proof. We have

$$\begin{aligned} \left| \mathbb{E}\left(\frac{\partial \tilde{\zeta}_i^n(\theta^*)}{\partial \theta} \middle| \mathcal{F}_{i-1}^n \right) \right| &\leq \frac{C}{nA(X_i^n)^4} \left| 2\psi_n(\theta^*, X_i^n) - \frac{\pi}{\Delta_n} \mathbb{E}(|\Delta_i^n Z \Delta_{i+1}^n Z| |\mathcal{F}_{i-1}^n) \right| \\ &\leq \frac{C}{nA(X_i^n)^4} \bigg(\left| \psi_n(\theta^*, X_i^n) - \sigma(\theta^*, X_i^n)^2 \right| \\ &\qquad + \left| \sigma(\theta^*, X_i^n)^2 - \frac{\pi}{2\Delta_n} \mathbb{E}(|\Delta_i^n Z \Delta_{i+1}^n Z| |\mathcal{F}_{i-1}^n) \right| \bigg) \\ &\leq C(\Delta_n)^{3/2}, \end{aligned}$$

hence it follows from Lemma 2.6 that

$$\mathbb{E}\bigg(\frac{\partial\tilde{U}_n(\theta^*)}{\partial\theta}\bigg)^2$$
$$\leq 2\mathbb{E}\bigg(\sum_{i=1}^{n-1}\bigg(\frac{\partial\tilde{\zeta}_i^n(\theta^*)}{\partial\theta}-\mathbb{E}\bigg(\frac{\partial\tilde{\zeta}_i^n(\theta^*)}{\partial\theta}\bigg|\mathcal{F}_{i-1}^n\bigg)\bigg)\bigg)^2+2\mathbb{E}\bigg(\sum_{i=1}^{n-1}\mathbb{E}\bigg(\frac{\partial\tilde{\zeta}_i^n(\theta^*)}{\partial\theta}\bigg|\mathcal{F}_{i-1}^n\bigg)\bigg)^2$$
$$\leq 2\sum_{i=1}^{n-1}\mathbb{E}\bigg(\frac{\partial\tilde{\zeta}_i^n(\theta^*)}{\partial\theta}\bigg)^2+(nC(\Delta_n)^{3/2})^2$$
$$\leq C\Delta_n.$$

Therefore,

$$\mathbb{E}\bigg(\frac{\partial\tilde{U}_n(\theta^*)}{\partial\theta}\bigg)^2\leq C\Delta_n.$$

Combining this relation with Lemma 2.8 yields the desired result. □

2.4.3. Proof of Theorem 2.1

We are now in position to give proof of the main theorem. First, we will show that

$$\hat{\theta}_n\xrightarrow{\mathbb{P}}\theta^*\quad as\quad n\to\infty. \tag{2.20}$$

Let us denote

$$\xi_\theta=\int_0^T\frac{(\sigma(\theta,X_s)^2-\sigma(\theta^*,X_s)^2)^2}{A(X_s)^6}ds,$$

and for each $\epsilon,\eta>0$,

$$C(\epsilon,\eta)=\{\omega:\inf_{\theta\in\Theta:|\theta-\theta^*|>\epsilon}\xi_\theta(\omega)\geq\eta\}.$$

Because of condition (A2), $\lim_{\eta\to 0}\mathbb{P}(C(\epsilon,\theta))=1$. Since

$$C(\epsilon,\eta)\cap\bigg\{\sup_{\theta\in\Theta}|U_n(\theta)-\frac{1}{T}\int_0^T F(\theta,X_s)ds|\leq\eta/2\bigg\}\ \subset\ \{|\hat{\theta}_n-\theta^*|\leq\epsilon\},$$

taking into account Lemma 2.10 we get (2.20).

Next, applying Taylor's expansion, we have

$$\frac{\partial U_n(\theta^*)}{\partial\theta}-\frac{\partial U_n(\hat{\theta}_n)}{\partial\theta}=(\theta^*-\hat{\theta}_n)\frac{\partial^2U_n(\mu_n)}{\partial\theta^2},$$

where μ_n is a random point between θ^* and $\hat{\theta}_n$. Since θ^* is in the interior of Θ, for any n large enough we have $\frac{\partial U_n(\hat{\theta}_n)}{\partial\theta}=0$, and

$$\frac{\partial U_n(\theta^*)}{\partial\theta}=(\theta^*-\hat{\theta}_n)\frac{\partial^2U_n(\mu_n)}{\partial\theta^2}. \tag{2.21}$$

It follows from (2.20) that $\mu_n \xrightarrow{\mathbb{P}} \theta^*$, and by virtue of Lemma 2.10, we have

$$\frac{\partial^2 U_n(\mu_n)}{\partial \theta^2} \xrightarrow{\mathbb{P}} \frac{1}{T}\int_0^T \frac{\partial^2 F(\theta^*, X_s)}{\partial \theta^2} > 0 \quad a.s., \tag{2.22}$$

where the last equality follows from condition (A2). By Lemma 2.11, for any $\delta \in (0, \delta_0)$, we have

$$n^\delta \, \frac{\partial U_n(\theta^*)}{\partial \theta} \xrightarrow{\mathbb{P}} 0.$$

Combining this fact with (2.21), (2.22) yields

$$n^\delta(\hat{\theta}_n - \theta^*) \xrightarrow{\mathbb{P}} 0, \quad \text{as} \quad n \to \infty,$$

for any $\delta \in (0, \delta_0)$, and this relation completes the proof.

3. Spot volatility estimator-based approach

3.1. Preliminary

In this section, we consider a process (X, Y) defined on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,T]}, \mathbb{P})$ and given by the following stochastic differential equation

$$dY(t) = b(\theta, X_t, Y_t)dt + \sigma(\theta, X_t)dW(t) + dJ(t), \quad 0 \le t \le T, \tag{3.1}$$

where W is a Brownian motion, J is a Lévy process without Brownian part, parameter θ belongs to Θ, which is a compact subset of $\mathbb{R}$. Assume that we observe X without micorstructure noise at time grid $t_i^n = iT/n$ for $i = 0, 1, \ldots, n$. Process Y is observed with microstructure noise at another time grid $t_j^m = jT/m$ for $j = 0, 1 \ldots, m$. More precisely, at each time t_j^m, we cannot observe $Y(t_j^m)$ but $\tilde{Y}(t_j^m) = Y(t_j^m) + \epsilon(t_j^m)$ with $\epsilon(.)$ being a microstructure noise. We suppose that $\epsilon(.)$ satisfies the following assumption.

(MN). i) For any $p > 0$, there exists a constant ϑ_p such that

$$\sup_{j,m} \mathbb{E}|\epsilon(t_j^m)|^p < \vartheta_p < \infty.$$

ii) For each m, the random variables $\{\epsilon(t_j^m); j = 0, \ldots, m\}$ are independent and have the same expectations.

In the following, we will denote $X_i^n = X(t_i^n)$. The following assumption plays a key role in the construction of our estimators.

(B1). There exist estimators of $\{\sigma^2(\theta, X_i^n); i = 0, \ldots, n\}$, called $\{\hat{\sigma}(t_i^n)^2;\ i = 0, \ldots, n\}$, which is based on $\{\tilde{Y}(t_j^m);\ j = 0, \ldots, m\}$, such that

$$\sup_{i,n} \mathbb{E}|\hat{\sigma}(t_i^n)^2 - \sigma(\theta^*, X_i^n)^2| \le L\left(\frac{1}{m}\right)^{\delta_0}, \tag{3.2}$$

and

$$\sup_{i,n} \mathbb{E}\left(\frac{\hat{\sigma}(t_i^n)^2}{\sigma(\theta^*, X_i^n)^2 + 1}\right)^3 \leq L, \tag{3.3}$$

for some positive constants L and δ_0. Here, θ^* is the true value of parameter θ. Assume further that $m = O(n^\kappa)$ for some $\kappa > 0$ and denote $\delta = \kappa\delta_0$.

In Section 4, under conditions on the integrability and Hölder continuity of coefficients b and σ, we will present several classes of estimator $\{\hat{\sigma}(t_i^n)^2\}$ which satisfy (B1). We will also make use of the following assumptions which are gathered here for easy reference:

(B2). Function $\sigma(\theta, x)$ is two times differential in θ, and there exists a function $A : \mathbb{R} \to [1, \infty)$ such that

$$\sup_{\theta \in \Theta} \left(|\sigma(\theta, x)| + |\sigma'_\theta(\theta, x)| + |\sigma''_{\theta\theta}(\theta, x)|\right) \leq A(x). \tag{3.4}$$

(B3). For any $\epsilon > 0$,

$$\liminf_{n\to\infty} \inf_{\theta\in\Theta:|\theta-\theta^*|>\epsilon} \frac{1}{n} \sum_{i=1}^n \left| \frac{\sigma(\theta, X_i^n)^2 - \sigma(\theta^*, X_i^n)^2}{A(X_i^n)^2} \right| > 0 \quad a.s. \tag{3.5}$$

(B4).

$$\liminf_{n\to\infty} \frac{1}{n} \sum_{i=1}^n \frac{1}{A(X_i^n)^4} \left(\frac{\partial\sigma(\theta^*, X_i^n)^2}{\partial\theta}\right)^2 > 0 \quad a.s. \tag{3.6}$$

3.2. Estimator

The contrast function is defined as follows

$$g_n(\theta) = \frac{1}{n} \sum_{i=1}^n \left(\frac{\sigma(\theta, X_i^n)^2 - \hat{\sigma}(t_i^n)^2}{A(X_i^n)^2}\right)^2. \tag{3.7}$$

Since function $\theta \mapsto g_n(\theta)$ is continuous, it has a minimum on the compact set Θ, and due to the measurable selection theorem we can find a measurable (with respect to the observed sigma algebra at stage n) variable $\hat{\theta}_n$ satisfying

$$g_n(\hat{\theta}_n) = \inf_{\theta\in\Theta} g_n(\theta). \tag{3.8}$$

Proposition 3.1. *Suppose that Assumptions (B1) and (B3) hold. Then*

$$\hat{\theta}_n \xrightarrow{\mathbb{P}} \theta^*.$$

Proof. For each n and $\epsilon > 0$, we denote

$$\zeta_n(\epsilon) = \inf_{\theta\in\Theta:|\theta-\theta^*|>\epsilon} \frac{1}{n} \sum_{i=1}^n \left| \frac{\sigma(\theta, X_i^n)^2 - \sigma(\theta^*, X_i^n)^2}{A(X_i^n)^2} \right|.$$

It follows from Assumption (B3) that, for any $\epsilon > 0$, there exists $\epsilon_1 > 0$ such that the event

$$A_\infty = \big\{ \liminf_{n\to\infty} \zeta_n(\epsilon) > \epsilon_1 \big\}$$

has probability large than $1 - \epsilon/3$. We denote

$$A_n = \left\{ \inf_{k\geq n} \zeta_k(\epsilon) > \frac{\epsilon_1}{2} \right\}.$$

Since the sequence $\{A_k; k \geq 1\}$ is nondecreasing and $A_\infty \subset \bigcup_{n=1}^\infty A_n$, there exists $n_0 \in \mathbb{Z}^+$ such that

$$\inf_{n\geq n_0} \mathbb{P}(A_n) \geq 1 - \frac{\epsilon}{2}. \tag{3.9}$$

On the other hand, we have

$$\begin{aligned}
\mathbb{P}(|\hat{\theta}_n - \theta^*| > \epsilon) \leq & \mathbb{P}\big(\inf_{\theta\in\Theta:|\theta-\theta^*|>\epsilon} g_n(\theta) < g_n(\theta^*) \big) \\
\leq & \mathbb{P}\Big(\zeta_n(\epsilon) < 4 g_n(\theta^*) \Big) \\
\leq & \mathbb{P}\Big(\zeta_n(\epsilon) < \frac{\epsilon_1}{2} \Big) + \mathbb{P}\Big(g_n(\theta^*) > \frac{\epsilon_1}{8} \Big).
\end{aligned}$$

Hence, it follows from (3.9) and Markov's inequality that

$$\limsup_{n\to\infty} \mathbb{P}(|\hat{\theta}_n - \theta^*| > \epsilon) \leq \frac{\epsilon}{2} + \limsup_{n\to\infty} 8\epsilon_1 \mathbb{E} g_n(\theta^*). \tag{3.10}$$

Furthermore, it follows from Hölder's inequality and Assumption (B1) that

$$\begin{aligned}
\mathbb{E} g_n(\theta^*) \leq & \frac{1}{n} \sum_{i=1}^n \left(\mathbb{E}|\sigma(\theta, X_i^n)^2 - \hat{\sigma}(t_i^n)^2| \; \mathbb{E}\left| \frac{\sigma(\theta, X_i^n)^2 - \hat{\sigma}(t_i^n)^2}{A(X_i^n)^2} \right|^3 \right)^{1/2} \\
\leq & \big(8L(1+8L) \big)^{1/2} \frac{1}{n^{\delta/2}}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\limsup_{n\to\infty} \mathbb{P}(|\hat{\theta}_n - \theta^*| > \epsilon) \leq & \frac{\epsilon}{2} + 8\epsilon_1 \big(8L(1+8L) \big)^{1/2} \limsup_{n\to\infty} \frac{1}{n^{\delta/2}} \\
= & \frac{\epsilon}{2},
\end{aligned}$$

for any $\epsilon > 0$. This relation yields the desired result. □

Now we are able to state the main theorem of this section. It tells us that the parametric estimator $\hat{\theta}_n$ converges at the same rate as the estimator for spot volatility does. More discussion about the rate of convergence will be provided at the end of Section 4.

Theorem 3.2. *Suppose that Assumptions (B1)–(B4) hold. Then if the true parameter θ^* is in the interior of Θ, the estimators $\hat{\theta}_n$ are n^δ-consistent, in the sense that the sequence $n^\delta(\hat{\theta}_n - \theta^*)$ is tight.*

Proof. Applying Taylor's expansion, we have

$$\frac{\partial g_n(\theta^*)}{\partial\theta} - \frac{\partial g_n(\hat{\theta}_n)}{\partial\theta} = \frac{1}{2}(\theta^* - \hat{\theta}_n)\frac{\partial^2 g_n(w_n)}{\partial\theta^2}, \tag{3.11}$$

where w_n is a random point between θ^* and $\hat{\theta}_n$. Let us denote $S(\theta, x) = \sigma(\theta, x)^2$. We write $\frac{\partial^2 g_n(w_n)}{\partial\theta^2} = \sum_{k=1}^4 Z_k^n$ with

$$Z_1^n = \frac{1}{n}\sum_{i=1}^n S''_{\theta\theta}(\theta^*, X_i^n)(S(\theta^*, X_i^n) - \hat{\sigma}(t_i^n)^2)A(X_i^n)^{-4},$$

$$Z_2^n = \frac{1}{n}\sum_{i=1}^n (S'_\theta(w_n, X_i^n)^2 - S'_\theta(\theta^*, X_i^n)^2)A(X_i^n)^{-4},$$

$$Z_3^n = \frac{1}{n}\sum_{i=1}^n S''_{\theta\theta}(w_n, X_i^n)(S(w_n, X_i^n) - S(\theta^*, X_i^n))A(X_i^n)^{-4},$$

$$Z_4^n = \frac{1}{n}\sum_{i=1}^n S'_\theta(\theta^*, X_i^n)^2 A(X_i^n)^{-4}.$$

For the first term Z_1^n, it follows from Assumptions (B1) and (B2) that

$$\mathbb{E}|Z_1^n| \le \frac{3}{n}\sum_{i=1}^n \mathbb{E}|\sigma(\theta^*, X_i^n)^2 - \hat{\sigma}(t_i^n)^2| \le 3L\frac{1}{n^\delta},$$

hence

$$n^\delta \mathbb{E}|Z_1^n| \le 3L. \tag{3.12}$$

To estimate the second term, we apply Taylor's expansion again,

$$Z_2^n = \frac{1}{n}\sum_{i=1}^n \frac{\partial S'_\theta(w_n^i, X_i^n)^2}{\partial\theta} A(X_i^n)^{-4}(w_n - \theta^*),$$

where w_n^i is a random point between θ^* and w_n for each $i = 1, \ldots, n$. By virtue of Assumption (B2), we get

$$\frac{\partial S'_\theta(\theta, x)^2}{\partial\theta} \le 8A(x)^4,$$

hence

$$|Z_2^n| \le \frac{8}{n}\sum_{i=1}^n |w_n^i - \theta^*| \le 8|w_n - \theta^*|. \tag{3.13}$$

Similarly, we have

$$|Z_3^n| \le 6|w_n - \theta^*|, \tag{3.14}$$

and

$$n^{\delta}\mathbb{E}\left|\frac{\partial g_n(\theta^*)}{\partial\theta}\right| \le 4L. \tag{3.15}$$

Furthermore, thank to Assumption (B4), for any $\epsilon > 0$, there exist constants $\kappa_1 > 0$ and $n_0 \in \mathbb{Z}^+$ such that

$$\mathbb{P}\big(\inf_{n\ge n_0} Z_4^n \ge 15\kappa_1\big) \ge 1 - \frac{\epsilon}{4}. \tag{3.16}$$

Since θ^* is in the interior of Θ, there exists a positive constant κ_2 such that $(\theta^* - \kappa_2, \theta^* + \kappa_2) \subset \Theta$ and $\kappa_2 < \kappa_1$. On the other hand, for any $\epsilon > 0$, it follows from Proposition 3.1 that

$$\sup_{n\ge n_1} \mathbb{P}(|\theta^* - \hat{\theta}_n| \ge \kappa_2) < \frac{\epsilon}{4}, \tag{3.17}$$

for some $n_1 \in \mathbb{Z}^+,\ n_1 \ge n_0$.
For each $n \ge n_1$, we denote

$$A_n = \{Z_4^n - 14|\hat{\theta}_n - \theta^*| \ge \kappa_2\} \cap \{|\theta^* - \hat{\theta}_n| < \kappa_2\}.$$

It follows from (3.16) and (3.17) that

$$\inf_{n\ge n_1} \mathbb{P}(A_n) \ge 1 - \frac{\epsilon}{2}. \tag{3.18}$$

Equation (3.11), together with (3.13) and (3.14), yields

$$n^{\delta}\left(\left|\frac{\partial g_n(\theta^*)}{\partial\theta}\right| + \left|\frac{\partial g_n(\hat{\theta}_n)}{\partial\theta}\right|\right) \ge n^{\delta}|\theta^* - \hat{\theta}_n|(Z_4^n - 14|\theta^* - \hat{\theta}_n|) - |\theta^* - \hat{\theta}_n|n^{\delta}|Z_1^n|.$$

Hence, by the virtue of (3.18), for any constant $\kappa_3 > 0$, we have

$$\begin{aligned}\mathbb{P}(n^{\delta}|\theta^* - \hat{\theta}_n| > \kappa_3) &\le \frac{\epsilon}{2} + \mathbb{P}\left(\kappa_2^{-1}n^{\delta}\left|\frac{\partial g_n(\theta^*)}{\partial\theta}\right| + n^{\delta}|Z_1^n| \ge \kappa_3; A_n\right)\\ &\le \frac{\epsilon}{2} + \frac{1}{\kappa_2\kappa_3}\mathbb{E}\left(n^{\delta}\left|\frac{\partial g_n(\theta^*)}{\partial\theta}\right| + \kappa_2 n^{\delta}|Z_1^n|\right)\end{aligned}$$

for any $n \ge n_1$. Therefore, it follows from (3.12) and (3.15) that

$$\sup_{n\ge n_1} \mathbb{P}(n^{\delta}|\theta^* - \hat{\theta}_n| > \kappa_3) \le \frac{\epsilon}{2} + \frac{(4 + 3\kappa_2)L}{\kappa_2\kappa_3},$$

which yields the desired result. □

3.3. Numerical Examples

3.3.1. Example: X is a deterministic process

Let $(X, Y, \tilde{Y})$ defined by

$$\begin{cases} X(t) = 2t - t^3,\ 0 \le t \le 1,\\ dY(t) = (2 + \sin(\pi\theta X(t)))dW(t) + dJ(t),\ Y(0) = 1\\ \tilde{Y}(t) = Y(t) + \epsilon(t), \end{cases}$$

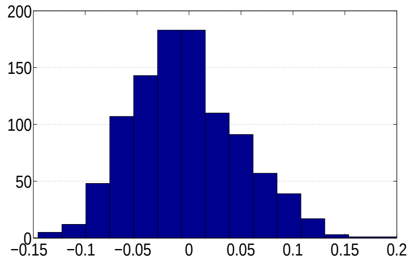

Fig 3. *Histogram of Error.*

where J is a stable Lévy process with stable index $\beta = 0.7$, and $\epsilon(.)$ has normal distribution $\mathcal{N}(0, \varepsilon^2)$.

Here we have $\sigma(\theta, x) = 2 + \sin(\pi\theta x)$ and this function satisfies assumptions (B2)–(B4) with $A(x) = O(1 + x^2)$ for any value of parameter θ.

Let us take $\theta = 4, \varepsilon = 10^{-3}$ and simulate a discrete sample path of $\{X(t), Y(t), \tilde{Y}(t)\}$ by using Euler's method with very small time-discretization step. We choose $m = 10^5, n = 10^3, L = M = 100$ and use the estimator (4.2) to approximate spot volatility $\sigma(\theta, X(t_i))^2$ for $t_i = i/n,\ i = 0, \ldots, n$. Then we calculate the estimator $\hat{\theta}$ with $A(x) = 1 + x^2$.

After iterating the simulation 1000 times, we get the histogram for the error $\hat{\theta} - \theta$ in Figure 3.

3.3.2. Example: X is a stochastic process

Let $(X, Y, \tilde{Y})$ defined by

$$\begin{cases} dX(t) = X(t)dt + X(t)d\tilde{W}(t),\ X(0) = 1,\ 0 \le t \le 1, \\ dY(t) = \theta X(t)dW(t) + dJ(t),\ Y(0) = 1 \\ \tilde{Y}(t) = Y(t) + \epsilon(t), \end{cases}$$

where W and $\tilde{W}$ are two Brownian motion, J is a stable Lévy process with stable index $\beta = 0.7$, and $\epsilon(.)$ has normal distribution $\mathcal{N}(0, \varepsilon^2)$.

Here we have $\sigma(\theta, x) = \theta x$ and this function satisfies assumptions (B2)–(B4) with $A(x) = O(1 + |x|)$ for any value of parameter $\theta > 0$.

Like in the previous example, let us take $\theta = 4, \varepsilon = 10^{-3}$ and simulate a discrete sample path of $\{X(t), Y(t), \tilde{Y}(t)\}$ by using Euler's method with very small time-discretization step. We choose $m = 10^6, n = 10^3, M = 10^3, L = 50$ and use the estimator (4.3) to approximate spot volatility $\sigma(\theta, X(t_i))^2$ for $t_i = i/n,\ i = 0, \ldots, n$. Then we calculate the estimator $\hat{\theta}$ with $A(x) = 1 + |x|$.

After iterating the simulation 1000 times, we get the histogram for the error $\hat{\theta} - \theta$ in Figure 4.

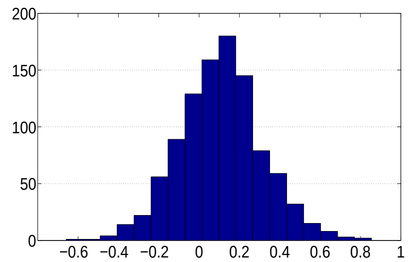

FIG 4. *Histogram of Error.*

Remark. In practice, it is necessary and very important to find an effective concrete scheme to compute $\hat{\theta}_n$ in formulae (2.6) and (3.8). This problem will be discussed in future work.

4. Spot volatility estimators

In this section, we present some simple estimation schemes for the spot volatility and study their rate of convergence in L^1-sense. A further discussion about these schemes can be found in [23], and an improvement of them was presented in [27]. In [29, 30], similar schemes were proposed for estimating integrated volatility. Other estimation schemes, which use the Fourier series method, were introduced in [15–17, 19, 21].

We consider a stochastic process $(Y(t))_{t\geq 0}$ defined on the filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ given by

$$dY(t) = A(t)dt + B(t)dW(t) + dJ(t), \quad 0 \leq t \leq T, \tag{4.1}$$

where W is a standard Brownian motion, J a Lévy process with no Brownian part, A a measurable drift process, and B a process which is adapted to the sigma algebra $(\mathcal{F}_t)_{t\geq 0}$. Let α denote the Blumenthal-Getoor index of J.

Before stating our estimators, we introduce the following assumptions:

(C1).
$$\forall q > 0, \quad \sup_{0\leq t\leq T} \mathbb{E}(|A(t)|^q + |B(t)|^q) < \infty,$$

and $\mathcal{F}_t$ is independent of the sigma algebra $\sigma(W_s - W_t, s > t)$ for all $t \geq 0$.

(C2). The volatility coefficient B satisfies the following Hölder continuity condition
$$\mathbb{E}|B(t) - B(s)|^2 \leq L|t-s|^{2\beta}, \quad \forall\ 0 \leq t < s \leq T,$$

for some $\beta \in (0, 1]$, where L is a constant.

4.1. Ideal case

Let $k_m = O(m^{\frac{2\beta}{2\beta+1}})$. For $i = k_m + 1, \ldots, m - k_m - 1$, we denote

$$\hat{B}(t_i^m)^2 = \frac{\pi}{2(2k_m+1)\Delta_n} \sum_{j=-k_m}^{k_m} |\Delta_{i+j}^m Y \Delta_{i+j+1}^m Y| I_{[|\Delta_{i+j}^m Y| < \Delta_m^\gamma, |\Delta_{i+j+1}^m Y| < \Delta_m^\gamma]},$$
$$\hat{B}(t)^2 = \hat{B}(t_i^m)^2 \quad \text{for} \quad t_i^m \le t < t_{i+1}^m, \tag{4.2}$$

with some $\gamma \in [0, 1/2)$ and $\Delta_i^m Y = Y(i\Delta_m) - Y((i-1)\Delta_m),\ \Delta_m = T/m$.

Proposition 4.1. *We assume that (C1), (C2) hold, the Blumenthal-Getoor index $\alpha < 2$. Then for any $p \in (\alpha, 2)$, there exist a constant c_p such that*

$$\sup_{0 \le t \le T} \mathbb{E}|\hat{B}(t)^2 - B(t)^2| \le c_p\, m^{-\min(\frac{\beta}{2\beta+1}, \frac{1}{p} - \frac{1}{2})}.$$

This proposition can be proved by following the arguments in the proofs of Theorem 3.1[23] and Proposition 2 [25].

4.2. Noisy case

In this section, we consider the case that the observation of Y is corrupted by noise. In other words, the observed data is not $Y(t_i^m)$ but rather $\tilde{Y}(t_i^m) = Y(t_i^m) + \epsilon(t_i^m)$. We call $\epsilon(.)$ microstructure noise. It is visibly clear that the estimator $\hat{B}(.)^2$ will explode as the number of observations increase if we replace Y with $\tilde{Y}$.

Let (M_m) and (L_m) be some nondecreasing sequences of positive integers. For each $L_m M_m \le i \le m - (L_m + 5)M_m$, we denote

$$\overline{Y}_{i,k}^m = \frac{1}{M_m} \sum_{j=0}^{M_m - 1} \tilde{Y}(t_{i+M_m k+j}^m),\ k = -L_m, \ldots, L_m + 4,$$

and

$$\Delta\overline{Y}_{i,k}^m = \overline{Y}_{i,k+1}^m - \overline{Y}_{i,k}^m,\ k = -L_m, \ldots, L_m + 3.$$

Furthermore, we denote

$$\chi = \left(\frac{1}{2\pi} \int_{\mathbb{R}} e^{-x^2/2} |x|^{2/3} dx \right)^3.$$

The estimator is defined by

$$\tilde{B}(t_i^m)^2 = \frac{3M_m}{2L_m(2M_m^2+1)\Delta_m \chi} \sum_{k=-L_m}^{L_m - 1} |\Delta\overline{Y}_{i,k}^m\, \Delta\overline{Y}_{i,k+2}^m\, \Delta\overline{Y}_{i,k+4}^m|^{2/3}$$
$$\tilde{B}(t)^2 = \tilde{B}(t_i^m)^2 \quad \text{for} \quad t_i^m \le t < t_{i+1}^m. \tag{4.3}$$

Proposition 4.2. *We assume that (C1), (C2) and (MN) hold, the Blumenthal-Getoor index $\alpha \le 1$. Then there exists a constants C such that*

$$\sup_{0\le t\le T} \mathbb{E}|\tilde{B}(t)^2 - B(t)^2| \le C\big((M_m\Delta_m)^{1/3} + (M_m^2\Delta_m)^{-1/3} + (M_mL_m\Delta_m)^{2\beta/3} + L_m^{-1/2}\big) \quad (4.4)$$

In particular:

i) If $\beta = 1$, we choose $M_m = O(m^{13/20}),\ L_m = O(m^{1/5})$, then

$$\sup_{0\le t\le T} \mathbb{E}|\tilde{B}(t)^2 - B(t)^2| \le C\frac{1}{m^{1/10}}.$$

ii) If $\beta = 1/2$, we choose $M_m = O(m^{8/13}),\ L_m = O(m^{2/13})$, then

$$\sup_{0\le t\le T} \mathbb{E}|\tilde{B}(t)^2 - B(t)^2| \le C\frac{1}{m^{1/13}}.$$

When process Y does not contain a jump part, i.e., $J \equiv 0$, we propose another estimator, which has a better rate of convergence, as follows.

$$\breve{B}(t_i^m)^2 = \frac{3\pi M_m}{4L_m(2M_m^2+1)\Delta_m}\sum_{k=-L_m}^{L_m-1} |\Delta\overline{Y}_{i,k}^m\,\Delta\overline{Y}_{i,k+2}^m|,$$
$$\breve{B}(t)^2 = \breve{B}(t_i^m)^2 \quad \text{for} \quad t_i^m \le t < t_{i+1}^m. \quad (4.5)$$

Proposition 4.3. *We assume that (C1), (C2) and (MN) hold and $J \equiv 0$. Then there exists a constant C such that*

$$\sup_{0\le t\le T} \mathbb{E}|\breve{B}(t)^2 - B(t)^2| \le C\big((M_m\Delta_m)^{1/2} + (M_m^2\Delta_m)^{-1/2} + (M_mL_m\Delta_m)^{\beta} + L_m^{-1/2}\big) \quad (4.6)$$

In particular:

i) If $\beta = 1$, we choose $M_m = O(m^{5/8}),\ L_m = O(m^{1/4})$, then

$$\sup_{0\le t\le T} \mathbb{E}|\breve{B}(t)^2 - B(t)^2| \le C\frac{1}{m^{1/8}}.$$

ii) If $\beta = 1/2$, we choose $M_m = O(m^{3/5}),\ L_m = O(m^{1/5})$, then

$$\sup_{0\le t\le T} \mathbb{E}|\breve{B}(t)^2 - B(t)^2| \le C\frac{1}{m^{1/10}}.$$

Propositions 4.2 and 4.3 can be proven by closely following the argument in [23].

Table 1 clarifies the rates of convergence of parametric estimator $(\hat{\theta}_n)$ in Theorem 3.2 in various situations mentioned above. It is worth to note that these rates of convergence may not be optimal. An improvement of the spot volatility estimation scheme will lead to a better rate for estimator $(\hat{\theta}_n)$. In particular, it has been shown in [21] that under more restricted assumptions on the model of Y, one can propose a scheme which has better rates of convergence in Proposition 4.3.

Table 1
Rates of convergence of parametric estimator $(\hat{\theta}_n)$

Ideal Case				Noisy Case			
$J \equiv 0$		$J \not\equiv 0$		$J \equiv 0$		$J \not\equiv 0, \alpha \leq 1$	
$\beta = 1$	$\beta = 1/2$	$\beta = 1$	$\beta = 1/2$	$\beta = 1$	$\beta = 1/2$	$\beta = 1$	$\beta = 1/2$
$n^{-\frac{1}{3}\kappa}$	$n^{-\frac{1}{4}\kappa}$	$n^{-(\frac{1}{3} \wedge \frac{2-\alpha}{2\alpha})\kappa}$	$n^{-(\frac{1}{4} \wedge \frac{2-\alpha}{2\alpha})\kappa}$	$n^{-\frac{1}{8}\kappa}$	$n^{-\frac{1}{10}\kappa}$	$n^{-\frac{1}{10}\kappa}$	$n^{-\frac{1}{13}\kappa}$

Acknowledgement

The author is grateful to Professor Shigeyoshi Ogawa for very useful discussions. He also would like to thank two anonymous Referees and an Associate Editor for their valuable comments and suggestions which were helpful while improving the early version of the paper.

References

[1] Alvarez, A., Panloup, F., Pontier, M., and Savy, N. (2010) Estimation of the instantaneous volatility. Preprint arXiv:0812.3538.

[2] Barndorff-Nielsen, O.E., Shephard, N. and Winkel, M. (2006). Limit theorems for multipower variation in the presence of jumps. *Stochastic Process. Appl.* **116**(5) 796–806. MR2218336

[3] Bodo, B. A. and Thomson, M. E. and Unny, T. E. (1987) A review on stochastic differential equations for applications in hydrology. *Stochastic Hydrol. Hydraul.* **1**(2) 81–100.

[4] Brigo, D. and Mercurio, F. (2007) *Interest rate models - Theory and Practice. With smile, inflation and credit.* 2nd. ed. Springer. MR2255741

[5] Cont, R. and Tankov, P. (2004). *Financial modeling with jump processes.* Chapman & Hall/CRS Financial Math. Ser. MR2042661

[6] Gukhal, C. R.(2001) Analytical valuation of American options on jump-diffusion processes. *Math. Finance* **11**(1) 97-115. MR1807850

[7] Guttrop, P. and Kulperger, R. (1984) Statistical inference for some Volterra population processes in a random environment. *Can. J. Statist.* **12** 289–302. MR0782951

[8] Hanson, F. B. and Tuckwell, H. C. (1981) Logistic growth with random density independent disasters. *Theor. Pop. Biol.* **19** 1–18.

[9] Iacus, S.M. and Yoshida, N. (2009) Estimation for the change point of the volatility in a stochastic differential equation. Preprint arXiv:0906.3108.

[10] Ibragimov, I.A. and Has'minskii, R.Z. (1981) *Statistical Estimation: Asymptotic Theory*. Springer-Verlag. MR0620321

[11] Jacod, J. (2006) Parametric inference for discretely observed non-ergodic diffusions. *Bernoulli* **12**(3) 383–401. MR2232724

[12] Kutoyants, Yu.A. (2004) *Statistical Inference for Ergodic Diffusion Processes.* Springer. MR2144185

[13] Liptser, R.S. and Shiryaev, A.N. (2001) *Statistic of Random Processes I: General Theory* 2nd ed. Springer-Verlag, Berlin. MR1800857

[14] Luschgy, H. and Pagès, G. (2008) Moment estimates for Lévy processes. *Electron. Commun. Probab.* **13** 422–434. MR2430710

[15] Malliavin, P. and Mancino, M. (2002) Fourier series method for measurement of multivariate volatilities. *Fin. Stoch.* **4** 49–61. MR1885583

[16] Malliavin, P. and Mancino, M. (2009) A Fourier transform method for nonparametric estimation of multivariate volatility. *Anal. Statist.* **4** 1983-2010 MR2533477

[17] Malliavin, P. and Thalmaier, A. (2006) *Stochastic calculus of variations in mathematical finance*. Springer Finance. MR2189710

[18] Mancini, C. (2004) Estimation of the characteristics of the jumps of a general Poisson-diffusion model. *Scand. Actuar. J.* **1** 42–52. MR2045358

[19] Mancino, M.E. and Sanfelici, S. (2008) Robustness of Fourier estimator of integrated volatility in the presence of microstructure noise. *Comp. Stat. & Data Analysis*, **52** 2966–2989.

[20] Mulinacci, S.(1996) An approximation of American option prices in a jump-diffusion model. *Stochastic Process. Appl.* **62**(1) 1-17. MR1388760

[21] Munk, A. and Schmidt-Hieber, J. (2010) Nonparametric estimation of the volatility function in a high-frequency model corrupted by noise. Preprint arXiv:0908.3163v3.

[22] Mtundu, N. D. and Koch, R. W. (1987) A stochastic differential equation approach to soil moisture. *Stochastic Hydrol. Hydraul.* **1**(2) 101–116.

[23] Ngo, H-L. and Ogawa, S. (2009) A central limit theorem for the functional estimation of the spot volatility. *Monte Carlo Method and Appl.* **15**(4) 353–380. MR2603470

[24] Ogawa, S. (2008) Real-time scheme for the volatility estimation in the presence of microstructure noise. *Monte Carlo Methods and Appl.* **14**(4) 331–342. MR2487774

[25] Ogawa, S. and Ngo, H-L. (2010) Real-time estimation scheme for the spot cross volatility of jump diffusion processes. *Math. and Comp. in Simulations.* **80** 1962–1976.

[26] Ogawa, S. and Ngo, H-L. (2008) Some remarks on the real-time scheme for the estimation of spot volatility. *Memoirs of the institute of Science and Engineering, Ritsumeikan University.* **67** 1–8. MR2543891

[27] Ogawa, S. and Sanfelici, S. (2009) An improved two-step regularization scheme for spot volatility estimation. Preprint.

[28] Ogawa, S. and Wakayama, K. (2007) On the real-time scheme for the estimation of volatility. *Monte Carlo Methods and Appl.* **13**(2) 99–116. MR2338084

[29] Podolskij, M. and Vetter, M. (2009) Estimation of volatility functionals in the simultaneous presence of microstructure noise and jumps. *Bernoulli.* **15**(3) 634–658. MR2555193

[30] Podolskij, M. and Vetter, M. (2009) Bipower-type estimation in a noisy diffusion setting. *Stochastic Process. Appl.* **119**(9) 2803-2831. MR2254029

[31] Prakasa Rao, B.L.S. (1999) *Statistical Inference for Diffusion Type Processes.* Oxford University Press, New York. MR1717690

[32] Scott, L. O. (1997) Pricing stock options in a jump-diffusion model with stochastic volatility and interest rates: applications of Fourier inversion methods. *Math. Finance* **7**(4) 413-426. MR1482711

[33] Shimizu, Y. and Yoshida, N. (2006) Estimation of parameters for diffusion processes with jumps from discrete observations. *Statist. Inference Stoch. Proc.* **9** 227–277. MR2252242

[34] Todorov, V. (2009) Estimation of continuous-time stochastic volatility models with jumps using high-frequency data. *J. Econometrics* **148**(2) 131–148. MR2500652

[35] van Schuppen, J. H. (2009) Mathematical control and system theory of stochastic systems in continuous time. Preprint. http://homepages.cwi.nl/~schuppen/courses/lecturenotes/controlstocct20090715.pdf