Finance Stoch (2007) 11: 495–519
DOI 10.1007/s00780-007-0048-2

Efficient estimation of drift parameters in stochastic volatility models

Arnaud Gloter

Received: 17 October 2005 / Accepted: 22 May 2007 / Published online: 8 August 2007

Abstract We study the parametric problem of estimating the drift coefficient in a stochastic volatility model $Y_t = \int_0^t \sqrt{V_s}\,\mathrm{d}W_s$, where Y is a log price process and V the volatility process. Assuming that one can recover the volatility, precisely enough, from the observation of the price process, we construct an efficient estimator for the drift parameter of the diffusion V. As an application we present the efficient estimation based on the discrete sampling $(Y_{i\delta_n})_{i=0,\dots,n}$ with $\delta_n \to 0$ and $n\delta_n \to \infty$. We show that our setup is general enough to cover the case of 'microstructure noise' for the price process as well.

Keywords Stochastic volatility model · Microstructure noise · Integrated volatility · Realized volatility · Efficient estimator

JEL Classification C13 · C15

Mathematics Subject Classification (2000) 62F12 · 62M09

1 Introduction

The aim of this paper is to estimate unknown parameters in the framework of the continuous time stochastic volatility models introduced by Hull and White [19]. Assume that the log-price process $Y_t = \log S_t$ of some asset S is given by the simple equation

$$\mathrm{d}Y_t = \mu(V_t)\,\mathrm{d}t + \sqrt{V_t}\,\mathrm{d}W_t,$$

A. Gloter (✉)
Laboratoire d'Analyse et de Mathématiques Appliquées, Université Paris-Est, UMR 8050,
5 boulevard Descartes, 77454 Marne-la-Vallée cedex 2, France
e-mail: arnaud.gloter@univ-mlv.fr

where the volatility process V_t is a solution of

$$\mathrm{d}V_t = b(V_t, \theta_0)\,\mathrm{d}t + a(V_t)\,\mathrm{d}B_t, \tag{1.1}$$

and (B, W) is a Brownian motion in $\mathbb{R}^2$.

A practical use of these models, for instance, for option pricing or volatility prediction, requires to estimate the unknown parameters governing the process V. However, this volatility process is unobservable; on the other hand, data coming from the log-price process (Y_t) are generally available. The problem of estimating the unknown parameters of the volatility process (V_t) from the discrete observation of the coordinate (Y_t) has been the subject of several recent contributions. As is always the case for discretely observed diffusion processes, a direct likelihood approach is hardly tractable.

Assuming that the volatility process is ergodic, several kinds of explicit estimators have been proposed. In the case of observations with fixed sampling step, empirical moment estimators are possible (see, e.g., [12, 25, 28]). However, such estimators may be strongly biased (see, e.g., [24]). Otherwise, the prediction-based estimating equations constructed by Sørensen [27] also provide explicit estimators. A maximum likelihood procedure is proposed in [1] where the unobserved components are reconstructed via observation of option prices. Filtering the unobserved components also yields a maximum likelihood procedure in [4].

Meanwhile, a growing literature in finance is devoted to the econometrics of high-frequency data with a view to reconstructing the unobserved volatility process (see [3, 21, 26, 31]).

A theoretical approach which is well-fitted to high-frequency data is to assume that we have some information on the price process at the instants $(t_{i,n})_{i=0,\dots,n}$, where the maximal sampling step $\delta_n = \sup_i (t_{i+1,n} - t_{i,n})$ converges to zero and $T_n = t_{n,n} \to \infty$. For instance, an explicit method based on the observation of a discrete sample $(Y_{j\delta_n}, j \le n)$ is proposed in [11]. This method is able to estimate the unknown parameters present in the stationary distribution of the unobserved diffusion V. Our purpose here is to infer the parameters of the (V_t) model present in the drift coefficient $b(., \theta_0)$, under the same ergodicity assumption on V. We focus mainly on the drift coefficient, for simplicity and since a positive answer can be obtained about the efficiency of the estimation procedure. Indeed, the problem of estimating the diffusion coefficient is known to be difficult in general (see [13, 18] in the case of non-ergodic volatility). Nevertheless, in the case where this diffusion coefficient is needed, we shall explain how to estimate it consistently before estimating the drift parameter.

The organization of the paper is as follows. In Sect. 2.1 we present our assumptions on the diffusion process V. Then we introduce an explicit contrast function in Sect. 2.2. The idea is that from the high frequency observations we know how to reconstruct the integrated volatility $\int_{i\Delta}^{(i+1)\Delta} V_s\,\mathrm{d}s$ over intervals of size Δ larger than δ_n. The problem of estimating the parameter in (1.1) from discrete sampling of its integral has been addressed in our two previous works [14, 15]. Relying on these works the study of a contrast function is feasible. The main result is given in Sect. 2.3. It states that, under suitable conditions on the quality of the reconstruction of the integrated volatility, our estimator for the drift parameter is consistent and asymptotically Gaussian. An interesting property is that our estimator has the same asymptotic behavior as the maximum likelihood estimator based on the direct observation of the volatility $(V_t, t \le T_n)$; especially, it converges with rate $\sqrt{T_n}$. Thus, no information about the drift of the volatility process is lost in the observation of the price process. Moreover, we can deduce that our estimator is asymptotically efficient (see Remark 2.5).

In Sect. 3 we apply this result to the case of a direct observation of the price process Y when the volatility is reconstructed with the help of the so-called 'realized volatility'. We study the estimator on explicit models and compare its behavior on finite sample with the estimator given in [11].

In Sect. 4 we show that our procedure applies even if the price process is contaminated by some microstructure noise provided the volatility is reconstructed by use of the 'two scales realized volatility' introduced by Zhang et al. [31].

We finally discuss in Sect. 5 the estimation of the diffusion coefficient of the volatility process.

Section 7 is devoted to the proofs of the results.

2 Main results

2.1 Assumptions on the model

Let (Y_t, V_t) be the two-dimensional diffusion process defined as the solution on a probability space $(\Omega, \mathcal{A}, P)$ of

$$\mathrm{d}Y_t = \mu(V_t)\,\mathrm{d}t + \sqrt{V_t}\,\mathrm{d}W_t, \qquad Y_0 = \eta', \tag{2.1}$$

$$\mathrm{d}V_t = b(V_t, \theta)\,\mathrm{d}t + a(V_t)\,\mathrm{d}B_t, \qquad V_0 = \eta, \tag{2.2}$$

where $(B_t, W_t)_{t\ge 0}$ is a two-dimensional Brownian motion with possible correlation between the components. We assume that the initial random variables (η, η') are independent of $(W_t, B_t)_{t\ge 0}$ and that η is positive.

We assume that the unknown parameter θ lies in a compact subset Θ of $\mathbb{R}^d$ and we shall denote by θ_0 the true value of the parameter. We make classical assumptions on a and b, which ensure that the solution of (2.2) is positive recurrent on $(0, \infty)$.

(A0) For all $\theta \in \Theta$, the equation (2.2) admits a unique strong solution $V > 0$.

(A1) The functions $a(.)$ and $b(., \theta)$ are $\mathcal{C}^2$ on $(0, \infty)$. Moreover, $b(x, \theta)$ is $\mathcal{C}^3$ with respect to the parameter θ, and the derivatives with respect to θ are $\mathcal{C}^2$ as functions of x. The following bounds hold for some $c > 0$:

$$\left|a(x)\right| + \left|b(x,\theta)\right| + \left|\frac{\partial^j}{\partial^j \theta} b(x,\theta)\right| \le c\left(1 + x^c\right), \quad \text{with } j \in \{1,2,3\}.$$

Moreover, we assume that if f stands for one of the functions a^{-1}, $\frac{\partial^{i+j}}{\partial^i x\,\partial^j \theta} b$, $\frac{\partial^i}{\partial^i x} a$, where $i = 1, 2$ and $j \le 3$, we have for some $c > 0$

$$\left|f(x)\right| \le c\left(x^c + x^{-c}\right). \tag{2.3}$$

Let us set

$$\mathcal{G}_t := \sigma\big((W_s, B_s), s \le t;\ \eta;\ \eta'\big). \tag{2.4}$$

(A2) For all $k \in \mathbb{R}$ there exists $c > 0$ such that for all $t \ge 0$

$$E\Big(\sup_{s\in[t,t+1]} V_s^k \mid \mathcal{G}_t\Big) \le c\big(1 + V_t^k\big).$$

Now we give conditions for ergodicity on $(0,\infty)$. For this let $x_0 \in (0,\infty)$, and we define the scale function $s(x,\theta) = \exp(-2\int_{x_0}^{x} \frac{b(u,\theta)}{a^2(u)}\,\mathrm{d}u)$ and the speed density $m(x,\theta) = a^{-2}(x)s^{-1}(x,\theta)$.

(A3) $\int_0 s(x,\theta_0)\,\mathrm{d}x = \int^{\infty} s(x,\theta_0)\,\mathrm{d}x = \infty,\ \int_0^{\infty} m(x,\theta_0)\,\mathrm{d}x = M_{\theta_0} < \infty.$

Let us recall that under (A3), it is known that the diffusion is positive recurrent with a stationary distribution given by

$$\nu_0(\mathrm{d}x) = \frac{1}{M_{\theta_0}} m(x,\theta_0)\mathbb{1}_{\{x\in(0,\infty)\}}\,\mathrm{d}x. \tag{2.5}$$

(A4) For all $c > 0$, we have $\int_0^{\infty}(x^{-c} + x^{c})\nu_0(\mathrm{d}x) < \infty$.
(A5) For all $c > 0$, we have $\sup_{t\ge 0} E(V_t^{-c} + V_t^{c}) < \infty$.

Let us introduce the identifiability conditions necessary for estimation of the parameter. We need the $d \times d$-matrix $I(\theta_0)$ defined by

$$I(\theta_0)_{i,j} = \int_0^{\infty} \frac{\partial}{\partial\theta_i} b(x,\theta_0) \frac{\partial}{\partial\theta_j} b(x,\theta_0) a^{-2}(x)\,\mathrm{d}\nu_0(x), \quad \text{for } 1 \le i,j \le d. \tag{2.6}$$

(I1) $b(x,\theta) = b(x,\theta_0)$ $\mathrm{d}x$ a.e. on $(0,\infty)$ implies $\theta = \theta_0$.
(I2) The information matrix $I(\theta_0)$ is invertible.

Note that the condition (A1) on the coefficients is natural, since it mainly states that the coefficients are smooth on $(0,\infty)$, with at most polynomial growth at ∞, and that they may have a singularity at the end point 0, which is standard for a diffusion restricted to be positive. The condition (A2) is a uniform control on the behavior of the diffusion near 0 and ∞, which is a useful tool in our proofs. It was introduced in [22] and it is shown in [14] that the condition holds for classical diffusion processes used in finance to model the volatility (see Sect. 2.4). The assumption (A4) is an integrability condition on the stationary measure and it immediately implies (A5) when the volatility process is stationary. It is shown in [15] that the condition (A5) may hold even for non-stationary processes.

2.2 Observations and the contrast function

We assume that we observe the price process Y on the interval $[0, T_n]$ at the sampling times $(t_{i,n})_{i=0,\dots,n}$ with $t_{0,n} = 0$ and $t_{n,n} = T_n \xrightarrow{n\to\infty} \infty$. As it happens in the high frequency setting, these observations might be contaminated by some noise, but we do not need to suppose yet an exact structure for the observations. We only assume that there exists some sequence of filtrations $(\tilde{\mathcal{G}}_t^n)_{0\le t\le T_n}$ generated by a family of variables independent of $\mathcal{G}_\infty$ (recall (2.4)). A possible situation is that $\tilde{\mathcal{G}}_t^n$ is generated by the microstructure noise in the observations up to time t. Then we set $\mathcal{G}_t^n = \tilde{\mathcal{G}}_t^n \vee \mathcal{G}_t$.

Let us assume now that from these observations we are able to recover the volatility process with some statistical error. Choose $0 < \Delta_n < 1$ as some sampling step (typically larger than the underlying sampling step on Y) and define the unobserved integrated volatility on an interval of length Δ_n,

$$\overline{V}_i = \overline{V}_{i,n} = \Delta_n^{-1} \int_{i\Delta_n}^{(i+1)\Delta_n} V_s \,\mathrm{d}s, \quad i = 0, \dots, N_n - 1, \tag{2.7}$$

where $N_n = \lfloor T_n/\Delta_n \rfloor$.

We denote by $\widehat{V}_i = \widehat{V}_{i,n}$ for $i = 0, \dots, N_n - 1$, the approximation of this integrated volatility based on the observations (see explicit examples in Sects. 3, 4). We specify the quality of the approximation in the following set of assumptions, where

$$E_i = E_{i,n} = \widehat{V}_i - \overline{V}_i, \quad i = 0, \dots, N_n - 1,$$

denotes the estimation error.

(V1) The variable $\widehat{V}_i$ is $\mathcal{G}^n_{(i+1)\Delta_n}$-measurable for all $0 \le i \le N_n - 1$.

(V2) For all $p \ge 1$, there exists $c(p) > 0$ such that $E(|\widehat{V}_i|^p \mid \mathcal{G}^n_{i\Delta_n}) \le c(p)(1 + V_{i\Delta_n}^{c(p)})$.

(V2′) The variables $\widehat{V}_i$ are a.s. positive and for all $p \ge 1$ there exists $c(p) > 0$ such that $E(\widehat{V}_i^{-p} \mid \mathcal{G}^n_{i\Delta_n}) \le c(p)(1 + V_{i\Delta_n}^{-c(p)})$.

(V3) There exists a sequence $b_n \xrightarrow{n\to\infty} 0$ such that

$$\forall i = 0, \dots, N_n - 1, \quad \left|E\left(E_i \mid \mathcal{G}^n_{i\Delta_n}\right)\right| \le b_n c\left(1 + V^c_{i\Delta_n}\right).$$

(V4) There exists a sequence $v_n \xrightarrow{n\to\infty} 0$ such that for all $p \ge 1$ there exists $c(p)$ with

$$\forall i = 0, \dots, N_n - 1, \quad E\left(|E_i|^{2p} \mid \mathcal{G}^n_{i\Delta_n}\right) \le v_n^p c(p)\left(1 + V_{i\Delta_n}^{c(p)}\right).$$

The condition (V1) is natural, since the true integrated volatility $\overline{V}_i$ is $\mathcal{G}_{(i+1)\Delta_n}$-measurable. Conditions analogous to (V2–V2′) are satisfied for $\overline{V}_i$ by direct application of (A2); hence, it is natural to state this control for the approximation. Actually, (V2) is immediately implied by (V4) and (A2), while (V2′) is more restrictive and might not hold for some choice of approximation. The condition (V3) controls the (conditional) bias between true and estimated integrated volatility, while (V4) is some control on the variance.

In the previous paper [15] we introduced a contrast function for the estimation of the parameters in (1.1) when we observe $(\overline{V}_i)$. This contrast function is based on the kind of Euler scheme formula for the integrated process $(\overline{V}_i)$ stated in [14] and recalled below in Sect. 7.1.1. Following the same approach we let, using positivity of $\widehat{V}_i$ under (V2′),

$$\mathcal{C}_n(\theta) = -\frac{1}{N_n \Delta_n} \sum_{i=1}^{N_n-2} (\widehat{V}_{i+1} - \widehat{V}_i) \frac{b(\widehat{V}_{i-1}, \theta)}{a^2(\widehat{V}_{i-1})} + \frac{1}{2N_n} \sum_{i=1}^{N_n-2} \frac{b^2(\widehat{V}_{i-1}, \theta)}{a^2(\widehat{V}_{i-1})}, \tag{2.8}$$

and set $\hat{\theta}_n = \operatorname{arginf}_{\theta \in \Theta} \mathcal{C}_n(\theta)$ for the minimum contrast estimator.

Formally the expression of $-\mathcal{C}_n(\theta)$ is very similar to the expression of the log-likelihood of the direct continuous observation of $(V_t,\ t \le T_n)$, which is (see [22])

$$\mathcal{L}_{T_n}(\theta) = \frac{1}{T_n} \int_0^{T_n} \frac{b(V_s, \theta)}{a^2(V_s)} \,\mathrm{d}V_s - \frac{1}{2T_n} \int_0^{T_n} \frac{b^2(V_s, \theta)}{a^2(V_s)} \,\mathrm{d}s.$$

However, the expression (2.8) is far from being immediate to obtain. Indeed, if one replaces $\frac{b}{a^2}(\widehat{V}_{i-1}, \theta)$ by $\frac{b}{a^2}(\widehat{V}_i, \theta)$ in the first sum then the contrast does not even produce a consistent estimator. It is important that the weight factor $\frac{b}{a^2}(\widehat{V}_{i-1}, \theta)$ in (2.8) depends only on past observations $(\widehat{V}_j,\ j < i)$. Other possible choices are given in Sects. 2.3 and 3.3.

2.3 Properties of the estimator

Let us define the quantity

$$\mathcal{C}(\theta_0, \theta) = \frac{1}{2} \int_0^{\infty} \big(b(x, \theta) - b(x, \theta_0)\big)^2 a^{-2}(x) \,\mathrm{d}\nu_0(x). \tag{2.9}$$

The following proposition justifies the choice of (2.8) as a contrast function.

Proposition 2.1 *Assume* (A0–A5), (V1–V4), (V2′), $b_n = o(\Delta_n)$, $v_n = o(\Delta_n T_n)$, $\Delta_n \to 0$. *Then*

$$\mathcal{C}_n(\theta) - \mathcal{C}_n(\theta_0) \xrightarrow{n \to \infty} \mathcal{C}(\theta_0, \theta), \quad \textit{in probability.} \tag{2.10}$$

We deduce the consistency property for the estimator.

Theorem 2.2 *Assume* (I1) *and the assumptions of Proposition* 2.1. *Then*

$$\hat{\theta}_n \xrightarrow{n \to \infty} \theta_0, \quad \textit{in probability.}$$

Remark 2.3 The conditions on Δ_n, b_n, and v_n are crucial points in the construction of the estimator. The fact that $\Delta_n \to 0$ implies that the integrated volatilities $\overline{V}_i$ are themselves in the framework of short sampling intervals. We shall see in the examples that these conditions for consistency are not very stringent and a proper choice for Δ_n and the $\widehat{V}_i$ will always be possible in our examples.

Theorem 2.4 *Assume* (A0–A5), (V1–V4), (V2′), (I1–I2), *and* $\theta_0 \in \overset{\circ}{\Theta}$. *Assume, moreover, that* $N_n\Delta_n^3 \to 0$, $b_n = o(\Delta_n^{1/2} N_n^{-1/2})$, $v_n = o(\Delta_n)$. *Then*

$$\sqrt{T_n}(\hat{\theta}_n - \theta_0) \xrightarrow[\mathcal{L}]{n\to\infty} \mathcal{N}\big(0, I(\theta_0)^{-1}\big). \tag{2.11}$$

Remark 2.5 (1) These conditions for asymptotic normality are rather stringent. However, $N_n\Delta_n^3 \to 0$ is the standard condition on the sampling step encountered for the estimation of the drift parameter from a direct observation of the sampling $(V_{i\Delta_n})_{i=0,\dots,N_n}$ (see [9, 29]). Actually, this condition suppresses a bias term of order $N_n^{1/2}\Delta_n^{3/2}$ while proving (2.11) (see details in Sect. 7.1.3). We benefit here from the fact that we only estimate the drift parameter: indeed, it is known that for the estimation of both drift and diffusion parameter from $V_{i\Delta_n}$ or $\overline{V}_i$ the classical condition is $N_n\Delta_n^2 \to 0$ (see [15, 22]).

(2) The estimator $\hat{\theta}_n$ has rate $\sqrt{T_n}$ with asymptotic covariance matrix $I(\theta_0)^{-1}$. It is remarkable that this is the same asymptotic behavior as any efficient estimator based on the direct observation of the hidden volatility $(V_t)_{t\in[0,T_n]}$. Since our data are randomizations of this hidden path, the estimator is efficient in our statistical model.

Remark 2.6 It may happen in some examples (see Sect. 4) that Assumption (V2′) does not hold. Then the contrast function (2.8) might not be defined if $\widehat{V}_i < 0$ for some $i \in \{1,\dots,N_n - 1\}$. In this case consider some smooth real function ψ equal to zero on some interval $(-\infty, \varepsilon]$ and equal to 1 on $[2\varepsilon, \infty)$ for a fixed $\varepsilon > 0$ and introduce the modified contrast function

$$\widetilde{\mathcal{C}}_n(\theta) = -\frac{1}{N_n\Delta_n}\sum_{i=1}^{N_n-2}(\widehat{V}_{i+1} - \widehat{V}_i)\frac{b(\widehat{V}_{i-1},\theta)}{a^2(\widehat{V}_{i-1})}\psi(\widehat{V}_{i-1})$$
$$+\frac{1}{2N_n}\sum_{i=1}^{N_n-2}\frac{b^2(\widehat{V}_{i-1},\theta)}{a^2(\widehat{V}_{i-1})}\psi(\widehat{V}_{i-1}),$$

and let $\tilde{\theta}_n$ be the corresponding minimum estimator.

If one replaces the assumption (I1) by the condition that the function $b(\cdot,\theta_0)$ is identified by its values on some interval $[\varepsilon,\infty)$, then we can show that under the assumptions of Theorem 2.2 except (V2′), consistency holds. Moreover, under the additional conditions of Theorem 2.4 we have

$$\sqrt{T_n}(\tilde{\theta}_n - \theta_0) \xrightarrow[\mathcal{L}]{n\to\infty} \mathcal{N}\big(0, \tilde{J}(\theta_0)^{-1}\tilde{I}(\theta_0)\tilde{J}(\theta_0)^{-1}\big),$$

with

$$\tilde{I}(\theta_0)_{i,j} = \int_\varepsilon^\infty \frac{\partial}{\partial\theta_i}b(x,\theta_0)\frac{\partial}{\partial\theta_j}b(x,\theta_0)a^{-2}(x)\psi^2(x)\,\mathrm{d}\nu_0(x), \quad \text{for } 1 \le i, j \le d,$$
$$\tilde{J}(\theta_0)_{i,j} = \int_\varepsilon^\infty \frac{\partial}{\partial\theta_i}b(x,\theta_0)\frac{\partial}{\partial\theta_j}b(x,\theta_0)a^{-2}(x)\psi(x)\,\mathrm{d}\nu_0(x), \quad \text{for } 1 \le i, j \le d.$$

In this situation the asymptotic variance depends on the choice of the truncation function ψ, but when $\varepsilon \to 0$ both matrices $\widetilde{I}(\theta_0)$ and $\widetilde{J}(\theta_0)$ converge to $I(\theta_0)$ since ψ converges a.s. to 1. Thus, the lack of efficiency can be made arbitrarily small by choosing ε close to zero.

2.4 Explicit volatility models

In this section we apply the results of Sect. 2.3 to some standard models for the volatility process.

2.4.1 Log-normal volatility model

Suppose the volatility process is log-normal, $V_t = \exp(Z_t)$, where Z_t is the stationary solution of $\mathrm{d}Z_t = \theta Z_t\,\mathrm{d}t + \sigma\,\mathrm{d}B_t$ (with $\theta < 0$). Thus, the equation for V is, by Itô's formula,

$$\mathrm{d}V_t = \left\{\theta V_t \ln V_t + \frac{\sigma^2}{2} V_t\right\}\mathrm{d}t + \sigma V_t\,\mathrm{d}B_t.$$

In [14] we have proved that (A2) holds for this model. The conditions (A3–A5) are simple to check, the stationary law being log-normal. Hence, the results of Sect. 2.3 apply, and we can find an explicit minimum for the contrast function (2.8), namely,

$$\hat{\theta}_n = \frac{\frac{1}{\Delta_n}\left\{\sum_{i=1}^{N-2} \frac{(\widehat{V}_{i+1}-\widehat{V}_i)(\ln \widehat{V}_{i-1})}{\widehat{V}_{i-1}}\right\} - \frac{\sigma^2}{2}\left\{\sum_{i=1}^{N-2} \ln \widehat{V}_{i-1}\right\}}{\sum_{i=1}^{N-2} (\ln \widehat{V}_{i-1})^2}. \tag{2.12}$$

Here an application of Theorem 2.4 gives $\sqrt{T_n}(\hat{\theta}_n - \theta) \xrightarrow{n\to\infty} \mathcal{N}(0, 2|\theta|)$.

2.4.2 GARCH diffusion model

Here the volatility process is the solution of

$$\mathrm{d}V_t = (\alpha V_t + \beta)\,\mathrm{d}t + \sigma V_t\,\mathrm{d}B_t, \qquad V_0 > 0,$$

with $\alpha < 0$, $\sigma, \beta > 0$. We have shown in [14] that (A2) holds for this model. Furthermore, the condition (A3) for ergodicity is satisfied, and we assume the diffusion is stationary with initial law given by

$$\nu_0(\mathrm{d}x) = \frac{\lambda^a}{\Gamma(a)} x^{-a-1} \exp\left(-\frac{\lambda}{x}\right) \mathbb{1}_{\{x>0\}}\,\mathrm{d}x,$$

where $a = -2\alpha/\sigma^2 + 1 > 1$ and $\lambda = 2\beta/\sigma^2 > 0$.

This distribution admits negative moments of any order $E[V_t^{-c}] < \infty$, but is heavy tailed near $+\infty$ and admits positive moments $E[V_t^c] < \infty$ only for $c < a$. Thus, conditions (A4–A5) do not exactly hold. Nevertheless, we can write down an explicit solution for the minimization of the contrast function (2.8):

$$\begin{bmatrix} \sum_{i=1}^{N-2} 1 & \sum_{i=1}^{N-2} \widehat{V}_{i-1}^{-1} \\ \sum_{i=1}^{N-2} \widehat{V}_{i-1}^{-1} & \sum_{i=1}^{N-2} \widehat{V}_{i-1}^{-2} \end{bmatrix} \begin{bmatrix} \hat{\alpha}_n \\ \hat{\beta}_n \end{bmatrix} = \begin{bmatrix} \frac{1}{\Delta_n} \sum_{i=1}^{N-2} \widehat{V}_{i-1}^{-1}(\widehat{V}_{i+1} - \widehat{V}_i) \\ \frac{1}{\Delta_n} \sum_{i=1}^{N-2} \widehat{V}_{i-1}^{-2}(\widehat{V}_{i+1} - \widehat{V}_i) \end{bmatrix}. \tag{2.13}$$

Then it can be shown directly that Theorem 2.4 holds for this estimator if the stationary law admits enough positive moments.

Theorem 2.7 *Assume* $a > 3$, $N_n\Delta_n^3 \to 0$, $b_n = o(\Delta_n^{1/2}{N_n}^{-1/2})$, *and* $v_n = o(\Delta_n)$. *Then*

$$\begin{bmatrix} \sqrt{T_n}(\hat{\alpha}_n - \alpha) \\ \sqrt{T_n}(\hat{\beta}_n - \beta) \end{bmatrix} \xrightarrow[\mathcal{D}]{n\to\infty} \mathcal{N}\left(0, \begin{bmatrix} 2\sigma^2 - 2\alpha & -2\beta \\ -2\beta & \frac{4\beta^2}{\sigma^2-2\alpha} \end{bmatrix}\right).$$

Proof We rewrite (2.13) as $\mathcal{M}_n \begin{bmatrix} \sqrt{N_n\Delta_n}(\hat{\alpha}_n-\alpha) \\ \sqrt{N_n\Delta_n}(\hat{\beta}_n-\beta) \end{bmatrix} = \mathcal{N}_n$, with

$$\mathcal{M}_n = \frac{1}{N_n}\begin{bmatrix} \sum_{i=1}^{N-2} 1 & \sum_{i=1}^{N-2} \widehat{V}_{i-1}^{-1} \\ \sum_{i=1}^{N-2} \widehat{V}_{i-1}^{-1} & \sum_{i=1}^{N-2} \widehat{V}_{i-1}^{-2} \end{bmatrix},$$

and

$$\mathcal{N}_n = \begin{bmatrix} (N_n\Delta_n)^{-1/2}\sum_{i=1}^{N-2} \widehat{V}_{i-1}^{-1}(\widehat{V}_{i+1} - \widehat{V}_i - (\alpha\widehat{V}_{i-1} + \beta)\Delta_n) \\ (N_n\Delta_n)^{-1/2}\sum_{i=1}^{N-2} \widehat{V}_{i-1}^{-2}(\widehat{V}_{i+1} - \widehat{V}_i - (\alpha\widehat{V}_{i-1} + \beta)\Delta_n) \end{bmatrix}.$$

Then the theorem follows from the expression of $\nu_0(\mathrm{d}x)$ and the two following results whose proof is detailed in Sect. 7.1.4:

$$\text{since } a > 1, \quad \mathcal{M}_n \xrightarrow[\mathbf{P}]{n\to\infty} \begin{bmatrix} 1 & \int_0^\infty x^{-1}\,\mathrm{d}\nu_0(x) \\ \int_0^\infty x^{-1}\,\mathrm{d}\nu_0(x) & \int_0^\infty x^{-2}\,\mathrm{d}\nu_0(x) \end{bmatrix}, \tag{2.14}$$

$$\text{since } a > 3, \quad \mathcal{N}_n \xrightarrow[\mathcal{L}]{n\to\infty} \mathcal{N}\left(0, \sigma^2 \begin{bmatrix} 1 & \int_0^\infty x^{-1}\,\mathrm{d}\nu_0(x) \\ \int_0^\infty x^{-1}\,\mathrm{d}\nu_0(x) & \int_0^\infty x^{-2}\,\mathrm{d}\nu_0(x) \end{bmatrix}\right). \tag{2.15}$$

□

2.5 The Heston square root model

In this popular model of finance [17] the volatility is the solution of the equation $\mathrm{d}V_t = (\alpha V_t + \beta)\,\mathrm{d}t + \sigma\sqrt{V_t}\,\mathrm{d}B_t$, $V_0 > 0$ with $\alpha < 0$, $\beta > 0$. If $c_0 = 2\beta/\sigma^2 > 1$, the process admits a stationary distribution $\mathrm{d}\nu_0(x) = \lambda^{c_0}\Gamma(c_0)^{-1}x^{c_0-1}e^{-\lambda x}\mathbb{1}_{\{x>0\}}\,\mathrm{d}x$, with $\lambda = -2\alpha/\sigma^2$. This stationary distribution admits inverse moments $\nu_0(x^{-c}) < \infty$ only for $c < c_0$. Actually, the model only satisfies the condition (A2) for $k > -c_0 + 1$, and (A4–A5) for $c < c_0$ (see [15] for details). However, a direct study of the explicit estimator obtained as the minimum of (2.8) is feasible and it can be shown that Theorem 2.4 is still valid if $c_0 > 13$.

3 Direct observation of the price process

We assume that we observe the price process directly at some equidistant sampling:

$$(Y_{t_{j,n}})_{j=0,\dots,n} = (Y_{j\delta_n})_{j=0,\dots,n}, \tag{3.1}$$

where $\delta_n \to 0$ is the sampling step and $n\delta_n = T_n$. Since we assume here no microstructure noise, we let $\mathcal{G}_t^n = \mathcal{G}_t$ for all $n \geq 1$ (recall (2.4)). Moreover, for the sake of simplicity we assume that the two Brownian motions W and B in (2.1–2.2) are uncorrelated.

3.1 The realized volatility

We split the data into successive blocks of size $k = k_n \leq n$, $k_n \to \infty$, let $\Delta_n = k_n\delta_n$, and $N = N_n = \lfloor n/k_n \rfloor$ is the number of blocks obtained. On each block we consider the quadratic variation of the observation divided by the observation time corresponding to one block

$$\widehat{V}_i = \widehat{V}_{i,k,n} = \Delta_n^{-1} \sum_{j=0}^{k-1} (Y_{(ik+j+1)\delta_n} - Y_{(ik+j)\delta_n})^2, \quad i = 0, \ldots, N_n - 1. \tag{3.2}$$

The relation between the realized variance (3.2) and the integrated one (2.7) has been used by several authors for statistical purposes in the context of stochastic volatility (see, e.g., [3, 5, 16]). In the next proposition we show that this choice satisfies the conditions (V1), (V3), (V4) (and thus (V2), too).

Proposition 3.1 *Assume* (A0–A2) *and* $\Delta_n \leq 1$. *Recall* $E_i = \widehat{V}_i - \overline{V}_i$. *Then* $\widehat{V}_i$, $\overline{V}_i$, *and* E_i *are* $\mathcal{G}^n_{(i+1)\Delta_n}$*-measurable. Moreover,*

$$\exists c,\ \forall i, n, k, \quad \left|E(E_i \mid \mathcal{G}_{i\Delta_n})\right| \leq c\delta_n\left(1 + V^c_{i\Delta_n}\right), \tag{3.3}$$

$$\textit{for } p \geq 1,\ \exists c(p),\ \forall i, n, k, \quad E\left(|E_i|^p \mid \mathcal{G}_{i\Delta_n}\right) \leq c(p)k^{-p/2}\left[1 + V^p_{i\Delta_n}\right]. \tag{3.4}$$

Proof The measurability condition is immediate by (2.4), (2.7), (3.2). Then denote by $\mathcal{G}_V$ the sigma field generated by the hidden path $(V_t)_{t\geq 0}$. Conditionally on $\mathcal{G}_V$ the process Y is a simple Gaussian diffusion. From this it can be shown (see Lemma 6 in [16] for details) that

$$\left|E(E_i \mid \mathcal{G}_V \vee \mathcal{G}_{i\Delta_n})\right| \leq c\delta_n \sup_{s\in[i\Delta_n,(i+1)\Delta_n]} \left(1 + V_s^c\right),$$

$$E\left(|E_i|^p \mid \mathcal{G}_V \vee \mathcal{G}_{i\Delta_n}\right) \leq c(p)k^{-p/2} \sup_{s\in[i\Delta_n,(i+1)\Delta_n]} \left(1 + V_s^p\right), \quad \text{for } p \geq 1.$$

Then we conclude taking conditional expectations with respect to $\mathcal{G}_{i\Delta_n}$ and using (A2) with $\Delta_n \leq 1$.

The fact that (V2′) holds is more delicate and is deferred to Sect. 7.2. □

Hence, the realized variances $(\widehat{V}_i)$ satisfy the conditions of Sect. 2.2. The following choices of k_n imply consistency and normality:

1. If $k_n \to \infty$ with $n^{-1/2}\delta_n^{-1} = o(k_n)$ and $k_n = o(\delta_n^{-1})$, then Theorem 2.2 holds.

2. Assume $n\delta_n^2 \to 0$ and choose k_n with $\delta_n^{-1/2} = o(k_n)$ and $k_n = o(n^{-1/2}\delta_n^{-3/2})$. Then Theorem 2.4 holds.

Let us stress that a choice of k_n that insures the consistency is always feasible, and the lower bound on k_n may even disappear if $n^{-1/2}\delta_n^{-1} \to \infty$. On the other hand, a choice that insures the asymptotic normality is only possible under a fast sampling assumption $n\delta_n^2 \to 0$, since the upper and lower bound conditions on k_n imply $\delta_n^{-1/2} = o(n^{-1/2}\delta_n^{-3/2})$.

3.2 Behavior on a finite sample

We made numerical simulations to study by a Monte Carlo method how the estimator performs on a finite sample in the model of Sect. 2.4.2. For the parameters, we choose the values $\alpha = -10$, $\beta = 4$, $\sigma = 2.5$ which are plausible for financial data if the time scale is expressed in years (see for instance [2]). We assume that we observe the price process during $T_n = 10$ years. We present three possible setups for the sampling of the price process:

- In case A, we let $n = 4 \times 10^6$. This corresponds to ultra-high frequency sampling with observations about every 10 seconds.
- In case B, we let $n = 5 \times 10^5$.
- In case C, we let $n = 1.33 \times 10^5$. This corresponds to data sampled about every 5 minutes.

To help the interpretation, we give numerical results also for the estimator $\overline{\theta}_n = (\overline{\alpha}_n, \overline{\beta}_n)$ obtained if we could observe directly the integrated volatility $\overline{V}_i$ and use it in the expression of contrast (2.8) in place of the $\widehat{V}_i$.

The results for empirical means (and standard deviations) with different choices for k_n are presented in the first 3 columns of Tables 1–3. The number of replications used in the Monte Carlo method is 1000.

We find that our estimators $\hat{\alpha}_n, \hat{\beta}_n$ perform well in case A and are very similar to the one based on the direct observation of $\overline{V}_i$. Here a proper choice for k_n seems $k_n \simeq 300$.

In case B the estimators $\hat{\alpha}_n, \hat{\beta}_n$ appear more biased and with larger standard deviation when compared to the estimation based on $\overline{V}_i$.

In case C the results worsen: either k is chosen small and $\hat{\alpha}_n, \hat{\beta}_n$ have large variances and are very different from $\overline{\alpha}_n, \overline{\beta}_n$; or k is chosen large and all the estimators are strongly biased.

3.3 Improvement on the finite sample

In case C we noticed that we cannot suppress both the bias and the additional variance due to the errors E_i. Actually, if Δ_n (or, equivalently, k_n) is chosen too small, even the property (2.14) does not seem to hold when we inspect the simulations.

We can circumvent this problem by a slight modification of the estimator. We let $L \geq 1$ and introduce a new approximation for $V_{i\Delta_n}$ using $L \geq 1$ lagged data, namely,

$$\widehat{V}_i^{(L)} = \frac{1}{L}\sum_{l=0}^{L-1} \widehat{V}_{i-l}.$$

Table 1 Case A, $\alpha = -10$, $\beta = 4$, $\sigma = 2.5$

k	$\hat{\alpha}_n$	$\hat{\beta}_n$	$\overline{\alpha}_n$	$\overline{\beta}_n$	L	$\hat{\alpha}_n^{(L)}$	$\hat{\beta}_n^{(L)}$
100	−9.58	3.80	−10.37	4.11	12	−10.27	4.07
	(3.53)	(1.14)	(1.83)	(0.50)		(1.86)	(0.51)
300	−10.08	4.00	−10.33	4.05	4	−10.22	4.04
	(1.90)	(0.53)	(1.86)	(0.51)		(1.85)	(0.51)
1000	−10.02	3.97	−10.09	4.01	1	−10.02	3.97
	(1.84)	(0.54)	(1.85)	(0.53)		(1.84)	(0.54)

Table 2 Case B, $\alpha = -10$, $\beta = 4$, $\sigma = 2.5$

k	$\hat{\alpha}_n$	$\hat{\beta}_n$	$\overline{\alpha}_n$	$\overline{\beta}_n$	L	$\hat{\alpha}_n^{(L)}$	$\hat{\beta}_n^{(L)}$
50	−9.01	3.57	−10.32	4.08	13	−10.06	3.98
	(3.66)	(1.17)	(1.80)	(0.49)		(1.85)	(0.53)
100	−9.45	3.76	−10.04	4.00	6	−9.79	3.90
	(2.13)	(0.66)	(1.83)	(0.51)		(1.88)	(0.54)
300	−9.54	3.78	−9.77	3.86	1	−9.54	3.78
	(1.72)	(0.51)	(1.75)	(0.50)		(1.72)	(0.51)

Table 3 Case C, $\alpha = -10$, $\beta = 4$, $\sigma = 2.5$

k	$\hat{\alpha}_n$	$\hat{\beta}_n$	$\overline{\alpha}_n$	$\overline{\beta}_n$	L	$\hat{\alpha}_n^{(L)}$	$\hat{\beta}_n^{(L)}$
30	−8.21	3.27	−10.12	4.03	12	−9.50	3.79
	(4.23)	(1.36)	(1.76)	(0.50)		(1.84)	(0.56)
50	−8.74	3.47	−9.86	3.91	7	−9.27	3.68
	(2.45)	(0.76)	(1.76)	(0.49)		(1.86)	(0.55)
100	−8.77	3.49	−9.40	3.74	3	−9.02	3.59
	(1.82)	(0.56)	(1.75)	(0.52)		(1.84)	(0.57)
150	−8.79	3.48	−9.20	3.64	2	−8.83	3.64
	(1.84)	(0.54)	(1.82)	(0.52)		(1.90)	(0.56)
300	−7.87	3.12	−8.06	3.20	1	−7.87	3.12
	(1.69)	(0.55)	(1.67)	(0.54)		(1.69)	(0.54)

Then we slightly change the weight factors in the expression of the contrast function (2.8) to

$$\mathcal{C}_n^{(L)}(\theta) = -\frac{1}{N_n \Delta_n} \sum_{i=L+1}^{N_n-2} (\widehat{V}_{i+1} - \widehat{V}_i) \frac{b}{a^2}\big(\widehat{V}_{i-1}^{(L)}, \theta\big) + \frac{1}{2N_n} \sum_{i=L+1}^{N_n-2} \frac{b^2}{a^2}\big(\widehat{V}_{i-1}^{(L)}, \theta\big), \tag{3.5}$$

Table 4 $\alpha = -10$, $\beta = 4$, $\sigma = 2.5$

Case A		Case B		Case C	
$\tilde{\alpha}_n$	$\tilde{\beta}_n$	$\tilde{\alpha}_n$	$\tilde{\beta}_n$	$\tilde{\alpha}_n$	$\tilde{\beta}_n$
−10.97	4.32	−10.86	4.38	−10.95	4.31
(2.35)	(0.74)	(2.31)	(0.72)	(2.45)	(0.76)

and we set $\hat{\theta}_n^{(L)} = \text{arginf}_{\theta \in \Theta} \mathcal{C}_n^{(L)}(\theta)$ for the associated minimum contrast estimator.

It can be shown that, for L fixed and with the same conditions on b_n, v_n as for Theorem 2.4, the asymptotic properties of $\hat{\theta}_n$ and $\hat{\theta}_n^{(L)}$ are the same. However, on a finite sample, if L is chosen different from 1, while k is not large enough, the behavior is significantly different.

In the case of the GARCH diffusion, numerical results are given in the two last columns of Tables 1–3 for choices of L roughly proportional to $1/k$.

The estimator is improved in all three cases, and it is much less sensitive to the choice of k. For applications to real data, it seems important to use this corrected version of the estimator.

3.4 Comparison with the estimator of [11]

In their paper [11], the authors presented contrast-based estimators for stochastic volatility models. If the volatility process is the same as in Sect. 2.4.2, they obtained the explicit contrast function

$$U_n(a,\lambda) = \frac{1}{n}\sum_{i=1}^{n}\left\{\left(a+\frac{1}{2}\right)\log\left(\lambda+\frac{X_{i,n}^2}{2}\right) - a\log(\lambda) + \log\left(\frac{\Gamma(a)}{\Gamma(a+1/2)}\right)\right\},$$

with $X_{i,n} = \delta_n^{-1/2}(Y_{i\delta_n} - Y_{(i-1)\delta_n})$, $a = -2\alpha/\sigma^2 + 1$ and $\lambda = 2\beta/\sigma^2 > 0$. Minimization with respect to (α, β) gives an estimator $(\tilde{\alpha}_n, \tilde{\beta}_n)$. Their contrast function is suggested by the approximation in law $X_{i,n} \simeq \epsilon\eta^{1/2}$ where ϵ is a standard normal variable and η is distributed as the stationary measure of the GARCH diffusion process V (namely as an inverse Gamma distribution).

This estimator is shown in [11] to be asymptotically normal with rate $\sqrt{T_n}$ as soon as $n\delta_n \to \infty$, $n\delta_n^2 \to 0$ and $a > 2$. But in contrast to ours, it is not asymptotically efficient. Results of numerical simulations are given in Table 4. Comparison with Tables 1–3 is unambiguous: the estimators of [11] are much less sensitive than ours to the conditions on the sampling step δ_n. However, in cases A and B where our estimators $(\alpha_n^{(L)}, \beta_n^{(L)})$ behave properly, they perform better than $(\tilde{\alpha}_n, \tilde{\beta}_n)$. Here we benefit from the efficiency property.

4 Observations with microstructure noise

When dealing with ultra-high frequency data, the assumption of observations given by (3.1) might not be realistic, since tick-by-tick data are not equally sampled. Moreover, empirical evidence shows that ultra-high frequency data seem contaminated by some error, commonly called 'microstructure noise' (see [31]).

We assume that we observe

$$X_{t_{j,n}} = Y_{t_{j,n}} + \epsilon_{j,n}, \quad \text{for } j = 0, \dots, n,$$

where $0 = t_{0,n} < \cdots < t_{n,n} = T_n$ and the random variables $(\epsilon_{j,n})_{n\geq 1; 0\leq j\leq n}$ are i.i.d. centered with moments of any order.

We denote by $\delta_n = \sup_{j=0,\dots,n-1}(t_{j+1,n} - t_{j,n})$ the maximum step between two observations and assume that for some fixed constant $0 < c < 1$ we have $t_{j+1,n} - t_{j,n} > c\delta_n$.

Let us choose $\Delta_n > \delta_n$. Then it is known that due to the noises $\epsilon_{j,n}$ the realized variance over some interval of size Δ_n is not a consistent estimator for the integrated variance. Instead we shall use the 'Two-Scales Realized Volatility' (TSRV) introduced by Zhang et al. [31]. Let us briefly recall their construction.

Define, for $i = 0, \dots, N_n - 1$, the set $B_{i,n} = \{j \mid t_{j,n} \in [i\Delta_n, (i+1)\Delta_n)\}$ and let $k_{i,n} = \sharp B_{i,n}$ be the number of data in the block i. Due to the non-equidistant sampling, this number is not constant for different blocks, but we have $c\Delta_n/\delta_n \leq k_{i,n} \leq \Delta_n/\delta_n$. Then Zhang et al. [31] introduce the estimated variance of the noise in the block i as

$$\widehat{\epsilon_i^2} = \frac{1}{2(k_{i,n}-1)} \sum_{\substack{j\in B_i \\ j+1\in B_i}} (X_{t_{j+1,n}} - X_{t_{j,n}})^2.$$

Then let $1 \leq M \leq k_{i,n}$ be some 'sub-sampling step' and define the averaged realized volatility obtained on this subgrid by

$$[X, X]_i^{\mathrm{avg},M} = \frac{1}{M} \sum_{\substack{j\in B_i \\ j+M\in B_i}} (X_{t_{j+M,n}} - X_{t_{j,n}})^2.$$

Following [31], we introduce the TSRV

$$\widehat{V_i} = \frac{1}{\Delta_n}\left[[X, X]_i^{\mathrm{avg},M} - \frac{2(k_{i,n} - M)}{M}\widehat{\epsilon_i^2}\right]. \tag{4.1}$$

The idea behind this construction is that the effect of the noises $\epsilon_{i,n}$ is predominant in the high frequency quadratic variation $\widehat{\epsilon_i^2}$, whereas it decreases in a medium frequency quadratic variation $[X, X]_i^{\mathrm{avg},M}$. The expression (4.1) is obtained by adjusting the bias effect due to the noises in the two quadratic variations.

The theoretical study of the asymptotic law, as the number of data in a block tends to infinity, of $\widehat{V_i} - \overline{V}_i$ is conducted in [31] (for $\Delta_n = \Delta$ fixed and a bounded volatility process V); the second order moment of $E_i = \widehat{V_i} - \overline{V}_i$ is bounded in [6]; a more general study for expressions like (4.1) is presented in [30]. Here we state the following result.

Proposition 4.1 *Assume* (A0–A2), $\Delta_n \leq 1$, *and that* B *and* W *are uncorrelated. Let* $\mathcal{G}_t^n = \mathcal{G}_t \vee \sigma(\epsilon_{j,n};\, j \text{ such that } t_{j,n} \leq t)$ *for* $t \geq 0$, $n \geq 1$. *Then* $\widehat{V}_i$ *is* $\mathcal{G}^n_{(i+1)\Delta_n}$-*measurable and*

$$\left|E\left(E_i \mid \mathcal{G}^n_{i\Delta_n}\right)\right| \leq c\left(\frac{1}{M} + \frac{M\delta_n}{\Delta_n}\right)\left(1 + V^c_{i\Delta_n}\right), \tag{4.2}$$

$$E\left(|E_i|^p \mid \mathcal{G}^n_{i\Delta_n}\right) \leq c(p)\left\{\frac{M\delta_n}{\Delta_n} + \frac{k_{i,n}}{\Delta_n^2 M^2}\right\}^{p/2}\left(1 + V^{c(p)}_{i\Delta_n}\right). \tag{4.3}$$

The detailed proof of this proposition is long and will not be given here. Equations (4.2) and (4.3) can be obtained along the lines of [31] where, in particular, we use assumptions (A0–A2) instead of the boundedness assumption for V stated in [31]. Using the terminology of [31] the term $M\delta_n/\Delta_n$ in (4.3) is the 'variance due to noise' and $k_{i,n}/(\Delta_n^2 M^2)$ is the 'variance due to discretization'.

Balancing these two terms via the selection of M yields, since $k_{i,n} \asymp \Delta_n/\delta_n$, the choice $M \sim c\delta_n^{-2/3}$. This is only feasible if $k_{i,n} > c\delta_n^{-2/3}$ which, in turn, yields $\delta_n^{1/3} = O(\Delta_n)$.

Then, with this choice for M, we see that the TSRV satisfies (V3–V4) with $b_n = v_n = \delta_n^{1/3}/\Delta_n$. The condition (V2′) is not satisfied, since it is clear that (4.1) can take negative values. By Remark 2.6 in Sect. 2.3 we can still find an estimator. It finally remains to calibrate Δ_n such that the conditions for consistency and normality in Theorems 2.2–2.4 hold. By an easy computation, we obtain:

1. If we choose $\Delta_n \to 0$ with $\delta_n = o(\Delta_n^6)$, then the consistency of the estimator holds.
2. Assume that $n\delta_n^{11/9} \to 0$; then it is possible to choose Δ_n with $\delta_n^{5/12} n^{1/4} = o(\Delta_n)$ and $\Delta_n = o(n^{-1/2}\delta_n^{-1/2})$ and the estimator is asymptotically normal.

Observe again that it is always possible to choose Δ_n such that consistency holds.

We made numerical simulations, using the improved version of the estimator (3.5) in the same framework as in Sects. 3.2–3.3 for the GARCH diffusion process. In Table 5 we present the results in the ultra-high frequency sampling case, where the microstructure noise is Gaussian with standard deviation equal to 0.002. In this situation, the microstructure noise has the same magnitude as a typical increment between two observations of the price process. The behavior of the estimator based on the TSRV appears rather independent of the presence of microstructure noise: the estimation of the parameter β is severely biased even in absence of noise, whereas the estimation of the mean reversion parameter α is very correct in all cases. On the other hand, the estimator based on the realized volatility (RV) was seriously affected by the presence of noise. Nevertheless, it seems that using the TSRV instead of the realized volatility is only useful in the presence of high noise.

5 Consistent estimation of the diffusion parameter

Estimating the diffusion coefficient in (2.2) is difficult. Several methods have been introduced [5, 13] but the problem of rate optimality is unsolved, in general (see [18]).

Table 5 $\alpha=-10$, $\beta=4$, $\sigma=2.5$, Case A ($T_n=10$, $n=4\cdot 10^6$), $k=300$, $L=12$, $M=10$

Based on T.S.R.V.	(without noise)	$\hat{\alpha}_n$	−9.99 (1.94)	$\hat{\beta}_n$	3.45 (0.47)
	(with noise)	$\hat{\alpha}_n$	−9.90 (2.15)	$\hat{\beta}_n$	3.42 (0.55)
Based on R.V.	(without noise)	$\hat{\alpha}_n$	−10.14 (1.84)	$\hat{\beta}_n$	4.02 (0.53)
	(with noise)	$\hat{\alpha}_n$	−9.98 (3.22)	$\hat{\beta}_n$	3.97 (1.03)

It is not the aim of this paper to consider this problem, but the minimization of the contrast function (2.8) may involve the knowledge of some parameter σ in the coefficient $a(\cdot,\sigma)$. Thus, to insure that our method is feasible we need to estimate this diffusion coefficient. In [15] one contrast function is introduced to estimate both parameters σ,θ from the direct observation of $\overline{V}_i$, but trying to replace $\overline{V}_i$ by $\widehat{V}_i$ in this contrast would lead to non-optimal conditions on b_n and v_n compared to the statement of Theorem 2.2. Instead, we use a two-stage method, as in [23], and estimate first σ by some specific contrast function.

Define

$$U_n(\sigma)=\frac{1}{N_n}\sum_{i=0}^{N_n-2}\frac{3(\widehat{V}_{i+1}-\widehat{V}_i)^2}{2\Delta_n a^2(\widehat{V}_i,\sigma)}+\frac{1}{N_n}\sum_{i=0}^{N_n-2}\log\bigl(a^2(\widehat{V}_i,\sigma)\bigr).$$

This contrast function is a modified version of the contrast function used in [10] where we take into account that the quadratic variation of $(\overline{V}_i)$ differs from the quadratic variation of $(V_{i\Delta_n})$ by a factor $2/3$ (see [15]).

Proposition 5.1 *Assume that* (A0–A5), (V1–V4), (V2′) *hold, that* $\frac{\partial a}{\partial\sigma}(.,\sigma)$ *exists and satisfies* (2.3), *and that the controls in* (A1) *are uniform in* σ. *Then, under the conditions* $\Delta_n\to 0$, $v_n=o(\Delta_n)$, *we have*

$$U_n(\sigma)\xrightarrow{n\to\infty}\int\left(\frac{a^2(x,\sigma_0)}{a^2(x,\sigma)}+\log\bigl(a^2(x,\sigma)\bigr)\right)\mathrm{d}\nu_0(x),$$

in probability, where σ_0 *is the true value of the parameter.*

Proof We know by Theorem 3 in [15] that $\frac{1}{N_n}\sum_{i=0}^{N_n-1}\frac{3(\overline{V}_{i+1}-\overline{V}_i)^2}{2\Delta_n a^2(\overline{V}_i,\sigma)}$ converges to $\int\frac{a^2(x,\sigma_0)}{a^2(x,\sigma)}\mathrm{d}\nu_0(x)$. Then, by computations analogous to the proof of Proposition 2.1, we can see that the control $v_n=o(\Delta_n)$ is sufficient to replace the $\overline{V}_i$ by the $\widehat{V}_i$ in this result. The proposition now follows from an application of Proposition 7.1 in Sect. 7. □

From this proposition it can be seen that, under suitable identifiability conditions, the minimum contrast estimator $\widehat{\sigma}_n$ associated to U_n is consistent in probability. Then one can replace in the contrast function (2.8) the unknown parameter σ by its estimated value and show that the results of Sect. 2.3 still hold. This is done by taking care that the convergence results on the contrast function $\mathcal{C}_n(\theta)$ are uniform with respect to the unknown parameter σ. We do not give more details here.

6 Conclusion

In this paper we have shown that the efficiency for the estimation of the drift parameter in stochastic volatility models can be achieved in a quite general setup for the observations. Moreover, the method appears suitable in practice for financial data.

7 Proof of the results

In this section we prove the results of Sects. 2 and 3.1.

7.1 Study of the contrast function

First, we need some ergodic property for the $(\widehat{V}_i)_i$.

Proposition 7.1 *Assume* (A0–A5), (V1–V4), (V2′), *and* $\Delta_n \to 0$. *Let a function* $f \in \mathcal{C}^1((0,\infty)\times\Theta)$ *satisfy*

$$\sup_{\theta\in\Theta}\left\{\left|f(x,\theta)\right| + \left|f'_x(x,\theta)\right| + \left|\nabla_\theta f(x,\theta)\right|\right\} \le c\left(x^c + x^{-c}\right).$$

Then

$$N_n^{-1}\sum_{i=0}^{N_n-1} f(\widehat{V}_i,\theta) \xrightarrow{n\to\infty} \int_0^\infty f(x,\theta)\,\mathrm{d}\nu_0(x), \tag{7.1}$$

uniformly in θ, *in probability.*

Proof It is shown in Proposition 2 of [15] that under (A0–A5) and $\Delta_n \to 0$ we have the uniform convergence of $N_n^{-1}\sum_{i=0}^{N_n-1} f(\overline{V}_i,\theta)$ to the limit term of (7.1). It remains to see that the difference between $f(\widehat{V}_i,\theta)$ and $f(\overline{V}_i,\theta)$ is controlled appropriately to make the substitution in the sum. But, by the assumptions on f and $E_i = \widehat{V}_i - \overline{V}_i$,

$$\sup_{\theta\in\Theta}\left|f(\widehat{V}_i,\theta) - f(\overline{V}_i,\theta)\right| \le c\left(\widehat{V}_i^c + \widehat{V}_i^{-c} + \overline{V}_i^c + \overline{V}_i^{-c}\right)|E_i|.$$

Clearly, we may assume $\Delta_n \le 1$. Then we deduce from 2.1, (V2), (V2′), (V4) that

$$E\Big(\sup_{\theta\in\Theta}\left|f(\widehat{V}_i,\theta) - f(\overline{V}_i,\theta)\right| \mid \mathcal{G}^n_{i\Delta_n}\Big) \le c\left(V^c_{i\Delta_n} + V^{-c}_{i\Delta_n}\right)v_n^{1/2}.$$

Then (A5) yields the control $E(\sup_{\theta\in\Theta}|f(\widehat{V}_i,\theta) - f(\overline{V}_i,\theta)|) \le c v_n^{1/2}$ and the proposition follows from $v_n \to 0$ by (V4). □

7.1.1 Proof of Proposition 2.1

First, we recall the Euler scheme given in [14] for the process $(\overline{V}_i)$.

Theorem 7.2 (Euler scheme) *We have for $i = 0, \dots, N-1$:*

$$\overline{V}_{i+1} - \overline{V}_i - \Delta_n b(\overline{V}_i, \theta_0) = a(V_{i\Delta_n})\Delta_n^{1/2} U_{i,n} + \varepsilon_{i,n}, \tag{7.2}$$

where $(U_{i,n})_{i=0,\dots,N-1}$ is a centered Gaussian stationary process with the MA(1) covariance structure $\operatorname{var}(U_i) = 2/3$, $\operatorname{cov}(U_i, U_{i+1}) = 1/6$. *Furthermore, the variable $U_{i,n}$ is independent of $\mathcal{G}_{i\Delta_n}$ and is $\mathcal{G}_{(i+2)\Delta_n}$-measurable. The remainder term $\varepsilon_{i,n}$ is $\mathcal{G}_{(i+2)\Delta_n}$-measurable, of order Δ_n, and almost centered*:

$$\left|E(\varepsilon_{i,n} \mid \mathcal{G}_{i\Delta_n})\right| \le c\Delta_n^2 \left(V_{i\Delta_n}^c + V_{i\Delta_n}^{-c}\right), \tag{7.3}$$

$$E\left(|\varepsilon_{i,n}|^p \mid \mathcal{G}_{i\Delta_n}\right) \le c\Delta_n^p \left(V_{i\Delta_n}^c + V_{i\Delta_n}^{-c}\right), \quad \textit{for } p \ge 1. \tag{7.4}$$

The exact values of the constants $c = c(p)$ are here unimportant and the condition on the moment of $\varepsilon_{i,n}$ was only stated for moments $p \le 4$ in [14], but a direct inspection of the proof shows that it holds for any $p \ge 1$.

Using this Euler decomposition we can split the contrast function (recall (2.8)) into three terms: $\mathcal{C}_n(\theta) = \sum_{l=1}^3 \sum_{i=1}^{N-2} C_i^{(l)}(\theta)$, where

$$C_i^{(1)}(\theta) = -\frac{b(\overline{V}_i, \theta_0) b(\widehat{V}_{i-1}, \theta)}{N_n a^2(\widehat{V}_{i-1})} + \frac{b^2(\widehat{V}_{i-1}, \theta)}{2N_n a^2(\widehat{V}_{i-1})},$$

$$C_i^{(2)}(\theta) = -\frac{1}{N_n \Delta_n}\left\{a(V_{i\Delta_n})\Delta_n^{1/2} U_{i,n} + \varepsilon_{i,n}\right\}\frac{b(\widehat{V}_{i-1}, \theta)}{a^2(\widehat{V}_{i-1})},$$

$$C_i^{(3)}(\theta) = -\frac{1}{N_n \Delta_n}\{E_{i+1,n,k} - E_{i,n,k}\}\frac{b(\widehat{V}_{i-1}, \theta)}{a^2(\widehat{V}_{i-1})}.$$

The study of these terms is rather long and very similar to computations in [15] for most of them. Thus, we shall only give the main steps of the proof.

In the proof of Proposition 2.1 only the first sum $\sum_i C_i^{(1)}(\theta)$ has a non-negligible contribution. A direct application of Proposition 7.1 gives the convergence

$$\sum_{i=1}^{N-2} \frac{b^2(\widehat{V}_{i-1}, \theta)}{2N_n a^2(\widehat{V}_{i-1})} \xrightarrow{n\to\infty} \int_0^\infty \frac{b^2(x, \theta)}{2a^2(x)}\, d\nu_0(x).$$

Then Proposition 7.1 again with $E(|\overline{V}_i - \overline{V}_{i-1}|) \le c\Delta_n^{1/2}$ and (V4) gives the convergence of $-\sum_{i=1}^{N-2} \frac{b(\overline{V}_i, \theta_0) b(\widehat{V}_{i-1}, \theta)}{N a^2(\widehat{V}_{i-1})}$ to $-\int_0^\infty \frac{b(x, \theta_0) b(x, \theta)}{a^2(x)}\, d\nu_0(x)$. From this we deduce (recall (2.9)) that

$$\sum_{i=1}^{N-2} C_i^{(1)}(\theta) - \sum_{i=1}^{N-2} C_i^{(1)}(\theta_0) \xrightarrow{n\to\infty} \mathcal{C}(\theta_0, \theta).$$

The negligibility of $\sum_{i=1}^{N-2} C_i^{(2)}(\theta)$ involves the almost centering condition on $\varepsilon_{i,n}$ and the centered property of $U_{i,n}$ and relies on a martingale limit theorem. We omit the proof, since an analogous computation is presented in detail in the proof of Theorem 2 of [15].

We detail the negligibility of $\sum_{i=1}^{N-2} C_i^{(3)}(\theta)$, since it governs the rate conditions for the convergence to zero of b_n and υ_n. Remark that, by (V1), the sequence

$$\widetilde{C}_{2i}^{(3)}(\theta) = C_{2i}^{(3)}(\theta) - E\big(C_{2i}^{(3)}(\theta) \mid \mathcal{G}_{2i\Delta_n}^n\big), \quad \text{for } i = 1, \dots, \lfloor N/2 \rfloor,$$

is a martingale increments array with respect to $(\mathcal{G}_{2i\Delta_n}^n)_i$, and let $M_i^n = \sum_{i' \le i} \widetilde{C}_{2i'}^{(3)}(\theta)$ be the corresponding martingale. From Assumptions (V3–V4) we easily bound its bracket by

$$\langle M^n \rangle_{\lfloor N/2 \rfloor} \le c \sum_{i=1}^{\lfloor N/2 \rfloor} \frac{1}{N_n^2 \Delta_n^2} \frac{b^2(\widehat{V}_{2i-1}, \theta)}{a^4(\widehat{V}_{2i-1})} \big(1 + |V_{2i\Delta_n}|^c\big)\big(\upsilon_n + b_n^2\big).$$

Now Assumptions (V2), (V2′), and (A5) give the control

$$E\big(\big(M_{\lfloor N/2 \rfloor}^n\big)^2\big) = E\big(\langle M^n \rangle_{\lfloor N/2 \rfloor}\big) \le c \frac{\upsilon_n + b_n^2}{N_n \Delta_n^2} \sim c \frac{\upsilon_n + b_n^2}{T_n \Delta_n}, \tag{7.5}$$

which converges to zero by the conditions $b_n = o(\Delta_n)$ and $\upsilon_n = o(\Delta_n T_n)$. Using (V3), the centering term $\sum_{i=1}^{\lfloor N/2 \rfloor} E(C_{2i}^{(3)}(\theta) \mid \mathcal{G}_{2i\Delta_n}^n)$ is easily bounded in L^1-norm by cb_n/Δ_n, and hence converges to zero.

Thus, we have shown the L^1-convergence to zero of terms of even index in $\sum_{i=1}^{N} C_i^{(3)}(\theta)$. The convergence of terms of odd index is obtained similarly.

7.1.2 *Proof of Theorem 2.2*

After having remarked that, by (I1), the quantity $\mathcal{C}(\theta_0, \theta)$ admits a unique minimum at $\theta = \theta_0$, it is classical (see [7, 22]) that the consistency property follows from the uniform convergence

$$\sup_{\theta \in \Theta} \big|\mathcal{C}_n(\theta) - \mathcal{C}_n(\theta_0) - \mathcal{C}(\theta_0, \theta)\big| \xrightarrow{n \to \infty} 0, \quad \text{in probability.} \tag{7.6}$$

For a uniformity in the parameter θ of the convergence (2.10) we use a Kolmogorov criterion (see Theorem 20 in Appendix 1 of [20], for example) by proving $E(|\sum_i \{C_i^{(\ell)}(\theta) - C_i^{(\ell)}(\theta')\}|^p) \le c|\theta - \theta'|^p$ for $\ell = 1, 2, 3$, and p large enough. We omit the details here (see the proof of Theorem 2 in [15] or Lemma 1 in [22], for example).

7.1.3 *Proof of Theorem 2.4*

First, remark that, since $N_n \Delta_n \sim T_n \to \infty$, the conditions of Theorem 2.4 imply the consistency.

Second, on the event $\{\hat{\theta}_n \in \overset{\circ}{\Theta}\}$, we classically write a Taylor expansion around $\hat{\theta}_n$ as

$$\int_0^1 \nabla^2 \mathcal{C}_n\big(\theta_0 + s(\hat{\theta}_n - \theta_0)\big)\,\mathrm{d}s(\hat{\theta}_n - \theta_0) = -\nabla \mathcal{C}_n(\theta_0).$$

Then the asymptotic normality may be deduced from the two following properties:

$$\sqrt{T_n}\nabla \mathcal{C}_n(\theta_0) \xrightarrow[\mathcal{L}]{n\to\infty} \mathcal{N}\big(0, I(\theta_0)\big), \tag{7.7}$$

$$\sup_{\theta\in\Theta}\big|\partial_{\theta_i}\partial_{\theta_j}\mathcal{C}_n(\theta) - I_{i,j}(\theta,\theta_0)\big| \xrightarrow[\mathbf{P}]{n\to\infty} 0, \quad i,j \in \{1,\dots,d\}, \tag{7.8}$$

where $I_{i,j}(\theta,\theta_0)$ is given by

$$I_{i,j}(\theta,\theta_0) = \int_0^\infty a^{-2}(x)\big\{\partial_{\theta_i,\theta_j} b(x,\theta)\big(b(x,\theta) - b(x,\theta_0)\big) + \partial_{\theta_i} b(x,\theta)\partial_{\theta_j} b(x,\theta)\big\}\,\mathrm{d}\nu_0(x).$$

We start with (7.7). We use the decomposition of Sect. 7.1.1, and we shall see that only $\sqrt{T_n}\sum_{i=1}^{N-2}\nabla C_i^{(2)}(\theta_0)$ has a non-negligible contribution.

We write $\sqrt{T_n}\sum_{i=1}^{N-2}\nabla C_i^{(2)}(\theta_0)$ as

$$-\frac{1}{N^{1/2}}\sum_{i=1}^{N-2} U_{i,n}\frac{\nabla b(\widehat{V}_{i-1},\theta_0)a(V_{i\Delta_n})}{a^2(\widehat{V}_{i-1})} + \frac{1}{N^{1/2}\Delta_n^{1/2}}\sum_{i=1}^{N-2}\varepsilon_{i,n}\frac{\nabla b(\widehat{V}_{i-1},\theta_0)}{a^2(\widehat{V}_{i-1})}.$$

Using that $(U_{i,n})_i$ is a Gaussian process with known covariance structure, the first sum in the expression above can be shown to converge in law to $\mathcal{N}(0, I(\theta_0))$ (recall (2.6)) with the help of a theorem for convergence in law for martingale increment arrays. Details on this convergence may be found in Theorem 4 in [15]. Then the second sum above can be shown to converge to zero by an application of Lemma 9 in [10], using (7.3–7.4) and $N_n^{1/2}\Delta_n^{3/2} \to 0$.

Again the negligibility of $\sqrt{T_n}\sum_{i=1}^{N-2}\nabla C_i^{(3)}(\theta_0)$ governs the conditions on b_n and v_n. By exactly the same computation as for (7.5) we show that the sum $\sqrt{T_n}\sum_{i=1}^{\lfloor N/2\rfloor -1}\{\nabla C_{2i}^{(3)}(\theta_0) - E(\nabla C_{2i}^{(3)}(\theta_0) \mid \mathcal{G}_{2i\Delta_n}^n)\}$, denoted by $\tilde{M}_{\lfloor N/2\rfloor}^n$, can be bounded as

$$E\big((\tilde{M}_{\lfloor N/2\rfloor}^n)^2\big) \le c\frac{T_n}{N\Delta_n^2}\big(b_n^2 + v_n\big) \sim c\frac{1}{\Delta_n}\big(b_n^2 + v_n\big) \xrightarrow{n\to\infty} 0.$$

Then, by (V3), the contribution of $\sqrt{T_n}\sum_{i=1}^{\lfloor N/2\rfloor -1}|E(\nabla C_{2i}^{(3)}(\theta_0) \mid \mathcal{G}_{2i\Delta_n}^n)|$ has an L^1-norm controlled by $\sqrt{T_n}b_n/\Delta_n \to 0$.

For the term $\sqrt{T_n}\sum_{i=1}^{N-2}\nabla C_i^{(1)}(\theta_0)$ we observe the following simple expression for $\nabla C_i^{(1)}(\theta_0)$:

$$\begin{aligned}\nabla C_i^{(1)}(\theta_0) &= \frac{\nabla_\theta b(\widehat{V}_{i-1},\theta_0)}{N_n a^2(\widehat{V}_{i-1})}\big[b(\widehat{V}_{i-1},\theta_0)-b(\overline{V}_i,\theta_0)\big]\\ &= \frac{\nabla_\theta b(\widehat{V}_{i-2},\theta_0)}{N_n a^2(\widehat{V}_{i-2})}\big[b(\widehat{V}_{i-1},\theta_0)-b(\overline{V}_i,\theta_0)\big]+r_{i,n},\end{aligned}$$

with $E|r_{i,n}|\le N_n^{-1}(\Delta_n+v_n)$, where we have used (V4), $E(|\overline{V}_{i-1}-\overline{V}_{i-2}|)\le c\Delta_n^{1/2}$ and (A5). Then, using the conditions $N_n\Delta_n^3\to 0$ and $v_n=o(\Delta_n)$, we deduce $\sqrt{T_n}\sum_{i=1}^{N-2} r_{i,n}\to 0$ in L^1-norm.

To end the proof of (7.7), it can be shown after rather long computations using Taylor expansions with Theorem 7.2, (V3–V4) and $v_n=o(\Delta_n)$ that

$$\big|E\big(b(\overline{V}_i,\mu_0)-b(\widehat{V}_{i-1},\mu_0)\mid \mathcal{G}^n_{(i-1)\Delta_n}\big)\big|\le c(b_n+\Delta_n)\big(V^c_{(i-1)\Delta_n}+V^{-c}_{(i-1)\Delta_n}\big).$$

This control, together with the condition on the second moment,

$$E\big(\big(b(\overline{V}_i,\theta_0)-b(\widehat{V}_{i-1},\theta_0)\big)^2\mid \mathcal{G}^n_{(i-1)\Delta_n}\big)\le c\Delta_n\big(V^c_{(i-1)\Delta_n}+V^{-c}_{(i-1)\Delta_n}\big),$$

is sufficient to apply Lemma 9 in [10] and deduce the convergence to zero of $\sqrt{T_n}\sum_{i=1}^{N-2}\nabla C_i^{(1)}(\theta_0)$.

Now we prove (7.8). Exactly as in Sect. 7.1.1 we show that

$$\partial^2_{\theta_u\theta_v}\mathcal{C}_n(\theta)=\sum_{i=1}^{N-2}\partial^2_{\theta_u\theta_v}C_i^{(1)}(\theta)+o_{\mathbf{P}}(1),$$

where $u,v\in\{1,\dots,d\}$. Then we differentiate $C_i^{(1)}(\theta)$ twice and obtain, uniformly in θ, the convergence as in Sect. 7.1.1:

$$\sum_{i=1}^{N-2}\partial^2_{\theta_u\theta_v}C_i^{(1)}(\theta)\xrightarrow{n\to\infty} I_{u,v}(\theta,\theta_0).$$

7.1.4 Details on the proof of Theorem 2.7

By applying Proposition 2 in [15], we have that $N_n^{-1}\sum_{i=1}^{N-2}\overline{V}_{i-1}^{-1}\to\int x^{-1}\,\mathrm{d}\nu_0(x)$ and $N_n^{-1}\sum_{i=1}^{N-2}\overline{V}_{i-1}^{-2}\to\int x^{-2}\,\mathrm{d}\nu_0(x)$, for $a>1$. We can deduce by a repetition of the proof of Proposition 7.1 the convergence for $\mathcal{M}_n$, where we replace the condition on the positive moments in (A5) by the fact that x^{-1}, x^{-2} and their derivatives are bounded near $+\infty$ and the condition on the first moment of the stationary law, i.e., $\sup_t E(V_t)<\infty$.

Before proving (2.15), we make precise the Euler scheme of [14] in the case of the GARCH model

$$\overline{V}_{i+1}-\overline{V}_i-\Delta_n(\alpha\overline{V}_i+\beta)=\sigma V_{i\Delta_n}\Delta_n^{1/2}U_{i,n}+\varepsilon_{i,n},$$

where $|E(\varepsilon_{i,n} \mid \mathcal{G}_i^n)| \le c\Delta_n^2(1 + V_{i\Delta_n}^2)$ and $E(|\varepsilon_{i,n}|^p \mid \mathcal{G}_i^n) \le c\Delta_n^p(1 + V_{i\Delta_n}^{3p/2})$, for $1 \le p \le 4$. Then the proof is a repetition of the proof of (7.7) where we carefully apply the condition (A5) only for positive moments, i.e., $\sup_t E(V_t^c) < \infty$ with $c \le 3$.

7.2 Proof of condition (V2′) for the realized variance

We intend to show the following result for the quantity $\widehat{V}_i$ given by (3.2).

Proposition 7.3 *Assume* (A0–A2) *and* $\Delta_n \le 1$. *Then, for all* $p \ge 1$, *we have for* n *large enough and* $i \in \{0, \dots, N_n - 1\}$ *the control*

$$E\big(\widehat{V}_i^{-p} \mid \mathcal{G}_{i\Delta_n}\big) \le c(p)\big(1 + V_{i\Delta_n}^{-p}\big).$$

Proof Conditionally on $\mathcal{G}_V \vee \mathcal{G}_{i\Delta_n}$ (see the notation in the proof of Proposition 3.1), the variable $\widehat{V}_i$ is equal in law to the sum $\sum_{j=1}^k Z_j^2$ with independent $Z_j \sim \mathcal{N}(\mu_j, v_j)$, where

$$v_j = \Delta_n^{-1} \int_{(ik+j-1)\delta_n}^{(ik+j)\delta_n} V_s \, ds \ge \frac{\delta_n}{\Delta_n} \Big(\inf_{s \in [i\Delta_n, (i+1)\Delta_n]} V_s \Big),$$

and the μ_j are some real constants. From Lemma 7.4 below (recall $k_n \to \infty$), we deduce for n large enough that

$$E\big(\widehat{V}_i^{-p} \mid \mathcal{G}_V \vee \mathcal{G}_{i\Delta_n}\big) \le c(p) \left(\frac{\Delta_n}{\delta_n}\right)^p \sup_{s \in [i\Delta_n, (i+1)\Delta_n]} V_s^{-p} \, k_n^{-p}.$$

Then using $\Delta_n/\delta_n = k_n$ and Assumption (A2) while taking conditional expectations in the expression above yields the result. □

Lemma 7.4 *Let* $Z_1, \dots, Z_k$ *be independent random variables such that the law of* Z_j *is* $\mathcal{N}(\mu_j, v_j)$. *Then, for all* $p \ge 0$, *there exists* $c(p)$, *a constant depending only on* p, *such that for all* $k \ge 2p + 3$

$$E\big(\big(Z_1^2 + \cdots + Z_k^2\big)^{-p}\big) \le c(p) \Big(\min_{j=1,\dots,k} v_j \Big)^{-p} k^{-p}.$$

Proof If there exists j such that $v_j = 0$, the result is obvious. Otherwise, for all $j = 1, \dots, k$ we set $\tilde{Z}_j = \frac{Z_j}{\sqrt{v_j}}$. Then

$$E\big(\big(Z_1^2 + \cdots + Z_k^2\big)^{-p}\big) \le \Big(\min_{j=1,\dots,k} v_j \Big)^{-p} E\big(\big(\tilde{Z}_1^2 + \cdots + \tilde{Z}_k^2\big)^{-p}\big).$$

Since the density of $\tilde{Z}_j$ is bounded on $\mathbb{R}$ by $\frac{1}{\sqrt{2\pi}}$ we can apply Lemma 7.5 given in Sect. 7.3 and we get the result. □

7.3 Inverse moments for positive variables

In this section we present a result on the behavior near zero for a sum of positive random variables that was needed for the proof of Lemma 7.4.

Lemma 7.5 *Let $Z_1, \dots, Z_k$ be k independent random variables. We suppose that Z_j has a density $\phi_j(x)$ and that there exist $R > 0$, $M > 0$ such that for all $x \in [-R, R]$ and for all $j \in \{1, \dots, k\}$, $\phi_j(x) \le M$. Then, for all $p \ge 0$, there exists $c(p, M, R)$, a constant depending on p, M, R, such that for all $k \ge 2p + 3$*

$$E\left(\left(\sum_{j=1}^{k} Z_j^2\right)^{-p}\right) \le c(p, M, R)k^{-p}. \tag{7.9}$$

Proof By increasing M, we may assume that $\frac{1}{2M} \le R$. Let $(U_j)_{j=1,\dots,k}$ be k independent real variables with uniform distribution on $[0, \frac{1}{2M}]$. First, we show that

$$M_p := E\left(\left(\sum_{j=1}^{k} Z_j^2\right)^{-p}\right) \le E\left(\left(\sum_{j=1}^{k} U_j^2\right)^{-p}\right) =: N_p. \tag{7.10}$$

Denote by ϕ_j the distribution function of $|Z_j|$ and by ψ the distribution function of U_j. Then, for $x \ge 0$,

$$\phi_j(x) = P(|Z_j| \le x) = \int_{-x}^{x} \phi_j(s)\, \mathrm{d}s \le (2Mx \wedge 1) = \psi(x). \tag{7.11}$$

This implies that, for all $j = 1, \dots, k$ and $t \in (0, 1)$,

$$\psi^{-1}(t) \le \phi_j^{-1}(t). \tag{7.12}$$

Now consider k independent real variables, $A_1, \dots, A_k$, with uniform law on $(0, 1)$, and set, for all j, $Z_j^* = \phi_j^{-1}(A_j)$, $\quad U_j^* = \psi^{-1}(A_j)$. We know that $(Z_j^*)_{j=1,\dots,k}$ and $(U_j^*)_{j=1,\dots,k}$ have, respectively, the same law as $(|Z_j|)_{j=1,\dots,k}$ and $(U_j)_{j=1,\dots,k}$. Furthermore, by (7.12), we have $\forall j$, $U_j^* \le Z_j^*$ a.s., and thus $M_p = E((\sum_{j=1}^k Z_j^{*2})^{-p}) \le E((\sum_{j=1}^k U_j^{*2})^{-p}) = N_p$. Hence, (7.10) is proved. It remains to obtain an upper bound for N_p. For this note that

$$\begin{aligned} P\big(U_1^2 + \cdots + U_k^2 \le \varepsilon\big) &= \int_{[0, \frac{1}{2M}]^k} \mathbb{1}_{\{u_1^2+\cdots+u_k^2 \le \varepsilon\}} (2M)^k \, \mathrm{d}u_1 \cdots \mathrm{d}u_k \\ &\le (2M)^k \int_{\mathbb{R}^k} \mathbb{1}_{\{u_1^2+\cdots+u_k^2 \le \varepsilon\}} \, \mathrm{d}u_1 \cdots \mathrm{d}u_k. \end{aligned}$$

We denote by $\sigma_k = 2\frac{\pi^{\frac{k}{2}}}{\Gamma(\frac{k}{2})}$ the area of the unit sphere in the Euclidean space $\mathbb{R}^k$. By a change of variable, we get

$$P\big(U_1^2 + \cdots + U_k^2 \le \varepsilon\big) \le (2M)^k \int_0^{\sqrt{\varepsilon}} \rho^{k-1} \, \mathrm{d}\rho \sigma_k = (2M)^k \frac{\varepsilon^{\frac{k}{2}}}{k} \sigma_k. \tag{7.13}$$

Let ρ_k be a positive constant depending on k that will be specified later. Define

$$N'_p = E\big(\big(U_1^2 + \cdots + U_k^2\big)^{-p} \mathbb{1}_{\{U_1^2+\cdots+U_k^2 \le \rho_k^2\}}\big).$$

We have $N'_p \le \sum_{q=1}^{\infty} u_q$, where $u_q = (q+1)^p \rho_k^{-2p} P(U_1^2 + \cdots + U_k^2 \in [\frac{\rho_k^2}{q+1}, \frac{\rho_k^2}{q}])$. Using (7.13) with $\varepsilon = \rho_k^2 q^{-1}$, we get $u_q \le (q+1)^p q^{-k/2} (2M)^k \rho_k^{k-2p} \sigma_k k^{-1}$. Since $k/2 - p \ge 3/2$, we can set $s = \sum_{q=1}^{\infty} (q+1)^p q^{-\frac{k}{2}} < \infty$, and $N'_p \le s(2M)^k \rho_k^{k-2p} \sigma_k k^{-1}$. Let us recall the exact inequality connected with the Stirling formula (see p. 54 [8]):

$$\Gamma\left(\frac{k}{2}\right) > \left(\frac{\frac{k}{2}-1}{e}\right)^{\frac{k}{2}-1} \sqrt{2\pi\left(\frac{k}{2}-1\right)}.$$

Using this inequality and the value of σ_k yields

$$N'_p \le \frac{s(2M)^k \rho_k^{k-2p} \sqrt{2\pi}}{k(\frac{\frac{k}{2}-1}{e\pi})^{\frac{k}{2}-1} \sqrt{\frac{k}{2}-1}}.$$

Now we set $\rho_k = (\frac{\frac{k}{2}-1}{4e\pi M^2})^{\frac{1}{2}}$, and the previous inequality reduces to

$$N'_p \le \frac{s\sqrt{2\pi}(4M)^p}{(\frac{\frac{k}{2}-1}{e\pi})^{p-1} k \sqrt{\frac{k}{2}-1}} \le Ck^{-p},$$

where C only depends on p, M, R.

Since $E((U_1^2 + \cdots + U_k^2)^{-p} \mathbb{1}_{\{U_1^2+\cdots+U_k^2 > \rho_k^2\}}) \le \rho_k^{-2p} = (\frac{8e\pi M^2}{k-2})^p$, we have

$$E\big(\big(U_1^2 + \cdots + U_k^2\big)^{-p}\big) \le c(p, M, R) k^{-p}.$$

Hence, Lemma 7.5 is proved. □

Acknowledgements The author would like to thank the referees and the associate editor for many suggestions that greatly improved the paper. He also wishes to thank Professor V. Genon-Catalot for constant encouragement.

References

1. Aït-Sahalia, Y., Kimmel, R.: Maximum likelihood estimation of stochastic volatility models. J. Finance Econ. **83**, 413–452 (2007)
2. Andersen, T.G., Bollerslev, T., Meddahi, N.: Correcting the errors: volatility forecast evaluation using high-frequency data and realized volatilities. Econometrica **73**, 279–296 (2005)
3. Barndorff-Nielsen, O.E., Shephard, N.: Econometric analysis of realized covariation: high frequency based covariance, regression, and correlation in financial economics. Econometrica **72**, 885–925 (2004)
4. Bates, D.S.: Maximum likelihood estimation of latent affine processes. Rev. Finance Stud. **19**, 909–965 (2006)

5. Bollerslev, T., Zhou, H.: Estimating volatility diffusion using conditional moments of integrated volatility. J. Econom. **109**, 33–65 (2002)
6. Corradi, V., Distaso, W.: Semi-parametric comparison of stochastic volatility models using realized measures. Rev. Econom. Stud. **73**, 635–667 (2006)
7. Dacunha-Castelle, D., Duflo, M.: Probabilité et Statistiques. Tome 2. Problèmes à Temps Mobile. Masson, Paris (1983)
8. Feller, W.: An Introduction to Probability Theory and its Application, vol. 1, 3rd edn. Wiley, New York (1969)
9. Florens-Zmirou, D.: Approximate discrete-time schemes for statistics of diffusion processes. Statistics **20**, 547–557 (1989)
10. Genon-Catalot, V., Jacod, J.: On the estimation of the diffusion coefficient for multidimensional diffusion processes. Ann. Inst. Henri Poincaré, Probab. Stat. **29**, 119–151 (1993)
11. Genon-Catalot, V., Jeantheau, T., Laredo, C.: Parameter estimation for discretely observed stochastic volatility models. Bernoulli **5**, 855–872 (1999)
12. Genon-Catalot, V., Jeantheau, T., Laredo, C.: Stochastic volatility models as hidden Markov models and statistical applications. Bernoulli **6**, 1051–1079 (2000)
13. Gloter, A.: Estimation du coefficient de diffusion de la volatilitité d'un modèle à volatilité stochastique. [Estimation of the volatility diffusion coefficient for a stochastic volatility model]. C.R. Acad. Sci. Paris Sér. I Math. **330**, 243–248 (2000)
14. Gloter, A.: Discrete sampling of an integrated diffusion process and parameter estimation of the diffusion coefficient. ESAIM Probab. Stat. **4**, 205–224 (2000)
15. Gloter, A.: Parameter estimation for a discretely observed integrated diffusion process. Scand. J. Stat. **33**, 83–104 (2006)
16. Gloter, A., Hoffmann, M.: Stochastic volatility and fractional Brownian motion. Stoch. Process. Appl. **113**, 143–172 (2004)
17. Heston, S.: A closed-form solution for options with stochastic volatility with applications to bonds and currency options. Rev. Finance Stud. **6**, 327–343 (1993)
18. Hoffmann, M.: Rate of convergence for parametric estimation in a stochastic volatility model. Stoch. Process. Appl. **97**, 147–170 (2002)
19. Hull, J., White, A.: The pricing of options on assets with stochastic volatilities. J. Finance **42**, 281–300 (1987)
20. Ibragimov, I.A., Khas'minskii, R.Z.: Statistical Estimation, Asymptotic Theory. Springer, Berlin (1981)
21. Jacod, J.: Limit of random measures associated with the increments of a Brownian semimartingale. Preprint 120, Laboratoire de Probabilité, Université Pierre et Marie Curie, Paris (1994)
22. Kessler, M.: Estimation of an ergodic diffusion from discrete observations. Scand. J. Stat. **24**, 211–229 (1997)
23. Phillips, P.C.B., Yu, J.: A two stage realized volatility approach to the estimation for diffusion processes from discrete observations. Cowles Foundation Discussion Paper 1523, Cowles Foundation, Yale University (2005). http://cowles.econ.yale.edu/P/cd/d15a/d1523.pdf
24. Ruiz, E.: Quasi-maximum likelihood estimation of stochastic volatility models. J. Econom. **63**, 289–306 (1994)
25. Scott, L.O.: Option pricing when the variance changes randomly: theory, estimation, and an application. J. Finance Quant. Anal. **22**, 419–438 (1987)
26. Shephard, N.: Stochastic Volatility (Advanced Texts in Econometric). Oxford University Press, London (2005)
27. Sørensen, M.: Prediction-based estimating functions. Econom. J. **3**, 123–147 (2000)
28. Wiggins, J.B.: Option values under stochastic volatility: theory and empirical estimates. J. Finance Econ. **19**, 351–372 (1987)
29. Yoshida, N.: Estimation for diffusion processes from discrete observation. J. Multivar. Anal. **41**, 220–242 (1992)
30. Zhang, L.: Efficient estimation of stochastic volatility using noisy observations: a multi-scale approach. Bernoulli **12**, 1019–1043 (2006)
31. Zhang, L., Mykland, P.A., Aït-Sahalia, Y.: A tale of two time scales: determining integrated volatility with noisy high-frequency data. J. Am. Stat. Assoc. **100**, 1394–1411 (2005)