Bernoulli **5**(5), 1999, 855–872

Parameter estimation for discretely observed stochastic volatility models

VALENTINE GENON-CATALOT[1*], THIERRY JEANTHEAU[1**] and CATHERINE LAREDO[2]

[1]*Université de Marne-la-Vallée, Equipe d'Analyse et de Mathématiques Appliquées, Cité Descartes 5, Bd Descartes, Champs-sur-Maine, 77454, Maine-La-Vallée, Cedex 2, France. E-mail: *genon@univ-mlv.fr **jeantheau@univ-mlv.fr*
[2]*Institut National de la Recherche Agronomique, Laboratoire de Biométrie, 78350 Jouy-en-Josas, France. E-mail: Catherine.Laredo@jouy.inra.fr*

This paper deals with parameter estimation for stochastic volatility models. We consider a two-dimensional diffusion process (Y_t, V_t). Only (Y_t) is observed at n discrete times with a regular sampling interval. The unobserved coordinate (V_t) rules the diffusion coefficient (volatility) of (Y_t) and is an ergodic diffusion depending on unknown parameters. We build estimators of the parameters present in the stationary distribution of (V_t), based on appropriate functions of the observations. Consistency is proved under the asymptotic framework that the sampling interval tends to 0, while the number of observations and the length of the observation time tend to infinity. Asymptotic normality is obtained under an additional condition on the rate of convergence of the sampling interval. Examples of models from finance are treated, and numerical simulation results are given.

Keywords: diffusion processes; discrete time observations; mathematical finance; parametric inference; stochastic volatility

1. Introduction

Many recent contributions in the field of finance are devoted to modelling stock prices. The seminal Black–Scholes (1973) model assumes that stock prices follow a geometric Brownian motion with constant volatility, but several empirical works have rejected the shortcomings of the Black–Scholes model, and new models have been introduced (see, for example, for a survey, Ghysels *et al*. (1996)). Among these, the so-called stochastic volatility models have been proposed (see, for example, Hull and White (1987), Chesney and Scott (1989) and Heston (1993)). These models include stochastic volatility in such a way that the couple (stock price, volatility) behaves as a two-dimensional diffusion process, depending on unknown parameters. The problem of estimating these parameters from stock prices is not clearly addressed in all these studies. Indeed, this raises a difficulty since only the stock price is observable, while the volatility is unobservable. It leads to a new problem in the statistics of diffusion processes.

In the case where all coordinates are observed, the estimation of drift and diffusion coefficients is now classical. Among many references, let us quote Kutoyants (1984),

1350–7265 © 1999 ISI/BS

Dacunha-Castelle and Florens-Zmirou (1986), Donahl (1987), Larédo (1990), Genon-Catalot and Jacod (1993), Bibby and Sørensen (1995) and Kessler (1997). These papers do not answer the estimation problem raised by stochastic volatility models.

Let us now come to the precise model investigated here. Assume that (Y_t) is the logarithm of the price process of some stock and that it is ruled by

$$\mathrm{d}Y_t = \mu(\sigma_t^2)\,\mathrm{d}t + \sigma_t\,\mathrm{d}B_t, \qquad Y_0 = 0.$$

Here, (σ_t) is the positive volatility of the stock price, (B_t) a Brownian motion and μ some real function. Assume also that $V_t = \sigma_t^2$ is another diffusion process, whose coefficients depend on an unknown multidimensional parameter θ, and which is defined by

$$\mathrm{d}V_t = b(\theta, V_t)\,\mathrm{d}t + a(\theta, V_t)\,\mathrm{d}W_t, \qquad V_0 = \eta.$$

We assume that $(B_t, W_t)_{t\geqslant 0}$ is a two-dimensional standard Brownian motion, (V_t) is a positive diffusion process and η is a positive random variable independent of $(B_t, W_t)_{t\geqslant 0}$. The sample path (Y_t) is discretely observed at regularly spaced times $t_i = i\Delta$, $i = 1, \ldots, n$. Our concern here is to give a general methodology for estimating θ from the observations $(Y_{t_1}, \ldots, Y_{t_n})$ only. For the statistical study, the key assumption is that the diffusion (V_t) has to be a strictly stationary positive process with stationary distribution π_θ having at least a second-order moment. Apart from this assumption, the diffusion model for (V_t) can be quite general. Thus, our results will apply to most of the specific models proposed in the financial literature.

As is always the case in the statistics of diffusion processes based on discrete observations, the likelihood function is untractable. Our approach is to construct appropriate and explicit functions of the observations to replace either the log-likelihood or the score function. We study the asymptotic properties of the estimators built using these functions. Our results rely on limit theorems proved in a previous paper (Genon-Catalot *et al.* 1998) which concern the empirical distribution of the increments $(Y_{t_i} - Y_{t_{i-1}}, i = 1, \ldots, n)$, and are obtained in the following asymptotic framework. The number of observations n tends to infinity, the sampling interval $\Delta = \Delta_n$ tends to zero and the length of the observation time $n\Delta_n$ tends to infinity. As a result of the previous paper, the function μ cannot be estimated in this framework and behaves as a nuisance parameter. Moreover, it can be set equal to 0 without modifying the estimators (see Section 7). Therefore, this paper is devoted to the case $\mu = 0$. Also, the only relevant parameters in finance are those of the volatility.

The paper is organized as follows. Assumptions and preliminary results are presented in Section 2. The essential property is that the random variables $\Delta_n^{-1/2}(Y_{t_i} - Y_{t_{i-1}})$ behave asymptotically like a sample of Q_θ, defined as the distribution of $\epsilon\eta^{1/2}$, where (ϵ, η) has the law $\mathcal{N}(0, 1) \otimes \pi_\theta$. Thus, Q_θ is a variance mixture of Gaussian laws, the mixing distribution being the stationary distribution of the unobservable diffusion (V_t). In Section 3, we study estimators based on the minimization of functions such as

$$U_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} u(\theta, \Delta_n^{-1/2}(Y_{t_i} - Y_{t_{i-1}})),$$

where $u(\theta, x)$ is a real function satisfying appropriate conditions. The minimum-contrast

estimators are proved to be consistent and, under the additional condition $n\Delta_n^2 \to 0$, asymptotically normal with rate $(n\Delta_n)^{1/2}$. Section 3.3 is devoted to the special case $u(\theta, x) = -\log q(\theta, x)$ where $q(\theta, x)$ is the probability density of the mixture distribution Q_θ. It is especially well fitted to the problem and possesses the following noteworthy feature: all the assumptions required to get identifiability, consistency and asymptotic normality can be expressed in terms of conditions on the stationary distribution π_θ of the (V_t) model. Another classical approach in situations where the likelihood is untractable is to define estimators from estimating equations. This is done in Section 4. A consequence is that empirical moment estimators are included in this approach. Both methods only provide estimators of the parameters present in the stationary distribution of (V_t). However, these parameters may come from its drift or diffusion coefficient. In Section 5, examples of parametric models commonly used in finance are fully treated. In Section 6, numerical results based on simulations are presented. Section 7 discusses extensions and gives concluding remarks.

2. Assumptions and preliminary results

2.1. The model

Let $(Y_t, V_t)_{t\geqslant 0}$ be a two-dimensional diffusion process defined by

$$\mathrm{d}Y_t = \sigma_t \,\mathrm{d}B_t, \qquad Y_0 = 0, \tag{1}$$

$$V_t = \sigma_t^2 \qquad \text{and} \qquad \mathrm{d}V_t = b(\theta, V_t)\,\mathrm{d}t + a(\theta, V_t)\,\mathrm{d}W_t, \qquad V_0 = \eta. \tag{2}$$

We assume that

(a) $(B_t, W_t)_{t\geqslant 0}$ is a standard Brownian motion of $\mathbb{R}^2$ defined on a probability space $(\Omega, \mathscr{A}, \mathbf{P})$ and

(b) η is a random variable defined on Ω, independent of $(B_t, W_t)_{t\geqslant 0}$.

Equation (2) defines a one-dimensional diffusion process whose coefficients depend on an unknown parameter $\theta \in \Theta \subset \mathbb{R}^d$. We make now the standard assumptions on functions $b(\theta, v)$ and $a(\theta, v)$ ensuring that the solution of (2) is a positive recurrent diffusion on $(0, +\infty)$ and a strictly stationary ergodic process.

(A1) For all $\theta \in \Theta$, $b(\theta, v)$ and $a(\theta, v)$ are continuous (in v) real functions on $\mathbb{R}$, and C^1 functions on $(0, +\infty)$ such that

$$\exists k > 0, \qquad \forall v > 0, \qquad b^2(\theta, v) + a^2(\theta, v) \leqslant k(1 + v^2) \qquad \text{and} \qquad \forall v > 0, \qquad a(\theta, v) > 0.$$

For $v_0 > 0$, define the derivative of the scale function of diffusion (V_t):

$$s(\theta, v) = \exp\left(-2\int_{v_0}^{v} \frac{b(\theta, u)}{a^2(\theta, u)}\,\mathrm{d}u\right). \tag{3}$$

(A2) For all $\theta \in \Theta$,

$$\int_{0+} s(\theta, v)\,\mathrm{d}v = +\infty, \qquad \int^{+\infty} s(\theta, v)\,\mathrm{d}v = +\infty, \qquad \int_0^{+\infty} \frac{\mathrm{d}v}{a^2(\theta, v)s(\theta, v)} = M_\theta < +\infty.$$

Let

$$\pi(\theta, v) = \frac{1}{M_\theta}\frac{1}{a^2(\theta, v)s(\theta, v)}1_{(v>0)}. \tag{4}$$

(A3) For all $\theta \in \Theta$, the initial random variable η has distribution $\pi_\theta(\mathrm{d}v) = \pi(\theta, v)\,\mathrm{d}v$.
(A4) There exists $\gamma \geqslant 2$ such that, $\forall\theta \in \Theta$, $\int_0^{+\infty} v^\gamma \pi(\theta, v)\,\mathrm{d}v < +\infty$.

Assumptions (A1) and (A2) ensure existence and uniqueness of the solution of (2) together with the positive recurrence on $(0, +\infty)$. In particular, $V_t > 0$ almost surely, for all $t \geqslant 0$. Assumption (A3) provides the strict stationarity property and the ergodicity (see, for example, Rogers and Williams (1987)). Moreover, note that, for each $t \geqslant 0$, V_t has the distribution π_θ and, under (A4), $\mathrm{E}V_t^\gamma < \infty$.

From now on, let $(C, \mathscr{C}, (Y_t, V_t)_{t\geqslant 0}, P_\theta)$ be the canonical diffusion solution of (1), (2), where $C = C(\mathbb{R}^+, \mathbb{R}^2)$ is the space of continuous functions on $\mathbb{R}^+$ and $\mathbb{R}^2$-valued, $\mathscr{C}$ is the Borel σ-field associated with the uniform topology on each compact subset of $\mathbb{R}^+$, $(Y_t, V_t)_{t\geqslant 0}$ is the canonical process, and P_θ is the distribution on $(C, \mathscr{C})$ of the solution of (1), (2).

2.2. Observations

We assume that (Y_t) is observable but at n discrete times only with regular sampling interval Δ. We denote by $t_i = i\Delta$, $i = 1, \ldots, n$, these observation times. Our results are obtained under the assumption that

$$\Delta = \Delta_n \to 0 \qquad \text{and} \qquad n\Delta_n \to +\infty. \tag{5}$$

Let us set

$$X_i = X_i^n = \frac{1}{\Delta_n^{1/2}}(Y_{t_i} - Y_{t_{i-1}}) \qquad \text{and} \qquad t_i = t_i^n = i\Delta_n, \qquad i = 1, \ldots, n. \tag{6}$$

Conditionally on $\mathscr{G} = \sigma(V_t, t \geqslant 0)$, the random variables X_i are independent and X_i has distribution $\mathscr{N}(0, \overline{V}_i)$ with

$$\overline{V}_i = \frac{1}{\Delta}\int_{t_{i-1}}^{t_i} V_s\,\mathrm{d}s. \tag{7}$$

Thus, under the model (1), (2), the X_i are variance mixtures of Gaussian distributions, and the likelihood function is untractable because the joint distribution of $(\overline{V}_1, \ldots, \overline{V}_n)$ is not explicitly known. So, the alternative approach is to construct appropriate and explicit functions of the observations to replace either the log-likelihood or the score function. The statistical study relies on limit theorems concerning the empirical distribution of the sample $(X_i, i \leqslant n)$, which are presented below.

Remark 2.1. The problem of estimating θ from a discrete observation of the sample path (V_t) itself has been investigated under different asymptotics. To obtain consistent estimators of parameters in the diffusion coefficient of (V_t), it is assumed that the sampling interval $\Delta = \Delta_n$ tends to 0 while the time interval $T = n\Delta_n$ where observations are available, remains fixed (Donahl 1987; Genon-Catalot and Jacod 1993). It is worth noting that these results cannot be applied to obtain estimators of θ in our model. Indeed, a discrete Δ-sampling observation of (Y_t) on a fixed length time interval $[0, T]$ leads to a pointwise estimation $(\hat{V}_t)$ of (V_t), $t \in [0, T]$, which is not uniform on $[0, T]$. This is not enough to derive consistent estimators of θ. To go further, we have to use the ergodicity properties of (V_t), and so let $T = n\Delta_n$ tend to infinity.

2.3. Preliminary results

We recall some results proved by Genon-Catalot *et al.* (1998). We have investigated, for φ belonging to a class of functions, the asymptotic behaviour of

$$\hat{P}_n(\varphi) = \frac{1}{n}\sum_{i=1}^{n} \varphi(X_i^n).$$

For φ: $\mathbb{R} \to \mathbb{R}$, let us consider when defined the function h_φ: $\mathbb{R}^+ \to \mathbb{R}$

$$h_\varphi(v) = \int_{\mathbb{R}} \varphi(xv^{1/2})n(x)\,\mathrm{d}x, \tag{8}$$

where $n(x) = \{1/(2\pi)^{1/2}\}\exp(-x^2/2)$ is the $\mathcal{N}(0, 1)$ density.

Let $p \geqslant 0$, and $\mathcal{F}_p$ be the class of functions satisfying the following two conditions (C1) and (C2).

(C1) $\varphi \in C^1(\mathbb{R})$ and $\exists K_1 > 0\ \forall x \in \mathbb{R}\ |\varphi(x)| \leqslant K_1(1 + |x|^p)$.
(C2) $h_\varphi \in C^1[0, +\infty)$ and $\exists K_2 > 0, \forall v \geqslant 0,\ |h'_\varphi(v)| \leqslant K_2(1 + v^{p/2})$.

Let us consider the probability density and the distribution defined on $\mathbb{R}$ by

$$q(\theta, x) = \int_0^{+\infty} \frac{1}{(2\pi v)^{1/2}} \exp\left(-\frac{x^2}{2v}\right)\pi(\theta, v)\,\mathrm{d}v \qquad \text{and} \qquad Q_\theta(\mathrm{d}x) = q(\theta, x)\,\mathrm{d}x. \tag{9}$$

The probability Q_θ is the distribution of $\varepsilon\eta^{1/2}$ where (ε, η) follows a $\mathcal{N}(0, 1) \otimes \pi_\theta$. For $k \leqslant \gamma$, let us define the kth moment of the stationary distribution π_θ by $m_\theta(k) = \mathrm{E}(\eta^k)$. Note that

$$\int_{\mathbb{R}} |x|^\alpha Q_\theta(\mathrm{d}x) = C_\alpha m_\theta\left(\frac{\alpha}{2}\right), \tag{10}$$

is well defined for $\alpha/2 \leqslant \gamma$, where C_α is the αth absolute moment of the law $\mathcal{N}(0, 1)$.

Remark 2.2. Set $\varphi(x) = \psi(x^2)$; φ belongs to $\mathcal{F}_p$ if $\psi \in C^1[0, +\infty)$ and satisfies

$$\exists K > 0, \forall u \geqslant 0, \qquad |\psi(u)| + |\psi'(u)| \leqslant K(1 + u^{p/2}).$$

For φ belonging to $\mathscr{F}_\gamma$, with γ given in (A4), let $Q_\theta(\varphi) = \int_{\mathbb{R}} \varphi(x) Q_\theta \, \mathrm{d}x$, which is well defined (see (10)). For functions φ, φ_1 and φ_2, let us introduce (see (8) and (A2))

$$A_\varphi(\theta, v) = \int_0^v \{h_\varphi(u) - Q_\theta(\varphi)\} \pi(\theta, u) \, \mathrm{d}u, \qquad v \geqslant 0, \tag{11}$$

$$V_\theta(\varphi_1, \varphi_2) = 4M_\theta \int_0^{+\infty} s(\theta, v) A_{\varphi_1}(\theta, v) A_{\varphi_2}(\theta, v) \, \mathrm{d}v. \tag{12}$$

We can give now the main result concerning the asymptotic behaviour of $\hat{P}_n(\varphi)$.

Theorem 2.1. *Assume* (A1)–(A4)*, and let* $\Delta_n \to 0$ *and* $n\Delta_n \to +\infty$.

(a) *Let* φ *belong to* $\mathscr{F}_\gamma$*, with* γ *given in* (A4)*, for all* $\theta \in \Theta$*,*

$$\hat{P}_n(\varphi) \underset{n \to +\infty}{\longrightarrow} Q_\theta(\varphi) \text{ in } P_\theta \text{ probability}.$$

(b) *Let* $(\varphi_1, \ldots, \varphi_k)$ *belong to* $\mathscr{F}_\gamma$*. Let* $\theta \in \Theta$ *and assume that* $V_\theta(\varphi_i, \varphi_i) < \infty$*, for* $i = 1, \ldots, k$*. If moreover* $n\Delta_n^2 \to 0$*, then*

$$(n\Delta_n)^{1/2} \begin{pmatrix} \hat{P}_n(\varphi_1) - Q_\theta(\varphi_1) \\ \vdots \\ \hat{P}_n(\varphi_k) - Q_\theta(\varphi_k) \end{pmatrix} \underset{n \to +\infty}{\overset{\mathscr{D}}{\longrightarrow}} \mathscr{N}_k(0, V(\theta)) \text{ under } P_\theta,$$

where $V(\theta)$ *is the matrix given by* $V(\theta)_{ij} = V_\theta(\varphi_i, \varphi_j)$*,* $1 \leqslant i, j \leqslant k$.

3. Minimum-contrast estimation

3.1. The method

Hereafter, we propose explicit estimators of θ, based on the minimization of suitable functions of the observations, called contrasts. We refer to Dacunha-Castelle and Duflo (1983, Chapter 3) for a general account on this notion. Set (see (6))

$$U_n(\theta) = \frac{1}{n} \sum_{i=1}^n u(\theta, X_i), \tag{13}$$

where $u(\theta, x)$ is a real function defined on $\Theta \times \mathbb{R}$ satisfying conditions to be defined precisely below. Let us stress the fact that $U_n(\theta)$ is equal to $\hat{P}_n(\varphi)$ with $\varphi(x) = u(\theta, x)$. We study the associated minimum-contrast estimators $\hat{\theta}_n$ defined as any solution of

$$\hat{\theta}_n = \arg \inf_{\theta \in \Theta} U_n(\theta). \tag{14}$$

3.2. Asymptotic properties of minimum-contrast estimators

3.2.1. Consistency

In what follows, we denote by θ_0 the true value of the parameter, i.e. the value of the parameter which rules the observation. According to the classical scheme, the following assumptions are needed. The exponent γ appearing below is that given in (A4).

(H1) Θ is a compact subset of $\mathbb{R}^d$.
(H2) (Identifiability assumption) For all $\theta \in \Theta$, the function $\theta' \to \int_{\mathbb{R}} u(\theta', x) Q_\theta(\mathrm{d}x)$ has a strict minimum at $\theta' = \theta$.
(H3) For all $\theta \in \Theta$, $u(\theta, .) \in \mathscr{F}_\gamma$.
(H4) For all $x \in \mathbb{R}$, $u(., x)$ is a continuous function on Θ.
(H5) For $\eta \geqslant 0$, there exists a function $\beta(\eta)$ such that

$$\forall \theta, \theta' \in \Theta, \qquad \forall x \in \mathbb{R}, \qquad \sup_{\|\theta-\theta'\| \leqslant \eta} |u(\theta, x) - u(\theta', x)| \leqslant \beta(\eta)(1 + |x|^\gamma)$$

and $\beta(\eta) \to 0$ as $\eta \to 0$.

Theorem 3.1. *Assume* (A1)–(A4) *and* (H1)–(H5). *Let* $\Delta_n \to 0$, *and* $n\Delta_n \to +\infty$. *Then,* $\widehat{\theta}_n \underset{n\to+\infty}{\to} \theta_0$ *in* P_{θ_0} *probability, where* $\widehat{\theta}_n$ *is the minimum-contrast estimator defined in* (14).

Proof. The proof follows Dacunha-Castelle and Duflo (1983, Chapter 3). Recall that $U_n(\theta) = \widehat{P_n}(u(\theta, .))$. Assumption (H3) ensures that, for all $\theta \in \Theta$, $\int |u(\theta, x)| q(\theta_0, x)\, \mathrm{d}x < +\infty$ (see (10)). Applying Theorem 2.1 to the functions $u(\theta, .)$ and $u(\theta_0, .)$, we obtain

$$U_n(\theta) - U_n(\theta_0) \underset{n\to+\infty}{\longrightarrow} K(\theta_0, \theta) \qquad \text{in } P_{\theta_0} \text{ probability,}$$

where

$$K(\theta_0, \theta) = \int_{\mathbb{R}} \{u(\theta, x) - u(\theta_0, x)\} q(\theta_0, x)\, \mathrm{d}x. \tag{15}$$

The identifiability assumption (H2) implies that the function $\theta \to K(\theta_0, \theta)$ has a unique minimum at $\theta = \theta_0$. The continuity of $\theta \to K(\theta_0, \theta)$ must hold and follows from (H4) and (H5). It remains to study the continuity modulus of $U_n(\theta)$. Let $w(U_n, \eta) = \sup_{\|\theta-\theta'\| \leqslant \eta} |U_n(\theta) - U_n(\theta')|$. Using (H5), we get

$$w(U_n, \eta) \leqslant \beta(\eta) \left(1 + \frac{1}{n} \sum_{i=1}^n |X_i|^\gamma \right).$$

Now, $E_{\theta_0} |X_i|^\gamma = C_\gamma E_{\theta_0} \overline{V}_i^{\gamma/2}$. Since $\gamma/2 \geqslant 1$, the Hölder inequality implies that $\overline{V}_i^{\gamma/2} \leqslant 1/(\Delta_n) \int_{t_{i-1}}^{t_i} V_s^{\gamma/2}\, \mathrm{d}s$. So, by the strict stationarity of (V_t), we get (see (10))

$$E_{\theta_0} |X_i|^\gamma \leqslant C_\gamma m_{\theta_0}\left(\frac{\gamma}{2}\right). \tag{16}$$

Following the usual standard proof, this leads to the consistency of the minimum-contrast estimator. □

3.2.2. Asymptotic normality

The next step of the statistical analysis is to study the asymptotic distribution of $\widehat{\theta}_n - \theta_0$ suitably normalized. The additional assumptions are as follows (u'_{θ_i} denotes the partial derivative of u with respect to θ_i).

(K1) The true value θ_0 of the parameter belongs to $\mathring{\Theta}$.
(K2) For all $x \in \mathbb{R}$, $u(., x) \in C^2(\mathring{\Theta})$.
(K3) For all $\theta \in \mathring{\Theta}$, for all i, $u'_{\theta_i}(\theta, .) \in \mathscr{F}_\gamma$, and $\int u'_{\theta_i}(\theta, x) q(\theta, x)\,\mathrm{d}x = 0$.
(K4) For all $\theta \in \mathring{\Theta}$,

$$\int_{0^+} \mathrm{d}v\, s(\theta, v)\left(\int_0^v \pi(\theta, u)\,\mathrm{d}u\right)^2 < +\infty \text{ and } \int^{\infty} \mathrm{d}v\, s(\theta, v)\left(\int_v^{+\infty} u^{\gamma/2}\pi(\theta, u)\,\mathrm{d}u\right)^2 < +\infty.$$

(K5) For all $\theta \in \mathring{\Theta}$, for all i, j, $u''_{\theta_i\theta_j}(\theta, .) \in \mathscr{F}_\gamma$, the matrix $J(\theta)$ defined by

$$J(\theta)_{ij} = \int_{\mathbb{R}} u''_{\theta_i\theta_j}(\theta, x) q(\theta, x)\,\mathrm{d}x \tag{17}$$

is positive definite.
(K6) For $\eta \geqslant 0$, there exists a function $\varepsilon(\eta)$ such that, for all $i, j \in \{1, \ldots, d\}$,

$$\forall \theta, \theta' \in \mathring{\Theta}, \qquad \forall x \in \mathbb{R}, \qquad \sup_{\|\theta-\theta'\| \leqslant \eta} |u''_{\theta_i\theta_j}(\theta, x) - u''_{\theta_i\theta_j}(\theta', x)| \leqslant \varepsilon(\eta)(1 + |x|^\gamma)$$

and $\varepsilon(\eta) \to 0$ as $\eta \to 0$.

Among these assumptions, some are those usually used to prove asymptotic normality ((K1), (K2), K(6) and the second parts of (K3) and (K5)). The others are required for applications of Theorem 2.1. Now, let us introduce the matrix

$$\Sigma(\theta) = (V_\theta(u'_{\theta_i}(\theta, .), u'_{\theta_j}(\theta, .)))_{i,j=1,\ldots,d}. \tag{18}$$

It has been proved by Genon-Catalot *et al.* (1998, Section 3.4) that, under (K4), the quantity $V_\theta(\varphi, \varphi)$ is finite for all $\varphi \in \mathscr{F}_\gamma$. Therefore, under (K3), (K4), $\Sigma(\theta)$ is well defined for all θ.

Theorem 3.2. *Assume* (A1)–(A4), (H1)–(H5) *and* (K1)–(K6), *and let* $\Delta_n \to 0$ *and* $n\Delta_n \to \infty$. *If moreover* $n\Delta_n^2 \to 0$, *then*

$$(n\Delta_n)^{1/2}(\widehat{\theta}_n - \theta_0) \xrightarrow{\mathscr{D}} \mathscr{N}_d(0, J^{-1}(\theta_0)\Sigma(\theta_0)J^{-1}(\theta_0)) \text{ under } P_{\theta_0}.$$

Proof. By the consistency of $\widehat{\theta}_n$ and (K1), we have $P_{\theta_0}(\widehat{\theta}_n \in \mathring{\Theta}) \to 1$. Let $\widehat{\theta_{n,j}}$ and $\theta_{0,j}$ be the coordinates of $\widehat{\theta}_n$, θ_0, and denote by $U'_{n,i}$ (or $U''_{n,ij}$) the partial derivatives of U_n with respect to θ_i (or $\theta_i\theta_j$). For $i \in \{1, \ldots, d\}$,

$$0 = U'_{n,i}(\widehat{\theta}_n) = U'_{n,i}(\theta_0) + \sum_{j=1}^{d} (\widehat{\theta_{n,j}} - \theta_{0,j}) \int_0^1 U''_{n,ij}(\theta_0 + s(\widehat{\theta}_n - \theta_0))\,\mathrm{d}s.$$

By (K3) and Theorem 2.1,

$$U'_{n,i}(\theta_0) = \widehat{P_n}(u'_{\theta_i}(\theta_0, .)) \to \int_{\mathbb{R}} u'_{\theta_i}(\theta_0, x) q(\theta_0, x)\,\mathrm{d}x = 0 \qquad \text{as } n \to \infty \text{ in } P_{\theta_0} \text{ probability.}$$

Using (K3), (K4), Theorem 2.1(b) may be applied to the set of functions $u'_{\theta_i}(\theta_0, x)$, $i \in \{1, \ldots, d\}$. So, we get (see (18) for $\Sigma(\theta)$)

$$(n\Delta_n)^{1/2}(U'_{n,i}(\theta_0))_{i\in\{1,\ldots,d\}} \xrightarrow{\mathscr{D}} \mathscr{N}_d(0, \Sigma(\theta_0)) \text{ under } P_{\theta_0}.$$

Applying Theorem 2.1(1) to $U''_{n,ij}(\theta_0)$ gives

$$U''_{n,ij}(\theta_0) = \widehat{P_n}(u''_{\theta_i\theta_j}(\theta_0, .)) \to J(\theta_0)_{ij} \qquad \text{in } P_{\theta_0} \text{ probability}$$

(J is defined in (K5)). Finally, a bound for the remainder term is obtained using (K6), for all $i, j \in \{1, \ldots, d\}$,

$$\left| \int_0^1 U''_{n,ij}(\theta_0 + s(\widehat{\theta}_n - \theta_0))\,\mathrm{d}s - U''_{n,ij}(\theta_0) \right| \leqslant \varepsilon(\|\widehat{\theta}_n - \theta_0\|) \left(1 + \frac{1}{n} \sum_{k=1}^{n} |X_k|^\gamma \right).$$

By the consistency, $\varepsilon(\|\widehat{\theta}_n - \theta_0\|) \to 0$, in P_{θ_0} probability. Using (16), the left-hand side of the previous inequality tends to 0 under P_{θ_0}. Assumption (K5) leads to the asymptotic distribution $\mathscr{N}_d(0, J^{-1}(\theta_0)\Sigma(\theta_0)J^{-1}(\theta_0))$ for $(n\Delta_n)^{1/2}(\widehat{\theta}_n - \theta_0)$. □

Examples of contrasts can be derived from the moments of Q_θ. For instance,

$$u(\theta, x) = (x^2 - m_\theta(1))^2 \qquad \text{or} \qquad u(\theta, x) = \log v(\theta) + \frac{(x^2 - m_\theta(1))^2}{v(\theta)},$$

with $m_\theta(1) = \int x^2 Q_\theta\,\mathrm{d}x$ and $v(\theta) = \int \{x^2 - m_\theta(1)\}^2 Q_\theta(\mathrm{d}x)$. We shall not give more details of these examples, but we have rather restricted our attention to an example which takes more into account the distribution of the X_i.

3.3. A contrast fitting the statistical model well

To build an appropriate contrast, in view of Theorem 2.1, we use the idea that X_i has approximately, for small Δ, the distribution Q_θ. Thus, let us set

$$u(\theta, x) = -\log q(\theta, x), \tag{19}$$

where $q(\theta, x)$ is the density of Q_θ (see (9)). In this case, the following theorem links the conditions on $u(\theta, x)$ with the properties of the stationary distribution π_θ of diffusion (V_t).

Theorem 3.3. *Recall that η has distribution π_θ.*

(a) *If, $\forall \theta, \theta' \in \Theta$, $\theta \neq \theta' \Rightarrow \pi_\theta \neq \pi_{\theta'}$, then* (H2) *holds.*
(b) *Let $\theta \in \Theta$. If $\mathrm{E}_\theta(\eta^{-3/2}) < \infty$, then $u(\theta, .)$ belongs to $\mathscr{F}_2$.*

Proof. For (a), let us remark that (see (15))

$$K(\theta_0, \theta) = \int \log\left(\frac{q(\theta_0, x)}{q(\theta, x)}\right) q(\theta_0, x)\,\mathrm{d}x = K(Q_{\theta_0}, Q_\theta),$$

which is the Kullback information of Q_{θ_0} with respect to Q_θ. Recall that $K(Q_{\theta_0}, Q_\theta)$ is positive if and only if $Q_\theta \neq Q_{\theta_0}$. One can calculate the characteristic function of Q_θ:

$$\varphi_\theta(t) = \int_{\mathbb{R}} \mathrm{e}^{itx}\, Q_\theta(\mathrm{d}x) = \int_0^{+\infty} \mathrm{e}^{-vt^2/2}\, \pi(\theta, v)\,\mathrm{d}v.$$

Therefore, the characteristic function of Q_θ at point t is equal to the Laplace transform of π_θ at point $t^2/2$, and, for all $(\theta, \theta') \in \Theta \times \Theta$, $\pi_\theta \neq \pi_{\theta'} \Leftrightarrow Q_\theta \neq Q_{\theta'}$. This gives the result.

For (b), let us set, for $t \geqslant 0$, $g(\theta, t) = \int_0^{+\infty} \{1/(2\pi v)^{1/2}\} \exp(-t/2v)\, \pi(\theta, v)\,\mathrm{d}v$ and $\varphi(\theta, t) = -\log g(\theta, t)$. So $\varphi(\theta, x^2) = u(\theta, x)$. Using Remark 2.2, we just need to prove that $\varphi(\theta, .) \in C^1([0, +\infty))$ and that

$$\exists K_\theta > 0, \forall t \geqslant 0, \qquad |\varphi(\theta, t)| + |\varphi'_t(\theta, t)| \leqslant K_\theta(1 + t). \tag{20}$$

Note that $g(\theta, 0) < \infty$ if and only if $\mathrm{E}_\theta(\eta^{-1/2}) < \infty$. Under this condition, since $g(\theta, .)$ is a Laplace transform, we have, for all $t \geqslant 0$, $0 < g(\theta, t) < \infty$ and $g(\theta, .)$ is continuous on $[0, +\infty)$. Similarly, $g(\theta, .) \in C^1([0, +\infty))$ if and only if $\mathrm{E}_\theta(\eta^{-3/2}) < \infty$. So, $\varphi(\theta, .) \in C^1([0, +\infty))$. It remains to prove (20). Let $\alpha \in [0, 1]$. By the Hölder inequality, for $t, t' \geqslant 0$,

$$g(\theta, \alpha t + (1-\alpha)t') \leqslant \left(\int_0^{+\infty} \exp\left(-\frac{t}{2v}\right) \frac{\pi(\theta, v)}{(2\pi v)^{1/2}}\,\mathrm{d}v\right)^{\alpha} \left(\int_0^{+\infty} \exp\left(-\frac{t'}{2v}\right) \frac{\pi(\theta, v)}{(2\pi v)^{1/2}}\,\mathrm{d}v\right)^{1-\alpha}.$$

Thus, $\varphi(\theta, \alpha t + (1-\alpha)t') \geqslant \alpha\varphi(\theta, t) + (1-\alpha)\varphi(\theta, t')$. Therefore, $\varphi(\theta, .)$ is a concave function and its derivative is given by

$$\varphi'_t(\theta, t) = \tfrac{1}{2}\left(\int_0^{+\infty} \frac{1}{v} \exp\left(-\frac{t}{2v}\right) \frac{\pi(\theta, v)}{(2\pi v)^{1/2}}\,\mathrm{d}v\right) \Big/ \left(\int_0^{+\infty} \exp\left(-\frac{t}{2v}\right) \frac{\pi(\theta, v)}{(2\pi v)^{1/2}}\,\mathrm{d}v\right). \tag{21}$$

So, for all $t \geqslant 0$,

$$0 < \varphi'_t(\theta, t) \leqslant \varphi'_t(\theta, 0) = \tfrac{1}{2}\frac{\mathrm{E}_\theta(\eta^{-3/2})}{\mathrm{E}_\theta(\eta^{-1/2})}. \tag{22}$$

This clearly implies (20), and the proof is complete. □

Theorem 3.3(a) means that all the parameters present in the stationary distribution π_θ can be consistently estimated. However, some parameters of $b(\theta, .)$ and $a(\theta, .)$ may no longer be present in the expression of π_θ. They must be estimated by another method. Further work is in progress in this direction. It is indeed a drawback, already encountered in other papers (see, for example, Kessler (1996)). On the other hand, this method provides simple and tractable computations of the estimators and requires weak assumptions easy to check. The second part corresponds to the weakest version of assumption (A4). For what concerns

assumptions implying asymptotic normality, let us point out that, for this contrast, $\int u'_{\theta_i} Q_\theta(\mathrm{d}x) = 0$ is immediate and that $J(\theta)$ is now the Fisher information matrix of Q_θ.

4. Estimating functions

In situations where explicit likelihood in untractable as in the case here, another approach is to construct an estimating function $G_n(\theta)$ to be a substitute for the score function (see, for example, Barndorff-Nielsen and Sørensen (1994)). Let us consider $\psi\colon \Theta \times \mathbb{R} \to \mathbb{R}^d$, and define

$$G_n(\theta) = \sum_{i=1}^{n} \psi(\theta, X_i). \tag{23}$$

We define estimators $\widetilde{\theta}_n$ as the solutions of

$$G_n(\widetilde{\theta}_n) = 0. \tag{24}$$

Let ψ_i, $i = 1, \ldots, d$, denote the coordinates of ψ. Clearly, a class of estimating functions is obtained by setting $\psi_i(\theta, x) = u'_{\theta_i}(\theta, x)$ and $u(\theta, x)$ a function leading to a contrast $U_n(\theta)$ (see (13)). However, other estimating functions can be used. Then, assumptions concern the functions (ψ_i). For instance, another class of estimating functions is derived from power functions and moment properties of the distribution Q_θ. In particular, the moment method corresponds to the choice

$$\psi_i(x) = C_{2i} m_\theta(i) - x^{2i}, \qquad i \geqslant 1. \tag{25}$$

Clearly, this leads to estimate $m_\theta(i)$ by $\sum_{k=1}^{n} X_k^{2i}/(nC_{2i})$.

In the contrast approach, we are able to prove that all the solutions of (14) are consistent. Afterwards, asymptotic normality is proved. In the estimating functions method, one proves that a solution of (24) exists which is consistent and asymptotically normal, under a different set of assumptions given below.

($\tilde{\mathrm{K}}$1) This is the same as (K1)
($\tilde{\mathrm{K}}$2) For all $x \in \mathbb{R}$, $\psi(., x) \in C^1(\mathring{\Theta})$.
($\tilde{\mathrm{K}}$3) For all $\theta \in \mathring{\Theta}$, for all i, $\psi_i(\theta, .) \in \mathscr{F}_\gamma$ and $\int \psi_i(\theta, x) q(\theta, x)\,\mathrm{d}x = 0$.
($\tilde{\mathrm{K}}$4) This is the same as (K4).
($\tilde{\mathrm{K}}$5) For all $\theta \in \mathring{\Theta}$, for all i, j, $(\partial/\partial\theta_j)\psi_i(\theta, .) \in \mathscr{F}_\gamma$ and the $d \times d$ matrix $\tilde{J}(\theta)$ defined by

$$\tilde{J}(\theta)_{ij} = \int_{\mathbb{R}} \frac{\partial}{\partial \theta_j} \psi_i(\theta, x) q(\theta, x)\,\mathrm{d}x \tag{26}$$

is positive definite.
($\tilde{\mathrm{K}}$6) For all $\theta \in \mathring{\Theta}$, for $\eta \geqslant 0$, there exists a function $\delta_\theta(\eta)$ such that, for all $i, j \in \{1, \ldots, d\}$,

$$\forall x \in \mathbb{R}, \qquad \sup_{\theta' : \|\theta' - \theta\| \leqslant \eta} \left| \frac{\partial}{\partial \theta_j} \psi_i(\theta', x) - \frac{\partial}{\partial \theta_j} \psi_i(\theta, x) \right| \leqslant \delta_\theta(\eta)(1 + |x|^\gamma)$$

with $\delta_\theta(\eta) \to 0$ as $\eta \to 0$.

The notation is chosen on purpose to establish a correspondence between assumptions (K1–K6) of Section 3 and assumptions ($\tilde{K}1$–$\tilde{K}6$) above. Note that the matrix $\tilde{J}(\theta)$ appearing in ($\tilde{K}5$) is not necessarily symmetric. The property $\tilde{J}(\theta)$ positive definite means that the inner product $\langle \tilde{J}(\theta)x, x \rangle$ is positive for all $x \neq 0$. Using the notation introduced in (12), let us define the matrix $\tilde{\Sigma}(\theta)$ by

$$\tilde{\Sigma}(\theta)_{ij} = V_\theta(\psi_i(\theta, .), \psi_j(\theta, .)). \tag{27}$$

Theorem 4.1. *Assume* (A1)–(A4) *and* ($\tilde{K}1$)–($\tilde{K}6$)*, and let* $n \to \infty$, $\Delta_n \to 0$, $n\Delta_n \to \infty$. *If moreover* $n\Delta_n^2 \to 0$*, then an estimator* $\widetilde{\theta}_n$*, which solves* (24)*, exists with a probability tending to one as* $n \to \infty$ *under* P_{θ_0} *and is weakly consistent. Moreover,*

$$(n\Delta_n)^{1/2}(\widetilde{\theta}_n - \theta_n) \xrightarrow{\mathscr{D}} \mathscr{N}_d(0, \tilde{J}^{-1}(\theta_0)\tilde{\Sigma}(\theta_0)\tilde{J}^{-1}(\theta_0)) \text{ under } P_{\theta_0}.$$

Proof. Our assumptions imply that

(a) $G_n(.) \in C^1(\mathring{\Theta})$,
(b) under P_{θ_0}, $(\Delta_n n)^{1/2}(1/n)G_n(\theta_0) \xrightarrow{\mathscr{D}} \mathscr{N}_d(0, \tilde{\Sigma}(\theta_0))$,
(c) $(1/n)\dot{G}_n(\theta_0) \xrightarrow{P_{\theta_0}} \tilde{J}(\theta_0)$, where $\dot{G}_n(\theta)$ denotes the matrix $((\partial/\partial\theta_j)G_{n,i}(\theta))$ and
(d) $\tilde{J}(\theta_0)$ is positive definite.

In addition, a uniformity condition is required. Set $M_{c,n} = \{\theta \colon \|\theta - \theta_0\| \leqslant c(n\Delta_n)^{-1/2}\}$; we must have

$$\sup_{\theta \in M_{c,n}} \left| \frac{1}{n} \dot{G}_n(\theta) - \tilde{J}(\theta_0) \right| \xrightarrow{P_{\theta_0}} 0. \tag{28}$$

This property will be proved later. Now, we can follow step by step the proof of Theorem A1 (Appendix) of Barndorff-Nielsen and Sørensen (1994) to get the existence and consistency of $\widetilde{\theta}_n$ and also that, for all convex combination θ_n^* of $\widetilde{\theta}_n$ and θ_0, $(1/n)\dot{G}_n(\theta_n^*) \xrightarrow{P_{\theta_0}} \tilde{J}(\theta_0)$. Then, the asymptotic normality is a straightforward consequence of the standard expansion:

$$0 = \left(\frac{\Delta_n}{n}\right)^{1/2} G_n(\theta_0) + (\Delta_n n)^{1/2}(\widetilde{\theta}_n - \theta_0)\frac{1}{n}\dot{G}_n(\theta_n^*).$$

We now check condition (28), following the proof of Theorem 3.4 and Lemma 3.5 of Bibby and Sørensen (1995). Let $f(\theta) = \int \dot{\psi}(\theta, x) Q_{\theta_0}(\mathrm{d}x)$, where $\dot{\psi} = ((\partial/\partial\theta_j)\psi_i)$. We have

$$\sup_{\theta \in M_{c,n}} \left| \frac{1}{n} \dot{G}_n(\theta) - \tilde{J}(\theta_0) \right| \leqslant \sup_{\theta \in K} \left| \frac{1}{n} \dot{G}_n(\theta) - f(\theta) \right| + \sup_{\theta \in M_{c,n}} |f(\theta) - \tilde{J}(\theta_0)|,$$

where K is a compact set such that $M_{c,n} \subset K \subset \mathring{\Theta}$. From ($\tilde{K}2$), ($\tilde{K}6$) and the Lebesgue

theorem, it follows that f is continuous. Thus, the second term in the inequality above tends to 0. Let us choose a finite covering of K by balls of radius η and centres θ_j. For θ such that $\|\theta - \theta_j\| \leqslant \eta$, we have, using ($\tilde{K}6$),

$$\left|\frac{1}{n}\sum_{i=1}^{n}\dot{\psi}(\theta, X_i) - f(\theta)\right| \leqslant \delta_{\theta_j}(\eta)\frac{1}{n}\sum_{i=1}^{n}|X_i|^{\gamma} + \left|\frac{1}{n}\sum_{i=1}^{n}\dot{\psi}(\theta_j, X_i) - f(\theta_j)\right| + |f(\theta_j) - f(\theta)|.$$

The proof may be ended by an appropriate choice of η and noting that $\dot{\psi}(\theta_j, .) \in \mathscr{F}_{\gamma}$. □

5. Application to some classical models in finance

In this section, we develop the previous estimation methods for parametric models on the hidden diffusion (V_t) commonly used in finance. We study below successively the contrast based on $\log q(\theta, x)$ (see Section 3.3) and the moment method (see (25)).

Model 1. The diffusion approximation of a GARCH(1,1)-M model (Nelson 1990) is given by

$$\mathrm{d}Y_t = \sigma_t\,\mathrm{d}B_t, \qquad Y_0 = 0,$$

$$\mathrm{d}V_t = \alpha(\beta - V_t)\,\mathrm{d}t + cV_t\,\mathrm{d}W_t, \qquad V_0 = \eta, \qquad V_t = \sigma_t^2,$$

where α, β and c are real numbers. Let us set $a = 1 + 2\alpha/c^2$ and $\lambda = 2\beta\alpha/c^2$. Assumption (A2) holds if and only if $\lambda > 0$ and $a > 0$, and the stationary distribution π_θ is the inverse Gamma distribution, with density

$$\pi(\theta, v) = \frac{\lambda^a}{\Gamma(a)}v^{-a-1}\exp\left(-\frac{\lambda}{v}\right)1_{\{v>0\}}, \tag{29}$$

where $\Gamma(a)$ is the usual gamma function. The moments of π_θ are given by

$$m_\theta(p) = \mathrm{E}_\theta(\eta^p) = \lambda^p\frac{\Gamma(a-p)}{\Gamma(a)} \qquad \text{if } p < a, \tag{30}$$

and $+\infty$ if $p \geqslant a$. Hence, (A4) is satisfied with $\gamma = 2$ if and only if $a > 2$. To sum up, we get the conditions $a > 2$, $\lambda > 0$, which are equivalent to $\beta > 0$, $\alpha > c^2/2$. From the expression for π_θ and Theorem 3.3, we see that only the parameters a and λ can be identified in our asymptotic framework; so the unknown parameter to estimate is $\theta = (a, \lambda)$. This means that either α or c is known. Applying (9), we remark that the mixture distribution Q_θ is completely explicit:

$$q(\theta, x) = \int_0^{+\infty}\frac{1}{(2\pi v)^{1/2}}\exp\left(-\frac{x^2}{2v}\right)\frac{\lambda^a}{\Gamma(a)}v^{-a-1}\exp\left(-\frac{\lambda}{v}\right)\mathrm{d}v$$

$$= \frac{1}{(2\pi)^{1/2}}\frac{\Gamma(a+\frac{1}{2})}{\Gamma(a)}\frac{\lambda^a}{(\lambda + x^2/2)^a + 1/2}.$$

This is the distribution of $(\lambda/a)^{1/2}T$ when T has the Student t distribution of $2a$ degrees of freedom. Note that the Student distribution has been already proposed to take into account the observed "fat tails" of stock returns (see, for example, Blattberg and Gonedes (1974)). Hence, we consider the contrast built with

$$u(\theta, x) = \left(a + \frac{1}{2}\right) \log\left(\lambda + \frac{x^2}{2}\right) - a\log(\lambda) + \log\left(\frac{\Gamma(a)}{\Gamma(a+\frac{1}{2})}\right). \tag{31}$$

Let us assume that $\theta = (a, \lambda) \in \Theta$, where Θ is a compact subset of $(2, +\infty) \times (0, +\infty)$ and let $\widehat{\theta}_n$ be the minimum-contrast estimator associated with (31). Let us remark that

$$\mathrm{E}_\theta(\eta^{-3/2}) = \frac{a+\frac{1}{2}}{\lambda^{3/2}} \frac{\Gamma(a)}{\Gamma(a+\frac{1}{2})}.$$

Theorem 3.3 implies that (H3) is satisfied with $\gamma = 2$. Note also that $\theta \to u(\theta, x)$ is continuously differentiable (the derivatives can be explicitely computed) and satisfy (H5). Therefore, the weak consistency follows from Theorem 3.1.

For the asymptotic normality, the main difficulty is to prove that the asymptotic covariance matrix is well defined and positive definite. An elementary proof shows that assumption (K4) of Theorem 3.2 is satisfied. For (K5), we can compute explicitly $J(\theta)$:

$$J(\theta)_{11} = \Psi'(a) - \Psi'(a+\tfrac{1}{2}), \qquad J(\theta)_{12} = \frac{-1}{\lambda(2a+1)} \qquad \text{and} \qquad J(\theta)_{22} = \frac{a}{\lambda^2(2a+3)},$$

where $\Psi(z)$ is the derivative of $\log\Gamma(z)$ (the so-called digamma function). To show that the determinant is positive, we use the following equality (Luke, 1969, Chapter 2):

$$\Psi(z) = \log(z) - \frac{1}{2z} - h(z),$$

with

$$h(z) = \int_0^\infty \{(\mathrm{e}^t - 1)^{-1} - t^{-1} + \tfrac{1}{2}\}\, \mathrm{e}^{-tz}\, \mathrm{d}t.$$

Let us stress the fact that asymptotic normality holds without any other additional restriction on the parameter set.

For the moment method, we use the estimating functions based on the moments of π_θ. A natural choice to estimate (a, λ) is to consider the first two moments of π_θ, which are finite if $a > 2$, and are given by $m_\theta(1) = \lambda/(a-1)$ and $m_\theta(2) = \lambda^2/(a-1)(a-2)$. The correspondence $(a, \lambda) \to (m_\theta(1), m_\theta(2))$ is a C^1 diffeomorphism of $(2, +\infty) \times (0, +\infty)$ onto $(0, +\infty)^2$. Therefore, we consider the estimation of $m = (m_\theta(1), m_\theta(2))$, and set $\psi = (\psi_1, \psi_2)$, with $\psi_1(m, x) = m_\theta(1) - x^2$ and $\psi_2(m, x) = 3m_\theta(2) - x^4$. Looking at ψ_2, we must choose $\gamma = 4$, which implies the condition $a > 4$ to get (A4) (see (30)). Thus, compared with the contrast method, the application of the moment method adds a restriction on the parameter set. Let us remark that assumptions $(\tilde{K}1)$–$(\tilde{K}6)$ are easy to check, and that the asymptotic covariance matrix of the estimators can be explicitly computed.

Model 2. (Heston 1993). We consider for (V_t) the classical square-root process used by Cox *et al*. (1985) for interest rates. Let

$$\mathrm{d}V_t = \alpha(\beta - V_t)\,\mathrm{d}t + cV_t^{1/2}\,\mathrm{d}W_t, \qquad V_0 = \eta,$$

where α, β and c are real numbers. Let us set $a = 2\beta\alpha/c^2$ and $\lambda = 2\alpha/c^2$. If $\lambda > 0$ and $a \geqslant 1$, the stationary distribution π_θ is the Gamma distribution with density

$$\pi(\theta, v) = \frac{\lambda^a}{\Gamma(a)} v^{a-1}\,\mathrm{e}^{-\lambda v}\,1_{\{v>0\}}. \tag{32}$$

Thus, this model has similarities with the variance-gamma model of Madan and Seneta (1990). The moments of π_θ are finite for all positive p:

$$m_\theta(p) = \mathrm{E}_\theta(\eta^p) = \frac{\Gamma(a+p)}{\Gamma(a)\lambda^p}. \tag{33}$$

In particular, $E_\theta(\eta) = a/\lambda$ and $\mathrm{E}_\theta(\eta^2) = a(a+1)/\lambda^2$. Here again, the parameter to estimate is $\theta = (a, \lambda)$. The mixture distribution Q_θ can be computed and has the density

$$q(\theta, x) = \frac{(2\lambda)^{1/2}}{\Gamma(a)\pi^{1/2}} \left(\frac{|x|(2\lambda)^{1/2}}{2}\right)^{a-1/2} K_{a-1/2}(|x|(2\lambda)^{1/2}),$$

where K_ν is the modified Bessel function of the second kind with index ν (see, for example, Luke (1969)). This distribution belongs to the Barndorff-Nielsen (1978) generalized hyperbolic distributions.

For the contrast method, we must check $\mathrm{E}_\theta(\eta^{-3/2}) < \infty$. Since

$$\mathrm{E}_\theta(\eta^{-3/2}) = \lambda^{3/2}\frac{\Gamma(a-\frac{3}{2})}{\Gamma(a)} < \infty \qquad \text{if} \qquad a > \tfrac{3}{2},$$

and $+\infty$ otherwise, this leads to an additional restriction. Assumption (K4) holds for all $\gamma \geqslant 2$ by an elementary proof. As for (H5), (K3) and (K6), the technical tool is to use that, at $+\infty$, $K_\nu(t) \sim \mathrm{e}^{-t}\pi^{1/2}2t^{-1/2}$, to control the polynomial growth of the functions involved. For the moment method, the restriction $a > \frac{3}{2}$ is no longer needed. Here also, the asymptotic covariance matrix of the moments estimators can be explicitly computed.

Model 3. The diffusion approximation of the exponential ARCH model (Nelson 1990) is given by

$$\mathrm{d}Y_t = \mathrm{e}^{Z_t/2}\,\mathrm{d}B_t,$$

$$\mathrm{d}Z_t = \alpha(\beta - Z_t)\,\mathrm{d}t + \rho\,\mathrm{d}W_t.$$

Both methods can be applied. The moment estimators have been used by Chesney and Scott (1989).

6. Numerical results

The aim of this section is to investigate the qualities of the estimators in Model 1 and Model 2 for finite sample, using 150 simulated paths of (Y_t).

We consider first Model 1, with $c = 2^{1/2}$. To simulate the diffusion process V_t, we have used a Euler scheme with sampling interval $\Delta/50$ and get from this an approximation of the integral $\overline{V}_i$, for each i. Then, we have computed the moment and the minimum-contrast estimators for each path. Table 1 gives the mean and the standard deviation of the estimates of the parameter (β, α) for both methods and for different values of n and Δ.

First, let us recall that, when $\alpha = 1.5$, the moment method does not provide an asymptotically normal estimator for this parameter. So, they are not present in Table 1. Let us remark that the contrast method provides better results than the moment method does, as can be seen from the means as well as the standard deviations. This is in accordance with the fact that the contrast method takes the distribution of the observations into account more. The accuracy of the estimates of β is better than those of α. This may be explained by the properties of the Student distribution; it is well known that, even for independent observations of the Student distribution, the parameter involving the number of degrees of freedom is badly estimated (see the discussion by Blattberg and Gonedes (1974)). Finally, these numerical results confirm that n and Δ must be in accordance with the asymptotic conditions $n\Delta$ large and $n\Delta^2$ small to get good estimates.

For Model 2, we have chosen values for the parameters which enable an exact simulation of V_t using the explicit transition densities $(\beta = 2, \alpha = 1, c = 2^{1/2})$. As suggested by Table 2, the same conclusions hold, and the results are even slightly better than for Model 1.

Table 1. Simulation results for Model 1 (150 replications)

			Moment method		Contrast method	
Asymptotic	Parameter	True value	Mean	Standard deviation	Mean	Standard deviation
$n = 1500$	β	2	2.02	0.13	2.02	0.13
$\Delta = 0.1$	α	3.5	3.97	1.53	3.72	1.26
$n = 3000$	β	2	2.04	0.09	2.03	0.08
$\Delta = 0.1$	α	3.5	3.52	1.05	3.54	0.84
$n = 3000$	β	2	2.02	0.11	2.02	0.11
$\Delta = 0.05$	α	3.5	3.83	1.29	3.66	1.06
$n = 1500$	β	2	2.07	0.34	2.08	0.33
$\Delta = 0.1$	α	1.5	—	—	1.58	0.53
$n = 1500$	β	2	2.12	0.23	2.1	0.18
$\Delta = 0.3$	α	1.5	—	—	1.54	0.41

Table 2. Simulation results for Model 2 (150 replications)

			Moment method		Contrast method	
Asymptotic	Parameter	True value	Mean	Standard deviation	Mean	Standard deviation
$n = 1000$	β	2	1.95	0.19	1.95	0.19
$\Delta = 0.1$	α	1	1.33	0.56	1.16	0.28
$n = 1500$	β	2	1.99	0.17	1.99	0.16
$\Delta = 0.1$	α	1	1.16	0.37	1.10	0.26

7. Extensions and concluding remarks

Stochastic models for stock prices and the instantaneous standard deviation of the stock returns are generally given by

$$\frac{\mathrm{d}S_t}{S_t} = r\,\mathrm{d}t + \sigma_t\,\mathrm{d}B_t, \qquad V_t = \sigma_t^2,$$

$$\mathrm{d}V_t = b(\theta, V_t)\,\mathrm{d}t + a(\theta, V_t)\,\mathrm{d}W_t,$$

where only (S_t) is observed. Thus, $Y_t = \log S_t$ has a drift term depending only on the instantaneous conditional variance V_t. This leads us to consider models such that

$$\mathrm{d}Y_t = \mu(\sigma_t^2)\,\mathrm{d}t + \sigma_t\,\mathrm{d}B_t.$$

Genon-Catalot *et al*. (1998) proved that, under appropriate conditions on $\mu(.)$, the results of Theorem 2.1 are identical. Therefore, the estimation methods apply in the same way, that is to say as if μ were identically null. This is consistent with the usual results concerning the different rates of convergence for the drift and diffusion coefficients estimation, as $\Delta_n \to 0$.

To include more models of the financial area, a relevant extension would be to assume that the two Brownian motions B_t and W_t are correlated and to study the estimation of their instantaneous correlation. Another possible extension is to study models where the coefficients of (V_t) depend on the state (Y_t). More generally, multidimensional processes including observed and unobserved coordinates can be encountered and justify further investigations.

References

Barndorff-Nielsen, O.E. (1978) Hyperbolic distribution and distribution on hyperbolae. *Scand. J. Statist.*, **5**, 151–157.

Barndorff-Nielsen, O.E. and Sørensen, M. (1994) A review of some aspects of asymptotic likelihood theory for stochastic processes. *Int. Statist. Rev.*, **62**, 133–165.

Bibby, B. and Sørensen, M. (1995) Martingale estimation functions for discretely observed diffusion processes. *Bernoulli*, **1**, 17–39.

Black, F. and Scholes, M. (1973) The pricing of options and corporate liabilities. *J. Political Economy*, **81**, 637–659.

Blattberg, R. and Gonedes, N. (1974) A comparison of the stable and Student distributions as statistical models for stock prices. *J. Business*, **47**, 244–280.

Chesney, M. and Scott, L. (1989) Pricing European currency options: A comparison of the modified Black–Scholes model and a random variance model. *J. Financial Quant. Anal.*, **24**, 267–289.

Cox, J.C., Ingersoll, J.E. and Ross, S.A. (1985) A theory of term structure of interest rates. *Econometrica*, **53**, 385–407.

Dacunha-Castelle, D. and Duflo, M. (1983) *Probabilités et Statistiques*, Vol. 2, *Problèmes à Temps Mobile*. Paris: Masson.

Dacunha-Castelle, D. and Florens-Zmirou, D. (1986) Estimation of the coefficient of a diffusion from discrete observations. *Stochastics*, **19**, 263–284.

Donahl, G. (1987) On estimating the diffusion coefficient. *J. Appl. Probab.*, **24**, 105–114.

Genon-Catalot, V. and Jacod, J. (1993) On the estimation of the diffusion coefficient for multidimensional diffusion processes. *Ann. Inst. Henri Poincaré, Probab.-Statist.*, **29**, 119–151.

Genon-Catalot, V., Jeantheau, T. and Larédo, C. (1998) Limit theorems for discretely observed stochastic volatility models. *Bernoulli*, **4**, 283–303.

Ghysels, E., Harvey, A. and Renault, E. (1996) Stochastic volatility. *Handbook Statist.*, **14**, 119–192.

Heston S.L. (1993) A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Rev. Financial Stud.*, **6**, 327–343.

Hull, J. and White, A. (1987) The pricing of options on assets with stochastic volatilities. *J. Finance*, **42**, 281–300.

Kessler, M. (1996) Simple and explicit estimating functions for a discretely observed diffusion process. Prépublication du Laboratoire de Probabilités de l'Université de Paris VI.

Kessler, M. (1997) Estimation of an ergodic diffusion from discrete observations. *Scand. J. Statist.*, **24**(2), 211–229.

Kutoyants, Yu.A. (1984) *Parameter Estimation for Stochastic Processes*. Berlin: Heldermann.

Larédo, C. (1990) A sufficient condition for asymptotic sufficiency of incomplete observations of a diffusion process. *Ann. Statist.*, **18**, 1158–1171.

Luke, Y.L. (1969) *The Special Functions and Their Approximations*. Math. Sci. Engng, **53**. New York: Academic Press.

Madan, D.B. and Seneta, E. (1990) The variance gamma model for share market returns. *J. Business*, **63**, 511–524.

Nelson, D.B. (1990) ARCH models as diffusion approximations. *J. Econometrics*, **45**, 7–38.

Rogers, L.C.G. and Williams, D. (1987) *Diffusions, Markov Processes and Martingales*. Vol. 2, Ito Calculus. Wiley Ser. Probab. Math. Statist. New York: Wiley.

Received March 1997 and revised October 1997