Econometric Theory, **24**, 2008, 1174–1206. Printed in the United States of America.
doi:10.1017/S026646660808047X

NONPARAMETRIC ESTIMATION OF THE DIFFUSION COEFFICIENT OF STOCHASTIC VOLATILITY MODELS

ROBERTO RENÒ
Università di Siena

In this paper, new fully nonparametric estimators of the diffusion coefficient of continuous time models are introduced. The estimators are based on Fourier analysis of the state variable trajectory observed and on the estimation of quadratic variation between observations by means of realized volatility. The estimators proposed are shown to be consistent and asymptotically normally distributed. Moreover, the Fourier estimator can be iterated to get a fully nonparametric estimate of the diffusion coefficient in a bivariate model in which one state variable is the volatility of the other. The estimators are shown to be unbiased in small samples using Monte Carlo simulations and are used to estimate univariate and bivariate models for interest rates.

1. INTRODUCTION

The last decade has witnessed an increase in the amount of literature in the field of diffusion coefficient estimation. The main motivation behind this branch of research is that the diffusion coefficient, most commonly known as volatility, plays a fundamental role in every financial application.

In this paper, we mostly deal with univariate models of the kind

$$dX_t = \mu(X_t)\,dt + \sigma(X_t)\,dW_t, \tag{1.1}$$

where W_t is a standard real Brownian motion and the real functions $\mu(x)$ and $\sigma(x)$ are such that a single solution X_t of the stochastic differential equation (1.1) exists. Our specific problem is to estimate the diffusion term $\sigma(x)$ when we observe a discrete realization of the process X_t, namely, $n+1$ observation $\hat{X}_0, \ldots, \hat{X}_n$ at times $t_0 = 0 < t_1 < \cdots < t_n = T$ in the interval $[0, T]$.

I acknowledge Torben Andersen, Paul Malliavin, Mavira Mancino, Vanessa Mattiussi, Angelo Melino, Maurizio Pratelli, Neil Shephard, Antonio Roma, Rosario Rizza, and especially Marzia De Donno for pointing out an error in an older version, and also Emma Thorley for revising the manuscript. Two anonymous referees and the editor substantially contributed to improving the paper. All errors and omissions are my own responsibility. Address correspondence to Roberto Renò, Dipartimento di Economia Politica, Piazza S. Francesco 7, 53100 Siena, Italy; e-mail: reno@unisi.it.

 © 2008 Cambridge University Press 0266-4666/08 $15.00

https://doi.org/10.1017/S026646660808047X Published online by Cambridge University Press

We build our theory on Nadaraya–Watson estimators of the kind

$$\hat{\sigma}^2(x) = \frac{\sum_{i=1}^{n} K\left(\frac{x - \hat{X}_i}{h}\right)\tilde{\sigma}_i^2}{\sum_{i=1}^{n} K\left(\frac{x - \hat{X}_i}{h}\right)}, \tag{1.2}$$

where $\tilde{\sigma}_i$ is a consistent estimate of the volatility at time t_i and h is a bandwidth parameter (Härdle, 1990). We propose a two-step procedure. In the first step, we estimate the time series $\tilde{\sigma}_i$, which is implemented in the second step with equation (1.2).

The contribution of this paper is twofold. First, we introduce two new estimators in the class (1.2) for the univariate diffusion (1.1) and derive their asymptotic properties. The first estimator is based on results on the Fourier coefficients of volatility in Malliavin and Mancino (2002) and the second on realized volatility. The proposed estimators of the diffusion coefficient are fully nonparametric in the sense that we impose very loose restrictions on the functional form of the drift term. Moreover, they are developed under mild regularity conditions for the stochastic differential equation (1.1). We only need a solution for the process to exist, plus technical conditions on the analicity of the diffusion coefficient. To assess the asymptotic properties, we borrow from the limit theory for semimartingales and in particular from the convergence of a semimartingale to a process with independent increments.

Second, but perhaps of more importance, we propose a theory for the estimation of the diffusion coefficient of a state variable, which is the latent volatility of an observable variable X_t. The model is of the kind

$$\begin{cases} dX_t = \mu_t\, dt + \sigma_t\, dZ_t \\ d\sigma_t^2 = m(\sigma_t^2)\, dt + \Lambda(\sigma_t^2)\, dW_t, \end{cases}$$

and we show that estimation of the function $\Lambda^2(x)$ (the variance of volatility) is feasible using observations of X_t alone. We also provide a feasible theory for asymptotic distribution. To our knowledge, this is the first contribution in this sense,[1] even though, since the contribution of Hull and White (1987), stochastic volatility models have been common tools for pricing financial derivatives.

We compare the proposed estimators with those in the literature. The most popular fully nonparametric estimator of the diffusion coefficient is that proposed by Florens-Zmirou (1993). Her estimator is obtained by setting $\tilde{\sigma}_i^2 = (n/T)(\hat{X}_{i+1} - \hat{X}_i)^2$ in (1.2), thus getting

$$\hat{\sigma}_{FZ}^2(x) = \frac{n\sum_{i=1}^{n-1} K\left(\frac{x - \hat{X}_i}{h}\right)(\hat{X}_{i+1} - \hat{X}_i)^2}{T\sum_{i=0}^{n-1} K\left(\frac{x - \hat{X}_i}{h}\right)}. \tag{1.3}$$

https://doi.org/10.1017/S026646660808047X Published online by Cambridge University Press

This estimator has been used by Stanton (1997) and by Jiang and Knight (1997) and for estimating univariate models for the short-term interest rate; see also Fan and Zhang (2003) and Fan (2005). An estimator similar to that of Florens-Zmirou (1993) has been proposed by Bandi and Phillips (2003) and implemented in Bandi (2002). In the framework of (1.2), the estimator proposed in Bandi and Phillips (2003) is obtained by setting

$$\tilde{\sigma}_i^2 = \frac{n}{T}\frac{1}{m_i}\sum_{j=0}^{m_i}[\hat{X}_{t_{i,j+1}} - \hat{X}_{t_{i,j}}]^2, \tag{1.4}$$

where $t_{i,j}$ is a subset of indexes such that

$$t_{i,0} = \inf\{k \geq 0 : |\hat{X}_k - \hat{X}_i| \leq \varepsilon_s\}$$

and

$$t_{i,j+1} = \inf\{k \geq t_{i,j} + \Delta t : |\hat{X}_k - \hat{X}_i| \leq \varepsilon_s\},$$

m_i is the number of times that $|\hat{X}_k - \hat{X}_i| \leq \varepsilon_s$, ε_s is a parameter to be selected, and Δt is the time step between adjacent observations. The difference between the Florens-Zmirou and Bandi and Phillips estimators is that whereas the Florens-Zmirou estimator weights the observation X_t with the quadratic variation at time t, the Bandi and Phillips estimator weights the observation X_t with the average quadratic variation at all observations that are "close" to X_t. We refer the reader to the paper cited for details. It is important to remark that Bandi and Phillips do not require the process (1.1) to be stationary but only impose the weaker condition that it is recurrent. Full asymptotic properties are derived in Bandi and Phillips (2003) for both the drift and diffusion estimators; see also Jeffrey, Kristensen, Linton, Nguyen, and Phillips (2004) and Bandi and Moloche (2004) for multivariate extensions.

There are related studies in the literature. See, for example, the work of Jacod (1999), who generalizes the Florens-Zmirou estimator, whereas Hoffmann (1999) proposes a wavelet estimator that is consistent in $\mathcal{L}^p$. Both of these authors study convergence rate properties for their estimators, showing that the Florens-Zmirou (1993) estimator is not optimal in this sense.

The comparison between the newly proposed estimators and, essentially, estimator (1.3) is done by means of Monte Carlo simulations. Moreover, because the consistency of the estimators is of asymptotic nature, we use simulated data to assess the unbiased nature of the estimators in small samples and also to discuss their implementation. Monte Carlo simulations show that the diffusion coefficient estimator based on Fourier analysis provides almost identical results with respect to the classical estimator (1.3), whereas the estimator based on realized variance is more precise. We also show that, by iterating the Fourier estimation on the trajectory of volatility, we get an unbiased estimator of the volatility of volatility.

https://doi.org/10.1017/S026646660808047X Published online by Cambridge University Press

Finally, we use the proposed estimators to estimate univariate and bivariate models for the short-term interest rate, using the time series of the three-month Treasury-bill interest rate as a proxy for the short rate, and we compare our results with those in the literature.

This paper is structured as follows. Section 2 borrows from Malliavin and Mancino (2002) and shows how to estimate volatility from discrete observations in an interval $[0,T]$. Section 3 introduces the estimator based on high-frequency data. Asymptotic results for the univariate and bivariate models are derived in Sections 4 and 5, respectively, using the limit theory for semimartingales. In Section 6 we discuss the implementation of the proposed estimators and test their performance on simulated data. In Section 7 we implement the estimators to measure the variance of the diffusion of short interest rates and compare our results to those in the literature and in particular with the results of Jiang (1998) and Andersen, Benzoni, and Lund (2004). Section 8 concludes.

2. LOW-FREQUENCY ESTIMATION OF $\tilde{\sigma}_i^2$: THE FOURIER METHOD

We start with a result stated by Malliavin and Mancino (2002), from which we borrow heavily in what follows. Here, and throughout the paper, proofs are postponed to the Appendix.

Consider, in a filtered probability space, the process (1.1), where $\sigma(\cdot), \mu(\cdot)$ are bounded, so that a single solution of the stochastic differential equation (1.1) exists in the interval $[0, 2\pi]$ and W_t is a standard Brownian motion. Define the Fourier coefficients of dX and σ^2 as follows:

$$
\begin{aligned}
a_0(dX) &= \frac{1}{2\pi}\int_0^{2\pi} dX_t \qquad a_0(\sigma^2) = \frac{1}{2\pi}\int_0^{2\pi} \sigma^2(X_t)\,dt,\\
a_k(dX) &= \frac{1}{\pi}\int_0^{2\pi} \cos(kt)\,dX_t \qquad a_k(\sigma^2) = \frac{1}{\pi}\int_0^{2\pi} \cos(kt)\sigma^2(X_t)\,dt, \qquad \textbf{(2.1)}\\
b_k(dX) &= \frac{1}{\pi}\int_0^{2\pi} \sin(kt)\,dX_t \qquad b_k(\sigma^2) = \frac{1}{\pi}\int_0^{2\pi} \sin(kt)\sigma^2(X_t)\,dt.
\end{aligned}
$$

It is worth noting that these integrals can be defined on each trajectory for almost every trajectory in Ω; see Föllmer (1981).

Then σ_t^2 can be computed using its Fourier coefficients:

$$
\sigma^2(t) = \lim_{M\to\infty} \sum_{k=0}^{M} [a_k(\sigma^2)\cos(kt) + b_k(\sigma^2)\sin(kt)]. \qquad \textbf{(2.2)}
$$

Convergence of Fourier sums is in $\mathcal{L}^2([0,2\pi])$ norm. The following result holds.

https://doi.org/10.1017/S026646660808047X Published online by Cambridge University Press

PROPOSITION 1 (Malliavin and Mancino, 2002). *Consider a process X_t satisfying (1.1) and define the Fourier coefficients of dX and σ^2 as in (2.1). Given an integer $n_0 > 0$, we almost surely have*

$$a_0(\sigma^2) = \lim_{N\to\infty} \frac{\pi}{N+1-n_0} \sum_{k=n_0}^{N} \frac{1}{2}\,(a_k^2(dX) + b_k^2(dX)), \tag{2.3}$$

$$a_q(\sigma^2) = \lim_{N\to\infty} \frac{\pi}{N+1-n_0} \sum_{k=n_0}^{N} (a_k(dX)a_{k+q}(dX) + b_k(dX)b_{k+q}(dX)), \tag{2.4}$$

$$b_q(\sigma^2) = \lim_{N\to\infty} \frac{\pi}{N+1-n_0} \sum_{k=n_0}^{N} (a_k(dX)b_{k+q}(dX) - b_k(dX)a_{k+q}(dX)). \tag{2.5}$$

An estimate of $\tilde{\sigma}_i^2$ in (1.2) can then be achieved using a truncation of the expansion (2.2), as discussed in Section 6. Usage of the Fourier estimator is very limited in the literature, and so far it has been employed as an alternative to realized volatility to estimate integrated variance; see Barucci and Renò (2002a, 2002b), Hansen, Lunde, and Nason (2003), Nielsen and Frederiksen (2008), and Hansen and Lunde (2006).

3. HIGH-FREQUENCY ESTIMATION OF $\tilde{\sigma}_i^2$: REALIZED VOLATILITY

A different estimator of the diffusion coefficient with the same asymptotic properties can be devised as follows: in the estimator (1.3), we can replace $(n/T)(\hat{X}_{i+1} - \hat{X}_i)^2$ with the integrated volatility in the interval $[t_i, t_{i+1}]$; that is, in (1.2) we set $\tilde{\sigma}_i^2 = (n/T)\int_{t_i}^{t_{i+1}} \sigma^2(t)\,dt$. Such an estimator would be infeasible, but we can substitute the integrated variance with a consistent estimate of it, namely, realized volatility; see Barndorff-Nielsen and Shephard (2002), Meddahi (2002), and Andersen, Bollerslev, Diebold, and Labys (2003), among others. The Florens-Zmirou estimator is a special case in this framework, because $(X_{i+1} - X_i)^2$ is itself an estimator of the integrated variance in the interval $[iT/n, (i+1)T/n]$. Andersen, Bollerslev, and Diebold (2003) offer a review of estimation techniques for the integrated variance.[2]

Thus, if we have, for example, daily observations of the state variable, we could estimate the quadratic variation using intraday data. The advantage of estimating quadratic variation with intraday data is obviously the gain in precision, as we show in Section 6 on simulated time series. On the other hand, using intraday data to implement directly the estimator proposed by Florens-Zmirou (1993) could be misleading, because intraday data display pronounced seasonalities and microstructure effects that could seriously distort the estimate. On the other hand, these peculiarities of intraday data do not affect the estimation of integrated variance. For example, Barndorff-Nielsen and Shephard (2002) show that intraday deterministic components do not affect daily realized volatility estimates; microstructure effects can be filtered out, as in

https://doi.org/10.1017/S026646660808047X Published online by Cambridge University Press

Aït-Sahalia, Mykland, and Zhang (2005) and Bandi and Russell (2006) or simply eliminated using a cutoff frequency, as in Barucci and Renò (2002a).

4. ASYMPTOTIC RESULTS: THE UNIVARIATE CASE

4.1. Diffusion Coefficient Estimation

We consider the stochastic differential equation (SDE)

$$\begin{cases} dX_t = \mu(X_t)\,dt + \sigma(X_t)\,dW_t \\ X_0 = x_0 \end{cases} \tag{4.1}$$

defined over the time interval $[0,T]$, in the filtered probability space $(\Omega,(\mathcal{F}_t)_{0\le t\le T},P)$ satisfying the usual conditions; W_t is a standard Brownian motion. We make the following assumption.

Assumption 1. Given the SDE (4.1), we have that

(a) $x_0 \in \mathcal{L}^2(\Omega)$ is independent of W_t, $t \in [0,T]$ and measurable with respect to $\mathcal{F}_0$;
(b) $\mu(x)$ and $\sigma(x)$ are defined on a compact interval I, and $\mu(x)$ is once continuously differentiable, whereas $\sigma(x)$ is thrice continuously differentiable;
(c) a constant K exists such that $0 < \sigma(x) \le K$ and $|\mu(x)| \le K$;
(d) (Feller condition for nonexplosion) given

$$S(\alpha) = \int_0^\alpha e^{\int_0^y -(2\mu(x)/\sigma^2(x))dx}\,dy, \tag{4.2}$$

$$V(\alpha) = \int_0^\alpha S'(y) \int_0^y \frac{2}{S'(x)\sigma^2(x)}\,dx dy, \tag{4.3}$$

then $V(\alpha)$ diverges at the boundaries of I.

Assumption 1 insures the existence and uniqueness of a strong solution. Requiring the Feller condition allows us to deal with models that do not satisfy global Lipschitz and growth conditions (e.g., CIR, as noted by Aït-Sahalia, 1996). Moreover, the Feller condition is both necessary and sufficient for recurrence in I; see the discussion in Bandi and Phillips (2003). Alternatively, one can demand global Lipschitz and growth conditions for μ and σ (Karatzas and Shreve, 1988).

Requiring bounded volatility (and drift) is harmless from an econometric point of view, because we always observe a finite, and therefore bounded, set of observations. For example, in the CIR model the variance is unbounded and proportional to $\sqrt{X_t}$. However, because estimation is on a finite sample, the observations $\hat{X}_i$ are bounded, and $\sqrt{X_t}$ is indistinguishable from

https://doi.org/10.1017/S026646660808047X Published online by Cambridge University Press

$\min(\sqrt{X_t}, \sqrt{\max \hat{X}_i})$. We consider a kernel function for nonparametric estimation with the following properties.

Assumption 2. We define a kernel $K(\cdot)$ to be a bounded function in $\mathcal{L}^2(\mathbb{R})$ that is continuously differentiable, with a positive bounded first derivative, with $\int_{\mathbb{R}} K(s)\, ds = 1$, and such that $\lim_{s \to \pm\infty} K(s) = 0$ faster than any inverse polynomial.

The Gaussian density function is a possible kernel:

$$K(s) = \frac{1}{\sqrt{2\pi}}\, e^{-(s^2/2)}. \tag{4.4}$$

We also define a sequence of bandwidths h_n such that we can make the following assumption.

Assumption 3. $(h_n)_{n \in \mathbb{N}}$ is a real sequence such that, as $n \to \infty$, we have $h_n \to 0$ and $nh_n \to \infty$.

An example that is very popular in applications (Scott, 1992) is as follows:

$$h_n = h_s \hat{\sigma} n^{-1/5}, \tag{4.5}$$

where h_s is a real constant to be tuned and $\hat{\sigma}$ is the sample standard deviation.

Consider the solution process X_t with $t \in [0, T]$. We now consider the fact that the process X_t is usually recorded in the form of $n + 1$ equally spaced observations. When subdividing the interval in $n + 1$ steps of equal length, we use $X_i = X_{iT/n}$ as a shorthand notation, so that $X_0 = x_0$, $X_n = X_T$. Moreover, we set $t_i = iT/n$. Assumption 1 implies that X_t is a continuous semimartingale; thus we can define its (chronological) local time (Revuz and Yor, 1998) as

$$L_t(x) = \lim_{\varepsilon \to 0} \frac{1}{2\varepsilon} \int_0^t I_{]x-\varepsilon, x+\varepsilon[}(X_\tau)\, d\tau. \tag{4.6}$$

We can estimate the local time of a diffusion via the following approximation:

$$L_t^n(x) = \frac{T}{nh_n} \sum_{i=1}^{[nt/T]} K\left(\frac{X_i - x}{h_n}\right), \tag{4.7}$$

where $[x]$ is the integer part of x. We have indeed the following result.

PROPOSITION 2 (Florens-Zmirou, 1993). *If as $n \to \infty$ we have $nh_n^4 \to 0$, then $L_t^n(x) \to L_t(x)$ in the $\mathcal{L}^2$ sense. The convergence is almost sure if $(\log n / nh_n^2) \to 0$.*

https://doi.org/10.1017/S026646660808047X Published online by Cambridge University Press

The diffusion coefficient estimator proposed in Florens-Zmirou (1993), Stanton (1997), and Jiang and Knight (1997) is based on the following quantity:

$$V^n(x) = \frac{1}{h_n}\sum_{i=0}^{n-1} K\left(\frac{X_i - x}{h_n}\right)(X_{i+1} - X_i)^2. \tag{4.8}$$

We have indeed the following result.

PROPOSITION 3 (Florens-Zmirou, 1993). *If* $nh_n^4 \to 0$ *as* $n \to \infty$, *then* $V^n(x)$ *converges to* $\sigma^2(x)L_T(x)$ *in the* $\mathcal{L}^2$ *sense.*

Thus, dividing $V^n(x)$ by $L_T^n(x)$, we get a consistent estimator of $\sigma^2(x)$.

Now we introduce the new estimators. When dealing with Fourier estimation, we need an additional sequence.

Assumption 4. $(M_n)_{n\in\mathbb{N}}$ is a real sequence such that, as $n \to \infty$, we have $M_n \to +\infty$.

The cutoff frequency M_n is the highest coefficient used in the volatility reconstruction. By $\sigma^2_{M_n}(t)$ we denote the partial Fourier sum up to frequency M_n:

$$\sigma^2_{M_n}(t) := \sum_{k=0}^{M_n} [a_k(\sigma^2)\cos(kt) + b_k(\sigma^2)\sin(kt)]. \tag{4.9}$$

The two estimators are then defined as

$$S^n_{LF}(x) = \frac{\displaystyle\sum_{i=1}^{n} K\left(\frac{X_i - x}{h_n}\right)\sigma^2_{M_n}(t_i)}{\displaystyle\sum_{i=1}^{n} K\left(\frac{X_i - x}{h_n}\right)}, \tag{4.10}$$

$$S^n_{HF}(x) = \frac{n\displaystyle\sum_{i=1}^{n} K\left(\frac{X_i - x}{h_n}\right)\int_{iT/n}^{(i+1)T/n} \sigma^2(X_s)\,ds}{T\displaystyle\sum_{i=1}^{n} K\left(\frac{X_i - x}{h_n}\right)}. \tag{4.11}$$

Our first result states the consistency of both estimators in $\mathcal{L}^2$.

THEOREM 1. *If* $nh_n^4 \to 0$ *as* $n \to \infty$, *then* $S^n_{LF}(x)$ *and* $S^n_{HF}(x)$ *are consistent estimators of* $\sigma^2(x)$ *in the* $\mathcal{L}^2$ *sense.*

We finally state the results on the asymptotic properties.

https://doi.org/10.1017/S026646660808047X Published online by Cambridge University Press

THEOREM 2. *Let* $n \to \infty$.

(i) If $nh_n^3 \to 0$ *and* $M_n^{-1}\sqrt{nh_n} \to 0$, *then*

$$\sqrt{nh_n}\left(\frac{S_{LF}^n(x)}{\sigma^2(x)} - 1\right) \xrightarrow{\mathcal{L}} \frac{1}{\sqrt{L_T(x)}}\,\mathcal{N}(0,1). \tag{4.12}$$

(ii) If $nh_n^3 \to 0$, *then*

$$\sqrt{nh_n}\left(\frac{S_{HF}^n(x)}{\sigma^2(x)} - 1\right) \xrightarrow{\mathcal{L}} \frac{1}{\sqrt{L_T(x)}}\,\mathcal{N}(0,1). \tag{4.13}$$

The preceding convergence is in law, and $\mathcal{N}(0,1)$ *is a standard normal variable.*

The preceding results are for equally spaced observations. They can be generalized to the case in which the observations X_i are not equally spaced, but if they are such that, as $n \to \infty$, we have $\sup|t_i - t_{i-1}| \to 0$.

5. ASYMPTOTIC RESULTS: THE BIVARIATE CASE

Suppose now that X_t follows the diffusion process

$$\begin{cases} dX_t = \tilde{\mu}_t\,dt + \sigma_t\,dZ_t \\ d\sigma_t^2 = m(\sigma_t^2)\,dt + \Lambda(\sigma_t^2)\,dW_t, \\ X_0 = \tilde{x}_0, \sigma_0^2 = s_0^2 \end{cases} \tag{5.1}$$

where W_t, Z_t are independent Brownian motions.

In what follows, we substitute Assumption 1 with the following assumption.

Assumption 5. Given the SDE (5.1), we have

(a) $s_0^2 \in \mathcal{L}^2(\Omega)$, $m(\sigma^2)$, $\Lambda(\sigma^2)$ satisfy Assumption 1 (in place of x_0, $\mu(x)$, $\sigma(x)$, respectively).
(b) $\tilde{\mu}_t$ is a bounded adapted stochastic process; $\tilde{x}_0 \in \mathcal{L}^2(\Omega), \tilde{\mu}_t, m(\sigma^2), \Lambda(\sigma^2)$ are such that a unique strong solution of equation (5.1) exists.

The first part of Assumption 5 provides the existence of a strong solution for the latent process σ_t^2; the second part provides the existence of a strong solution for the state variable X_t. We iterate Fourier estimation, that is, we estimate the trajectory of Λ_t^2 with the technique described in the preceding sections, via

$$\Lambda_M^2(X_t) := \sum_{k=0}^{M} [a_k(\Lambda^2)\cos(kt) + b_k(\Lambda^2)\sin(kt)]. \tag{5.2}$$

https://doi.org/10.1017/S026646660808047X Published online by Cambridge University Press

We introduce two additional sequences: the first is needed to truncate the expansion of Λ_t^2, and the second is the bandwidth parameter for nonparametric estimation of $\Lambda^2(\sigma^2)$.

Assumption 6. $(\widetilde{M}_n)_{n\in\mathbb{N}}$ is a real sequence such that, as $n \to \infty$, we have $\widetilde{M}_n \to +\infty$. Also, $(\tilde{h}_n)_{n\in\mathbb{N}}$ is a real sequence such that, as $n \to \infty$, we have $\tilde{h}_n \to 0$ and $n\tilde{h}_n \to \infty$.

First of all, it is possible to estimate the local time of σ_t^2 using the observations of X_t. Using a slight abuse of notation, we denote the local time of σ_t^2 in (5.1) by $L_t(\sigma^2)$.[3] The estimator is

$$\tilde{L}_t^n(\sigma^2) = \frac{T}{n\tilde{h}_n} \sum_{i=1}^{[nt/T]} K\left(\frac{\sigma_{M_n}^2(t_i) - \sigma^2}{\tilde{h}_n}\right). \tag{5.3}$$

THEOREM 3. *Assume that, as $n \to \infty$, $M_n\tilde{h}_n^2 \to \infty$. If, as $n \to \infty$, we have $n\tilde{h}_n^4 \to 0$, then $\tilde{L}_t^n(\sigma^2) \to L_t(\sigma^2)$ in the $\mathcal{L}^2$ sense. The convergence is almost sure if $(\log n)/n\tilde{h}_n^2 \to 0$.*

The natural estimator for $\Lambda^2(\sigma^2)$ is then

$$\tilde{\Lambda}_n^2(\sigma^2) = \frac{\sum_{i=1}^n K\left(\frac{\sigma_{M_n}^2(t_i) - \sigma^2}{\tilde{h}_n}\right)\Lambda_{\widetilde{M}_n}^2(t_i)}{\sum_{i=1}^n K\left(\frac{\sigma_{M_n}^2(t_i) - \sigma^2}{\tilde{h}_n}\right)}. \tag{5.4}$$

THEOREM 4. *Assume that, as $n \to \infty$, $M_n\tilde{h}_n^2 \to \infty$. If $n\tilde{h}_n^4 \to 0$ as $n \to \infty$, then $\tilde{\Lambda}_n^2(\sigma^2)$ is a consistent estimator of $\Lambda^2(\sigma^2)$ in the $\mathcal{L}^2$ sense. If, additionally, $n\tilde{h}_n^3 \to 0$ and $\widetilde{M}_n^{-1}\sqrt{n\tilde{h}_n} \to 0$, then*

$$\sqrt{n\tilde{h}_n}\left(\frac{\tilde{\Lambda}_n^2(\sigma^2)}{\Lambda^2(\sigma^2)} - 1\right) \xrightarrow{\mathcal{L}} \frac{1}{\sqrt{\tilde{L}_T(\sigma^2)}}\,\mathcal{N}(0,1). \tag{5.5}$$

The preceding convergence is in law, and $\mathcal{N}(0,1)$ is a standard normal variable.

Note that the estimator (5.4) and the asymptotic distribution (5.5) depend on three sequences: $M_n, \tilde{h}_n, \widetilde{M}_n$. It is important to remark that the estimator of the local time (5.3) and of $\Lambda^2(\sigma^2)$, together with the estimation of asymptotic confidence bands (5.5), are all feasible starting from the observations $\hat{X}_0, \ldots, \hat{X}_n$ of the state variable.

Further computational time saving can be achieved when computing the coefficients $a_k(\Lambda^2), b_k(\Lambda^2)$ because, as a result of basic properties of the trigonometric system and stochastic integration by parts (Karatzas and Shreve, 1988), we have

$$a_k(d\sigma^2) = kb_k(\sigma^2), \qquad b_k(d\sigma^2) = -ka_k(\sigma^2). \tag{5.6}$$

https://doi.org/10.1017/S026646660808047X Published online by Cambridge University Press

Although it is possible, in principle, to devise an estimator for $\Lambda^2(\sigma^2)$ that is of the form (1.3) with $\hat{X}_i = \sigma^2_{M_n}(t_i)$, we experienced many difficulties in implementing such an estimator on simulated data, the main problem being the strong autocorrelation in $\sigma^2_{M_n}(t_i)$ due to the smoothing of the Fourier procedure. This is the reason why we limit ourselves to the study of the estimator (5.4).

6. IMPLEMENTATION OF THE METHOD

In this section, we discuss the implementation of the proposed estimators on simulated data. We start with the Fourier method. In finite samples, two main sources of error arise: the fact that we compute the Fourier coefficients approximately and the fact that we cannot achieve the limits $N, M \to \infty$ in (2.2) and (2.3). To compute the Fourier coefficients of X_t we proceed as in Barucci and Renò (2002a). Because we observe the process X_t only at discrete times $t_0, \ldots, t_n$, we set $X_t = \hat{X}_i$ in the interval $t_i \leq t < t_{i+1}$.[4] Then the Fourier coefficients of dX can be computed as

$$a_k(dX) = \frac{1}{\pi}\int_0^{2\pi} \cos(kt)\, dX_t = \frac{X_{2\pi} - X_0}{\pi} + \frac{k}{\pi}\int_0^{2\pi} \sin(kt) X_t\, dt, \tag{6.1}$$

which can be computed using

$$\frac{k}{\pi}\int_{t_i}^{t_{i+1}} \sin(kt) X_t\, dt = X_{t_i}\frac{k}{\pi}\int_{t_i}^{t_{i+1}} \sin(kt)\, dt = X_{t_i}\frac{1}{\pi}\left[\cos(kt_i) - \cos(kt_{i+1})\right]. \tag{6.2}$$

Before computing (6.1), we add a linear trend such that we get $X_{2\pi} = X_0$, which does not affect the volatility estimate. Then we stop the expansions (2.3)–(2.5) at a properly selected frequency N to get the coefficients of σ^2. For equally spaced data, the maximum N that prevents aliasing effects is $N = n/2$; see Priestley (1979). We use $n_0 = 1$. When actually implementing the estimator, we use the technique suggested in Malliavin and Thalmaier (2005) to achieve positive volatility even if we use a finite number of coefficients.

Precision in measuring the volatility coefficients of order q deteriorates as q increases, because computation of $a_q(\sigma^2), b_q(\sigma^2)$ is done with $N - q$ terms. Thus, instead of using (4.9) we filter the larger frequencies using the following approximation of the volatility:

$$\sigma_t^2 = \sum_{k=0}^{M} \varphi_F(\delta_1 k)\left[a_k(\sigma^2)\cos(kt) + b_k(\sigma^2)\sin(kt)\right], \tag{6.3}$$

where φ_F is the Fejer kernel (Priestley, 1979; Malliavin and Thalmaier, 2005), defined as

$$\varphi_F(x) = \frac{\sin^2(x)}{x^2}, \tag{6.4}$$

https://doi.org/10.1017/S026646660808047X Published online by Cambridge University Press

and $\varphi_F(0) = 1$. The parameter δ_1 is an additional smoothing parameter. The larger δ_1 is, the smoother is the volatility function that is estimated, but less detailed. Instead of choosing M and δ_1 separately in (6.3), for volatility reconstruction it is convenient to use the maximum M that can be computed, that is, $M = N - 1$, and then tune δ_1 to filter high frequencies. Thus the procedure is a double-smoothing estimation, in which the parameter δ_1 controls the smoothness of the volatility function and the parameter h_n controls the smoothness of the nonparametric estimate. Because the Fourier algorithm performs badly at boundaries, we skip the first and last 1% of observations.

The bandwidth parameter used, h_n, is of the form (4.5). The choice of the bandwidth parameter h_s has been debated at length in the literature. As using $h_s = 1.06$ is optimal for density estimation with respect to the mean integrated standard error criterion (Silverman, 1986; Scott, 1992; Wand and Jones, 1995), we use this value for simulations because we find that larger bandwidths lead to biased estimates.

Parametric models are simulated to evaluate the small-sample performance of the proposed estimators. We concentrate on models for the short interest rate and consider the parametric model of Chan, Karolyi, Longstaff, and Sanders (1992), which nests many popular one-factor models, including the Cox, Ingersoll, and Ross (1985) and Vasicek (1977) models:

$$dr_t = \mu(\alpha - r_t)\,dt + \sigma r_t^{\gamma}\,dW_t. \tag{6.5}$$

This model has an affine drift, and the variance is stochastic because it depends on the state variable itself, with the parameter γ measuring the degree of dependence of the variance on the interest rate level. This model has been estimated by Jiang (1998) (among others) using indirect inference. Parameter estimates are $\hat{\alpha} = 0.079(0.044)$, $\hat{\mu} = 0.093(0.100)$, $\hat{\gamma} = 1.474(0.008)$, $\hat{\sigma} = 0.794(0.019)$, where standard errors are in parentheses. We simulate the model (6.5) with these parameters, adopting a second-order Euler discretization scheme to get daily values of the short rate r_t; that is, the time step between adjacent observations is set to $\frac{1}{252}$.

Figure 1 shows one example of the simulated trajectory of 5,000 observations of r_t and $\sigma_t^2 = \hat{\sigma}^2 r_t^{2\hat{\gamma}}$, together with their spectral reconstruction. The reconstruction of the path of r_t is almost perfect. The smoothing parameter δ_1 for Fourier estimation is set to $\frac{1}{50}$.[5] The most intriguing feature of the Fourier method is that it is able to provide a consistent estimate of the trajectory of the diffusion coefficient, as shown by the figure.

The estimator (4.11) is implemented as follows. The distance between two adjacent observations is $\frac{1}{252}$. For each observation at time $t_i = i$, we divide this interval into m steps, and we estimate the integrated volatility using realized volatility as follows:

$$\hat{\sigma}^2_{i,HF} = \frac{n}{T} \sum_{j=0}^{m-1} (r_{i+((j+1)/m)} - r_{i+(j/m)})^2. \tag{6.6}$$

https://doi.org/10.1017/S026646660808047X Published online by Cambridge University Press

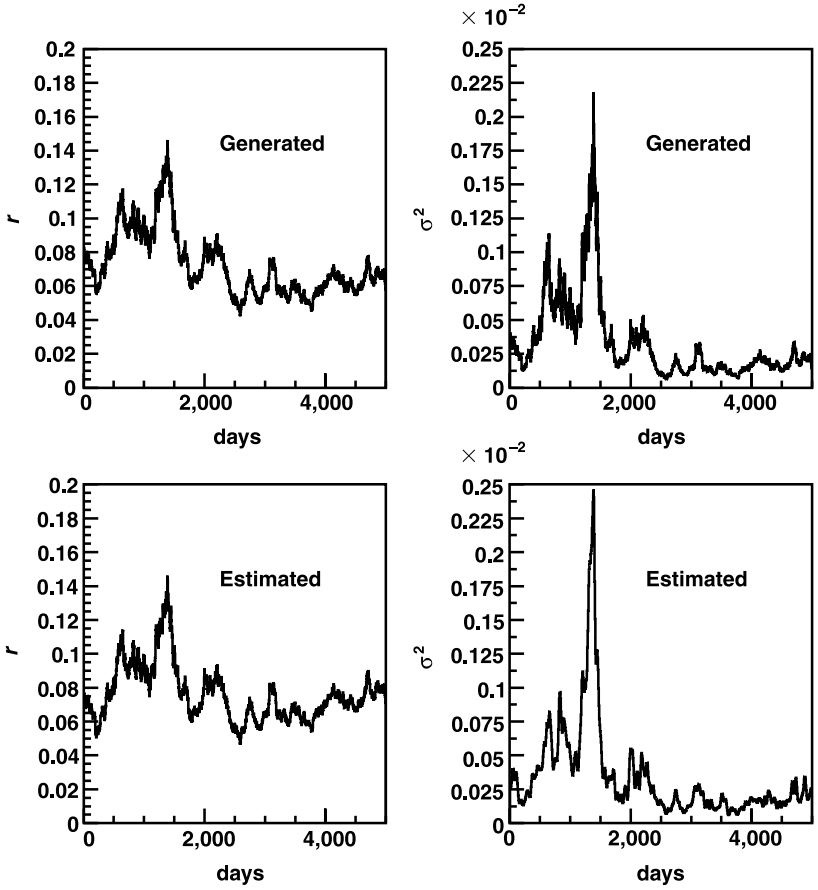


FIGURE 1. Top panel: one simulated trajectory of r_t (left) and the squared diffusion coefficient $\sigma^2 r_t^{2\gamma}$ (right) as generated by model (6.5). Bottom panel: the trajectory of r_t (left) and the squared diffusion coefficient (right) as estimated with the Fourier method. The figures in the top and bottom panels are on the same scale.

We then simulate 1,000 trajectories, and we compute the Fourier estimator and the estimator (4.11) on these simulations. For the latter we compute (6.6) with $m = 200$.

Last, we compare the newly proposed estimators with the Florens-Zmirou (1993) estimator. We do not compare them to the estimator proposed by Bandi and Phillips (2003), because Monte Carlo simulations show that it is almost indistinguishable from the Florens-Zmirou (1993) estimator in estimating the diffusion coefficient; see Renò, Roma, and Schaefer (2006). Thus the Florens-Zmirou (1993) estimator will be our benchmark throughout the paper. It is important to remark that the Florens-Zmirou and Fourier estimator use low-frequency

https://doi.org/10.1017/S026646660808047X Published online by Cambridge University Press

data, whereas the realized variance estimator uses the full sample of high-frequency data.

Results are shown in Figure 2. We find that there is no substantial difference between the Fourier estimator and the Florens-Zmirou estimator. All three estimators implemented are unbiased, but the high-frequency estimator is much more precise as shown by the narrower confidence bands. Further Monte Carlo simulations show that the precision of the high-frequency estimator increases with m, as expected. However, the disadvantage of the high-frequency estimator is that it requires much more data. Moreover, microstructure effects are absent in our simulated data but may be relevant in practice.

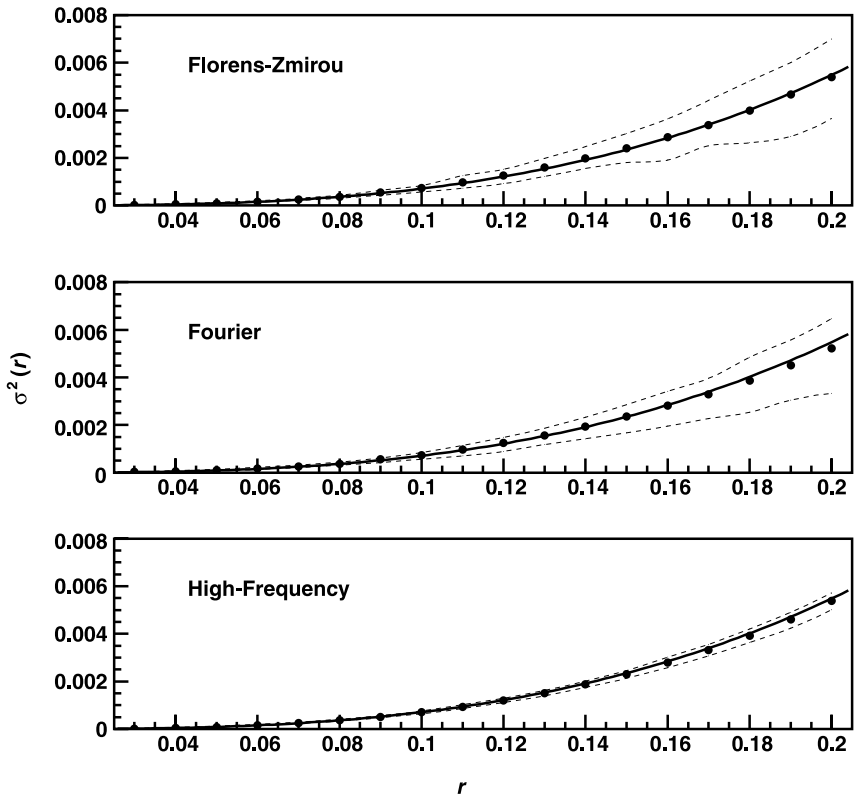


FIGURE 2. The generated Monte Carlo squared diffusion coefficient according to the CEV model (solid thick line) together with the average estimate (black dots), using 1,000 replications, with three methods: the classical estimator (1.3), top; the Fourier estimator (4.10), center; and the high-frequency estimator (4.11), bottom. The dashed lines are 5% and 95% confidence intervals, as computed with the simulations. All estimators are unbiased, but the high-frequency estimator is more precise.

https://doi.org/10.1017/S026646660808047X Published online by Cambridge University Press

Finally, the estimation methodology is implemented on the following bivariate stochastic volatility model:

$$\begin{aligned} dr_t &= k_1(\mu - r_t)\,dt + \sqrt{V_t}\,dW_{1,t}, \\ dV_t &= k_2(\alpha - V_t)\,dt + \eta_1\sqrt{V_t}\,dW_{2,t}, \end{aligned} \tag{6.7}$$

which is of the form (5.1) with $\Lambda^2(\sigma^2) = \eta_1^2\sigma^2$. This model has been estimated by Andersen et al. (2004) on interest rate data, and we use their estimated parameters, that is, $\hat{\mu} = 0.0512$, $\hat{k}_1 = 0.2593$, $\hat{\alpha} = 0.000054$, $\hat{k}_2 = 6.4995$, $\hat{\eta}_1 = 0.0195$. In the computation of Λ_t, the Fourier coefficients of $d\sigma^2$ are needed and also two additional bandwidths; see Theorem 4. Similarly to (6.3), we compute

$$\Lambda_t^2 = \sum_{k=0}^{\widetilde{M}} \varphi_F(\delta_2 k)[a_k(\Lambda^2)\cos(kt) + b_k(\Lambda^2)\sin(kt)]. \tag{6.8}$$

The coefficients of $d\sigma^2$ are measured with much larger error than that on the coefficients of dr, which is almost negligible. Thus, even if we use all the coefficients of dr up to the Nyquist frequency[6] $M/2$ to compute the coefficients of σ^2, it is advisable to use fewer coefficients of $d\sigma^2$ to compute the coefficients of Λ^2. Coherently with the choice of δ_1, we use $\widetilde{M} = 1/\delta_1$. When reconstructing Λ_t^2, we skip the first and last 5% of observations. Both the bandwidths $h_n, \tilde{h}_n$ are of the form (4.5) with $h_s = 1.06$.

Figure 3 shows one simulated trajectory of V_t and $\Lambda_t^2 = \hat{\eta}_1^2 V_t$ of length 5,000, together with the trajectory estimated with the Fourier method using the simulated observations of r_t alone. We use $\delta_1 = \frac{1}{50}$, $\widetilde{M} = 50$, $\delta_2 = \frac{1}{25}$. As for the constant elasticity of variance (CEV) model (6.5), the volatility is well estimated for the stochastic volatility model (6.7). The reconstruction of the volatility of the volatility trajectory is less precise but still quite reasonable.

We then simulate 1,000 trajectories of length 5,000 of the model (6.7), and we compute the Fourier estimator of $\Lambda^2(\sigma^2)$ on these simulations. Figure 4 shows the result. The estimator is slightly upward-biased for the lowest values of volatility, but for most values of volatility the generated Λ^2 is well inside the estimated confidence intervals.

7. APPLICATION TO INTEREST RATE MODELING

In this section, we turn to the analysis of actual interest rate data. Our aim is to estimate the diffusion coefficient $\sigma(\cdot)$ of the univariate model

$$dr_t = \mu(r_t)\,dt + \sigma(r_t)\,dW_t \tag{7.1}$$

https://doi.org/10.1017/S026646660808047X Published online by Cambridge University Press

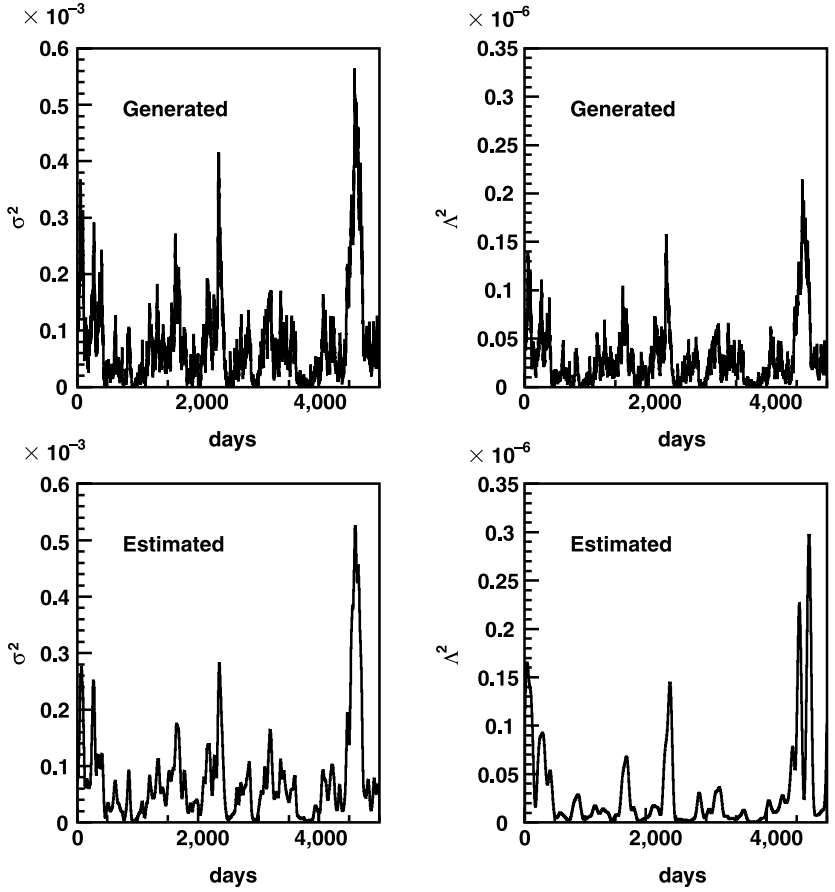


FIGURE 3. Top panel: one simulated trajectory of $\sigma_t^2 = V_t$ (left) and the squared volatility diffusion coefficient $\Lambda_t^2 = \eta_1^2 V_t$ (right) as generated by model (6.7). Bottom panel: the trajectory of the volatility (left) and of the squared volatility diffusion coefficient (right) as estimated with the Fourier method. The figures in the top and bottom panel are on the same scale.

and of $\Lambda(\cdot)$ in the stochastic volatility model

$$\begin{aligned} dr_t &= \mu_t\, dt + \sigma_t\, dZ_t, \\ d\sigma_t^2 &= m(\sigma_t^2)\, dt + \Lambda(\sigma_t^2)\, dW_t, \end{aligned} \tag{7.2}$$

while discretely observing the short rate r_t in a time interval $[0, T]$. Because the spot rate is inherently unobservable, we use proxies such as the three-month rate for it, as is common in the literature: see Duffee (1996) and Chapman,

https://doi.org/10.1017/S026646660808047X Published online by Cambridge University Press

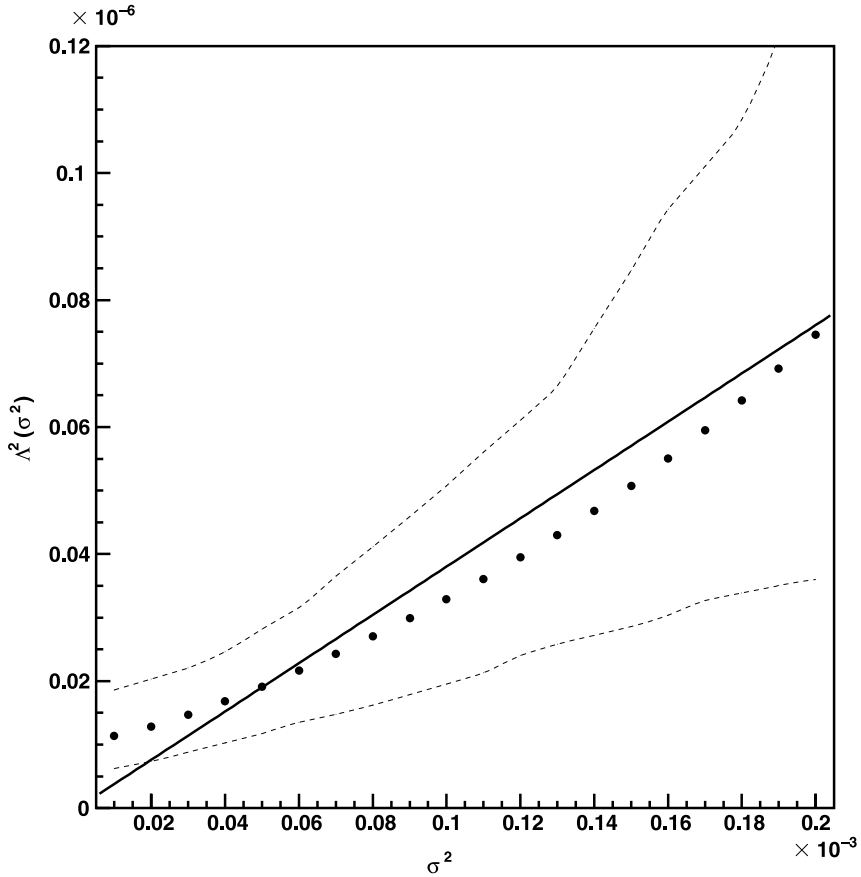


FIGURE 4. Shows the generated Monte Carlo volatility squared diffusion coefficient according to model (6.7) (solid line) together with its average estimate (black dots), using 1,000 replications. The dashed lines are 5% and 95% confidence intervals, as computed with the simulations.

Long, and Pearson (1999). Alternatively, one can regard models (7.1) and (7.2) as models for the three-month rate itself.

7.1. The Univariate Model: Fourier Estimation

We start by implementing the Fourier estimator on the same data set used in Jiang (1998), that is, the daily time series of the annualized yields of the three-month U.S. Treasury bill, from January 1962 to January 1996, for a total of 8,503 observations. The minimum and maximum of the yield over this time

https://doi.org/10.1017/S026646660808047X Published online by Cambridge University Press

span are 2.61% and 17.14%; thus estimates of the diffusion coefficient extrapolated outside this interval are an artifact of the nonparametric estimation procedure. Figure 5 shows the estimation results. The Fourier estimator is implemented with the $\delta_1 = \frac{1}{50}$ and $h_s = 4$. We use a larger value for h_s than that used for simulations, because this is typical in estimating continuous time models for the short interest rate. For example, Stanton (1997) uses $h_s = 4$, whereas Aït-Sahalia (1996) uses $h_s = 5$.

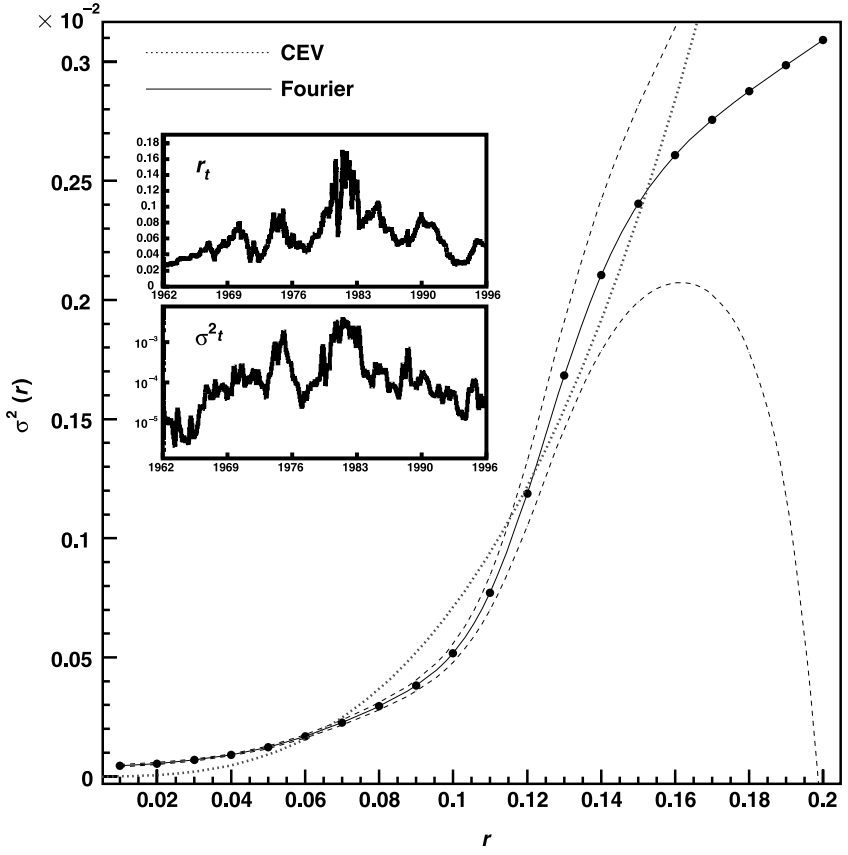


FIGURE 5. Estimate of the diffusion coefficient $\sigma^2(r)$ on Jiang (1998) data set. Solid line: Fourier estimate. Dashed lines: 5% and 95% confidence intervals, computed using (4.12). Dotted line: estimate obtained with the parametric model (6.5). Inset: the time series of the yields on three-month Treasury bill under study, from January 1966 to January 1996 (top), and the reconstructed time series of the variance according to the Fourier method (bottom).

https://doi.org/10.1017/S026646660808047X Published online by Cambridge University Press

The estimate obtained confirms a departure from standard parametric models, such as the Vasicek variance $\sigma^2(r) = k$ or the CIR variance $\sigma^2(r) = kr$. However, it does not look very different from the CEV model, as estimated by Jiang (1998) on the same data set using indirect inference. Figure 5 shows the estimated diffusion function $\hat{\sigma}^2 r^{2\hat{\gamma}}$ for comparison with the nonparametric estimate. Standard errors of the nonparametric estimate are computed via (4.12). We see that it is hard to reject the CEV model on the data set under analysis, even though we find a larger volatility for interest rates under 6% and a lower volatility in the interval 7%–12%.

7.2. The Univariate Model: Estimation Based on Realized Variance

The implementation of the estimator (4.11) is much more demanding from the point of view of data requirements. To properly estimate quadratic variation, we need high-frequency data; daily sampling is not enough. The idea is to sample observations of the short rate at a daily level, using closing prices, while using intraday data to estimate the integrated volatility between daily closing short rates.

To this purpose, we use a data set that contains all the transactions of the three-month Eurodollar futures as traded at the Chicago Mercantile Exchange (CME). We select this contract because it is the most liquid short-rate instrument in the U.S. interest rate market. The data at our disposal start with the opening of this market, on 11 February 1982, and end on 30 September 2004. The market is open from 7:20 to 14:00, CME time. For each day, after converting the Eurodollar futures price to an interest rate, we estimate quadratic variation using realized volatility, computed as the sum of eighty 5-minute squared returns, plus the squared close-to-open overnight return. Therefore on day $i + 1$ we compute

$$\int_{iT/n}^{((i+1)T)/n} \sigma^2(r_s)\, ds \simeq \sum_{j=0}^{79} (r_{((i+1)T/n)-(j-1)\Delta} - r_{((i+1)T/n)-j\Delta})^2 + (r_{OPEN,\,i+1} - r_{iT/n})^2,$$

where $r_{OPEN,\,i+1}$ denotes the opening rate at day $i + 1$ and $\Delta = T/80n$. The time series of the closing interest rate and of realized volatility is shown in the inset of Figure 6.[7] In total, we have 5,669 days.

Using integrated volatility instead of the squared close-to-close return as in Florens-Zmirou (1993) potentially could be economically significant. Indeed, even if volatility is high, so that the rate changes during the day, it could close at the same rate as the previous day, which would lead to an estimate of null volatility if using the squared close-to-close return.

In Figure 6, we show the estimated $\sigma^2(r)$ using realized volatility according to estimator (4.11), and we compare it with the classical estimator of

https://doi.org/10.1017/S026646660808047X Published online by Cambridge University Press

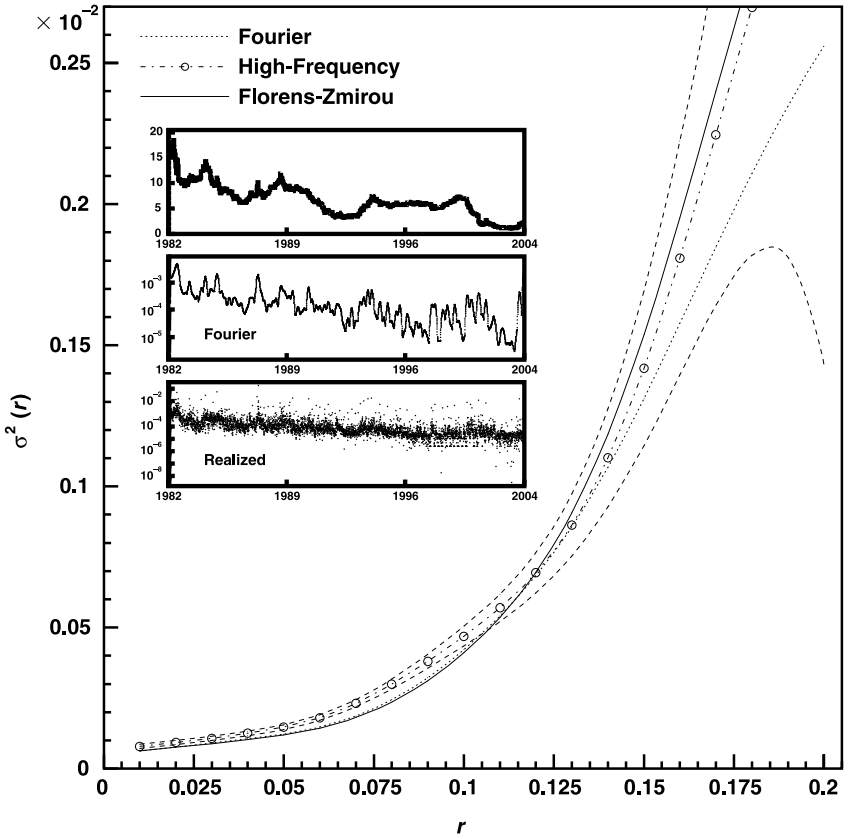


FIGURE 6. Estimate of the diffusion coefficient $\sigma^2(r)$ on the Eurodollar futures data set. Solid line: Florens-Zmirou estimate. Dashed line: Fourier estimate. Dashed-dotted line with dots: estimate using all high-frequency transactions to estimate realized volatilities. Dashed lines are 5% and 95% confidence intervals for the latter estimator, computed using (4.12). Inset: the time series of the yields on three-month Eurodollar futures under study, from February 1982 to September 2004 (top); the reconstructed time series of the variance according to the Fourier method (center); the time series of realized volatilities (bottom).

Florens-Zmirou (1993) and to the estimator (4.10) obtained with Fourier analysis. We use $h_s = 4$ and, for Fourier estimation, $\delta_1 = \frac{1}{50}$.

The Fourier, classical, and high-frequency estimators provide close estimates, the latter of which is supposed to be most precise. For low interest rates (below 10%) the high-frequency estimator provides a slightly greater variance. When making use of all the information available, that is, including high-frequency data, our result suggests that the volatility of the short rate is still substantial when the short rate is low.

https://doi.org/10.1017/S026646660808047X Published online by Cambridge University Press

7.3. The Bivariate Model

Last, we estimate the function $\Lambda^2(\sigma^2)$ of model (7.2). We use the daily yields on the three-month Treasury bill from 4 February 1960 to 11 December 2003, for a total of 10,944 observations (minimum 0.79% on 19 June 2003, maximum 17.14% on 11 December 1980). The bandwidth parameters h_n and $\tilde{h}_n$ are both of the form (4.5), but we use $h_s = 4$ and $h_s = 1.06$, respectively. There is no need to use a larger bandwidth when estimating $\Lambda^2(\sigma^2)$, because volatility is not as sparse as the interest rate data. For Fourier estimation, we use $\delta_1 = \frac{1}{50}$ and $\delta_2 = \frac{1}{25}$. Figure 7 shows the results. To gauge our estimate, we also plot the estimated (parametric) value of Andersen et al. (2004).[8] We compute confidence intervals using (5.3) and (5.5), but in the latter equation we use $2\widetilde{M} = 2/\delta_1$ instead of $n = 10{,}944$ to get more realistic confidence intervals, because if M is the Nyquist frequency, the number of data points would be $2M$. We find that the volatility of volatility is increasing in an almost linear fashion with respect to volatility, but this result has to be taken cautiously, because the confidence intervals are quite broad and the estimate is quite sensitive to the choice of $\widetilde{M}$. Thus, the estimate shown in Figure 7 has to be regarded as a first step toward nonparametric estimation of the diffusion coefficient of the volatility equation, even though it may be extremely useful already at this level, compared to parametric models.

8. CONCLUSIONS AND DIRECTIONS FOR FUTURE WORK

In this paper new nonparametric estimators for the diffusion coefficient based on discretely sampled observations are devised. The first is based on a volatility estimation result obtained in Malliavin and Mancino (2002). The second is based on the idea of estimating quadratic variation between adjacent observations using high-frequency data. Both estimators are proved to be consistent and asymptotically normally distributed, and feasible asymptotic confidence intervals are provided.

The two estimators are important in different respects. By making use of a larger quantity of data, the high-frequency estimator of volatility is more precise than the Fourier estimator and the estimators used so far in the literature, namely, those introduced by Florens-Zmirou (1993) and Bandi and Phillips (2003). The Fourier estimator, instead, provides similar results for volatility to those obtained with the Florens-Zmirou estimator. But, as shown using simulated data, the Fourier estimator has the advantage that it can be iterated to get an estimate of the diffusion coefficient of a bivariate model, in which the second state variable is the latent volatility of the first.

We then use these estimators for interest rate modeling. We estimate the diffusion coefficient of univariate and bivariate models in which the observable quantity is the instantaneous short rate, which we proxy with the three-month Treasury bill rate, and compare them to each other and to some well-known parametric estimates.

https://doi.org/10.1017/S026646660808047X Published online by Cambridge University Press

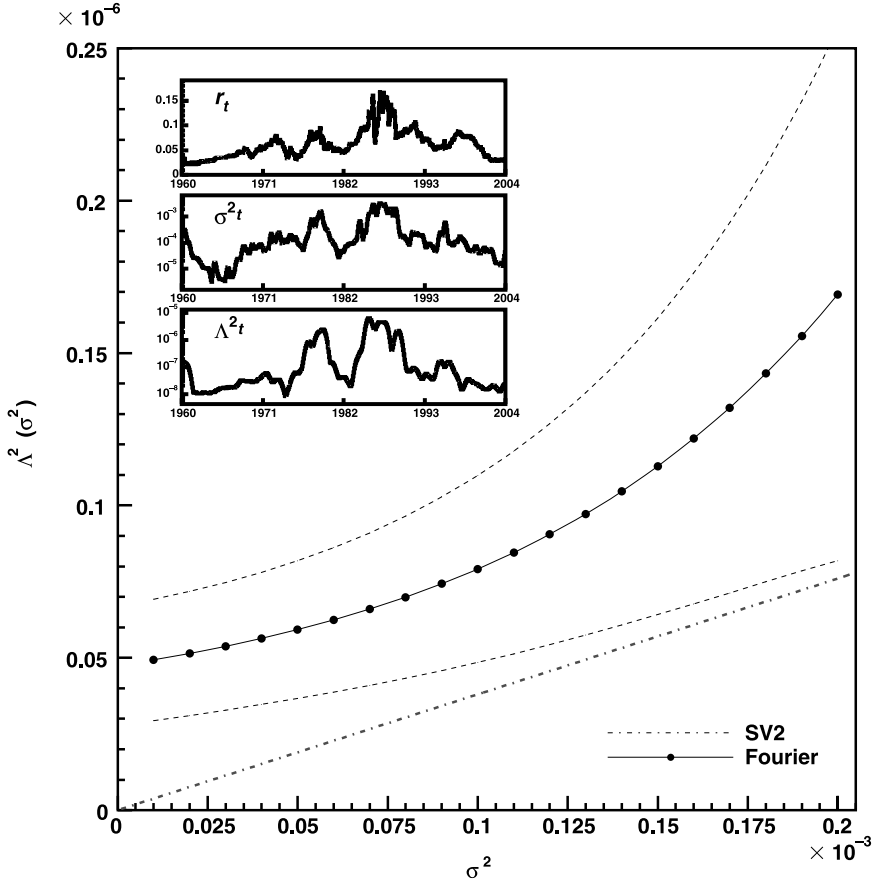


FIGURE 7. Estimate of the squared volatility diffusion coefficient $\Lambda^2(\sigma^2)$ on the three-month Treasury bill rate time series. Solid line: Fourier estimate. Dashed lines are 5% and 95% confidence intervals, computed using (5.3) and (5.5). Inset: the interest rate time series (top); the estimation of the variance time series obtained with the Fourier method (center); the estimation of the variance of the variance time series, obtained with the Fourier method (bottom). The dashed-dotted line is the parametric estimate of the SV2 model of Andersen et al. (2004).

When estimating the diffusion coefficient of a univariate model with the Fourier estimator, we observe slight departures from a CEV model. When using the high-frequency estimator, we confirm the estimation of a not vanishing volatility for very low interest rate levels and a greater volatility for low interest rate levels than that obtained with the classical and the Fourier methods.

When estimating the diffusion coefficient of a bivariate model with the Fourier estimator, we find that the volatility of volatility increases in a linear

https://doi.org/10.1017/S026646660808047X Published online by Cambridge University Press

fashion with the volatility itself. Moreover, we find a slightly greater volatility of volatility than that estimated with a parametric model by Andersen et al. (2004), using a similar time span.

The nonparametric estimators introduced can be used in a variety of applications. The possibility of nonparametrically estimating the volatility of volatility on actual data is particularly interesting, and this paper is the first step in this direction. This can be of crucial importance, both theoretically and practically, when devising parametric models. This is also true for derivative pricing purposes, because volatility does not change its shape when changing probability measure, that is, when incorporating risk preferences in the model. However, in small samples, the estimator proposed is quite far from being fully nonparametric. Four bandwidths, at least, affect the estimation of the volatility of volatility $(h_n, \tilde{h}_n, \delta_1, \delta_2)$ and have to be selected. It is clear that optimal tuning of the bandwidths, as for any nonparametric application, is strongly model-dependent. However it is also clear that more light needs to be shed on this crucial issue, and this can be done via either simulation studies or theoretical computation.

On the other hand, a serious limitation to this approach is that the assumption of a continuous sample path for asset prices may be too restrictive. Including discontinuous variations, such as jumps, may be crucial to obtain correct specification. In this respect, the results of Mancini and Renò (2006), who propose a nonparametric estimator to filter the jump process, may be of help, and this topic is currently under investigation.

NOTES

1. A different direction for estimating nonparametrically stochastic volatility models has been pointed out in Renò (2006).

2. The Fourier analysis of Malliavin and Mancino (2002) also provides an estimator of the integrated volatility, via the coefficient $a_0(\sigma^2)$.

3. Given the nature of the estimator, and for clarity, we prefer to use σ^2 instead of x as the real variable.

4. Using different interpolation techniques we get a bias in the volatility measurement (Barucci and Renò, 2002b).

5. This value of δ_1 was found by minimizing the integrated square error between the generated and reconstructed trajectory, where the model (6.7) was used for generation.

6. The Nyquist frequency is defined as the larger frequency that prevents aliasing effects; see Priestley (1979). For N equally spaced data, it is equal to $N/2$.

7. Many days in the final part of our data set display no volatility at all; that is, the three-month Eurodollar rate is the same across the whole day, possibly because of the fact that the interest rate is at its lowest in the sample.

8. The two estimates can be only compared loosely, because the data sets do not span exactly the same time period, although they overlap substantially. Whereas Andersen et al. (2004) use weekly data, we use daily data to get better resolution on the coefficients.

REFERENCES

Aït-Sahalia, Y. (1996) Nonparametric pricing of interest rate derivative securities. *Econometrica* 64, 527–560.

https://doi.org/10.1017/S026646660808047X Published online by Cambridge University Press

Aït-Sahalia, Y., P. Mykland, & L. Zhang (2005) How often to sample a continuous-time process in the presence of microstructure noise. *Review of Financial Studies* 18, 351–416.

Andersen, T., L. Benzoni, & J. Lund (2004) Stochastic Volatility, Mean Drift and Jumps in the Short-Term Interest Rate. Working paper, Northwestern University.

Andersen, T., T. Bollerslev, & F.X. Diebold (2003) Parametric and Nonparametric Volatility Measurement. Working paper. Forthcoming in L.P. Hansen & Y. Aït-Sahalia (eds.), *Handbook of Financial Econometrics*. North-Holland.

Andersen, T., T. Bollerslev, F. Diebold, & P. Labys (2003) Modeling and forecasting realized volatility. *Econometrica* 71, 579–625.

Bandi, F. (2002) Short-term interest rate dynamics: A spatial approach. *Journal of Financial Economics* 65, 73–110.

Bandi, F. & G. Moloche (2004) On the Functional Estimation of Multivariate Diffusion Processes. Working paper, University of Chicago and MIT.

Bandi, F. & P. Phillips (2003) Fully nonparametric estimation of scalar diffusion models. *Econometrica* 71, 241–283.

Bandi, F. & J. Russell (2006) Separating microstructure noise from volatility. *Journal of Financial Economics* 79, 655–692.

Barndorff-Nielsen, O.E. & N. Shephard (2002) Econometric analysis of realised volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society, Series B* 64, 253–280.

Barucci, E. & R. Renò (2002a) On measuring volatility and the GARCH forecasting performance. *Journal of International Financial Markets, Institutions and Money* 12, 183–200.

Barucci, E. & R. Renò (2002b) On measuring volatility of diffusion processes with high frequency data. *Economics Letters* 74, 371–378.

Chan, K., A. Karolyi, F. Longstaff, & A. Sanders (1992) An empirical comparison of alternative models of the short-term interest rate. *Journal of Finance* 47, 1209–1227.

Chapman, D., J. Long, & N. Pearson (1999) Using proxies for the short rate: When are three months like an instant? *Review of Financial Studies* 12, 763–806.

Cox, J., J. Ingersoll, & S. Ross (1985) A theory of the term structure of interest rates. *Econometrica* 53, 385–406.

Duffee, G. (1996) Idiosyncratic variation of Treasury bill yield spread. *Journal of Finance* 51, 527–552.

Fan, J. (2005) A selective overview of nonparametric methods in finance. *Statistical Science* 20, 317–337.

Fan, J. & C. Zhang (2003) A re-examination of Stanton's diffusion estimations with applications to financial model validation. *Journal of the American Statistical Association* 98, 118–134.

Florens-Zmirou, D. (1993) On estimating the diffusion coefficient from discrete observations. *Journal of Applied Probability* 30, 790–804.

Föllmer, H. (1981) Calcul d'Itô sans probabilités. *Seminaire de Probabilites Strasbourg XV*, Lecture Notes in Mathematics 850, pp. 143–150. Springer.

Hansen, P. & A. Lunde (2006) Consistent ranking of volatility models. *Journal of Econometrics* 131, 97–121.

Hansen, P., A. Lunde, & J.M. Nason (2003) Choosing the best volatility models: The model confidence set approach. *Oxford Bulletin of Economics and Statistics* 65, 939–961.

Härdle, W. (1990) *Applied Nonparametric Regression*. Cambridge University Press.

Hoffmann, M. (1999) L_p estimation of the diffusion coefficient. *Bernoulli* 5, 447–481.

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

Jacod, J. (1999) Non-parametric kernel estimation of the coefficient of a diffusion. *Scandinavian Journal of Statistics* 27, 83–96.

Jacod, J. & A.N. Shiryaev (1987) *Limit Theorems for Stochastic Processes*. Springer-Verlag.

Jeffrey, A., D. Kristensen, O. Linton, T. Nguyen, & P.C.B. Phillips (2004) Nonparametric estima-

https://doi.org/10.1017/S026646660808047X Published online by Cambridge University Press

tion of a multifactor Heath-Jarrow-Morton model: An integrated approach. *Journal of Financial Econometrics* 2, 251–289.
Jiang, G. (1998) Nonparametric modeling of U.S. interest rate term structure dynamics and implications on the prices of derivative securities. *Journal of Financial and Quantitative Analysts* 33, 465–497.
Jiang, G. & J. Knight (1997) A nonparametric approach to the estimation of diffusion processes, with an application to a short-term interest rate model. *Econometric Theory* 13, 615–645.
Karatzas, I. & E. Shreve (1988) *Brownian Motion and Stochastic Calculus*. Springer-Verlag.
Knight, F.B. (1970) A reduction of continuous square integrable martingale to Brownian motion. In H. Dinges (ed.), *Lecture Notes in Mathematics*, Volume 190, pp. 19–31. Springer-Verlag.
Malliavin, P. & M. Mancino (2002) Fourier series method for measurement of multivariate volatilities. *Finance and Stochastics* 6, 49–61.
Malliavin, P. & A. Thalmaier (2005) *Stochastic Calculus of Variations in Mathematical Finance*. Springer Finance series. Springer.
Mancini, C. & R. Renò (2006) Threshold Estimation of Jump-Diffusion Models and Interest Rate Modeling. Manuscript, University of Florence and University of Siena.
Meddahi, N. (2002) A theoretical comparison between integrated and realized volatility. *Journal of Applied Econometrics* 17, 479–508.
Nielsen, M.O. & P.H. Frederiksen (2008) Finite sample accuracy and choice of sampling frequency in integrated volatility estimation. *Journal of Empirical Finance* 15, 265–286.
Priestley, M. (1979) *Spectral Time Series Analysis*. Wiley.
Renò, R. (2006) Nonparametric estimation of stochastic volatility models. *Economics Letters* 90, 390–395.
Renò, R., A. Roma, & S. Schaefer (2006) A comparison of alternative nonparametric estimators of the diffusion coefficient. *Economic Notes* 35, 227–252.
Revuz, D. & M. Yor (1998) *Continuous Martingales and Brownian Motion*. Springer-Verlag.
Scott, D. (1992) *Multivariate Density Estimation*. Wiley.
Silverman, B.W. (1986) *Density Estimation for Statistics and Data Analysis*. Chapman and Hall/CRC.
Stanton, R. (1997) A nonparametric model of term structure dynamics and the market price of interest rate risk. *Journal of Finance* 52, 1973–2002.
Vasicek, O. (1977) An equilibrium characterization of the term structure. *Journal of Financial Economics* 5, 177–188.
Wand, M.P. & M.C. Jones (1995) *Kernel Smoothing*. Chapman and Hall.
Wiener, N. (1924) The quadratic variation of a function and its Fourier coefficients. *Massachusetts Journal of Mathematics* 3, 72–94.

APPENDIX: Proofs

Proof of Proposition 1. Proposition 1 is a slightly extended version of the main result which can be found in Malliavin and Mancino (2002). ■

Proof of Propositions 2 and 3. These are Propositions 1, 2, and 3 in Florens-Zmirou (1993), with the kernel given by the indicator function. The extension to a general kernel is studied in Jiang and Knight (1997). ■

Proof of Theorem 1. Suppose $\mu(x) = 0$ in (4.1). For all $\omega \in \Omega$, consider the solution $X_t(\omega)$ of equation (4.1) and define the process Y_t^ω defined by

$$dY_t^\omega = \sigma(X_t(\omega))\, dW_t', \tag{A.1}$$

https://doi.org/10.1017/S026646660808047X Published online by Cambridge University Press

where W_t' is a standard Brownian motion in an auxiliary probability space $(\Omega', (\mathcal{F}_t')_{0\le t\le T}, P')$ and $\sigma(X_t(\omega))$ is the realization of $\sigma(X_t)$. Assumption 1 guarantees the existence of the solution of (A.1). Then we can construct random variables in Ω by taking expectations in Ω'. We denote by E' the expected value in Ω' and by $\mathbb{E}$ the expected value in Ω.

Rewrite equation (2.2) as

$$\sigma^2(X_t) = \lim_{M\to\infty} \sum_{k=-M}^{M} A_k(\sigma^2) e^{ikt}, \tag{A.2}$$

where

$$A_k(\sigma^2) = \begin{cases} \dfrac{1}{2}\,(a_k(\sigma^2) - ib_k(\sigma^2)), & k \ge 1 \\ \dfrac{1}{2}\,a_0(\sigma^2), & k = 0. \\ \dfrac{1}{2}\,(a_{|k|}(\sigma^2) + ib_{|k|}(\sigma^2)) & k \le -1 \end{cases} \tag{A.3}$$

Using equation (A.2), we get for $0 \le t < s \le T$ and almost surely in Ω

$$\begin{aligned} \frac{1}{s-t}\,\mathrm{E}'[(Y_t^\omega - Y_s^\omega)^2] &= \frac{1}{s-t}\int_t^s \sigma^2(X_u(\omega))\,du \\ &= \frac{1}{s-t}\int_t^s \sum_{q=-\infty}^{+\infty} A_q(\sigma^2) e^{iqu}\,du \\ &= \sum_{q=-\infty}^{+\infty} A_q(\sigma^2) e^{iqt}\,\frac{e^{iq(s-t)}-1}{iq(s-t)} \\ &= \sigma^2(X_t(\omega)) + \sum_{q=-\infty}^{+\infty} A_q(\sigma^2) e^{iqt}\left(\frac{e^{iq(s-t)}-1}{iq(s-t)} - 1\right) \\ &:= \sigma^2(X_t(\omega)) + F(s,t) \\ &:= \sigma_{M_n}^2(X_t(\omega)) + R_{M_n}(s,t) + F(s,t) \end{aligned} \tag{A.4}$$

because the Fourier series can be integrated term by term. The remainder R_{M_n} is defined as the approximation error of the truncated Fourier sum, $\sigma^2(X_t) - \sigma_{M_n}^2(X_t)$, and it has the property that, given the smoothness of $\sigma(\cdot)$ and the fact that X_t is a continuous function limited quadratic variation, we have (Wiener, 1924) that there exists $C \in \mathbb{R}$ such that

$$|R_{M_n}(s,t)| = |\sigma^2(X_t) - \sigma_{M_n}^2(X_t)| \le \frac{C}{M_n}, \quad \forall t,s \in [0,T]; \tag{A.5}$$

https://doi.org/10.1017/S026646660808047X Published online by Cambridge University Press

C does not depend on ω, because $\sigma(\cdot)$ is bounded. Moreover, because the integral of a Fourier series converges uniformly, we have almost surely in Ω

$$\lim_{s\to t} F(s,t) = 0. \tag{A.6}$$

Now it is simple to prove that, almost surely,

$$\mathbb{E}[\mathrm{E}'[(Y_t^\omega - Y_s^\omega)^2]|\mathcal{F}_s] = \mathbb{E}[(X_t - X_s)^2|\mathcal{F}_s], \tag{A.7}$$

where X is the solution of (4.1), because both are equal to $\int_s^t \mathbb{E}[\sigma^2(X_u)|\mathcal{F}_s]\,du$. Moreover, using $\mathbb{E}[(X_t - X_s)^4|\mathcal{F}_s] = 3\mathbb{E}^2[\int_s^t \sigma^2(X_u)\,du|\mathcal{F}_s]$ and Cauchy–Schwartz inequality, we get almost surely

$$\mathbb{E}[\mathrm{E}'^2[(Y_t^\omega - Y_s^\omega)^2]|\mathcal{F}_s] \le \mathbb{E}[(X_t - X_s)^4|\mathcal{F}_s]. \tag{A.8}$$

Define

$$U^n(x) = \frac{T}{nh_n}\sum_{i=0}^{n-1} K\left(\frac{X_i - x}{h_n}\right)\sigma^2_{M_n}(t_i), \tag{A.9}$$

where $\sigma^2_{M_n}(t_i)$ is computed via truncation of the expansion (2.2) at a properly selected frequency M_n fulfilling Assumption 4. We then have $S^n_{LF}(x) = U^n(x)/L^n_T(x)$.

Now, let us denote the $\mathcal{L}^2(\Omega)$ norm of X by $\|X\|^2 = \mathbb{E}[X^2]$. We then use almost sure identity (A.4) and get

$$\begin{aligned}\|U^n(x) - \sigma^2(x)L^n_T(x)\| = \Bigg\| &\frac{1}{h_n}\sum_{i=0}^{n-1} K\left(\frac{X_i - x}{h_n}\right)\mathrm{E}'[(Y^\omega_{i+1} - Y^\omega_i)^2] \\ &- \sigma^2(x)L^n_T(x) - \frac{T}{nh_n}\sum_{i=0}^{n-1} K\left(\frac{X_i - x}{h_n}\right)F(t_{i+1},t_i) \\ &- \frac{T}{nh_n}\sum_{i=0}^{n-1} K\left(\frac{X_i - x}{h_n}\right)R_{M_n}(t_{i+1},t_i)\Bigg\|.\end{aligned} \tag{A.10}$$

When expanding the square in the norm (A.10), we can use the fact that $\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X|\mathcal{F}_{t_i}]]$ and then use (A.7) and (A.8) to get

$$\begin{aligned}&\|U^n(x) - \sigma^2(x)L^n_T(x)\| \\ &\quad= \Bigg\| V^n(x) - \sigma^2(x)L^n_T(x) - \frac{T}{nh_n}\sum_{i=0}^{n-1} K\left(\frac{X_i - x}{h_n}\right)F(t_{i+1},t_i) \\ &\qquad - \frac{T}{nh_n}\sum_{i=0}^{n-1} K\left(\frac{X_i - x}{h_n}\right)R_{M_n}(t_{i+1},t_i)\Bigg\| \\ &\quad\le \|V^n(x) - \sigma^2(x)L^n_t(x)\| + \Bigg\|\frac{T}{nh_n}\sum_{i=0}^{n-1} K\left(\frac{X_i - x}{h_n}\right)F(t_{i+1},t_i)\Bigg\| \\ &\qquad + \Bigg\|\frac{T}{nh_n}\sum_{i=0}^{n-1} K\left(\frac{X_i - x}{h_n}\right)R_{M_n}(t_{i+1},t_i)\Bigg\|.\end{aligned} \tag{A.11}$$

https://doi.org/10.1017/S026646660808047X Published online by Cambridge University Press

The three terms converge to zero: the first, because of Proposition 3; the second given Proposition 2 and because $F(t_{i+1}, t_i) = o(1/n)$; the third, given Proposition 2, equation (A.5), and the fact that $M_n \to +\infty$.

If $\mu(x) \neq 0$, then (A.7) becomes

$$\mathbb{E}[\mathrm{E}'[(Y_t^\omega - Y_s^\omega)^2]] = \mathbb{E}[(X_t - X_s)^2] - \mathbb{E}\left[\left(\int_s^t \mu(X_u)\, du\right)^2\right], \tag{A.12}$$

and the second term of the right-hand side vanishes as $s \to t$, that is, $n \to \infty$.

The consistency for S_{HF}^n is obtained in the same way, after noting that

$$\mathrm{E}'[(Y_t^\omega - Y_s^\omega)^2] = \int_t^s \sigma^2(X_u(\omega))\, du. \tag{A.13}$$

■

LEMMA 1. *Let $T = 1$. If $nh_n^3 \to 0$ when $n \to \infty$, then*

$$\sum_{i=1}^{[nt]} \mathbb{E}\left[\frac{n}{h_n}\left(K\left(\frac{X_{i-1} - x}{h_n}\right)[(X_i - X_{i-1})^2 - \sigma^2(x)/n]\right)^2 \middle| \mathcal{F}_{i-1}\right] \to \sigma^4(x) L_t(x), \tag{A.14}$$

where the preceding convergence is in probability.

Proof of Lemma 1. This is Lemma 2(b) in Florens-Zmirou (1993), extended for a smooth kernel in Jiang and Knight (1997). ■

LEMMA 2. *Consider the process Y_t^n defined, on $(\Omega, (\mathcal{F}_i)_{i\in\mathbb{N}}, P)$, as*

$$Y_t = \sum_{i=1}^{[nt/T]} \Theta_i, \tag{A.15}$$

where $[x]$ is the integer part of x, with Θ_i bounded and adapted to $\mathcal{F}_{i-1}$, and make the following assumptions:

(i) $\sum_{i=1}^{[nt/T]} \mathbb{E}[\Theta_i | \mathcal{F}_{i-1}] \to 0$ *in probability;*
(ii) $\sum_{i=1}^{[nt/T]} \mathbb{E}[\Theta_i^2 | \mathcal{F}_{i-1}] \to V_t$ *in probability;*
(iii) $\forall \varepsilon > 0$, $\sum_{i=1}^{[nt/T]} \mathbb{E}[\Theta_i^2 I_{\{|\Theta_i| > \varepsilon\}} | \mathcal{F}_{i-1}] \to 0$ *in probability (conditional Lindeberg condition).*

Then Y_t converges in distribution to the continuous martingale M_t with quadratic variation $[M, M]_t = V_t$.

Proof of Lemma 2. We have just to verify that the assumptions of Theorem VIII.2.17 in Jacod and Shiryaev (1987) hold. We compute the characteristics (B^n, C^n, ν^n) of Y^n by Theorem II.3.11 in Jacod and Shiryaev (1987), with $h(x) = x \wedge \sup \Theta_i$ and $\tilde{C}$ defined as

$$\tilde{C}_t = C_t + h^2 * \nu_t - \sum_{s \leq t} (\Delta B_s)^2. \tag{A.16}$$

The characteristics of M_t are $(0, V_t, 0)$. The conditions to be fulfilled are as follows.

(a) $\sup_{s \leq t} |B_s^n - B_s| \to 0$ *in probability,* $\forall t \in [0, T]$.
We have $B'^n = B^n$. Because Θ_i is bounded, this follows directly.

https://doi.org/10.1017/S026646660808047X Published online by Cambridge University Press

(b) $\tilde{C}_t^n \to \tilde{C}_t$ *in probability,* $\forall t \in [0,T]$.
This comes directly from the definition of C'^n and the fact that $\sum_{s\le t}(\Delta B_s)^2 = \sum_{i=1}^{[nt/T]} \mathbb{E}[\Theta_i^2 | \mathcal{F}_{i-1}] \to V_t$.

(c) $g * \nu_t^n \to g * \nu_t$ *in probability,* $\forall t \in [0,T], g \in \mathcal{C}_1(\mathbb{R})$.
If $g \in \mathcal{C}_1$ there exist real numbers k, K such that $|g(x)| \le Kx^2 I_{\{|x|>k\}}$; thus the conditional Lindeberg condition implies $g * \nu^n \to 0$.

Thus the proof is complete. ■

LEMMA 3. *Let $g(x): \mathbb{R} \to \mathbb{R}$ be a continuously differentiable bounded function, with bounded first derivative. Let $nh_n^3 \to 0$ when $n \to \infty$. Consider*

$$G_t(x) = \frac{1}{\sqrt{nh_n}} \sum_{i=1}^{[nt/T]} K\left(\frac{X_i - x}{h_n}\right)[g(X_i) - g(x)]. \tag{A.17}$$

Then, as $n \to \infty$, $G_t(x) \to 0$ in the $\mathcal{L}^1$ sense and thus in probability.

Proof of Lemma 3. We have

$$\mathbb{E}[|G_t(x)|] \le \mathbb{E}\left[\frac{1}{\sqrt{nh_n}} \sum_{i=1}^{[nt/T]} K\left(\frac{X_i - x}{h_n}\right)|g(X_i) - g(x)|\right]. \tag{A.18}$$

We now divide the sum into terms such that $|X_i - x| \le n^{-1/3}$ and the terms for which the opposite holds. Then

$$\begin{aligned}
\mathbb{E}[|G(x)|] &\le \mathbb{E}\left[\frac{1}{\sqrt{nh_n}} \sup_{|X_i - x| \le n^{-1/3}} |g(X_i) - g(x)| \sum_{|X_i - x| \le n^{-1/3}} K\left(\frac{X_i - x}{h_n}\right)\right. \\
&\qquad \left. + \frac{1}{\sqrt{nh_n}} \sum_{|X_i - x| > n^{-1/3}} K\left(\frac{X_i - x}{h_n}\right)|g(X_i) - g(x)|\right] \\
&\le \mathbb{E}\left[\sqrt{nh_n} \sup_{|X_i - x| \le n^{-1/3}} |g(X_i) - g(x)| \sum_{i=1}^{[nt/T]} \frac{1}{nh_n} K\left(\frac{X_i - x}{h_n}\right)\right. \\
&\qquad \left. + \frac{1}{\sqrt{nh_n}} \sup_{|X_i - x| > n^{-1/3}} |g(X_i) - g(x)| \sum_{|X_i - x| > n^{-1/3}} K\left(\frac{X_i - x}{h_n}\right)\right].
\end{aligned} \tag{A.19}$$

Now, using Taylor's rule, and given the boundedness of the first derivative of $g(\cdot)$, we get that $\sup_{|X_i - x| \le n^{-1/3}} |g(X_i) - g(x)| = o(n^{-1/3})$. Then, using Proposition 2, we have that the first term goes as $(nh_n^3)^{1/6}$; then it goes to zero as $n \to \infty$. The second term goes to zero given the boundedness of g, the fact that the argument of the kernel is greater than $1/n^{1/3}h_n = 1/(nh_n^3)^{1/3}$, and the fact that the kernel goes to zero faster than inverse polynomials when its argument goes to infinity, as in this case. ■

https://doi.org/10.1017/S026646660808047X Published online by Cambridge University Press

Proof of Theorem 2. Consider the discrete filtration $\mathcal{F}_i = \mathcal{F}_{t_i}$, $i = 0, \ldots, n$. Define the following term:

$$\Theta_i = \sqrt{\frac{T}{nh_n}} K\left(\frac{X_{i-1} - x}{h_n}\right)[\sigma^2_{M_n}(X_{i-1}) - \sigma^2(x)]. \tag{A.20}$$

Because $\sigma^2_{M_n}(\cdot)$ is bounded for every n, and $\sigma^2(\cdot)$ is bounded, Θ_i is bounded, and it is adapted to $\mathcal{F}_{i-1}$ (it is actually a constant with respect to $\mathcal{F}_{i-1}$). We want now to verify the conditions of Lemma 2. We have the following expression:

(a)

$$\begin{aligned}\sum_{i=1}^{[nt/T]} \mathbb{E}[\Theta_i|\mathcal{F}_{i-1}] &= \sum_{i=1}^{[nt/T]} \Theta_i \\ &= \sqrt{T/nh_n} \sum_{i=1}^{[nt/T]} K\left(\frac{X_{i-1} - x}{h_n}\right)[\sigma^2(X_{i-1}) - \sigma^2(x)] \\ &\quad + \sqrt{T/nh_n} \sum_{i=1}^{[nt/T]} K\left(\frac{X_{i-1} - x}{h_n}\right)[\sigma^2_{M_n}(X_{i-1}) - \sigma^2(X_{i-1})].\end{aligned}$$

The first term tends to zero in probability, given Lemma 3. The second term tends to zero, given Proposition 2 and because the approximation of the Fourier series is $o(1/M_n)$ and $M_n^{-1}\sqrt{nh_n} \to \infty$.

(b) We prove that $\sum_{i=1}^{[nt/T]} \mathbb{E}[\Theta_i^2|\mathcal{F}_{i-1}] \to \sigma^4(x)L_t(x)$. Given Lemma 1, this is equivalent to proving that

$$\begin{aligned}\Gamma_t := \sum_{i=1}^{[nt/T]} \mathbb{E}\Bigg[\frac{T}{nh_n} \Bigg\{ K\left(\frac{X_{i-1} - x}{h_n}\right) \\ \times \left[\sigma^2_{M_n}(X_{i-1}) - \frac{n}{T}(X_i - X_{i-1})^2\right]\Bigg\}^2 \Bigg| \mathcal{F}_{i-1} \Bigg] \to 0.\end{aligned}$$

We can use the derivation (A.4) and use the same reasoning leading to the proof of Theorem 1, together with inequality (A.8), to get

$$\begin{aligned}\Gamma_t &= \sum_{i=1}^{[nt/T]} \mathbb{E}\Bigg[\frac{T}{nh_n} \Bigg\{ K\left(\frac{X_{i-1} - x}{h_n}\right)\Bigg[\frac{n}{T} \mathrm{E}'[(Y_i^\omega - Y_{i-1}^\omega)^2] - \frac{n}{T}(X_i - X_{i-1})^2 \\ &\qquad - F(t_i, t_{i-1}) - R_{M_n}(t_i, t_{i-1})\Bigg]\Bigg\}^2 \Bigg| \mathcal{F}_{i-1} \Bigg] \\ &\le \sum_{i=1}^{[nt/T]} \mathbb{E}\Bigg[\frac{T}{nh_n} \Bigg\{ K\left(\frac{X_{i-1} - x}{h_n}\right)[F(t_i, t_{i-1}) + R_{M_n}(t_i, t_{i-1})]\Bigg\}^2 \Bigg| \mathcal{F}_{i-1} \Bigg].\end{aligned} \tag{A.21}$$

From this inequality, Proposition 2, and equation (A.5) we get the result.

https://doi.org/10.1017/S026646660808047X Published online by Cambridge University Press

(c) We have to prove the conditional Lindeberg condition. We have

$$\mathbb{E}\left[\left|\mathbb{E}\left[\Theta_i^2 I_{\{|\Theta_i|>\varepsilon\}}|\mathcal{F}_{i-1}\right]\right|\right] = \sum_{|\Theta_i|>\varepsilon} \Theta_i^2. \tag{A.22}$$

Now, the sum (A.22) is bounded by $\sigma^4(x)L_t(x)$; moreover, we can rewrite $|\Theta_i| > \varepsilon$ as

$$K\left(\frac{X_{i-1}-x}{h_n}\right)\left|\sigma^2_{M_n}(X_{i-1}) - \sigma^2(x)\right| > \varepsilon\sqrt{\frac{nh_n}{T}}. \tag{A.23}$$

The left-hand side of equation (A.23) is bounded; thus as $n \to \infty$ we have $\varepsilon\sqrt{nh_n} \to \infty$, and the sum (A.22) vanishes in probability.

Thus, we fulfill the assumptions of Lemma 2. Then if we define $Y_t^n(x)$ as

$$Y_t^n(x) = \sum_{i=1}^{[nt/T]} \Theta_i(x), \tag{A.24}$$

we have that $Y_t^n(x)$ converges in law to the continuous martingale M_t with quadratic variation $[M,M]_t = \sigma^4(x)L_t(x)$. We then set $M_t = B(\sigma^4(x)L_t(x))$, where $B(t)$ is a standard Brownian motion. Now consider

$$Z_t^n(x) = \sum_{i=1}^{[nt/T]} (W_{t_i} - W_{t_{i-1}}), \tag{A.25}$$

where W_t is the standard Brownian motion in (4.1). It is clear that $Z_t^n(x)$ converges in law to the standard Brownian motion W_t. Moreover we have

$$\sum_{i=1}^{[nt/T]} \mathbb{E}\left[\Theta_i(W_{t_i} - W_{t_{i-1}})|\mathcal{F}_{i-1}\right] = 0. \tag{A.26}$$

By equation (A.26), we get that M_t and W_t are orthogonal. We can also write $B(t) = M_{\tau(t)}$, where $\tau(t) = \inf_s(s/\sigma^4(x)L_s(x))$. Then, by Knight's theorem (Knight, 1970), we get that $B(t)$ and W_t are independent Brownian motions. Then $B(t)$ and $L_t(x)$ are independent, because the filtration generated by X_t is included in the filtration generated by W_t. We then have that $Y_t^n(x) \to \sqrt{L_t(x)}\sigma^2(x)\mathcal{N}(0,1)$, where $\mathcal{N}(0,1)$ is a standard normal random variable independent of $L(x)$. Because $L_t^n(x)$ converges in probability to $L_t(x)$, we have the desired result.

The asymptotic properties for S_{HF}^n are obtained in the same way, using (A.13). ■

https://doi.org/10.1017/S026646660808047X Published online by Cambridge University Press

Proof of Theorem 3. Assumption 5 implies that, by Proposition 2,

$$\frac{T}{n\tilde{h}_n}\sum_{i=1}^{[nt/T]} K\left(\frac{\sigma^2(t_i)-x}{\tilde{h}_n}\right) \to L_t(\sigma^2).$$

Now, from the properties of the kernel, we have that, for each t_i, there exists ξ_i such that

$$K\left(\frac{\sigma^2_{M_n}(t_i)-x}{\tilde{h}_n}\right) = K\left(\frac{\sigma^2(t_i)-x}{\tilde{h}_n}\right) + K'(\xi_i)\frac{\sigma^2_{M_n}(t_i)-\sigma^2(t_i)}{\tilde{h}_n}. \tag{A.27}$$

From the boundedness of K' and (A.5) we get

$$\frac{1}{n\tilde{h}_n}\left|\sum_{i=1}^{n} K\left(\frac{\sigma^2_{M_n}(t_i)-x}{\tilde{h}_n}\right) - \sum_{i=1}^{n} K\left(\frac{\sigma^2(t_i)-x}{\tilde{h}_n}\right)\right|$$

$$\le C\frac{1}{n\tilde{h}_n}\sum_{i=1}^{n}\left|\frac{\sigma^2_{M_n}(t_i)-\sigma^2(t_i)}{\tilde{h}_n}\right| \le C\frac{1}{M_n\tilde{h}_n^2},$$

where C is a suitable constant. Because $M_n\tilde{h}_n^2 \to \infty$ we get the result. ■

Proof of Theorem 4. Assumption 5, together with Theorem 1, implies that

$$\frac{\sum_{i=1}^{n} K\left(\frac{\sigma^2(t_i)-\sigma^2}{\tilde{h}_n}\right)\Lambda^2_{\bar{M}_n}(t_i)}{\sum_{i=1}^{n} K\left(\frac{\sigma^2(t_i)-\sigma^2}{\tilde{h}_n}\right)} \to \Lambda^2(\sigma^2) \tag{A.28}$$

in the $\mathcal{L}^2$ sense. Define ξ_i as in (A.27) and denote by $\delta_n = (1/n\tilde{h}_n)\sum_{i=1}^{n} K'(\xi_i)[(\sigma^2_{M_n}(t_i) - \sigma^2(t_i))/\tilde{h}_n]$. As in the proof of Proposition 3, we have that almost surely $\delta_n = o(1/M_n\tilde{h}_n^2)$, so that we can write, by Taylor's rule and for every trajectory, $1/(x+\delta_n) = (1/x) - (1/x^2)\delta_n + o(\delta_n^2)$.

Using this fact and equation (A.27) and neglecting second-order terms in δ_n, we have

$$\|\tilde{\Lambda}^2_n(\sigma^2) - \Lambda^2(\sigma^2)\|$$

$$= \left\| \frac{\frac{1}{n\tilde{h}_n}\sum_{i=1}^{n}\left(K\left(\frac{\sigma^2(t_i)-\sigma^2}{\tilde{h}_n}\right) + K'(\xi_i)\frac{\sigma^2_{M_n}(t_i)-\sigma^2(t_i)}{\tilde{h}_n}\right)\Lambda^2_{\bar{M}_n}(t_i)}{\frac{1}{n\tilde{h}_n}\sum_{i=1}^{n} K\left(\frac{\sigma^2(t_i)-\sigma^2}{\tilde{h}_n}\right) + \delta_n} - \Lambda^2(\sigma^2) \right\|$$

https://doi.org/10.1017/S026646660808047X Published online by Cambridge University Press

$$\leq \left\| \frac{\sum_{i=1}^{n} K\left(\frac{\sigma^2(t_i)-\sigma^2}{\tilde{h}_n}\right)\Lambda^2_{\widetilde{M}_n}(t_i)}{\sum_{i=1}^{n} K\left(\frac{\sigma^2(t_i)-\sigma^2}{\tilde{h}_n}\right)} - \Lambda^2(\sigma^2) \right\|$$

$$+ \left\| \frac{\frac{1}{n\tilde{h}_n}\sum_{i=1}^{n} K'(\xi_i)\frac{\sigma^2_{M_n}(t_i)-\sigma^2(t_i)}{\tilde{h}_n}\Lambda^2_{\widetilde{M}_n}(t_i)}{\frac{1}{n\tilde{h}_n}\sum_{i=1}^{n} K\left(\frac{\sigma^2(t_i)-\sigma^2}{\tilde{h}_n}\right)} \right\|$$

$$+ \left\| \frac{\frac{\delta_n}{n\tilde{h}_n}\sum_{i=1}^{n} K\left(\frac{\sigma^2(t_i)-\sigma^2}{\tilde{h}_n}\right)\Lambda^2_{\widetilde{M}_n}(t_i)}{\left[\frac{1}{n\tilde{h}_n}\sum_{i=1}^{n} K\left(\frac{\sigma^2(t_i)-\sigma^2}{\tilde{h}_n}\right)\right]^2} \right\|$$

$$+ \left\| \frac{\frac{\delta_n}{n\tilde{h}_n}\sum_{i=1}^{n} K'(\xi_i)\frac{\sigma^2_{M_n}(t_i)-\sigma^2(t_i)}{\tilde{h}_n}\Lambda^2_{\widetilde{M}_n}(t_i)}{\left[\frac{1}{n\tilde{h}_n}\sum_{i=1}^{n} K\left(\frac{\sigma^2(t_i)-\sigma^2}{\tilde{h}_n}\right)\right]^2} \right\|.$$

The first and the third terms go to zero, given (A.28) and Proposition 2. The second and fourth terms go to zero using the boundedness of K' and $\Lambda_{\widetilde{M}_n}$, equation (A.5), the fact that $M_n\tilde{h}_n^2 \to \infty$, and Proposition 2.

To get the asymptotic properties, we follow the same reasoning of the proof of Theorem 2, using Proposition 3 to estimate the local time and substituting (A.20) with

$$\tilde{\Theta}_i = \sqrt{\frac{T}{n\tilde{h}_n}} K\left(\frac{\sigma^2_{M_n}(t_{i-1})-\sigma^2}{\tilde{h}_n}\right)[\Lambda^2_{\widetilde{M}_n}(X_{i-1}) - \Lambda^2(\sigma^2)] \qquad \textbf{(A.29)}$$

and using repeatedly (A.27) and the fact that $M_n\tilde{h}_n^2 \to \infty$. ■

https://doi.org/10.1017/S026646660808047X Published online by Cambridge University Press