*Econometric Theory*, **19**, 2003, 754–777. Printed in the United States of America.
DOI: 10.1017/S0266466603195035

# BIAS REDUCTION IN NONPARAMETRIC DIFFUSION COEFFICIENT ESTIMATION

JOÃO NICOLAU
*Instituto Superior de Economia e Gestão*
*Universidade Técnica de Lisboa*

In this paper, we quantify the asymptotic bias of the Florens-Zmirou (1993, *Journal of Applied Probability* 30, 790–804) and Jiang and Knight (1997, *Econometric Theory* 13, 615–645) estimator for the diffusion coefficient when the step of discretization is fixed, and then we propose a bias adjustment that partially compensates for the distortion. Also, we show that our estimators have all the asymptotic properties of the Florens-Zmirou and Jiang and Knight estimator when the step of discretization goes to zero. We provide some examples.

## 1. INTRODUCTION

Stochastic continuous-time processes, especially diffusion processes, have been widely used in physical and biological sciences and, more recently, in financial economics. In mathematical finance the success of the diffusion continuous-time approach can be attributed to its many attractive properties (see Merton, 1990).

However, all models involve unknown parameters or functions, which need to be estimated from observations of the process. The estimation of diffusion processes is therefore a crucial step in all applications, in particular, in applied finance. Nonparametric estimation based on continuous sampling observations has been considered in the literature for many years (see references in Rao, 1999). However, as has been stressed by various authors, the continuous sampling observations hypothesis is unreasonable because, in practice, it is obviously impossible to observe a process continuously over any given interval, as a result, for instance, of the limitations on the precision of the measuring instrument or of the unavailability of observations at every point in time (Rao, 1999). It is therefore natural that the most recent research in diffusion processes estimation has been concerned with discrete time observations, where some progress has been made, in both parametric and nonparametric estima-

I thank the editor Peter C.B. Phillips and the two referees for comments and suggestions that led to considerable improvement of the paper. I am also grateful to Carlos Braumann and Tom Kundert for helpful comments. This research was supported by the Fundação para a Ciência e a Tecnologia (FCT) and by POCTI. Address correspondence to: João Nicolau, Instituto Superior de Economia e Gestão, Rua do Quelhas 6, 1200-781 Lisboa, Portugal; e-mail: nicolau@iseg.utl.pt.

  0266-4666/03 $12.00

tion. In nonparametric estimation based on discrete-time observations Florens-Zmirou (1993) proposes an estimator for the diffusion coefficient using a uniform kernel, without imposing any restrictions on the drift coefficient. Aït-Sahalia (1996) proposes a semiparametric estimation procedure for the diffusion term based on a parametric drift coefficient. Stanton (1997) builds approximations to the infinitesimal coefficients. Jiang and Knight (1997) use the Florens-Zmirou estimator with a Gaussian kernel for the diffusion coefficient and propose a drift estimator based on the diffusion coefficient estimator and the stationary (or marginal) density. Hansen, Scheinkman, and Touzi (1998) propose spectral methods to recover the infinitesimal coefficients from the stationary density and an eigenvalue-eigenfunction pair of the conditional expectation operator. More recently, Bandi and Phillips (2003) have developed an asymptotic theory for nonparametric estimates of the drift and diffusion coefficients, under broad assumptions on the data generating process. In particular, they do not require the existence of a time-invariant marginal data density, and thus stationarity is not needed.

We consider the stochastic differential equation (1) subsequently with unknown coefficients $a$ and $b$. The $X$ process is observed at instants $\{t_i = \Delta_n, i = 0,1,\dots,n\}$, where $\Delta_n$ is the step of discretization. To estimate the diffusion coefficient $b^2(x)$ consistently it is well known that, in principle, it is not necessary that the time span $t_n$ go to infinity (see Bandi and Phillips, 2003). Nevertheless, a crucial but to some extent unrealistic assumption, in both cases of estimation (drift and diffusion coefficients), is that $\Delta_n \to 0$ as $n$ increases. In fact, in most empirical applications the step of discretization is generally fixed, i.e., $\Delta_n = \Delta$ as $n$ increases. For example, the series could be daily, weekly, or monthly.

In this paper, we quantify the asymptotic bias of the Florens-Zmirou (1993) and Jiang and Knight (1997) estimator for the diffusion coefficient when the step of discretization is fixed (as $t_n = n\Delta \to +\infty$), and then we propose a bias adjustment that partially compensates for the distortion. This is done by including the drift coefficient or its nonparametric estimates in the diffusion estimator, which, as we will see, enables a reduction of the asymptotic bias. In addition, when $\Delta_n \to 0$ we will show that our estimators have all the asymptotic properties of the Florens-Zmirou and Jiang and Knight estimator. We conjecture that other bias adjustments can be entertained.

The rest of the paper is organized as follows. Section 2 defines the diffusion process and the main assumptions. Section 3 presents our principal results. Section 4 illustrates the proposed nonparametric estimator with some examples. Section 5 outlines some other possible developments and concludes.

## 2. THE DIFFUSION PROCESS

We assume that $X = \{X_t, t \geq 0\}$ is a diffusion process, with state space $I = (l, r)$, governed by the stochastic differential equation

$$dX_t = a(X_t)\,dt + b(X_t)\,dW_t, \qquad X_0 = x, \tag{1}$$

https://doi.org/10.1017/S0266466603195035 Published online by Cambridge University Press

where $\{W_t, t \geq 0\}$ is a (standard) Wiener process, $a$ and $b$ are unknown coefficients or functions, and $x$ is either a constant value or a random value $\mathcal{F}_0$-measurable independent of $W_t$. We assume that $a$ and $b$ have continuous derivatives.

Let $s(z) = \exp\{-\int_{z_0}^{z} 2a(u)/b^2(u)\,du\}$ be the scale density function ($z_0$ is an arbitrary point inside $I$) and $m(u) = (b^2(u)s(u))^{-1}$ the speed density function. Let $S(l,x] = \lim_{x_1 \to l} \int_{x_1}^{x} s(u)\,du$ and $S[x,r) = \lim_{x_2 \to r} \int_{x}^{x_2} s(u)\,du$ where $l < x_1 < x < x_2 < r$. We now present a set of five assumptions that are used throughout the paper.

A1. $S(l,x] = S[x,r) = +\infty$ for $x \in I$.

According to Arnold (1974, p. 114), if the infinitesimal coefficients $a$ and $b$ have continuous derivatives with respect to $x$, then there exists a unique continuous process that is defined up to a random explosion time $\eta$ in the interval $t_0 < \eta \leq +\infty$. The A1 condition assures that $P[\eta = +\infty | X_0 = x] = 1$ (Ikeda and Watanabe, 1981, pp. 362–363). Furthermore, the boundaries $l$ and $r$ are neither attracting nor attainable (see Karlin and Taylor, 1981, Ch. 15), and the process is recurrent, i.e., $P[T_y < \infty | X_0 = x] = 1$ for every $x, y \in I$ where $T_y = \inf\{t \geq 0, X_t = y\}$ (Ikeda and Watanabe, 1981, Theorem 3.1, Ch. VI). Roughly speaking, the boundaries $l$ and $r$ are never attained although every finite point can be reached with probability one in finite time. Global Lipschitz and growth conditions, which fail to be satisfied for many interesting models in economics in finance (Aït-Sahalia, 1996), are not needed in the presence of the previous assumptions. The A1 condition is not very strong: e.g., the standard Brownian motion satisfies the A1 condition (Karlin and Taylor, 1981, p. 228). Actually, every process with zero drift $a(x) = 0$ (and $b(x) > 0$) satisfies the A1 condition.

A2. $\int_l^r m(x)\,dx < +\infty$.

The A1 and A2 conditions assure that $X$ is ergodic and the invariant distribution $P^0$ has density $\bar{p}(x) = m(x)/\int_l^r m(u)\,du$ with respect to the Lebesgue measure (Skorokhod, 1989, Theorem 16). The expression $\bar{p}(x)$ is usually denoted as stationary density.

A3. $X_0 = x$ has distribution $P^0$.

Assumption A3 together with A1 and A2 implies that $X$ is stationary (Arnold, 1974). Assumption A3 can, in some cases, be replaced by the following assumption: $X_0$ is a random variable with mean $\mu$ and variance $\sigma^2$ such that $\int x\bar{p}(x)\,dx = \mu < +\infty$ and $\int (x-\mu)^2 \bar{p}(x)\,dx = \sigma^2 < +\infty$. In this case, $X$ is a covariance-stationary process. Even if the A3 condition does not hold, it is known that when $t$ is sufficiently large, the distribution of the ergodic process $X_t$ is well approximated by the distribution with density $\bar{p}(x)$. For simplicity, in our nonparametric estimation framework, we assume stationarity, although more broad assumptions can be considered (see Bandi and Phillips, 2003). As we will see, our estimators have important advantages over the Florens-Zmirou (1993) and Jiang and Knight (1997) estimator when the process exhibits strong reversion effects, which occurs typically in a stationary framework.

A4. $\lim_{x \to r} \sup\big([a(x)/b(x)] - [b'(x)/2]\big) < 0$, $\lim_{x \to l} \sup\big([a(x)/b(x)] - [b'(x)/2]\big) > 0$.

These conditions are discussed in Chen, Hansen, and Carrasco (1998) and are similar to ones proposed by Hansen and Scheinkman (1995). Under Assumption A4 the process is $\rho$-mixing (see Chen et al., 1998). Technically, for a Markov process, the notion of $\rho$-mixing requires the conditional expectations operator for any interval of time to be a strong contraction for all functions with zero mean and finite variance. As a consequence, the $j$th autocovariance of $f(X_t)$ tends to zero at exponential rate as $j \to +\infty$, for all functions $f$ such that $\int f(x)\bar{p}(x)\,dx = 0$ and $\int f^2(x)\bar{p}(x)\,dx < +\infty$ (see Hansen and Scheinkman, 1995, Proposition 8; Florens-Zmirou, 1989). We notice that even if the drift is zero or converges to zero, Assumption A4 can hold, provided that volatility grows at least linearly.

The kernel function $K$ that we will use to define our estimators satisfies the following assumption.

A5. $K(.)$ is symmetric and continuously differentiable and $\int_{\mathbb{R}} K(u)\,du = 1$, $\int_{\mathbb{R}} uK(u)\,du = 0$, $\int_{\mathbb{R}} K^2(u)\,du = K_2 < \infty$.

## 3. BIAS REDUCTION IN NONPARAMETRIC ESTIMATION

Florens-Zmirou (1993) proposes the estimator

$$S_n(x) = \frac{\displaystyle\sum_{i=0}^{n-1} K\left(\frac{x - X_{t_i}}{h_n}\right) \frac{(X_{t_{i+1}} - X_{t_i})^2}{\Delta_n}}{\displaystyle\sum_{i=0}^{n-1} K\left(\frac{x - X_{t_i}}{h_n}\right)} \tag{2}$$

for $b^2(x)$ where $K$ is the uniform kernel. Under some conditions, including $\Delta_n \to 0$, he proves that $S_n(x)$ converges in mean square error (MSE) to $b^2(x)$ and that the asymptotic distribution of $S_n(x)$ (suitably standardized) is normal. Jiang and Knight (1997) extend these results for the Gaussian kernel, and Bandi and Phillips (2003) consider a more general estimator under broad assumptions on the data generating process. Although $S_n(x)$ requires only mild regularity conditions to be consistent—in particular the time span $t_n$ can be fixed—it is necessary that the step of discretization go to zero (see, e.g., Bandi and Phillips, 2003). However, in most applications the step of discretization is generally fixed (i.e., $\Delta_n = \Delta$ as $n$ increases). For example, the series could be daily, weekly, or monthly. Even when $\Delta_n$ decreases as $n$ increases (which is possible if one passes, e.g., from weekly to daily observations), the condition $\Delta_n \to 0$ is too strong as it requires continuous sampling observations at the limit. When high frequency data are available (i.e., $\Delta_n$ is small), the advantages of using all the available information are not clear. In fact, high frequency data are replete with empirical anomalies, such as heteroskedasticity, heterokurtosis, non-synchronous trading, and many other market microstructure "frictions," which can obviously contaminate the estimates (see Sawyer, 1993; Andersen, Bollerslev, Diebold, and Labys, 2001).

### 3.1. Behavior of the Nonparametric Estimators When the Step of Discretization Is Fixed

In this section we quantify the bias of the Florens-Zmirou (1993) and Jiang and Knight (1997) estimator for the diffusion coefficient when the step of discretization is fixed and then we propose a bias adjustment that partially compensates for the distortion. We assume the A1–A5 conditions in all theorems.

THEOREM 1. *If* $\Delta_n = \Delta$ *(constant),* $t_n \to +\infty$*,* $h_n \to 0$*,* $nh_n \to +\infty$ *as* $n \to +\infty$*, and* $E[(X_\Delta - X_0)^4] < +\infty$ *then*

$$S_n(x) \xrightarrow{p} b^2(x) + a^2(x)\Delta + f(x)\Delta + O(\Delta^2), \tag{3}$$

*where* $f(x) = b^2(x)a'(x) + a(x)b(x)b'(x) + \frac{1}{2}b^2(x)(b'(x))^2 + \frac{1}{2}b^3(x)b''(x)$.

It is clear that the consistency of $S_n$ depends crucially on whether $\Delta_n$ goes to zero or not. When $\Delta_n = \Delta$ is constant the bias is of order $O(\Delta)$. This bias can vary considerably in the state space $I$. In general, it is minimum at point $x_0$ such that $a(x_0) = 0$, because in this case several terms associated with the bias in the stochastic limit (3) vanish. On the contrary, the bias increases in those points $x$ such that $|a(x)|$ is high, which in stationary framework corresponds to the interval where the process exhibits stronger reversion effects.

Based on these results we propose a bias correction with the following estimator:

$$V_n(x) = \frac{\displaystyle\sum_{i=0}^{n-1} K\left(\frac{x - X_{t_i}}{h_n}\right) \frac{(X_{t_{i+1}} - X_{t_i} - a(X_{t_i})\Delta_n)^2}{\Delta_n}}{\displaystyle\sum_{i=0}^{n-1} K\left(\frac{x - X_{t_i}}{h_n}\right)}.$$

We have the following theorem.

https://doi.org/10.1017/S0266466603195035 Published online by Cambridge University Press

THEOREM 2. *If* $\Delta_n = \Delta$ *(constant),* $t_n \to +\infty$, $h_n \to 0$, $nh_n \to +\infty$ *as* $n \to +\infty$, *and* $E[(X_\Delta - X_0 - a(X_0)\Delta)^4] < +\infty$ *then*

$$V_n(x) \xrightarrow{p} b^2(x) + f(x)\Delta + O(\Delta^2), \tag{4}$$

*where* $f(x)$ *is given in Theorem 1.*

Therefore, with $\Delta_n = \Delta$ constant, the $V_n(x)$ estimator generally has smaller bias than the $S_n$ estimator. The bias reduction occurs because of the elimination of the $a^2(x)\Delta$ term in the stochastic limit in equation (3). Actually, when $\Delta_n$ is constant, the evaluation of the errors $(X_{t_{i+1}} - X_{t_i} - a(X_{t_i})\Delta_n)^2/\Delta_n$ enables a more accurate estimation of the infinitesimal conditional variance than the errors $(X_{t_{i+1}} - X_{t_i})^2/\Delta_n$. However, the $V_n$ estimator is unfeasible because the function $a(x)$ is unknown. As in Aït-Sahalia (1996), we can assume $a(x)$ linear, i.e., of type $a(x) = \theta_1 - \theta_2 x$, $\theta_1, \theta_2 > 0$ where $\theta_1$ and $\theta_2$ can easily be estimated using semiparametric methods. However, if we do not want to rely on the parametric specification of the drift, we should use a nonparametric estimator for $a(x)$. In this case, we propose

$$V_n^*(x) = \frac{\sum_{i=0}^{n-1} K\left(\frac{x - X_{t_i}}{h_n}\right) \frac{(X_{t_{i+1}} - X_{t_i} - T_n(X_{t_i})\Delta_n)^2}{\Delta_n}}{\sum_{i=0}^{n-1} K\left(\frac{x - X_{t_i}}{h_n}\right)}, \tag{5}$$

where $T_n(x)$ is a nonparametric estimator of $a(x)$. We consider

$$T_n(x) = \frac{\sum_{i=0}^{n-1} K\left(\frac{x - X_{t_i}}{h_n}\right) \frac{(X_{t_{i+1}} - X_{t_i})}{\Delta_n}}{\sum_{i=0}^{n-1} K\left(\frac{x - X_{t_i}}{h_n}\right)}.$$

This estimator is a simpler version of a more general estimator proposed in Bandi and Phillips (2003). The Jiang and Knight (1997) estimator for $a(x)$ cannot, in principle, be used here because it depends on the diffusion coefficient estimator (see however Section 5). The following theorem characterizes the bias of the $T_n$ estimator.

THEOREM 3. *If* $\Delta_n = \Delta$ *(constant),* $t_n \to +\infty$, $h_n \to 0$, $nh_n \to +\infty$ *as* $n \to +\infty$, *and* $E[(X_\Delta - X_0)^2] < +\infty$ *then*[1]

$$T_n(x) \xrightarrow{p} a(x) + \left(\frac{1}{2}a(x)a'(x) + \frac{1}{4}b^2(x)a''(x)\right)\Delta + O(\Delta^2).$$

Therefore, when $\Delta$ is fixed, the drift estimator $T_n$ has a bias of order $O(\Delta)$. Because $V_n^*$ depends explicitly on $T_n$, it is important to identify the impact that the bias of the drift estimator has on the bias of the diffusion coefficient estimator. The following theorem addresses this question.

THEOREM 4. *If $\Delta_n = \Delta$ (constant), $t_n \to +\infty$, $h_n \to 0$, $nh_n \to +\infty$ as $n \to +\infty$, and $E[(X_\Delta - X_0 - a(X_0)\Delta)^4] < +\infty$ then*

$$V_n^*(x) \xrightarrow{p} b^2(x) + f(x)\Delta + O(\Delta^2), \tag{6}$$

*where $f(x)$ is given in Theorem 1.*

Theorem 4 shows that the impact of the bias of the $T_n$ estimator on the bias of the $V_n^*$ estimator is negligible. In fact, equation (6) is basically the same expression obtained with the $V_n$ estimator (see expression (4)). With the $V_n^*$ estimator we obtain a bias adjustment gain that is similar to that of the $V_n(x)$ estimator (ignoring the terms higher than $O(\Delta^2)$). A simple justification for this result is the following. Although the bias of the $a(x)$ estimation is of order $O(\Delta)$ (by Theorem 3), the bias involving the estimation of the $a^2(x)\Delta$ function, which is the term we correct and eliminate in equation (3), is of order $O(\Delta^3)$ (see the proof of Theorem 4 in Appendix B). Thus, the bias in estimating the $a(x)$ coefficient only appears in the stochastic limit of $V_n^*$ in the terms of order $O(\Delta^3)$.

The advantages of using the $V_n^*(x)$ (and $V_n(x)$) estimator over $S_n(x)$ are more evident when $x$ belongs to the interval where $|a(x)|$ is high, which in a stationary framework (with a nonzero drift) corresponds to the interval where the process exhibits stronger reversion effects. This is clear from equations (3) and (6). Obviously, if stationarity is volatility-induced with zero drift then there are no advantages of using the proposed estimators.

### 3.2. Limiting Properties as the Frequency of Observations Increases

We now address the limiting properties of $V_n$ and $V_n^*$ estimators as $\Delta_n \to 0$.

THEOREM 5.

(a) *If $\Delta_n \to 0$, $h_n \to 0$, $nh_n \to +\infty$ as $n \to +\infty$, and $E[(X_{\Delta_n} - X_0 - a(X_0)\Delta_n)^4] < +\infty$ then*

$$V_n(x) \xrightarrow{p} b^2(x).$$

*Let $B_n(x) = (nh)^{-1} \sum_{i=0}^{n-1} K((x - X_{t_i})/h_n)$.*

(b) *If, in addition to the previous conditions, $nh_n^3 \to 0$ and $\sqrt{nh_n}\Delta_n \to 0$ as $n \to +\infty^2$ then*

$$\sqrt{\frac{nh_n}{2K_2} B_n}\left(\frac{V_n(x)}{b^2(x)} - 1\right) \xrightarrow{d} N(0,1),$$

*where $K_2 = \int_{\mathbb{R}} K^2(u)\,du$.*

https://doi.org/10.1017/S0266466603195035 Published online by Cambridge University Press

Theorem 5 shows that our asymptotic results are essentially equal to those of Florens-Zmirou (1993) and Jiang and Knight (1997). In fact, we get $S_n(x) - V_n(x) \xrightarrow{p} 0$, and, if we use the uniform kernel ($K(u) = 1/2$ if $|u| < 1$ and $K(u) = 0$ if $|u| \geq 1$), as in Florens-Zmirou (1993), we conclude that, by the asymptotic equivalence theorem, $\sqrt{(nh_n/2)B_n}\big([S_n(x)/b^2(x)] - 1\big)$ and $\sqrt{(nh_n/2)B_n}\big([V_n(x)/b^2(x)] - 1\big)$ have the same asymptotic distribution.[3] Thus, in asymptotic framework with $\Delta_n \to 0$, $V_n$ and $S_n$ are equivalent estimators. Our claim, however, is that our estimator is preferable when $\Delta_n = \Delta$ is constant, as we have pointed out.

We observe that in Theorem 5 it is not necessary that the observation period go to infinity, i.e., $t_n$ can be constant[4] as $n \to +\infty$ (in this case, the condition $\sqrt{nh_n}\Delta_n \to 0$ is redundant). In fact, it is known that every finite interval of continuous observations contains all the information needed about the diffusion coefficient $b^2(x)$ to estimate it with probability one, provided that $x$ is visited by $X$ (see Brown and Hewitt, 1975; Bandi and Phillips, 2003).

The next theorem analyzes the asymptotic behavior of the $V_n^*$ estimator.

THEOREM 6. *Let the assumptions of the previous theorem hold. In addition suppose that $t_n h_n \to +\infty$ (thereby $t_n \to +\infty$).*[5] *Then*

*(a)*

$$V_n^*(x) \xrightarrow{p} b^2(x)$$

*and*

*(b)*

$$\sqrt{\frac{nh_n}{2K_2} B_n}\left(\frac{V_n^*(x)}{b^2(x)} - 1\right) \xrightarrow{d} N(0,1),$$

*where* $K_2 = \int_{\mathbb{R}} K^2(u)\, du$.

In Theorem 6 one has the additional condition $t_n h_n \to +\infty$. Because $h_n \to 0$ must hold, we have to impose $t_n \to +\infty$. This is an obvious consequence of the fact that we use the $T_n$ drift estimator in the expression of $V_n^*$ and direct identification of the drift coefficient based on this kernel estimator is impossible over a fixed sampling period. Several authors (e.g., Bandi and Phillips, 2003; Phillips and Yu, 2000; Jiang and Knight, 1997) have stressed that the estimation of the drift is usually empirically and theoretically more difficult than the estimation of the diffusion coefficient. Therefore, there is no benefit in using the $V_n^*$ estimator (or even the $V_n$ estimator) when the step of discretization tend to zero. In fact, $S_n$ and $V_n^*$ are asymptotically equivalent estimators. Furthermore, unlike the $S_n$ estimator, $V_n^*$ requires that the time span tend to infinity. Nevertheless, when $\Delta_n = \Delta$ is fixed, the estimator $V_n^*$ is preferable because it reduces the bias of the $S_n$ estimator, as we have stressed.

## 4. EVALUATION OF THE BIAS USING A PARAMETRIC INTEREST RATE MODEL

In this section, we compare the bias of the $S_n$ and $V_n^*$ estimators from some specified (parametric) models in the framework $n \to +\infty$. Given $a(x)$ and $b^2(x)$ it is easy, from equations (3) and (6), to evaluate the bias of the $S_n$ and $V_n^*$ estimators as $n \to +\infty$ for a fixed step of discretization $\Delta$ (thereby $t_n = n\Delta \to +\infty$). In fact, assuming that $O(\Delta^2)$ is negligible, the biases are given by the expressions

$$Bias_1(x) = S(x) - b^2(x) = a^2(x)\Delta + f(x)\Delta,$$

$$Bias_2(x) = V^*(x) - b^2(x) = f(x)\Delta,$$

where $f(x) = b^2(x)a'(x) + a(x)b(x)b'(x) + \frac{1}{2}b^2(x)(b'(x))^2 + \frac{1}{2}b^3(x)b''(x)$ (note that we suppress the subscript $n$ associated with the estimators because we are dealing with the case $n \to +\infty$). It follows that $S(x) = b^2(x) + a^2(x)\Delta + f(x)\Delta$ and $V^*(x) = b^2(x) + f(x)\Delta$ (assuming $n \to +\infty$). We start by selecting the coefficients $a(x)$ and $b^2(x)$ based on some estimated models by Aït-Sahalia (1999) using monthly Fed funds data from January 1963 through December 1998. From Table VI of Aït-Sahalia (1999) we select two models:

$$dX_t = \underset{(0.12)}{0.261}(\underset{(0.014)}{0.0717} - X_t)dt + \underset{(0.00078)}{0.02237}dW_t,$$

$$dX_t = \underset{(0.10)}{0.219}(\underset{(0.016)}{0.0721} - X_t)dt + \underset{(0.0023)}{0.06665}\sqrt{X_t}dW_t,$$

where the parameters (expressed in annualized form) were estimated by the maximum likelihood method and the asymptotic standard errors are shown below the parameters. From these models we evaluate the functions $Bias_1(x)$, $Bias_2(x)$, $S(x) = b^2(x) + a^2(x)\Delta + f(x)\Delta$, and $V^*(x) = b^2(x) + f(x)\Delta$ running $x$ (interest rate expressed in annualized form) in the interval $]0,0.22]$ (the maximum value observed in the period was 0.2236). The constant $\Delta$ is equal to $\frac{1}{12}$. Before we present the results, it is worth mentioning that this simulation seemingly does not best serve our purposes because, in the interest rate estimation framework, the drift coefficient is numerically close to zero. As we have pointed out, this is the case where $V_n^*$ and $S_n$ present similar results. However, we will see, even in this situation, that the $V_n^*$ estimator seems to be better. Then to appreciate how $V_n^*$ can improve the $S_n$ estimates, we slightly increase the parameters associated with the reversion effects.

From Figures 1 and 2 one can conclude that in low/moderate values of interest rate both nonparametric estimators produce similar and good estimates whereas in high/very high values, the $S_n$ estimates are generally worse than the $V_n^*$ estimates. We give a short explanation for these results. Coherently with most empirical studies of interest rate series, we observe the following pattern.

https://doi.org/10.1017/S0266466603195035 Published online by Cambridge University Press

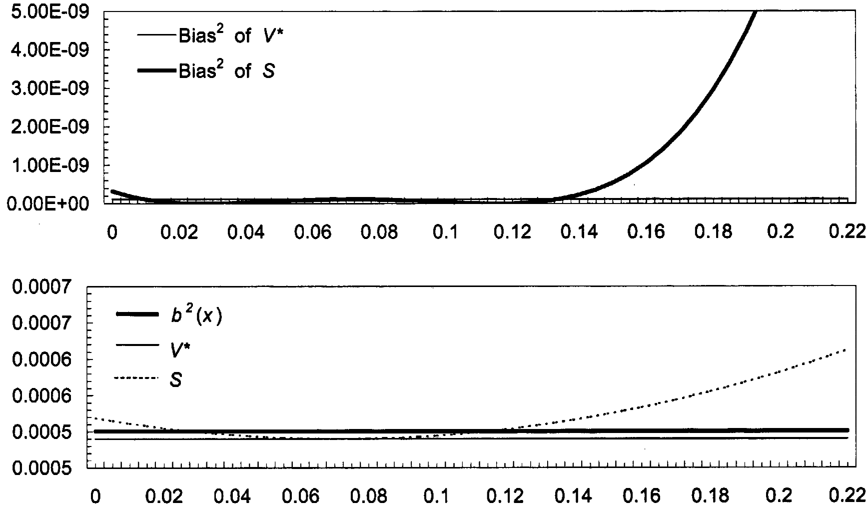


**FIGURE 1.** Behavior of the $S$ and $V^*$ estimators (as $n \to +\infty$) when $\Delta = \frac{1}{12}$ and the model is $dX_t = 0.261(0.0717 - X_t)\,dt + 0.02237 dW_t$.

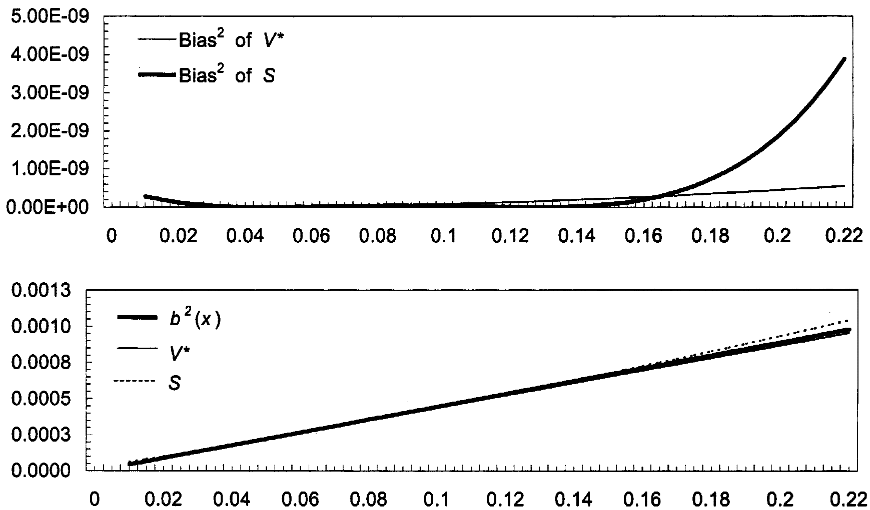


**FIGURE 2.** Behavior of the $S$ and $V^*$ estimators (as $n \to +\infty$) when $\Delta = \frac{1}{12}$ and the model is $dX_t = 0.219(0.0721 - X_t)\,dt + 0.06665\sqrt{X_t} dW_t$.

https://doi.org/10.1017/S0266466603195035 Published online by Cambridge University Press

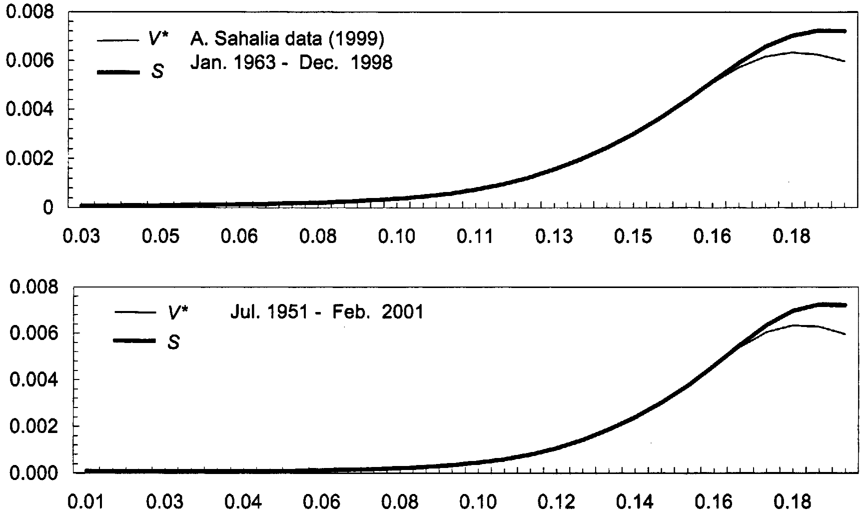


**FIGURE 3.** Nonparametric diffusion estimation—Fed fund data.

In the interval of moderate/low values not only is the volatility very low but also the process displays martingale behavior, in the sense that $a(x) \approx 0$ for $x$ belonging to the interval of moderate/low values. In this interval, the asymptotic bias of $S_n$ (and also of $V_n^*$) is in general very low. By contrast, in the interval of high/very high values of interest rates, the volatility increases significantly and the process exhibits stronger reversion effects. It is in this interval, where $|a(x)|$ is higher, that $S_n$ produces greater bias and, simultaneously, that $V_n^*$ shows significant improvement over $S_n$.

It is now interesting to compare the $S_n$ and $V_n^*$ estimates for the diffusion coefficient using the Fed fund data referred to previously.

From Figure 3 it seems that the volatility at high values of interest rate, although considerable, is not as great as the $S_n$ estimator suggests. However, no definitive conclusions can be made because in the areas where the drift coefficient is substantially negative (high in absolute value) there are not sufficient observations to draw precise statistical inference (for a discussion on this topic, see Bandi, 2002).

Let us consider a slight increase in the parameters of the preceding models associated with the reversion effects (everything else remains unchanged):

$$dX_t = 0.9(0.0717 - X_t)dt + 0.02237 dW_t,$$

$$dX_t = 0.9(0.0721 - X_t)dt + 0.06665\sqrt{X_t}dW_t$$

(in both cases, the parameter $\beta$ defined in the drift function $a(x) = \beta(\mu - x)$ has changed to 0.9). Although these models certainly no longer reflect the be-

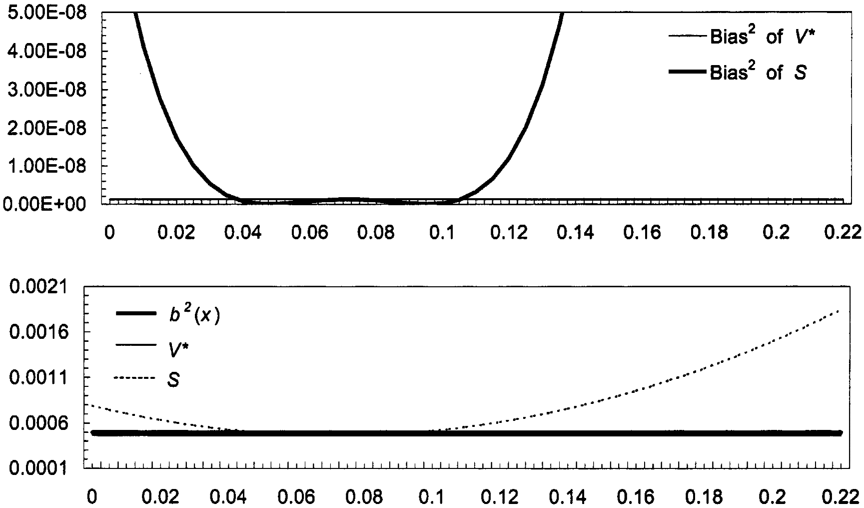


**FIGURE 4.** Behavior of the $S$ and $V^*$ estimators (as $n \to +\infty$) when $\Delta = \frac{1}{12}$ and the model is $dX_t = 0.9(0.0717 - X_t)dt + 0.02237dW_t$.

havior of the interest rate process, they allow us to appreciate the advantages of the proposed estimator when stronger reversion effects characterize the dynamics of the conditional mean. In fact, as can be seen from Figures 4 and 5, greater improvement is achieved by the $V_n^*$ estimates over the $S_n$ estimates.

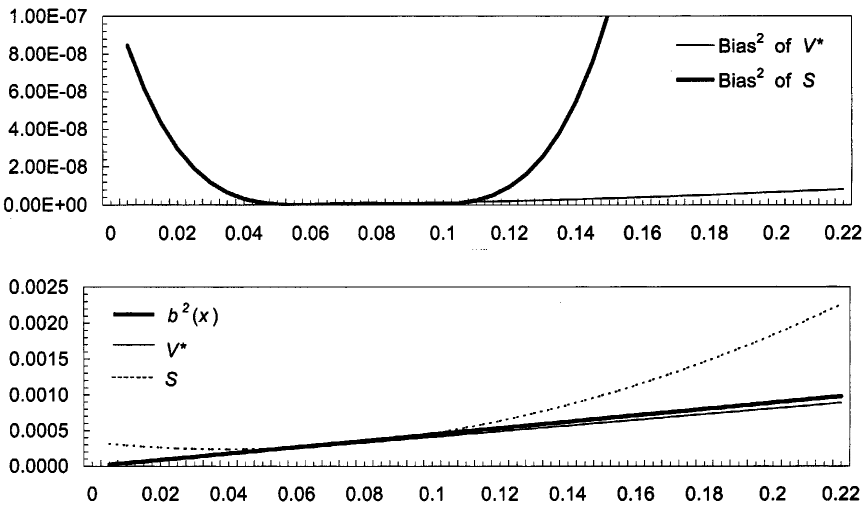


**FIGURE 5.** Behavior of the $S$ and $V^*$ estimators (as $n \to +\infty$) when $\Delta = \frac{1}{12}$ and the model is $dX_t = 0.9(0.0721 - X_t)\,dt + 0.06665\sqrt{X_t}dW_t$.

https://doi.org/10.1017/S0266466603195035 Published online by Cambridge University Press

## 5. CONCLUSIONS AND OTHER EXTENSIONS

In this paper, we quantified the asymptotic bias of the Florens-Zmirou (1993) and Jiang and Knight (1997) estimator for the diffusion coefficient when the step of discretization is fixed, and subsequently we proposed a bias adjustment that partially compensates for the distortion. Also, we showed that our estimators have all the asymptotic properties of the Florens-Zmirou and Jiang and Knight estimator when the step of discretization goes to zero. There are some improvements that can be considered. First, more general assumptions on the data generator process as in Bandi and Phillips (2003) can in principle be applied. Second, other drift estimators can be considered, instead of $T_n$. For instance, the Jiang and Knight (1997) drift estimator can be used once one has a preliminary estimate of the diffusion estimator. Subsequently, an iterative procedure can be applied. I conjecture that the asymptotic properties of the Jiang and Knight drift estimator, with $\Delta$ fixed, can be improved if one uses the $V_n^*$ estimates instead of the $S_n$ estimates. Third, a more accurate estimator for $b^2(x)$ can, in principle, be designed through the identification of the $f(x)$ function using standard methods for derivative estimation, albeit at a natural cost of a reduction in the rates of convergence. Fourth, as in the case of the diffusion estimator, similar bias adjustments can be designed for the usual drift estimator $T_n$. Fifth, an evaluation of the bias produced by other nonparametric estimators (in the case $\Delta_n = \Delta$ constant) such as the local linear or local quadratic regression can be analyzed (see Fan and Yao, 1998).

### *NOTES*

1. I thank a referee for pointing this out.

2. Let us assume $t_n = O(n^\gamma)$, $h_n = O(n^\beta)$, and consequently $\Delta_n = t_n/n = O(n^{\gamma-1})$. To assure the conditions $\Delta_n \to 0$, $h_n \to 0$, $nh_n \to +\infty$, $nh_n^3 \to 0$, and $\sqrt{nh_n}\Delta_n \to 0$ as $n \to +\infty$ it is necessary and sufficient that $\beta$ and $\gamma$ be such that $-1 < \beta < -\frac{1}{3}$ and $0 \le \gamma < \frac{1}{2} - \beta/2$.

3. With uniform kernel, we have $K_2 = \frac{1}{2}$ and $(nh_n/2K_2)B_n = \sum_{i=0}^{n-1} K((x - X_{t_i})/h_n) = \sum_{i=0}^{n-1} \frac{1}{2}\mathcal{I}_{\{|x-X_{t_i}|<h_n\}} = N_x/2$; thus, $\sqrt{(nh_n/2)B_n}\left([S_n(x)/b^2(x)]-1\right) = \sqrt{N_x/2}\left([S_n(x)/b^2(x)]-1\right) \xrightarrow{d} N(0,1)$, as in Florens-Zmirou (1993), Theorem 1.

4. From note 2 we can conclude that $\gamma$ can be zero where $\gamma$ is defined in $t_n = O(n^\gamma)$. That is, consistency can be obtained with $t_n$ constant. In this case, the condition $\sqrt{nh_n}\Delta_n \to 0$ is redundant.

5. To assure the conditions $\Delta_n \to 0$, $h_n \to 0$, $nh_n \to +\infty$, $nh_n^3 \to 0$, $\sqrt{nh_n}\Delta_n \to 0$, and $t_n h_n \to +\infty$ as $n \to +\infty$ it is necessary and sufficient that $-1 < \beta < -\frac{1}{3}$, $-\beta < \gamma < \frac{1}{2} - \beta/2$ (see notations in note 2), which defines a nonempty convex open set in $\mathbb{R}^2$. Obviously, the condition $\gamma > -\beta$ means that the time span cannot be fixed, i.e., we must have $\gamma > 0$.

### *REFERENCES*

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

Aït-Sahalia, Y. (1999) Transition densities for interest rate and other nonlinear diffusions. *Journal of Finance* 54, 1361–1395.

Andersen, T., T. Bollerslev, F. Diebold, & P. Labys (2001) The distribution of realized exchange rate volatility. *Journal of the American Statistical Association* 96, 42–51.

Arnold, L. (1974) *Stochastic Differential Equations: Theory and Application*. New York: Wiley.

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

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

Brown, B. & J. Hewitt (1975) Asymptotic likelihood theory for diffusion processes. *Journal of Applied Probability* 12, 228–238.

Chen, X., L. Hansen, & M. Carrasco (1998) Nonlinearity and Temporal Dependence. Working paper.

Davidson, J. (1994) *Stochastic Limit Theory*. Oxford, UK: Oxford University Press.

Fan, J. & Q. Yao (1998) Efficient estimation of conditional variance functions in stochastic regression. *Biometrika* 85, 645–660.

Florens-Zmirou, D. (1989) Approximate discrete-time schemes for statistics of diffusion processes. *Statistics* 4, 547–557.

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

Hansen, L. & J. Scheinkman (1995) Back to the future: Generating moment implications for continuous-time Markov processes. *Econometrica* 63, 767–804.

Hansen, L., J. Scheinkman, & N. Touzi (1998) Spectral methods for identifying scalar diffusions. *Journal of Econometrics* 86, 1–32.

Ikeda, N. & S. Watanabe (1981) *Stochastic Differential Equations and Diffusion Processes*. North-Holland Mathematical Library. Amsterdam: North-Holland.

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.

Karlin, S. & H. Taylor (1981) *A Second Course in Stochastic Processes*. New York: Academic Press.

Merton, R. (1990) *Continuous-Time Finance*. Oxford, UK: Blackwell.

Phillips, P. & J. Yu (2000) Exact Gaussian Estimation of Continuous Time Models of the Term Structure of Interest Rates. Working paper, Yale University.

Rao, B. (1983) *Nonparametrical Functional Estimation*. New York: Academic Press.

Rao, B. (1999) *Statistical Inference for Diffusion Type Processes*. New York: Oxford University Press.

Sawyer, K. (1993) Continuous Time Financial Models: Statistical Applications of Stochastic Processes. In G. Maddala, C. Rao, H. Vinod (eds.), *Handbook of Statistics*, pp. 573–610. Amsterdam: Elsevier Science Publishers.

Skorokhod, A. (1989) *Asymptotic Methods in the Theory of Stochastic Differential Equations*, Translation of Mathematical Monographs 78. Providence, RI: American Mathematical Society.

Stanton, R. (1997) A nonparametric model of term structure dynamics and the market price of interest rate risk. *Journal of Finance* 52, 1973–2000.

# APPENDIX A: PRELIMINARY RESULTS

LEMMA 7 (Florens-Zmirou, 1989). *Assume that $f \in \mathcal{C}^{2(s+1)}$ and $a, b^2 \in \mathcal{C}^{2s}$ ($\mathcal{C}^k$ is the space of k times continuously differentiable functions). Then,*

$$E[f(X_{t_i})|X_{t_{i-1}}] = \sum_{j=0}^{s} L^j f(X_{t_{i-1}}) \frac{\Delta^j}{j!} + R,$$

https://doi.org/10.1017/S0266466603195035 Published online by Cambridge University Press

*where* $L = a(d/dx) + \frac{1}{2}b^2(d^2/dx^2)$ *and*

$$R = \int_0^{\Delta}\int_0^{u_1}\dots\int_0^{u_s} E[L^{s+1}f(X_{t_{i-1}+u_{s+1}})|X_{t_{i-1}}]\,du_1\dots du_{s+1}.$$

See Hansen and Scheinkman (1995) for additional discussion. We assume in all applications involving Lemma 7 that

$$E[|L^{s+1}f(X_{t_{i-1}+u_{s+1}})||X_{t_{i-1}}] < +\infty.$$

For example, if $|L^{s+1}f(x)| \le k_1(1 + x^{2r})$ holds for $k_1 > 0$ and if, for some integer $r > 0$, we have $E[X_t^{2r}|X_s] \le (1 + |X_s^{2r}|)e^{A(t-s)}$ for $A > 0$, then $E[|L^{s+1}f(X_{t_{i-1}+u_{s+1}})||X_{t_{i-1}}] < +\infty$ (these conditions are generally easy to satisfy). It is worth mentioning that $R$ depends on $X_{t_{i-1}}$ and $\Delta^{s+1}$. We denote $R$ by $R = O(\Delta^{s+1})$.

The next lemmas deal with a central limit theorem version for discrete time processes $\{X_{t_i}, i \ge 0\}$ extracted from the diffusion process $\{X_t; t \ge 0\}$.

LEMMA 8 (Florens-Zmirou, 1989). *Let A1–A4 hold. Let* $f: E \subset \mathbb{R} \to \mathbb{R}$ *such that* $E[f(X)] = 0$ *and* $E[f^2(X)] < +\infty$ *where the expected values are evaluated with respect to the invariant distribution* $P^0$. *Then* $(1/\sqrt{n})\sum_{i=1}^n f(X_{t_i}) \xrightarrow{d} N(0, V_1(f))$ *where* $V_1(f) = E[f^2(X)] + 2\sum_{j=1}^{+\infty} E[f(X_0)f(X_{t_j})]$.

(We note that our A4 condition implies the $H^*$ condition defined in Florens-Zmirou, 1989.)

LEMMA 9 (Florens-Zmirou, 1989). *Let A1–A4 hold. Let* $g: E^2 \subset \mathbb{R}^2 \to \mathbb{R}$ *such that* $E[g(X_{t_{i-1}}, X_{t_i})] = 0$ *and* $E[g^2(X_{t_{i-1}}, X_{t_i})] < +\infty$ *where the expected values are evaluated with respect to the invariant distribution* $P^0$. *Then* $(1/\sqrt{n})\sum_{i=1}^n g(X_{t_{i-1}}, X_{t_i}) \xrightarrow{d} N(0, V_2(g))$ *where* $V_2(g) = E[g^2(X_{t_{i-1}}, X_{t_i})] + 2\sum_{j=1}^{+\infty} E[g(X_0, X_{t_1})g(X_{t_j}, X_{t_{j+1}})]$.

Note that

$$E[g^2(X_{t_{i-1}}, X_{t_i})] = E[E[g^2(X_{t_{i-1}}, X_{t_i})|X_{t_{i-1}}]]$$
$$= \int\left(\int g^2(x,y)p(\Delta,x,y)\,dy\right)\bar{p}(x)\,dx,$$

where $y \mapsto p(\Delta, x, y)$ is the transition (or conditional) density of $X_{t+\Delta}$ given $X_t = x$ and $\bar{p}$ is the stationary density. The following result is well known (see, e.g., Rao, 1983).

LEMMA 10. *Let A1–A5 hold and* $h_n \to 0$ *and* $nh_n \to +\infty$ *as* $n \to +\infty$. *Then* $B_n(x) \xrightarrow{L^2} \bar{p}(x)$ *where* $B_n(x) = (1/nh_n)\sum_{i=0}^{n-1} K((x - X_{t_i})/h_n)$ ($\xrightarrow{L^2}$ *denotes convergence in* MSE).

We discuss subsequently two distinct cases involving the estimator $B_n$. They are as follows: $\Delta_n \to 0$ and $\Delta_n = \Delta$ constant as $n \to +\infty$. In both cases, $B_n(x)$ is consistent in quadratic mean for $\bar{p}(x)$ provided that the other conditions in Lemma 10 are satisfied. We stress that the convergence of the density estimator in the presence of a constant $\Delta$ requires an increasing span $t_n$.

# APPENDIX B: PROOFS

The arguments we use to prove theorems in this Appendix differ from those of Florens-Zmirou (1993) and Jiang and Knight (1997) in that we do not consider an expansion of transition density.

**Proof of Theorem 1.** Let $S_n(x) = A_n(x)/B_n(x)$ where

$$A_n(x) = \frac{1}{nh_n}\sum_{i=0}^{n-1} K\left(\frac{x - X_{t_i}}{h_n}\right)\frac{(X_{t_{i+1}} - X_{t_i})^2}{\Delta_n}.$$

Given Lemma 10, it is enough to prove that $A_n(x) \xrightarrow{L^2} (b^2(x) + a^2(x)\Delta + f(x)\Delta + O(\Delta^2))\bar{p}(x)$. To simplify we write $A_n$ instead of $A_n(x)$. We have (with $\Delta_n = \Delta$)

$$\begin{aligned} E[A_n] &= \frac{1}{h_n}\int_{\mathbb{R}}\int_{\mathbb{R}} K\left(\frac{x-z}{h_n}\right)\frac{(y-z)^2}{\Delta}\, p(\Delta, z, y)\bar{p}(z)\, dydz \\ &= \frac{1}{h_n}\int_{\mathbb{R}} K\left(\frac{x-z}{h_n}\right) E\left[\frac{(X_\Delta - X_0)^2}{\Delta}\,\middle|\, X_0 = z\right]\bar{p}(z)\, dz \\ &= \int K(u) E\left[\frac{(X_\Delta - X_0)^2}{\Delta}\,\middle|\, X_0 = x - h_n u\right]\bar{p}(x - h_n u)\, du \end{aligned} \tag{B.1}$$

(with the change of variable $z = x - h_n u$). Using Lemma 7 (with $\Delta_n = \Delta$)

$$\begin{aligned} E&\left[\frac{(X_\Delta - X_0)^2}{\Delta}\,\middle|\, X_0\right] \\ &= b^2(X_0) + [a^2(X_0) + b^2(X_0)a'(X_0) + a(X_0)b(X_0)b'(X_0) \\ &\qquad + \tfrac{1}{2}b^2(X_0)(b'(X_0))^2 + \tfrac{1}{2}b^3(X_0)b''(X_0)]\Delta + O(\Delta^2) \\ &= b^2(X_0) + a^2(X_0)\Delta + f(X_0)\Delta + O(\Delta^2) \end{aligned}$$

equation (B.1) can be written as

$$\begin{aligned} E[A_n] = \int K(u)(b^2(x - h_n u) + a^2(x - h_n u)\Delta \\ + f(x - h_n u)\Delta + O(\Delta^2))\bar{p}(x - h_n u)\, du. \end{aligned}$$

Considering $h_n \to 0$, $\int K(u)\, du = 1$, and the fact that $a$, $b$, $f$, and $\bar{p}$ are continuous functions (these sort of arguments will be used throughout the proofs), one has

$$\begin{aligned} \lim E[A_n] &= (b^2(x) + a^2(x)\Delta + f(x)\Delta + O(\Delta^2))\bar{p}(x)\int K(u)\, du \\ &= (b^2(x) + a^2(x)\Delta + f(x)\Delta + O(\Delta^2))\bar{p}(x). \end{aligned}$$

https://doi.org/10.1017/S0266466603195035 Published online by Cambridge University Press

It is now necessary to prove $\text{Var}[A_n] \to 0$. We have

$$\begin{aligned}\text{Var}[A_n] &= \text{Var}\left[\frac{1}{nh_n}\sum_{i=0}^{n-1} K\left(\frac{x - X_{t_i}}{h_n}\right)\frac{(X_{t_{i+1}} - X_{t_i})^2}{\Delta}\right] \\ &= \frac{1}{nh_n}\text{Var}\left[\frac{1}{\sqrt{n}}\sum_{i=0}^{n-1}\frac{1}{\sqrt{h_n}} K\left(\frac{x - X_{t_i}}{h_n}\right)\frac{(X_{t_{i+1}} - X_{t_i})^2}{\Delta}\right] \\ &= \frac{1}{nh_n}\text{Var}\left[\frac{1}{\sqrt{n}}\sum_{i=0}^{n-1} f(X_{t_i}, X_{t_{i+1}})\right] \\ &= \frac{1}{nh_n}\text{Var}\left[\frac{1}{\sqrt{n}}\sum_{i=0}^{n-1} f^*(X_{t_i}, X_{t_{i+1}})\right],\end{aligned}$$

where $f^*(X_{t_i}, X_{t_{i+1}}) = f(X_{t_i}, X_{t_{i+1}}) - E[f(X_{t_i}, X_{t_{i+1}})]$ and

$$f(X_{t_i}, X_{t_{i+1}}) = \frac{1}{\sqrt{h_n}} K\left(\frac{x - X_{t_i}}{h_n}\right)\frac{(X_{t_{i+1}} - X_{t_i})^2}{\Delta}.$$

Note that $E[(f^*(X_{t_i}, X_{t_{i+1}}))^2] = E[f^2(X_{t_i}, X_{t_{i+1}})] - (E[f(X_{t_i}, X_{t_{i+1}})])^2$ and

$$\begin{aligned}E[f^2(X_{t_i}, X_{t_{i+1}})] &= \frac{1}{h_n} E\left[K^2\left(\frac{x - X_{t_i}}{h_n}\right)\frac{(X_{t_{i+1}} - X_{t_i})^4}{\Delta^2}\right] \\ &= \frac{1}{h_n} E\left[K^2\left(\frac{x - X_{t_i}}{h_n}\right) E\left[\frac{(X_{t_{i+1}} - X_{t_i})^4}{\Delta^2} \,|X_{t_i}\right]\right].\end{aligned}$$

By Lemma 7 it is possible to check that

$$E\left[\frac{(X_{t_{i+1}} - X_{t_i})^4}{\Delta^2}\,\middle|\, X_{t_i}\right] = 3b^4(X_{t_i}) + O(\Delta).$$

In this way,

$$\begin{aligned}E[f^2(X_{t_i}, X_{t_{i+1}})] &= \frac{1}{h_n}\int K^2\left(\frac{x - z}{h_n}\right)(3b^4(z) + O(\Delta))\bar{p}(z)\,dz \\ &= \int K^2(u)(3b^4(x - h_n u) + O(\Delta))\bar{p}(x - h_n u)\,du \\ &= (3b^4(x) + O(\Delta))\bar{p}(x)\int K^2(u)\,du.\end{aligned}$$

This expression is clearly finite. On the other hand,

$$\begin{aligned}E[f(X_{t_i}, X_{t_{i+1}})] &= E\left[\frac{1}{\sqrt{h_n}} K\left(\frac{x - X_{t_i}}{h_n}\right)\frac{(X_{t_{i+1}} - X_{t_i})^2}{\Delta}\right] \\ &= \sqrt{h_n} E[A_n]\end{aligned}$$

is finite and tends to zero.

https://doi.org/10.1017/S0266466603195035 Published online by Cambridge University Press

In conclusion, $E[(f^*(X_{t_i}, X_{t_{i+1}}))^2]$ is finite because $E[f^2(X_{t_i}, X_{t_{i+1}})]$ is finite and $E[f(X_{t_i}, X_{t_{i+1}})]$ tends to zero. Then, by Lemma 9, $\text{Var}[1/\sqrt{n}\sum_{i=0}^{n-1} f^*(X_{t_i}, X_{t_{i+1}})] \to V_2(f^*) < +\infty$, and so $\text{Var}[A_n] = (1/nh_n)(V_2(f^*) + o(1)) \to 0$. Therefore, $A_n \xrightarrow{p} (b^2(x) + a^2(x)\Delta + f(x)\Delta + O(\Delta^2))\bar{p}(x)$ and $A_n/B_n \xrightarrow{p} (b^2(x) + a^2(x)\Delta + f(x)\Delta + O(\Delta^2))$ because $B_n \xrightarrow{p} \bar{p}(x) > 0$ by Lemma 10. ■

The proofs of Theorems 2 and 3 are similar to the proof of Theorem 1 and thus are omitted.

**Proof of Theorem 4.** Let $V_n^*(x) = A_n^*(x)/B_n(x)$ where

$$A_n^*(x) = \frac{1}{nh_n}\sum_{i=0}^{n-1} K\left(\frac{x - X_{t_i}}{h_n}\right)\frac{(X_{t_{i+1}} - X_{t_i} - T(X_{t_i})\Delta_n)^2}{\Delta_n}.$$

To simplify we write $A_n^*$ instead of $A_n^*(x)$. We first note that (with $\Delta_n = \Delta$)

$$\begin{aligned}
A_n^* &= \frac{1}{nh_n}\sum_{i=0}^{n-1} K\left(\frac{x - X_{t_i}}{h_n}\right)\frac{(X_{t_{i+1}} - X_{t_i} - T_n(X_{t_i})\Delta)^2}{\Delta}\\
&= \frac{1}{nh_n}\sum_{i=0}^{n-1} K\left(\frac{x - X_{t_i}}{h_n}\right)\frac{((X_{t_{i+1}} - X_{t_i} - a(X_{t_i})\Delta) - (T_n(X_{t_i}) - a(X_{t_i}))\Delta)^2}{\Delta}\\
&= \frac{1}{nh_n}\sum_{i=0}^{n-1} K\left(\frac{x - X_{t_i}}{h_n}\right)\frac{(X_{t_{i+1}} - X_{t_i} - a(X_{t_i})\Delta)^2}{\Delta}\\
&\quad - \frac{2}{nh_n}\sum_{i=0}^{n-1} K\left(\frac{x - X_{t_i}}{h_n}\right)\frac{(X_{t_{i+1}} - X_{t_i} - a(X_{t_i})\Delta)(T_n(X_{t_i}) - a(X_{t_i}))\Delta}{\Delta}\\
&\quad + \frac{1}{nh_n}\sum_{i=0}^{n-1} K\left(\frac{x - X_{t_i}}{h_n}\right)(T_n(X_{t_i}) - a(X_{t_i}))^2\Delta\\
&= A_n^{\#} + D_{1,n} + D_{2,n}.
\end{aligned}$$

To simplify we write $D_{1,n}$ and $D_{2,n}$ instead of $D_{1,n}(x)$ and $D_{2,n}(x)$, respectively. We now need the following result: $T_n(X_t) \xrightarrow{p} a(X_t) + O(\Delta)$. Pointwise convergence, $T_n(x) \xrightarrow{p} a(x) + O(\Delta)$ (Theorem 3), does not imply $T_n(X_t) \xrightarrow{p} a(X_t) + O(\Delta)$ (see Davidson, 1994, Ch. 21). From Theorem 21.6 of Davidson (1994), suitably adapted to our problem, if $T_n(x)$ converges in probability to $a(x) + O(\Delta)$ uniformly on an open set containing all the possible values of $X_t$, then it follows that $T_n(X_t) \xrightarrow{p} a(X_t) + O(\Delta)$. In principle, the uniform convergence must hold in $\mathbb{R}$ although, in practice, because $X$ is stationary, it is enough to take an open set $V$ contained in some compact set $\bar{V}$ such that $P(X \in V) = 1$. Following Davidson (1994), Theorem 21.9, $T_n(x)$ converges in probability to $a(x)$ uniformly on a set $V$ if and only if $T_n(x) \xrightarrow{p} a(x)$ for each $x \in V$ and $T_n(x)$ is stochastically equicontinuous. To guarantee this last condition, it is sufficient that $\sup_{x \in V}|dT_n(x)/dx|$ be bounded in probability (Davidson, 1994, Theorem 21.10). Under the conditions established (note that we imposed that the kernel function and the infinitesimal coefficients are continuously differentiable) it follows that $T_n(X_t) \xrightarrow{p} a(X_t) + O(\Delta)$, i.e., $T_n(X_t) = a(X_t) + O(\Delta) + \varepsilon_n$, where $\varepsilon_n = o_p(1)$. Thus,

$$D_{1,n} = -\frac{2}{nh_n}\sum_{i=0}^{n-1} K\left(\frac{x - X_{t_i}}{h_n}\right)\frac{(X_{t_{i+1}} - X_{t_i} - a(X_{t_i})\Delta)(T_n(X_{t_i}) - a(X_{t_i}))\Delta}{\Delta}$$

$$= -\frac{2}{nh_n}\sum_{i=0}^{n-1} K\left(\frac{x - X_{t_i}}{h_n}\right)(X_{t_{i+1}} - X_{t_i} - a(X_{t_i})\Delta)(O(\Delta) + \varepsilon_n). \quad \textbf{(B.2)}$$

We have (with $\Delta_n = \Delta$)

$$E[D_{1,n}] = \frac{1}{h_n}\int_{\mathbb{R}}\int_{\mathbb{R}} K\left(\frac{x - z}{h_n}\right)(y - z - a(z)\Delta)(O(\Delta) + \varepsilon_n)p(\Delta, z, y)\bar{p}(z)\,dydz$$

$$= \frac{1}{h_n}\int_{\mathbb{R}} K\left(\frac{x - z}{h_n}\right)E[(X_\Delta - X_0 - a(X_0)\Delta)(O(\Delta) + \varepsilon_n)|X_0 = z]\bar{p}(z)\,dz$$

$$= \int_{\mathbb{R}} K(u)E[(X_\Delta - X_0 - a(X_0)\Delta)(O(\Delta) + \varepsilon_n)|X_0 = x - h_n u]\bar{p}(x - h_n u)\,du.$$

Using Lemma 7 (with $\Delta_n = \Delta$),

$$E[(X_\Delta - X_0 - a(X_0)\Delta)(O(\Delta) + \varepsilon_n)|X_0] = O(\Delta^3),$$

and the hypothesis $h_n \to 0$, one has $\lim E[D_{1,n}] = O(\Delta^3)$. To prove $\mathrm{Var}[D_{1,n}] \to 0$ we follow the same arguments we used in previous proofs; i.e., we define

$$\mathrm{Var}[D_{1,n}] = \frac{4}{nh_n}\mathrm{Var}\left[\frac{1}{\sqrt{n}}\sum_{i=0}^{n-1} f^*(X_{t_i}, X_{t_{i+1}})\right],$$

where $f^*(X_{t_i}, X_{t_{i+1}}) = f(X_{t_i}, X_{t_{i+1}}) - E[f(X_{t_i}, X_{t_{i+1}})]$ and

$$f(X_{t_i}, X_{t_{i+1}}) = \frac{1}{\sqrt{h_n}}K\left(\frac{x - X_{t_i}}{h_n}\right)(X_{t_{i+1}} - X_{t_i} - a(X_{t_i})\Delta)(O(\Delta) + \varepsilon_n).$$

Using Lemma 7 one concludes

$$E[f^2(X_{t_i}, X_{t_{i+1}})] = O(\Delta^3)\bar{p}(x)\int K^2(u)\,du,$$

$$E[f(X_{t_i}, X_{t_{i+1}})] = \sqrt{h_n}E[D_{1,n}],$$

and using Lemma 9 and Assumption A5 we have $\lim \mathrm{Var}[D_{1,n}] = 0$ and thus $D_{1,n}(x) \xrightarrow{p} O(\Delta^3)$.

Also, it is easy to see that

$$D_{2,n} = \frac{1}{nh_n}\sum_{i=0}^{n-1} K\left(\frac{x - X_{t_i}}{h_n}\right)(T_n(X_{t_i}) - a(X_{t_i}))^2\Delta$$

$$= \frac{1}{nh_n}\sum_{i=0}^{n-1} K\left(\frac{x - X_{t_i}}{h_n}\right)(\varepsilon_n + O(\Delta))^2\Delta$$

https://doi.org/10.1017/S0266466603195035 Published online by Cambridge University Press

converges in probability to $O(\Delta^3)$ (note that by Slutsky's theorem $(\varepsilon_n)^2 \xrightarrow{p} 0$). It follows from $D_{1,n}(x) \xrightarrow{p} O(\Delta^3)$, $D_{2,n}(x) \xrightarrow{p} O(\Delta^3)$, Theorem 2, and Lemma 10 that $A_n^*(x)/B(x) \xrightarrow{p} b^2(x) + f(x)\Delta + O(\Delta^2)$. ∎

**Proof of Theorem 5.**

(a) Let $V_n(x) = A_n^\#(x)/B_n(x)$ where

$$A_n^\#(x) = \frac{1}{nh_n}\sum_{i=0}^{n-1} K\left(\frac{x - X_{t_i}}{h_n}\right)\frac{(X_{t_{i+1}} - X_{t_i} - a(X_{t_i})\Delta_n)^2}{\Delta_n}.$$

Given Lemma 10, it is enough to prove that $A_n^\#(x) \xrightarrow{L^2} b^2(x)\bar{p}(x)$. From the proof of Theorem 1 it is straightforward to verify that

$$E[A_n^\#(x)] \to b^2(x)\bar{p}(x), \qquad \mathrm{Var}[A_n^\#(x)] \to 0$$

as $\Delta_n \to 0$, $h_n \to 0$, and $nh_n \to +\infty$. Therefore, $A_n^\#(x)/B_n(x) \xrightarrow{p} b^2(x)$. Let us see the second part of the theorem.

(b) To simplify we use the notation $E_i[\cdot] = E[\cdot|X_{t_i}]$, and we write $\sum$ instead of $\sum_{i=0}^{n-1}$. We note that

$$\sqrt{\frac{nh_n}{2K_2}B_n(x)}\left(\frac{V_n(x)}{b^2(x)} - 1\right) = \frac{\frac{1}{\sqrt{n}}\sum g(X_{t_i}, X_{t_{i+1}})}{b^2(x)\sqrt{B_n K_2}},$$

where

$$g(X_{t_i}, X_{t_{i+1}}) = \frac{1}{\sqrt{2h_n}} K\left(\frac{x - X_{t_i}}{h_n}\right)\eta_{i+1},$$

$$\eta_{i+1} = \Delta_n^{-1}(X_{t_{i+1}} - X_{t_i} - a(X_{t_i})\Delta_n)^2 - b^2(x).$$

First, we show the following:

(b1) $\lim E[(1/\sqrt{n})\sum g(X_{t_i}, X_{t_{i+1}})] = 0$;
(b2) $\lim E[g^2(X_{t_0}, X_{t_1})] = b^4(x)\bar{p}(x)K_2, \qquad (K_2 = \int K^2(u)\,du)$;
(b3) $\lim E[g(X_{t_0}, X_{t_1})g(X_{t_j}, X_{t_{j+1}})] = 0, \quad$ for $j \geq 1$.

In effect,

(b1) applying Lemma 7 we have

$$E_i[\eta_{i+1}] = E_i[\Delta_n^{-1}(X_{t_{i+1}} - X_{t_i} - a(X_{t_i})\Delta_n)^2 - b^2(x)]$$
$$= b^2(X_{t_i}) - b^2(x) + O(\Delta_n).$$

Thus,

$$E_i[g(X_{t_i}, X_{t_{i+1}})] = \frac{1}{\sqrt{2h_n}} K\left(\frac{x - X_{t_i}}{h_n}\right)(b^2(X_{t_i}) - b^2(x) + O(\Delta_n))$$

https://doi.org/10.1017/S0266466603195035 Published online by Cambridge University Press

and

$$E[g(X_{t_i},X_{t_{i+1}})] = E[E_i[g(X_{t_i},X_{t_{i+1}})]]$$

$$= \frac{1}{\sqrt{2h_n}}\int K\left(\frac{x-z}{h_n}\right)(b^2(z)-b^2(x)+O(\Delta_n))\bar{p}(z)\,dz$$

$$= \frac{h_n}{\sqrt{2h_n}}\int K(u)(b^2(x-h_nu)-b^2(x)+O(\Delta_n))\bar{p}(x-h_nu)\,du$$

$$= \frac{\sqrt{h_n}}{\sqrt{2}}(h_nO(1)+O(\Delta_n))$$

(using the change of variable, $z = x - h_n u$, and Taylor's formula, $b^2(x-h_nu)-b^2(x) = O(h_n)(b^2(x-\theta h_n u))'$, with $0 < \theta < 1$). Therefore,

$$\lim E\left[\frac{1}{\sqrt{n}}\sum g(X_{t_i},X_{t_{i+1}})\right] = \lim\sqrt{n}\left(\frac{\sqrt{h_n}}{\sqrt{2}}(h_nO(1)+O(\Delta_n))\right)$$

$$= \lim(\sqrt{nh_n^3}+\sqrt{nh_n}\Delta_n)O(1)$$

$$= 0$$

under the conditions of the theorem.

(b2) By Lemma 7,

$$E_i[\eta_{i+1}^2] = E_i[(\Delta_n^{-1}(X_{t_{i+1}}-X_{t_i}-a(X_{t_i})\Delta_n)^2-b^2(x))^2]$$

$$= b^4(x)-2b^2(x)b^2(X_{t_i})+3b^4(X_{t_i})+O(\Delta_n).$$

Thus

$$E_i[g^2(X_{t_i},X_{t_{i+1}})] = \frac{1}{2h_n}K^2\left(\frac{x-X_{t_i}}{h_n}\right)(b^4(x)-2b^2(x)b^2(X_{t_i})+3b^4(X_{t_i})+O(\Delta_n))$$

and

$$E[g^2(X_{t_i},X_{t_{i+1}})]$$

$$= E[E_i[g^2(X_{t_i},X_{t_{i+1}})]]$$

$$= \frac{1}{2h_n}\int\left(K^2\left(\frac{x-z}{h_n}\right)(b^4(x)-2b^2(x)b^2(z)+3b^4(z)+O(\Delta_n))\right)\bar{p}(z)\,dz$$

$$= \frac{1}{2}\int(K^2(u)(b^4(x)-2b^2(x)b^2(x-h_nu)$$

$$+3b^4(x-h_nu)+O(\Delta_n)))\bar{p}(x-h_nu)\,du.$$

With $n \to +\infty$, $h_n \to 0$, and $\Delta_n \to 0$ we have

$$\lim E[g^2(X_{t_i}, X_{t_{i+1}})]$$
$$= \frac{1}{2}\int \lim K^2(u)[b^4(x) - 2b^2(x)b^2(x - h_n u) + 3b^4(x - h_n u) + O(\Delta_n)]$$
$$\times \bar{p}(x - h_n u)\, du$$
$$= b^4(x)\bar{p}(x)\int K^2(u)\, du$$
$$= b^4(x)\bar{p}(x)K_2, \qquad \left(K_2 = \int K^2(u)\, du < +\infty\right).$$

(b3) We first show $\lim E[g(X_{t_0}, X_{t_1})g(X_{t_1}, X_{t_2})] = 0$. Given

$$E[g(X_{t_0}, X_{t_1})g(X_{t_1}, X_{t_2})] = E[E_0[E_1[g(X_{t_0}, X_{t_1})g(X_{t_1}, X_{t_2})]]]$$

we have by Lemma 7

$$E_1[g(X_{t_0}, X_{t_1})g(X_{t_1}, X_{t_2})]$$
$$= \frac{1}{2h_n} K\left(\frac{x - X_{t_0}}{h_n}\right) K\left(\frac{x - X_{t_1}}{h_n}\right) \eta_1 E_1[\eta_2]$$
$$= \frac{1}{2h_n} K\left(\frac{x - X_{t_0}}{h_n}\right) K\left(\frac{x - X_{t_1}}{h_n}\right) \eta_1 (b^2(X_{t_1}) - b^2(x) + O(\Delta_n)),$$

$$E_0[E_1[g(X_{t_0}, X_{t_1})g(X_{t_1}, X_{t_2})]]$$
$$= \frac{1}{2h_n} K\left(\frac{x - X_{t_0}}{h_n}\right) E_0\left[K\left(\frac{x - X_{t_1}}{h_n}\right) \eta_1 (b^2(X_{t_1}) - b^2(x) + O(\Delta_n))\right]$$
$$= \frac{1}{2h_n} K^2\left(\frac{x - X_{t_0}}{h_n}\right) \{(b^2(X_{t_0}) - b^2(x))^2 + O(\Delta_n)\},$$

and, finally,

$$E[g(X_{t_0}, X_{t_1})g(X_{t_1}, X_{t_2})]$$
$$= \int \frac{1}{2h_n} K^2\left(\frac{x - z}{h_n}\right) \{(b^2(z) - b^2(x))^2 + O(\Delta_n)\}\bar{p}(z)\, dz$$
$$= \frac{1}{2}\int K^2(u)\{(b^2(x - h_n u) - b^2(x))^2 + O(\Delta_n)\}\bar{p}(x - h_n u)\, dz$$
$$= (h_n^2 + \Delta_n)O(1).$$

Hence, with $h_n \to 0$ and $\Delta_n \to 0$ (as $n \to +\infty$), we get

$$\lim E[g(X_{t_0}, X_{t_1})g(X_{t_1}, X_{t_2})] = 0. \qquad \textbf{(B.3)}$$

https://doi.org/10.1017/S0266466603195035 Published online by Cambridge University Press

We now show $\lim E[g(X_{t_0}, X_{t_1}) g(X_{t_j}, X_{t_{j+1}})] = 0$ for $j \geq 1$. By Assumption A4, the $j$th autocovariance of any function (square $P^0$-integrable) tends to zero at exponential rate as $j \to +\infty$; thus, for $j \geq 1$,

$$|E[g(X_{t_0}, X_{t_1}) g(X_{t_j}, X_{t_{j+1}})]| \leq |E[g(X_{t_0}, X_{t_1}) g(X_{t_1}, X_{t_2})]|. \tag{B.4}$$

Given (B.3) and (B.4) we get (b3).

According to Lemma 9 (also see Florens-Zmirou, 1989, Theorem 3) we have

$$\frac{1}{\sqrt{n}} \sum g(X_{t_i}, X_{t_{i+1}}) \xrightarrow{d} N(0, b^4(x)\bar{p}(x)K_2),$$

and thus (after standardization) $(1/\sqrt{n}) \sum g(X_{t_i}, X_{t_{i+1}}) / (b^2(x)\sqrt{\bar{p}(x)K_2}) \xrightarrow{d} N(0,1)$. Using the asymptotic equivalence theorem, $(1/\sqrt{n}) \sum g(X_{t_i}, X_{t_{i+1}}) / (b^2(x)\sqrt{B_n K_2})$ and $(1/\sqrt{n}) \sum g(X_{t_i}, X_{t_{i+1}}) / (b^2(x)\sqrt{\bar{p}(x)K_2})$ have the same asymptotic distribution, and thus the result is proved. ■

**Proof of Theorem 6.**

(a) Let $V_n^*(x) = A_n^*(x)/B_n(x)$ where

$$A_n^*(x) = \frac{1}{nh_n} \sum_{i=0}^{n-1} K\left(\frac{x - X_{t_i}}{h_n}\right) \frac{(X_{t_{i+1}} - X_{t_i} - T_n(X_{t_i})\Delta_n)^2}{\Delta_n}.$$

In the proof of Theorem 4 we saw that $A_n^*(x)$ can be written as $A_n^* = A_n^\# + D_{1,n} + D_{2,n}$. A crucial step in this proof is that $T_n(x) \xrightarrow{p} a(x)$ as $\Delta_n \to 0$ and $n \to +\infty$. The estimator $T_n(x)$ does not converge to $a(x)$ as $\Delta_n \to 0$ unless we assume $t_n h_n \to +\infty$, thereby $t_n \to +\infty$ (see also Bandi and Phillips, 2003). With the conditions stated one has $T_n(x) \xrightarrow{p} a(x)$ and even $T_n(X_t) \xrightarrow{p} a(X_t)$, i.e., $T_n(X_t) = a(X_t) + \varepsilon_{\Delta_n, n}$, where $\operatorname{plim}_{\Delta_n \to 0, n \to +\infty} \varepsilon_{\Delta_n, n} = 0$ (see the proof of Theorem 4). Hence,

$$\begin{aligned} D_{1,n} &= -\frac{2}{nh_n} \sum_{i=0}^{n-1} K\left(\frac{x - X_{t_i}}{h_n}\right) \frac{(X_{t_{i+1}} - X_{t_i} - a(X_{t_i})\Delta_n)(T_n(X_{t_i}) - a(X_{t_i}))\Delta_n}{\Delta_n} \\ &= -\frac{2}{nh_n} \sum_{i=0}^{n-1} K\left(\frac{x - X_{t_i}}{h_n}\right) (X_{t_{i+1}} - X_{t_i} - a(X_{t_i})\Delta_n)\varepsilon_{\Delta_n, n}. \end{aligned} \tag{B.5}$$

It is straightforward to conclude that $D_{1,n} \xrightarrow{p} 0$ because $(2/nh_n)\sum_{i=0}^{n-1} K((x - X_{t_i})/h_n) \times (X_{t_{i+1}} - X_{t_i} - a(X_{t_i})\Delta_n) \xrightarrow{p} 0$ and $\varepsilon_{\Delta_n, n} \xrightarrow{p} 0$ under the assumptions stated (in fact, each term of the summation in (B.5) is of order $o_p(1)$). Using the same arguments, it is easy to verify that

$$\begin{aligned} D_{2,n} &= \frac{1}{nh_n} \sum_{i=0}^{n-1} K\left(\frac{x - X_{t_i}}{h_n}\right) (T_n(X_{t_i}) - a(X_{t_i}))^2 \Delta_n \\ &= \frac{1}{nh_n} \sum_{i=0}^{n-1} K\left(\frac{x - X_{t_i}}{h_n}\right) (\varepsilon_{\Delta_n, n})^2 \Delta_n \end{aligned}$$

converges in probability to zero (note that by Slutsky's theorem $(\varepsilon_{\Delta_n, n})^2 \xrightarrow{p} 0$).

(b) First, note that

$$\sqrt{\frac{nh_n}{2K_2}\,B_n(x)}\left(\frac{V_n^*(x)}{b^2(x)}-1\right)=\sqrt{\frac{nh_n}{2K_2}\,B_n(x)}\left(\frac{V_n(x)+\dfrac{D_{1,n}}{B_n(x)}+\dfrac{D_{2,n}}{B_n(x)}}{b^2(x)}-1\right)$$

$$=\sqrt{\frac{nh_n}{2K_2}\,B_n(x)}\left(\frac{V_n(x)}{b^2(x)}-1\right)$$

$$+\frac{1}{b^2(x)\sqrt{2K_2\,B_n(x)}}\,(\sqrt{nh_n}D_{1,n}+\sqrt{nh_n}D_{2,n}).$$

Now $\sqrt{nh}D_{i,n}=\sqrt{nh}\Delta_n(D_{i,n}/\Delta_n)\xrightarrow{p}0$ $(i=1,2)$ because $\sqrt{nh}\Delta_n\to 0$ (by hypothesis) and

$$\frac{D_{1,n}}{\Delta_n}=-\frac{2}{nh_n}\sum_{i=0}^{n-1}K\left(\frac{x-X_{t_i}}{h_n}\right)\frac{(X_{t_{i+1}}-X_{t_i}-a(X_{t_i})\Delta_n)}{\Delta_n}\,\varepsilon_{\Delta_n,n}\xrightarrow{p}0,$$

$$\frac{D_{2,n}}{\Delta_n}=\frac{1}{nh_n}\sum_{i=0}^{n-1}K\left(\frac{x-X_{t_i}}{h_n}\right)(\varepsilon_{\Delta_n,n})^2\xrightarrow{p}0.$$

These results are obvious and their proofs are omitted. It remains to observe that $b^{-2}(x)(2K_2B_n(x))^{-1/2}$ converge in probability to $b^{-2}(x)(2K_2\bar{p}(x))^{-1/2}$. Therefore, the result (b) follows from Theorem 5. ■

https://doi.org/10.1017/S0266466603195035 Published online by Cambridge University Press