*Econometric Theory*, **23**, 2007, 880–898. Printed in the United States of America.
DOI: 10.1017/S0266466607070375

# NONPARAMETRIC ESTIMATION OF SECOND-ORDER STOCHASTIC DIFFERENTIAL EQUATIONS

JOÃO NICOLAU
*School of Economics and Management (ISEG)*
*Universidade Técnica de Lisboa*

We propose nonparametric estimators of the infinitesimal coefficients associated with second-order stochastic differential equations. We show that under appropriate conditions, the proposed estimators are consistent. Also, we state conditions ensuring the asymptotic normality of these estimators. We conclude our paper with a Monte Carlo experiment in which we assess the response of the nonparametric estimators with respect to the step of discretization.

## 1. INTRODUCTION

In economics and finance many stochastic processes can be seen as integrated stochastic processes in the sense that the current observation behaves as the cumulation of all past perturbations. In a discrete-time framework the concept of integration and differentiation of a stochastic process plays an essential role in modern econometric analysis. For instance, the stochastic process $\{y_t;\ t = 0,1,2,\ldots\}$ where $y_t = \alpha + y_{t-1} + \varepsilon_t$ ($\varepsilon_t \sim i.i.d.\ N(0,1)$) is an example of an integrated process. Notice that $y$ can be written as $y_t = y_0 + t\alpha + \sum_{k=1}^{t} \varepsilon_k$, or

$$y_t = y_0 + \sum_{k=1}^{t} x_k, \tag{1}$$

where $x_t = \alpha + \varepsilon_t$. One way to deal with such processes is to use a differenced-data model (e.g., $\Delta y_t = \alpha + \varepsilon_t$, in the previous example). Differencing has been used mostly to solve nonstationary problems viewed as unit roots, although historically differenced-data models arose early in econometrics as a procedure to remove common trends between dependent and independent variables.

Integrated and differentiated diffusion processes in the same sense as integrated and differentiated discrete-time processes are almost absent in applied

I thank two anonymous referees who made valuable suggestions that led to considerable improvements in 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 FEDER/POCI 2010. Address correspondence to João Nicolau, Instituto Superior de Economia e Gestão, Universidade Técnica de Lisboa, Rua do Quelhas 6, 1200-781 Lisboa, Portugal; e-mail: nicolau@iseg.utl.pt.

 © 2007 Cambridge University Press 0266-4666/07 $15.00

https://doi.org/10.1017/S0266466607070375 Published online by Cambridge University Press

econometric analysis. One of the reasons continuous-time differentiated processes have not been considered in applied econometrics is perhaps related to the difficulties in interpreting the "differentiated" process. In fact, if $Z$ is a diffusion process driven by a Brownian motion, then all sample functions are of unbounded variation and nowhere differentiable; i.e., $dZ_t/dt$ does not exist with probability one (unless some smoothing effect of the measurement instrument is introduced, such as described in Arnold, 1974).

One way to model integrated and differentiated diffusion processes and overcome the difficulties associated with the nondifferentiability of the Brownian motion is through the representation

$$\begin{cases} dY_t = X_t\,dt \\ dX_t = a(X_t)\,dt + b(X_t)\,dW_t, \end{cases} \tag{2}$$

where $a$ and $b$ are the infinitesimal coefficients (respectively, the drift and the diffusion coefficient), $W$ is a (standard) Wiener process (or Brownian motion), and $X$ is (by hypothesis) a stationary process. In this model, $Y$ is a differentiable process, by construction. It represents the integrated process

$$Y_t = Y_0 + \int_0^t X_u\,du \tag{3}$$

(note the analogy with the corresponding expression in a discrete-time setting, $y_t = y_0 + \sum_{k=1}^{t} x_k$, eqn. (1)), and $X_t = dY_t/dt$ is the stationary differentiated process (which can be considered the equivalent concept to the first differences sequence in discrete-time analysis). If $X$ represents the continuously compounded return or log return of an asset, the first equation in system (2) should be rewritten as $d\log Y_t = X_t\,dt$.

We think that (2) can be a useful model in empirical finance for at least two reasons. First, the model accommodates nonstationary integrated stochastic processes ($Y$) that can be made stationary by differencing. Such transformation cannot be undertaken in common univariate diffusion processes used in finance (because all sample paths from univariate diffusion processes are nowhere differentiable with probability one). Yet, many processes in economics and finance (e.g., stock prices and nominal exchange rates) behave as the cumulation of all past perturbations (basically in the same sense as unit root processes in a discrete framework). Second, in the context of stock prices or exchange rates, the model suggests directly modeling the (instantaneous) returns, in contrast to the usual continuous-time models in finance, which directly model the prices. General properties for returns (stylized facts) are well known and documented (e.g., returns are generally stationary in mean, the distribution is not normal, the autocorrelations are weak, and the correlations between the magnitude of returns are positive and statistically significant, etc.). One advantage of directly modeling the returns ($X$) is that these general properties are easier to specify in a

https://doi.org/10.1017/S0266466607070375 Published online by Cambridge University Press

model like (2) than in a diffusion univariate process for the prices. In fact, several interesting models can be obtained by selecting $a(x)$ and $b^2(x)$ appropriately. For example, the choice $a(x) = \beta(\tau - x)$ and $b(x) = \sqrt{\sigma^2 + \lambda(X_t - \mu)^2}$ leads to an integrated process $Y$ whose returns, $X$, have an asymmetric leptokurtic stationary distribution (see the example that follows). This specification can be appropriated in financial time series data. Bibby and Sørensen (1997) have already noticed that a process similar to (2) could be a good model for stock prices.

We observe that the model defined in equation (2) can be written as a second-order stochastic differential equation, $d(dY_t/dt) = a(X_t)\,dt + b(X_t)\,dW_t$. Integrated diffusions like $Y$ in equation (3) arise naturally when only observations of a running integral of the process are available. For instance, this can occur when a realization of the process is observed after passage through an electronic filter. Another example is provided by ice-core data on oxygen isotopes used to investigate paleotemperatures (see Ditlevsen and Sørensen, 2004).

To illustrate continuous-time integrated processes, Figure 1 presents two simulated independent paths of $Y_t = Y_0 + \int_0^t X_u\,du$ where $X$ is governed by the stochastic differential equation

$$dX_t = 20(0.01 - X_t)\,dt + \sqrt{0.1 + 10(X_t - 0.05)^2}\,dW_t$$

($X$ is also represented in Figure 1). All paths are composed of 1,000 observations defined in the interval $t \in [0,10]$. In Section 4 we give more details on how we generate these observations. It is interesting to observe that $Y$ displays all the features of an integrated process (with a positive drift because $\mathrm{E}[X_t] = 0.01$): There is an absence of mean reversion; shocks are persistent; mean and variance depend on time, etc. On the other hand, the unconditional distribution of $X$ (return) is asymmetric and leptokurtic.

Estimation of second-order stochastic differential equations raises new challenges for two main reasons. On the one hand, only the integrated process $Y$ is observable at instants $\{t_i;\, i = 1,2,\ldots\}$, and thus $X$ in model (2) is a latent nonobservable process. In fact, for a fixed sampling interval, it is impossible to obtain the value of $X$ at time $t_i$ from the observation $Y_{t_i}$ that represents the integral $Y_0 + \int_0^{t_i} X_u\,du$. On the other hand, the estimation of model (2) cannot in principle be based on the observations $\{Y_{t_i};\, i = 1,2,\ldots\}$ because the conditional distribution of $Y$ is generally unknown, even if that of $X$ is known. An exception is the case where $X$ follows an Ornstein–Uhlenbeck process, which is analyzed in Gloter (2001).

However, with discrete-time observations $\{Y_{i\Delta};\, i = 1,2,\ldots\}$ (to simplify we use the notation $t_i = i\Delta$, where $\Delta = t_i - t_{i-1}$), and given that

$$Y_{i\Delta} - Y_{(i-1)\Delta} = \int_0^{i\Delta} X_u\,du - \int_0^{(i-1)\Delta} X_u\,du = \int_{(i-1)\Delta}^{i\Delta} X_u\,du,$$

https://doi.org/10.1017/S0266466607070375 Published online by Cambridge University Press

https://doi.org/10.1017/S0266466607070375 Published online by Cambridge University Press



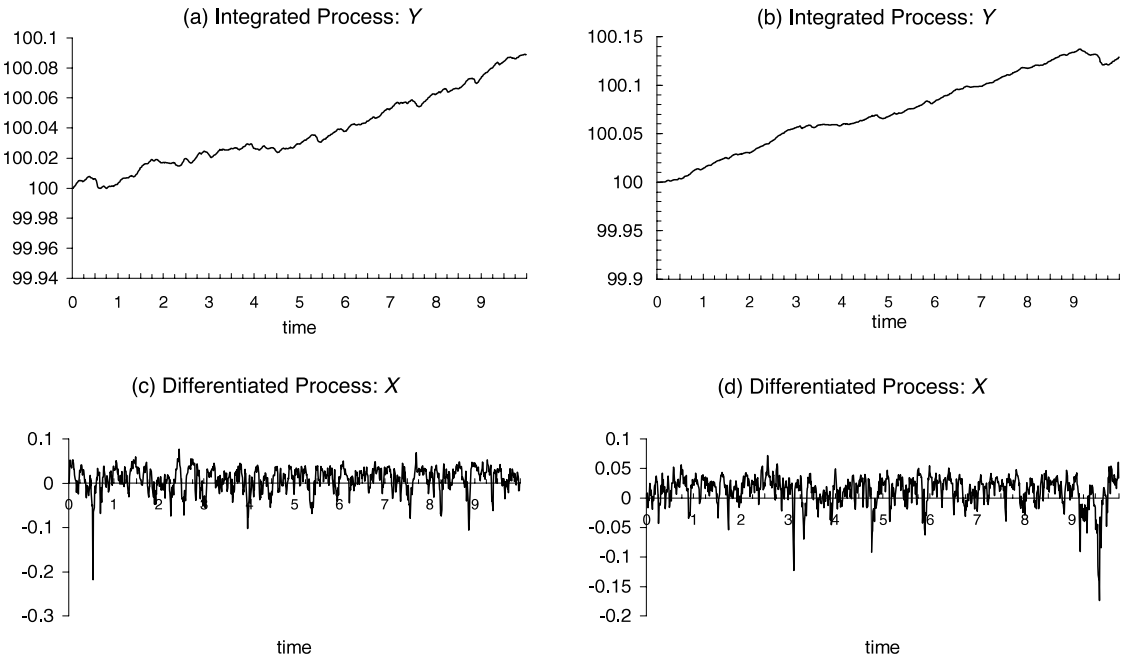



**FIGURE 1.** Simulation of two independent paths of the process $Y_t = Y_0 + \int_0^t X_u\,du$ where $X$ is governed by the stochastic differential equation $dX_t = 20(0.01 - X_t)\,dt + \sqrt{0.1 + 10(X_t - 0.05)^2}\,dW_t$.

we can obtain a measure of $X$ at instant $t_i = i\Delta$ using the formula

$$\widetilde{X}_{i\Delta} = \frac{Y_{i\Delta} - Y_{(i-1)\Delta}}{\Delta}. \tag{4}$$

Naturally, the accuracy of (4) as a proxy for $X_{i\Delta}$ depends on the magnitude of $\Delta$. Regardless of the magnitude of $\Delta$ we have in our sample, we should base our estimation procedures on the sample $\{\widetilde{X}_{i\Delta}\} := \{\widetilde{X}_{i\Delta}; i = 0,1,2,\ldots\}$ because $X$ is not observable.

Parametric and semiparametric estimation of integrated diffusions are analyzed in Gloter (2001, 2006) and Ditlevsen and Sørensen (2004).

We suppose that both infinitesimal coefficients $a$ and $b$, defined in model (2) are unknown, and our aim is their nonparametric functional estimation. We propose nonparametric estimators for the infinitesimal coefficients $a$ and $b$. Our analysis reveals that the standard estimators based on the sample $\{\widetilde{X}_{i\Delta}\}$ are inconsistent, even if we allow the step of discretization $\Delta$ to go to zero asymptotically. Introducing slight modifications to these estimators we provide consistent estimators. For a review of nonparametric estimation of (first-order) stochastic differential equations see, e.g., Rao (1983), Florens-Zmirou (1993), Jiang and Knight (1997), Bandi and Phillips (2003), Nicolau (2003), and Gobet, Hoffmann, and Reiß (2004). For a review of parametric and semiparametric estimation see, e.g., Aït-Sahalia (2002) and Kessler (1997) and the references therein.

The rest of the paper is organized as follows. In Section 2 we formulate the main hypotheses concerning the $X$ process, and we identify our estimators. In Section 3 we establish the main results. We show that under appropriate conditions, the proposed estimators are consistent and asymptotically normal. In Section 4 we perform a Monte Carlo experiment in which we assess the response of the nonparametric estimator with respect to the step of discretization $\Delta$. Section 5 concludes. Appendix A presents some auxiliary results. All proofs appear in Appendix B.

## 2. ESTIMATORS AND ASSUMPTIONS

The aim of this paper is to study the following estimators:

$$\hat{p}_n(x) = \frac{1}{nh_n}\sum_{i=1}^{n} K\left(\frac{x - \widetilde{X}_{(i-1)\Delta_n}}{h_n}\right), \qquad \hat{a}_n(x) = \frac{A_n(x)}{\hat{p}_n(x)}, \qquad \hat{b}_n^2(x) = \frac{B_n(x)}{\hat{p}_n(x)} \tag{5}$$

for, respectively, $p(x)$, $a(x)$, and $b^2(x)$, based on the observations $\{\widetilde{X}_{i\Delta_n}\}$, where

https://doi.org/10.1017/S0266466607070375 Published online by Cambridge University Press

$$A_n(x) = \frac{1}{nh_n}\sum_{i=1}^{n} K\left(\frac{x - \widetilde{X}_{(i-1)\Delta_n}}{h_n}\right)\frac{(\widetilde{X}_{(i+1)\Delta_n} - \widetilde{X}_{i\Delta_n})}{\Delta_n}, \tag{6}$$

$$B_n(x) = \frac{1}{nh_n}\sum_{i=1}^{n} K\left(\frac{x - \widetilde{X}_{(i-1)\Delta_n}}{h_n}\right)\frac{\frac{3}{2}(\widetilde{X}_{i\Delta_n} - \widetilde{X}_{(i-1)\Delta_n})^2}{\Delta_n}. \tag{7}$$

Consider the process $(Y, X)$ governed by the system (2). Let $I = (l, r)$ be the state space of $X$. 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 eight assumptions that are used throughout the paper.

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

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 $p(x) = m(x)/\int_l^r m(u)\,du$ with respect to the Lebesgue measure (Skorokhod, 1989, Thm. 16). The expression $p(x)$ is usually referred to as the stationary density.

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

Assumption A3 together with A1 and A2 implies that $X$ is stationary (Arnold, 1974). The hypothesis that $X$ is a stationary process is obviously a plausible assumption because for major integrated time series data a simple differentiation generally assures stationarity.

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

These conditions are discussed in Chen, Hansen, and Carrasco (1999) and are similar to conditions proposed by Hansen and Scheinkman (1995). Under Assumption A4 the process is $\rho$-mixing and $\alpha$-mixing (see Chen et al., 1999). 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 on the linear space $L_2^0 = \{f: \int f(x)p(x)\,dx = 0, \int f^2(x)p(x)\,dx < \infty\}$. As a consequence, the $j$th autocovariance of $f(X_t)$ tends to zero at an exponential rate as $j \to \infty$, for all $f \in L_2^0$ (see Hansen and Scheinkman, 1995, Prop. 8; Florens-Zmirou, 1989).

We notice that the discretized diffusion $\{X_{i\Delta}\} := \{X_{i\Delta};\ i = 0, 1, 2, \ldots\}$ with fixed $\Delta$ is an ergodic process because $X = \{X_t;\ t \geq 0\}$ is ergodic under Assumptions A1–A4 (see Dacunha-Castelle and Florens-Zmirou, 1986). The process $\{\widetilde{X}_{i\Delta}\}$ inherits some properties from $\{X_{i\Delta}\}$. As Ditlevsen and Sørensen (2004) point out, because $X$ is stationary, the distribution of $X_t$ is the same for all

https://doi.org/10.1017/S0266466607070375 Published online by Cambridge University Press

intervals $\{[(i-1)\Delta, i\Delta];\ i \geq 1\}$, and thus $\{\widetilde{X}_{i\Delta}\}$ is a stationary process. Also if $\mathrm{E}[X] < \infty$, we have by Chebyshev integral inequality and stationarity

$$\mathrm{E}[\widetilde{X}_{i\Delta}^2] = \mathrm{E}\left[\left(\frac{\int_{(i-1)\Delta}^{i\Delta} X_u\, du}{\Delta}\right)^2\right]$$

$$\leq \mathrm{E}\left[\frac{\Delta \int_{(i-1)\Delta}^{i\Delta} X_u^2\, du}{\Delta^2}\right] = \frac{\int_{(i-1)\Delta}^{i\Delta} \mathrm{E}[X_u^2]\, du}{\Delta} = \mathrm{E}[X_0^2].$$

On the other hand, $\{\widetilde{X}_{i\Delta}\}$ is $\alpha$-mixing ($\rho$-mixing) because measurable functions of mixing processes are mixing and of the same size.

The kernel function $K$ satisfies the following assumption.

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

A6($\alpha$). $\lim_{h\to 0} \mathrm{E}[(1/h)|K'(\xi_{n,i})|^{\alpha}] < \infty$ where $\alpha = 2$ or $\alpha = 4$ and $\xi_{n,i} = \theta((x - X_{(i-1)\Delta_n})/h) + (1-\theta)((x - \widetilde{X}_{i\Delta_n})/h)$, $0 \leq \theta \leq 1$.

Assumption A6($\alpha$) is generally satisfied under very weak conditions. For instance, with a Gaussian kernel and a Cauchy stationary density (which has heavy tails) we still have $\lim_{h\to 0} \mathrm{E}[(1/h)|K'(X/h)|^4] < \infty$. Notice that the expectation with respect to the distribution of $\xi_{n,i}$ depends on the stationary densities of $X$ and $\widetilde{X}$ because $\xi_{n,i}$ is a convex linear combination of $X$ and $\widetilde{X}$.

A7.

(i) The infinitesimal coefficients $a$ and $b$ have continuous derivatives of order 4 and satisfy $|a(x)| \leq C(1 + |x|)^{\lambda}$ and $|b(x)| \leq C(1 + |x|)^{\lambda}$ for some $\lambda > 0$.
(ii) $\mathrm{E}[X_0^r] < \infty$ where $r = \max\{4\lambda, 1 + 3\lambda, -1 + 5\lambda, -2 + 6\lambda\}$.

This assumption is related to the application of Lemma 4 (Appendix A). See the general discussion in Remark 1.

A8.

(a) $\Delta_n \to 0, h_n \to 0, nh_n \to \infty$, $\sqrt{\Delta_n}/h_n \to 0$ as $n \to \infty$;
(b1) $(n\Delta_n/h_n)(\Delta_n \log(1/\Delta_n))^{1/2} \to 0, h_n n\Delta_n \to \infty$;
(b2) $h_n^5 n\Delta_n \to 0$ as $n \to \infty$;
(c1) $(n\Delta_n/h_n)(\Delta_n \log(1/\Delta_n))^{1/2} \to 0$;
(c2) $h_n^5 n \to 0$ as $n \to \infty$.

https://doi.org/10.1017/S0266466607070375 Published online by Cambridge University Press

Assumptions A1–A5 and A8 assure consistency and asymptotic normality of the standard estimators for $p(x)$, $a(x)$, $b^2(x)$, based on the observations $\{X_{i\Delta_n}\}$, respectively,

$$\hat{p}_n^0(x) = \frac{1}{nh_n}\sum_{i=1}^{n} K\left(\frac{x - X_{(i-1)\Delta_n}}{h_n}\right), \qquad \hat{a}_n^0(x) = \frac{A_n^0(x)}{\hat{p}_n^0(x)},$$
$$(\hat{b}_n^2)^0(x) = \frac{B_n^0(x)}{\hat{p}_n^0(x)}, \tag{8}$$

where

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

In fact, it can be proved (see, e.g., Bandi and Phillips, 2003) that

$$\hat{p}_n^0(x) \xrightarrow{p} p(x) \quad \text{(convergence in probability)},$$
$$\hat{a}_n^0(x) \xrightarrow{p} a(x), \qquad \sqrt{h_n n\Delta_n}(\hat{a}_n^0(x) - a(x)) \xrightarrow{d} N\left(0, K_2\frac{b^2(x)}{p(x)}\right),$$
$$(\hat{b}_n^2)^0(x) \xrightarrow{p} b^2(x), \qquad \sqrt{nh_n}((\hat{b}_n^2)^0(x) - b^2(x)) \xrightarrow{d} N\left(0, \frac{4K_2 b^4(x)}{p(x)}\right).$$

## 3. MAIN RESULTS

We assume conditions A1–A7 in all theorems.

THEOREM 1. *Assume A6($\alpha = 2$) and A8(a). Then*

$$\hat{p}_n(x) = \frac{1}{nh_n}\sum_{i=1}^{n} K\left(\frac{x - \widetilde{X}_{(i-1)\Delta_n}}{h_n}\right) \xrightarrow{p} p(x).$$

It is well known that the standard $\hat{p}_n^0$ estimator (i.e., built from the observations $\{X_{i\Delta}\}$) does not require that the step of discretization $\Delta$ goes to zero as $n \to \infty$ (if $\Delta$ is fixed it is necessary that $n\Delta \to \infty$). However, as one would expect, the convergence of the density estimator based on the observations $\{\widetilde{X}_{i\Delta}\}$ requires $\Delta_n \to 0$.

https://doi.org/10.1017/S0266466607070375 Published online by Cambridge University Press

THEOREM 2.

*(i) Assume A6($\alpha = 2$) and A8(a) and (b1). Then*

$$\hat{a}_n(x) = \frac{\dfrac{1}{nh_n}\displaystyle\sum_{i=1}^{n} K\left(\frac{x - \widetilde{X}_{(i-1)\Delta_n}}{h_n}\right)\frac{(\widetilde{X}_{(i+1)\Delta_n} - \widetilde{X}_{i\Delta_n})}{\Delta_n}}{\dfrac{1}{nh_n}\displaystyle\sum_{i=1}^{n} K\left(\frac{x - \widetilde{X}_{(i-1)\Delta_n}}{h_n}\right)} \xrightarrow{p} a(x).$$

*(ii) If additionally A8(b2) and $h_n n\Delta_n^3 \to 0$ then*

$$\sqrt{h_n n\Delta_n}(\hat{a}_n(x) - a(x)) \xrightarrow{d} N\left(0, K_2 \frac{b^2(x)}{p(x)}\right).$$

Standard conditions for consistency of $\hat{a}_n^0(x)$ still assure $\hat{a}_n(x) \xrightarrow{p} a(x)$. It is implicit in the conditions of the theorem that $n\Delta_n \to \infty$, and this condition should also hold in the standard case.

The "natural" estimator for $a(x)$, based on the observations $\{\widetilde{X}_{i\Delta_n}\}$, is

$$\check{a}_n(x) = \left(\frac{1}{nh_n}\sum_{i=1}^{n} K\left(\frac{x - \widetilde{X}_{(i-1)\Delta_n}}{h_n}\right)\frac{(\widetilde{X}_{i\Delta_n} - \widetilde{X}_{(i-1)\Delta_n})}{\Delta_n}\right)/\hat{p}_n(x). \tag{9}$$

It can be proved, however, that this estimator is inconsistent. When $\Delta$ is fixed the advantages of the $\hat{a}_n$ over the $\check{a}_n$ are not clear. To shed light on this issue, we perform some Monte Carlo experiments in Section 4. One of the main conclusions is that the $\check{a}_n$ estimator may present a smaller mean squared error than the $\hat{a}_n$ estimator when the step of discretization is not small enough. On the contrary, when $\Delta_n \to 0$ the mean squared error of the $\hat{a}_n$ estimator converges to the mean squared error of the drift estimator based on the observations $\{X_{i\Delta}\}$.

THEOREM 3.

*(i) Assume A6($\alpha = 4$) and A8(a) and (c1). Then*

$$\hat{b}_n^2(x) = \frac{\dfrac{1}{nh_n}\displaystyle\sum_{i=1}^{n} K\left(\frac{x - \widetilde{X}_{(i-1)\Delta_n}}{h_n}\right)\frac{\frac{3}{2}(\widetilde{X}_{i\Delta_n} - \widetilde{X}_{(i-1)\Delta_n})^2}{\Delta_n}}{\dfrac{1}{nh_n}\displaystyle\sum_{i=1}^{n} K\left(\frac{x - \widetilde{X}_{(i-1)\Delta_n}}{h_n}\right)} \xrightarrow{p} b^2(x).$$

*(ii) If additionally A8(c2) and $h_n n\Delta_n^2 \to 0$ then*

$$\sqrt{nh_n}(\hat{b}_n^2(x) - b^2(x)) \xrightarrow{d} N\left(0, \frac{4K_2 b^4(x)}{p(x)}\right).$$

https://doi.org/10.1017/S0266466607070375 Published online by Cambridge University Press

The most important difference between this theorem and the standard counterpart result is that our estimator $\hat{b}_n^2$ involves a multiplicative factor of $\frac{3}{2}$. Gloter (2006), using different arguments and working in a semiparametric framework, has already foreseen the need for this correction. For more details see the proof of Theorem 3. Therefore, the "natural" estimator for $b^2(x)$, based on the observations $\{\widetilde{X}_{i\Delta_n}\}$,

$$\hat{b}^2_{n,*}(x) = \frac{\sum_{i=1}^{n} K\left(\frac{x - \widetilde{X}_{(i-1)\Delta_n}}{h_n}\right) \frac{(\widetilde{X}_{i\Delta_n} - \widetilde{X}_{(i-1)\Delta_n})^2}{\Delta_n}}{\sum_{i=1}^{n} K\left(\frac{x - \widetilde{X}_{(i-1)\Delta_n}}{h_n}\right)} \tag{10}$$

has an asymptotic bias of $-b^2(x)/3$ (when $\Delta_n \to 0$). This bias is confirmed through Monte Carlo simulations (Section 4).

## 4. MONTE CARLO SIMULATION

In this section we perform a Monte Carlo experiment aimed at assessing the response of the nonparametric estimators with respect to the step of discretization $\Delta$. Our Monte Carlo experiment is based on the simulation of $Y_t = Y_0 + \int_0^t X_u\,du$ for $t \in [0,T] = [0,1{,}000]$ where $X$ is an ergodic process governed by the stochastic differential equation $dX_t = -10X_t\,dt + \sqrt{0.1 + 0.1X_t^2}dW_t$. To simulate $Y$ and $X$ we consider a time discretization $0 < \delta < 2\delta < \cdots < n\delta = T$ of the interval $[0,T]$ and the Euler–Maruyama approximation. We found that the value $\delta = 0.001$ is reliable in the sense that the approximations $\hat{X}$ and $\hat{Y}$ are sufficiently close to the respective sample path of the Itô process. As such, we simplify the notation, and we write $X = \hat{X}$ and $Y = \hat{Y}$. We generate 5,000 replicas of $X$ and $Y$. We have thus in each replica a sequence of $n = 1{,}000{,}000$ values of $X$ and $Y$ (note that $T/\delta = 1{,}000/0.001 = 1{,}000{,}000$) from which we estimate the infinitesimal coefficients assuming different values for the step of discretization. Specifically, we consider the following samples: $\{Y_{i\times 0.01};\ i = 0,1,\ldots,100{,}000\}$, $\{Y_{i\times 0.02};\ i = 0,1,\ldots,50{,}000\},\ldots,\{Y_{i\times 0.2};\ i = 0,1,\ldots,5{,}000\}$; i.e., step of discretization ranges from $\Delta = 0.01$ to $\Delta = 0.2$ with a step of 0.01. In each sample (and for each replica) we estimate the infinitesimal coefficients $a(x) = -10x$ and $b^2(x) = 0.1 + 0.1x^2$. The values of $x$ are from the set $A = \{q_{0.20}, q_{0.30}, \ldots, q_{0.70}, q_{0.80}\}$ where $q_\alpha$ is the quantile of order $\alpha$ of the stationary distribution of $X$, i.e., $p(x) = 1.80656(1 + x^2)^{-11}$.

A measure of error is based on the Monte Carlo average of the "integrated" squared error (ISE) obtained in each replica. We denote this statistic as AISE. For example, the AISE of $\hat{a}_n$, based on a certain value for $\Delta$, is given by

$$AISE(\hat{a}_n, \Delta) = \sum_{j=1}^{S} \frac{ISE(\hat{a}_n, \Delta, j)}{S}, \qquad S = 5{,}000,$$

https://doi.org/10.1017/S0266466607070375 Published online by Cambridge University Press

where $ISE(\hat{a}_n, \Delta, j)$ is the "integrated" squared error over the set $A$ of the estimator $\hat{a}_n$, obtained in the $j$th replica, i.e.,

$$ISE(\hat{a}_n, \Delta, j) = \frac{\sum_{x \in A} (\hat{a}_n(x) - a(x))^2}{\#A}, \qquad a(x) = -10x$$

(#A: number of elements of the set $A$; in our case, $\#A = 7$). Although the $ISE(\hat{a}_n, \Delta, j)$ is only a rough approximation to the standard ISE measure, $\int_{\mathbb{R}} (\hat{a}_n(u) - a(u))^2 \, du$, it can provide valuable information on the error of approximation of the nonparametric estimator.

We study the drift estimators $\hat{a}_n$ and $\check{a}_n$ (see (5) and (9), respectively) and the diffusion estimators $\hat{b}_n^2$ and $\hat{b}_{n,*}^2$ (see (5) and (10), respectively). For comparison purposes we also analyze the drift and the diffusion estimators based on the observations $\{X_{i\Delta}\}$, $\hat{a}_n^0$ and $(\hat{b}_n^2)^0$ (see eqn. (8)). In all cases we consider a Gaussian kernel and the common bandwidth $h_n = \bar{h} \hat{\sigma}_X n^{-1/5}$, with $\bar{h} = 1.06$. The results are shown in Figure 2. The main conclusions are as follows. First, as expected, the estimators based on the observations $\{X_{i\Delta}\}$, $\hat{a}_n^0$ and $(\hat{b}_n^2)^0$, generally present smaller AISE when the step of discretization is moderate or high. These differences are, however, negligible when $\Delta$ is small. Second, as expected, the AISEs of the estimators $\hat{a}_n$ and $\hat{b}_n^2$ (dashed line) converge to zero as $\Delta \to 0$. Third, the $\check{a}_n$ estimator (dotted line) seems inconsistent because the AISE does not converge to zero as $\Delta$ shrinks. However, for moderate/high values of $\Delta$, the $\check{a}_n$ estimator performs reasonably well, and better than the $\hat{a}_n$ estimator. Consequently, in practical applications the $\hat{a}_n$ estimator should be used with some caution, especially when the step of discretization is not small. In these cases, the $\check{a}_n$ estimator may be better. Fourth, the $\hat{b}_{n,*}^2(x)$ estimator, as predicted

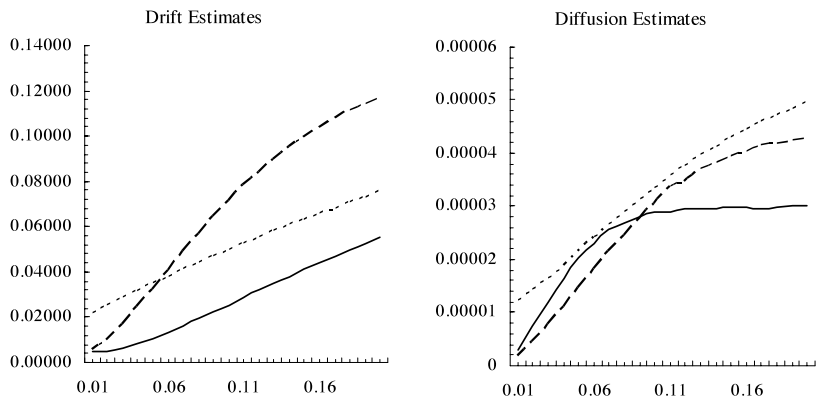



**FIGURE 2.** Drift and diffusion estimators: Monte Carlo average of ISE. Solid line: $\hat{a}^0$ and $(\hat{b}^2)^0$; dashed line: $\hat{a}$ and $\hat{b}^2$; dotted line: $\check{a}$ and $\hat{b}_*^2$.

https://doi.org/10.1017/S0266466607070375 Published online by Cambridge University Press

in Section 3, has the worst performance when compared with the other estimators. Finally, the AISEs associated with the diffusion estimators are much smaller for all values of $\Delta > 0$. This may be related to the data generating process selected and/or to the fact that the estimation of $a$ is harder than that of $b$. This is a point emphasized, e.g., by Gobet et al. (2004) and Bandi and Phillips (2003).

## 5. CONCLUSIONS AND OTHER EXTENSIONS

In this paper we showed that standard nonparametric estimators of the infinitesimal coefficients of second-order stochastic differential equations, based on the observations $\{\widetilde{X}_{i\Delta}\}$, are not appropriate, because they are inconsistent, even when $\Delta_n \to 0$. However, as shown, slight modifications to the standard estimators allow us to define consistent estimators.

There are some extensions that can be considered, such as nonequidistant observations, optimal bandwidth choice, data-driven estimation, and minimax rates. More general assumptions on the data generating process as in Bandi and Phillips (2003) can also be applied.

### *REFERENCES*

Aït-Sahalia, Y. (2002) Maximum likelihood estimation of discretely sampled diffusions: A closed-form approximation approach. *Econometrica* 70, 223–262.

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

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

Bibby, B. & M. Sørensen (1997) A hyperbolic diffusion model for stock prices. *Finance and Stochastics* 1, 25–41.

Chen, X., L. Hansen, & M. Carrasco (1999) Nonlinearity and Temporal Dependence. Unpublished.

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

Ditlevsen, S. & M. Sørensen (2004) Inference for observations of integrated diffusion processes. *Scandinavian Journal of Statistics* 31, 417–429.

Florens-Zmirou, D. (1989) Approximate discrete-time schemes for statistics of diffusion processes. *Statistics* 20, 547–557.

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

Gloter, A. (2001) Parameter estimation for a discrete sampling of an integrated Ornstein-Uhlenbeck process. *Statistics* 35, 225–243.

Gloter, A. (2006) Parameter estimation for a discretely observed integrated diffusion process. *Scandinavian Journal of Statistics* 33, 83–104.

Gobet, E., M. Hoffmann, & M. Reiß (2004) Nonparametric estimation of scalar diffusions based on low frequency data. *Annals of Statistics* 32, 2223–2253.

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

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.

Kessler, M. (1997) Estimation of an ergodic diffusion from discrete observations. *Scandinavian Journal of Statistics* 24, 211–229.

https://doi.org/10.1017/S0266466607070375 Published online by Cambridge University Press

Nicolau, J. (2003) Bias reduction in nonparametric diffusion coefficient estimation. *Econometric Theory* 19, 754–777.

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

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

# APPENDIX A: Auxiliary Results

In the following lemma we generalize a formula of Florens-Zmirou (1989) to multivariate diffusions. Lemma 4 is essential in the proofs of all the theorems presented in this paper.

LEMMA 4. *Z be a d-dimensional diffusion process governed by the stochastic integral equation*

$$Z_t = Z_0 + \int_0^t a(Z_s)\,ds + \int_0^t b(Z_t)\,dW_s,$$

*where* $a(z) = [a_i(z)]_{d\times 1}$ *is a* $d \times 1$ *vector,* $[b_{ij}(z)]_{d\times d}$ *is a* $d \times d$ *diagonal matrix, and* $W_t$ *is a* $d \times 1$ *vector of independent Brownian motions. One assumes that a and b have continuous partial derivatives of order* $2s$*. Let* $f(z)$ *denote a continuous function defined on* $\mathbf{R}^d$ *with values in* $\mathbf{R}^d$ *and with continuous partial derivatives of order* $2s + 2$*. Then*

$$\mathrm{E}[\,f(Z_{i\Delta})|Z_{(i-1)\Delta}] = \sum_{k=0}^{s} L^k f(Z_{(i-1)\Delta})\,\frac{\Delta^k}{k!} + R,$$

*where L is a second-order differential operator defined as*

$$L = \sum_{i=1}^{d} a_i(z)\,\frac{\partial}{\partial z_i} + \frac{1}{2}\sum_{j=1}^{d}\sum_{i=1}^{d} b_{ij}(z)\,\frac{\partial^2}{\partial z_i\,\partial z_j} \quad \left(\frac{\partial f}{\partial z_i} \text{ and } \frac{\partial^2 f}{\partial z_i\,\partial z_j} \text{ are } d\times 1 \text{ vectors}\right)$$

$$= \sum_{i=1}^{d} a_i(z)\,\frac{\partial}{\partial z_i} + \frac{1}{2}\left(b_1^2(z)\,\frac{\partial^2}{\partial z_1^2} + b_2^2(z)\,\frac{\partial^2}{\partial z_2^2} + \cdots + b_d^2(z)\,\frac{\partial^2}{\partial z_d^2}\right)$$

*and R is a stochastic function of order* $\Delta^{s+1}$ *(see Remark 1 in Appendix B for details on the R function).*

A particularly important case is $d = 2$ with

$$dY_t = X_t\,dt,$$

$$dX_t = a(X_t)\,dt + b(X_t)\,dW_t.$$

In this case, the second-order differential operator is $L = x(\partial/\partial y) + a(x)(\partial/\partial x) + \frac{1}{2}b^2(x)(\partial^2/\partial x^2)$. Lemma 4 allows us to calculate a variety of mathematical expectations involving the $\widetilde{X}$ process. For example, using Lemma 4 one can conclude that $\mathrm{E}[\Delta^{-1}(\widetilde{X}_{(i+1)\Delta} - \widetilde{X}_{i\Delta})] = \mathrm{E}[a(X_0)] + O(\Delta)$ and $\mathrm{E}[\Delta^{-1}(\widetilde{X}_{(i+1)\Delta} - \widetilde{X}_{i\Delta})^2] = \frac{2}{3}\mathrm{E}[b^2(X_0)] + O(\Delta)$.

https://doi.org/10.1017/S0266466607070375 Published online by Cambridge University Press

LEMMA 5. *Consider*

$$\varepsilon_{1,n} = \frac{1}{nh_n}\sum_{i=1}^{n} K\left(\frac{x - \widetilde{X}_{(i-1)\Delta_n}}{h}\right) g(\widetilde{X}_{(i-1)\Delta_n}, \widetilde{X}_{i\Delta_n}),$$

$$\varepsilon_{2,n} = \frac{1}{nh_n}\sum_{i=1}^{n} K\left(\frac{x - X_{(i-1)\Delta_n}}{h}\right) g(\widetilde{X}_{(i-1)\Delta_n}, \widetilde{X}_{i\Delta_n}).$$

*Assume* $\sqrt{\Delta_n}/h_n \to 0$, *A1–A3, and A5.*

*(i) If A6(*$\alpha = 4$*) and* $\mathrm{E}[(g(\widetilde{X}_{(i-1)\Delta_n}, \widetilde{X}_{i\Delta_n}))^2] < \infty$ *or*
*(ii) if A6(*$\alpha = 2$*) and* $h_n^{-1}\mathrm{E}[|(\widetilde{X}_{(i-1)\Delta_n} - X_{(i-1)\Delta_n})(g(\widetilde{X}_{(i-1)\Delta_n}, \widetilde{X}_{i\Delta_n}))|^2] < \infty$ *then* $|\varepsilon_{1,n} - \varepsilon_{2,n}| \xrightarrow{p} 0$.

This result simplifies the proofs of Theorems 1–3. With this lemma, we can obtain the limit in probability of $\varepsilon_{1,n}$ by calculating the limit in probability of $\varepsilon_{2,n}$ (which is easier to obtain).

# APPENDIX B: Proofs

**Proof of Lemma 4.** By Itô's formula we have

$$f(Z_{i\Delta}) = f(Z_{(i-1)\Delta}) + \int_{(i-1)\Delta}^{i\Delta} Lf(Z_{u_1})\, du_1 + \int_{(i-1)\Delta}^{i\Delta} \sum_{i=1}^{d} b_i(z)\, \frac{\partial f(Z_{u_1})}{\partial z_i}\, dW_{u_1}.$$

Applying Itô's formula several times to $Lf(Z_{u_1})$ we can conclude, given the properties of the Wiener process, that

$$\mathrm{E}[\, f(Z_{i\Delta}) | Z_{(i-1)\Delta}] = \sum_{k=0}^{s} L^k f(Z_{(i-1)\Delta})\, \frac{\Delta^k}{k!} + R, \tag{B.1}$$

where $R = \int_{(i-1)\Delta}^{i\Delta}\int_{(i-1)\Delta}^{u_1}\int_{(i-1)\Delta}^{u_2}\cdots\int_{(i-1)\Delta}^{u_s}\mathrm{E}[L^{s+1}f(Z_{u_{s+1}})|Z_{(i-1)\Delta}]\, du_{s+1}\, du_s \ldots du_1$. ■

**Remark 1.** Consider the $j$th row of the vector $R$:

$$R_j = \int_{(i-1)\Delta}^{i\Delta}\int_{(i-1)\Delta}^{u_1}\int_{(i-1)\Delta}^{u_2}\cdots\int_{(i-1)\Delta}^{u_s}\mathrm{E}[L^{s+1}f_j(Z_{u_{s+1}})|Z_{(i-1)\Delta}]\, du_{s+1}\, du_s \ldots du_1.$$

Given stationarity, it is obvious that $\mathrm{E}[L^{s+1}f_j(Z_0)] < \infty$ implies $\mathrm{E}[R_j] < \infty$. Assumption A7 guarantees that the expression $\mathrm{E}[L^{s+1}f_j(Z_0)]$ is finite in all applications of Lemma 4 in this paper. For example, in the proof of Lemma 5 we consider

https://doi.org/10.1017/S0266466607070375 Published online by Cambridge University Press

$$\mathrm{E}[|(\widetilde{X}_{(i-1)\Delta_n} - X_{(i-1)\Delta_n})|^4] = \frac{\Delta_n^2 \mathrm{E}[b^4(X_0)]}{3} + \mathrm{E}[R],$$

and $\mathrm{E}[R]$ is of order $\Delta_n^3$. Using Assumption A7, simple but tedious calculations enable us to find that $L^3(\widetilde{X}_0 - X_0)^4$ is bounded by a polynomial degree of order $r = \max\{4\lambda, 1 + 3\lambda, -1 + 5\lambda, -2 + 6\lambda\}$, where $\lambda$ is such that $|a(x)| \le C(1 + |x|)^\lambda$ and $|b(x)| \le C(1 + |x|)^\lambda$. Thus, $\mathrm{E}[X_0^r] < \infty$ guarantees that the expansion (B.1) is well defined. In this paper all expressions involving the application of Lemma 4 (in these cases $s$ is equal to or lower than 2) are bounded by a polynomial degree of order $r$. Therefore, Assumption A7 assures that Lemma 4 is used properly throughout the paper.

**Proof of Lemma 5.** By Taylor's formula

$$K\left(\frac{x - \widetilde{X}_{(i-1)\Delta_n}}{h_n}\right) = K\left(\frac{x - X_{(i-1)\Delta_n}}{h_n}\right) + K'(\xi_{n,i}) \frac{(\widetilde{X}_{(i-1)\Delta_n} - X_{(i-1)\Delta_n})}{h_n},$$

where $\xi_{n,i} = \theta((x - X_{(i-1)\Delta_n})/h_n) + (1 - \theta)((x - \widetilde{X}_{i\Delta_n})/h_n)$, $0 \le \theta \le 1$, stationarity, and Hölder's inequality, we have

$$\begin{aligned}
&\mathrm{E}[|\varepsilon_{1,n} - \varepsilon_{2,n}|] \\
&\quad = \mathrm{E}\left[\left|\frac{1}{nh_n}\sum_{i=1}^{n}\left(K\left(\frac{x - \widetilde{X}_{(i-1)\Delta}}{h}\right) - K\left(\frac{x - X_{(i-1)\Delta}}{h}\right)\right) g(\widetilde{X}_{(i-1)\Delta}, \widetilde{X}_{i\Delta})\right|\right] \\
&\quad \le \mathrm{E}\left[\left|\frac{1}{h_n}\left(K\left(\frac{x - \widetilde{X}_{(i-1)\Delta}}{h}\right) - K\left(\frac{x - X_{(i-1)\Delta}}{h}\right)\right) g(\widetilde{X}_{(i-1)\Delta}, \widetilde{X}_{i\Delta})\right|\right] \\
&\quad = \mathrm{E}\left[\left|\frac{1}{h_n} K'(\xi_{n,i}) \frac{(\widetilde{X}_{(i-1)\Delta_n} - X_{(i-1)\Delta_n})}{h_n} g(\widetilde{X}_{(i-1)\Delta}, \widetilde{X}_{i\Delta})\right|\right] \\
&\quad \le \left(\frac{1}{h_n}\mathrm{E}[|K'(\xi_{n,i})|^\alpha]^{1/\alpha}\right) \times \left(\frac{1}{h_n}\mathrm{E}[|(\widetilde{X}_{(i-1)\Delta_n} - X_{(i-1)\Delta_n})|^\beta]^{1/\beta}\right) \\
&\qquad \times (\mathrm{E}[|g(\widetilde{X}_{(i-1)\Delta}, \widetilde{X}_{i\Delta})|^\gamma]^{1/\gamma}),
\end{aligned}$$

where $1/\alpha + 1/\beta + 1/\gamma = 1$. Selecting $\alpha = 4$, $\beta = 4$, and $\gamma = 2$ we have by hypothesis

$$\left(\frac{1}{h_n}\mathrm{E}[|K'(\xi_{n,i})|^4]^{1/4}\right) < \infty, \qquad (\mathrm{E}[|g(\widetilde{X}_{(i-1)\Delta}, \widetilde{X}_{i\Delta})|^2]^{1/2}) < \infty,$$

and by Lemma 4

$$\frac{1}{h_n}\mathrm{E}[|(\widetilde{X}_{(i-1)\Delta_n} - X_{(i-1)\Delta_n})|^4]^{1/4} = \frac{1}{h_n}\left(\frac{\Delta_n^2 \mathrm{E}[b^4(X_0)]}{3} + O(\Delta_n^3)\right)^{1/4} \to 0.$$

Thus $\mathrm{E}[|\varepsilon_{1,n} - \varepsilon_{2,n}|] \to 0$. Part (ii) is immediate from Cauchy–Schwarz inequality (in this case it is sufficient to consider $\lim_{h\to 0}\mathrm{E}[(1/h)|K'(\xi_{n,i})|^2] < \infty$). ■

https://doi.org/10.1017/S0266466607070375 Published online by Cambridge University Press

**Proof of Theorem 1.** Let

$$\hat{p}_n^0(x) = \frac{1}{nh_n}\sum_{i=1}^{n} K\left(\frac{x - X_{(i-1)\Delta_n}}{h_n}\right).$$

It is well known that under the conditions of the theorem one has

$$\mathrm{E}[|\hat{p}_n^0(x) - p(x)|] \to 0$$

(see, e.g., Bandi and Phillips, 2003). It remains to show that $\mathrm{E}[|\hat{p}_n(x) - \hat{p}_n^0(x)|] \to 0$. This follows immediately from Lemma 5(ii) with $g \equiv 1$. Therefore, the right-hand side of the inequality

$$\mathrm{E}[|\hat{p}_n(x) - p(x)|] \le \mathrm{E}[|\hat{p}_n(x) - \hat{p}_n^0(x)|] + \mathrm{E}[|\hat{p}_n^0(x) - p(x)|]$$

goes to zero (under the conditions of the theorem). Thus, we have $\mathrm{E}[|\hat{p}_n(x) - p(x)|] \to 0$ and, by Chebyshev's inequality, $\hat{p}_n(x) \xrightarrow{p} p(x)$. ■

**Proof of Theorem 2.** (i) From the proof of Theorem 1 we know that $\hat{p}_n(x) \xrightarrow{p} p(x)$. Thus, to prove $\hat{a}_n(x) = A_n(x)/\hat{p}_n(x) \xrightarrow{p} a(x)$ it is sufficient to verify that $A_n(x) \xrightarrow{p} a(x)p(x)$. Let us consider

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

Under the conditions of the theorem it is known that $A_n^0(x) \xrightarrow{p} a(x)p(x)$ (see Bandi and Phillips, 2003). The idea is to prove that $A_n(x) - A_n^0(x) \xrightarrow{p} 0$, which implies that $A_n(x) \xrightarrow{p} a(x)p(x)$ and consequently $\hat{a}_n(x) \xrightarrow{p} a(x)$. By Lemma 5(ii) $A_n(x) - A_n^0(x)$ has the same limit in probability as

$$\begin{aligned}\delta_{1,n} &= \frac{1}{nh_n}\sum_{i=1}^{n} K\left(\frac{x - X_{(i-1)\Delta_n}}{h_n}\right)\frac{(\widetilde{X}_{(i+1)\Delta_n} - \widetilde{X}_{i\Delta_n})}{\Delta_n} \\ &\quad - \frac{1}{nh_n}\sum_{i=1}^{n} K\left(\frac{x - X_{(i-1)\Delta_n}}{h_n}\right)\frac{(X_{i\Delta_n} - X_{(i-1)\Delta_n})}{\Delta_n} \\ &= \frac{1}{nh_n}\sum_{i=1}^{n} K\left(\frac{x - X_{(i-1)\Delta_n}}{h_n}\right)\left(\frac{(\widetilde{X}_{(i+1)\Delta_n} - \widetilde{X}_{i\Delta_n})}{\Delta_n} - \frac{(X_{i\Delta_n} - X_{(i-1)\Delta_n})}{\Delta_n}\right).\end{aligned}$$

https://doi.org/10.1017/S0266466607070375 Published online by Cambridge University Press

We now show that $\delta_{1,n}(x) \xrightarrow{p} 0$. By Lemma 4, A7, and stationarity, we have

$$\begin{aligned}
&\mathrm{E}[\delta_{1,n}(x)] \\
&= \mathrm{E}\left[\frac{1}{h_n} K\left(\frac{x - X_{(i-1)\Delta_n}}{h_n}\right)\left(\frac{(\widetilde{X}_{(i+1)\Delta_n} - \widetilde{X}_{i\Delta_n})}{\Delta_n} - \frac{(X_{i\Delta_n} - X_{(i-1)\Delta_n})}{\Delta_n}\right)\right] \\
&= \mathrm{E}\left[\frac{1}{h_n} K\left(\frac{x - X_{(i-1)\Delta_n}}{h_n}\right)\mathrm{E}\left[\mathrm{E}\left[\left(\frac{(\widetilde{X}_{(i+1)\Delta_n} - \widetilde{X}_{i\Delta_n})}{\Delta_n} - \frac{(X_{i\Delta_n} - X_{(i-1)\Delta_n})}{\Delta_n}\right)\middle| \mathcal{F}_{i\Delta_n}\right]\middle| \mathcal{F}_{(i-1)\Delta_n}\right]\right] \\
&= \frac{\Delta_n}{2}\mathrm{E}\left[\frac{1}{h_n} K\left(\frac{x - X_{(i-1)\Delta_n}}{h_n}\right)\left(a(X_{(i-1)\Delta_n})a'(X_{(i-1)\Delta_n}) + \frac{1}{2} b^2(X_{(i-1)\Delta_n})a''(X_{(i-1)\Delta_n})\right)\right] \\
&= \frac{\Delta_n}{2}\int \frac{1}{h_n} K\left(\frac{x-u}{h_n}\right)\left(a(u)a'(u) + \frac{1}{2} b^2(u)a''(u)\right)p(u)\,du \\
&= O(\Delta_n).
\end{aligned}$$

On the other hand,

$$\begin{aligned}
&\mathrm{Var}[\delta_{1,n}(x)] \\
&= \frac{1}{n\Delta_n h_n}\mathrm{Var}\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n} \underbrace{\frac{1}{\sqrt{h_n}} K\left(\frac{x - X_{(i-1)\Delta_n}}{h_n}\right)\sqrt{\Delta_n}\left(\frac{(\widetilde{X}_{(i+1)\Delta_n} - \widetilde{X}_{i\Delta_n})}{\Delta_n} - \frac{(X_{i\Delta_n} - X_{(i-1)\Delta_n})}{\Delta_n}\right)}_{g_i}\right] \\
&= \frac{1}{n\Delta_n h_n}\mathrm{Var}\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n} g_i\right].
\end{aligned}$$

We have $\mathrm{Var}[(1/\sqrt{n})\sum_{i=1}^{n} g_i] = (1/n)\sum_{i=1}^{n}\mathrm{E}[(1/h)K^2((x - X_{(i-1)\Delta_n})/h_n)f^2] + \gamma_n$ where $\gamma_n$ represents the sum of $(2/n)\sum_{j=1}^{n-1}\sum_{i=j+1}^{n}$ terms involving the autocovariances. Under stationarity and Assumption A4 ($X$ and $\widetilde{X}$ are $\rho$-mixing and $\alpha$-mixing), the $j$th autocovariance of $g_i$ tends to zero at exponential rate as $j \to \infty$, for all functions $g$ such that $\mathrm{E}[g_i^2] < \infty$. Therefore, if $\mathrm{E}[g_i^2] < \infty$, the series $\gamma_\infty$ is absolutely convergent, and one has $\mathrm{Var}[(1/\sqrt{n})\sum_{i=1}^{n} g_i] < \infty$. Because $\int K^2(u)\,du < \infty$ and $X$ has stationary density, one can conclude that

$$\mathrm{E}[g_i^2] = \mathrm{E}\left[\frac{1}{h_n} K^2\left(\frac{x - X_{(i-1)\Delta_n}}{h_n}\right)\Delta_n\left(\frac{(\widetilde{X}_{(i+1)\Delta_n} - \widetilde{X}_{i\Delta_n})}{\Delta_n} - \frac{(X_{i\Delta_n} - X_{(i-1)\Delta_n})}{\Delta_n}\right)^2\right]$$

is finite because, by Lemma 4,

$$\mathrm{E}\left[\Delta_n\left(\frac{(\widetilde{X}_{(i+1)\Delta_n} - \widetilde{X}_{i\Delta_n})}{\Delta_n} - \frac{(X_{i\Delta_n} - X_{(i-1)\Delta_n})}{\Delta_n}\right)^2\right] = \frac{2}{3}\mathrm{E}[b^2(X_0)] + O(\Delta_n).$$

In conclusion, $\mathrm{Var}[\delta_{1,n}(x)] = (1/n\Delta_n h_n)\mathrm{Var}[(1/\sqrt{n})\sum_{i=1}^{n} g_i] \to 0$ as $n\Delta_n h_n \to \infty$.

(ii) By hypothesis (see Bandi and Phillips, 2003)

$$Z_n^0(x) = \sqrt{h_n n\Delta_n}(\hat{a}_n^0(x) - a(x)) \xrightarrow{d} N\left(0, K_2\frac{b^2(x)}{p(x)}\right), \tag{B.2}$$

https://doi.org/10.1017/S0266466607070375 Published online by Cambridge University Press

where $\hat{a}_n^0(x) = A_n^0(x)/\hat{p}_n(x)$. By the asymptotic equivalence theorem, it suffices to prove that $Z_n(x) - Z_n^0(x) \xrightarrow{p} 0$ where $Z_n(x) = \sqrt{h_n n\Delta_n}(\hat{a}_n(x) - a(x))$. From part (i) we know that

$$Z_n(x) - Z_n^0(x) = \sqrt{h_n n\Delta_n}\left(\frac{\delta_{1,n}(x)}{\hat{p}_n(x)}\right)$$

and $\delta_{1,n}(x)/\hat{p}_n(x) = O_p(\Delta_n)$. Thus, under the assumption $h_n n\Delta_n^3 \to 0$, the result holds. ■

**Proof of Theorem 3.** (i) From the proof of Theorem 1 we know that $\hat{p}_n(x) \xrightarrow{p} p(x)$. Thus, to prove $\hat{b}_n^2(x) = B_n(x)/\hat{p}_n(x) \xrightarrow{p} b^2(x)$ it is sufficient to verify that $B_n(x) \xrightarrow{p} b^2(x)p(x)$. Let us consider

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

Under the conditions of the theorem it is known that $B_n^0(x) \xrightarrow{p} b^2(x)p(x)$ (see Bandi and Phillips, 2003). The idea is to prove that $B_n(x) - B_n^0(x) \xrightarrow{p} 0$, which implies that $B_n(x) \xrightarrow{p} b^2(x)p(x)$ and consequently $\hat{b}_n^2(x) \xrightarrow{p} b^2(x)$. By Lemma 5(i) $B_n(x) - B_n^0(x)$ has the same limit in probability as

$$\begin{aligned}\delta_{2,n}(x) &= \frac{1}{nh_n}\sum_{i=1}^{n} K\left(\frac{x - X_{(i-1)\Delta_n}}{h_n}\right)\frac{\frac{3}{2}(\widetilde{X}_{i\Delta_n} - \widetilde{X}_{(i-1)\Delta_n})^2}{\Delta_n} \\ &\quad - \frac{1}{nh_n}\sum_{i=1}^{n} K\left(\frac{x - X_{(i-1)\Delta_n}}{h_n}\right)\frac{(X_{i\Delta_n} - X_{(i-1)\Delta_n})^2}{\Delta_n} \\ &= \frac{1}{nh_n}\sum_{i=1}^{n} K\left(\frac{x - X_{(i-1)\Delta_n}}{h_n}\right)\left(\frac{\frac{3}{2}(\widetilde{X}_{i\Delta_n} - \widetilde{X}_{(i-1)\Delta_n})^2}{\Delta_n} - \frac{(X_{i\Delta_n} - X_{(i-1)\Delta_n})^2}{\Delta_n}\right).\end{aligned}$$

We now prove that $\delta_{2,n}(x) \xrightarrow{p} 0$. Let us consider

$$\delta_{3,n}(x) = \frac{1}{nh_n}\sum_{i=1}^{n} K\left(\frac{x - X_{(i-1)\Delta_n}}{h_n}\right) f_i,$$

where $f_i = ((\frac{3}{2}(\widetilde{X}_{i\Delta_n} - \widetilde{X}_{(i-1)\Delta_n})^2/\Delta_n) - ((X_{i\Delta_n} - X_{(i-1)\Delta_n})^2/\Delta_n))$. Because by Lemma 4

$$\mathrm{E}\left[\mathrm{E}\left[\frac{\frac{3}{2}(\widetilde{X}_{i\Delta_n} - \widetilde{X}_{(i-1)\Delta_n})^2}{\Delta_n}\,\middle|\,\mathcal{F}_{i\Delta_n}\right]\middle|\,\mathcal{F}_{(i-1)\Delta_n}\right] = b^2(X_{(i-1)\Delta_n}) + O(\Delta_n),$$

$$\mathrm{E}\left[\frac{(X_{i\Delta_n} - X_{(i-1)\Delta_n})^2}{\Delta_n}\,\middle|\,\mathcal{F}_{i\Delta_n}\right] = b^2(X_{(i-1)\Delta_n}) + O(\Delta_n),$$

https://doi.org/10.1017/S0266466607070375 Published online by Cambridge University Press

one has

$$\mathrm{E}[f_i] = \mathrm{E}\left[\left(\frac{\frac{3}{2}(\widetilde{X}_{i\Delta_n} - \widetilde{X}_{(i-1)\Delta_n})^2}{\Delta_n} - \frac{(X_{i\Delta_n} - X_{(i-1)\Delta_n})^2}{\Delta_n}\right)\right] = O(\Delta_n).$$

On the other hand,

$$\begin{aligned}
&\mathrm{Var}[\delta_{2,n}(x)] \\
&= \frac{1}{nh_n}\mathrm{Var}\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\underbrace{\frac{1}{\sqrt{h_n}}K\left(\frac{x - X_{(i-1)\Delta_n}}{h_n}\right)\left(\frac{\frac{3}{2}(\widetilde{X}_{i\Delta_n} - \widetilde{X}_{(i-1)\Delta_n})^2}{\Delta_n} - \frac{(X_{i\Delta_n} - X_{(i-1)\Delta_n})^2}{\Delta_n}\right)}_{g_i}\right] \\
&= \frac{1}{nh_n}\mathrm{Var}\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n} g_i\right].
\end{aligned}$$

Using the same arguments as in the proof of Theorem 2(i), it is easy to conclude that the condition

$$\mathrm{E}\left[\left(\frac{\frac{3}{2}(\widetilde{X}_{i\Delta_n} - \widetilde{X}_{(i-1)\Delta_n})^2}{\Delta_n} - \frac{(X_{i\Delta_n} - X_{(i-1)\Delta_n})^2}{\Delta_n}\right)^2\right] = O(1)$$

assures $\mathrm{Var}[\delta_{2,n}(x)] \to 0$.

(ii) By hypothesis (see Bandi and Phillips, 2003)

$$Z_n^0(x) = \sqrt{nh_n}((\hat{b}_n^2)^0(x) - b^2(x)) \xrightarrow{d} N\left(0, \frac{4K_2 b^4(x)}{p(x)}\right),$$

where $(\hat{b}_n^2)^0(x) = B_n^0(x)/\hat{p}_n(x)$. By the asymptotic equivalence theorem it is sufficient to prove that $Z_n(x) - Z_n^0(x) \xrightarrow{p} 0$ where $Z_n(x) = \sqrt{nh_n}(\hat{b}_n^2(x) - b^2(x))$. From part (i) we know that

$$Z_n(x) - Z_n^0(x) = \sqrt{nh_n}\,\frac{\delta_{2,n}(x)}{\hat{p}_n(x)}$$

and $\delta_{3,n}(x) = O_p(\Delta_n)$. Thus, under the assumption $h_n n\Delta_n^2 \to 0$ the result holds. ■

https://doi.org/10.1017/S0266466607070375 Published online by Cambridge University Press