Accelerated Asymptotics for Diffusion Model Estimation*

Federico M. Bandi
GSB, The University of Chicago
and
Peter C. B. Phillips
Cowles Foundation for Research in Economics
Yale University

First Draft: October 17, 1999

Abstract

We propose a semiparametric estimation procedure for scalar homogeneous stochastic differential equations. We specify a parametric class for the underlying diffusion process and identify the parameters of interest by minimizing criteria given by the integrated squared difference between kernel estimates of drift and diffusion function and their parametric counterparts. The nonparametric estimates are simplified versions of those in Bandi and Phillips (1998). A complete asymptotic theory for the semiparametric estimates is developed. The limit theory relies on *infill* and *long span* asymptotics and the asymptotic distributions are shown to depend on the *chronological local time* of the underlying diffusion process. The estimation method and asymptotic results apply to both stationary and nonstationary processes. As is standard with semiparametric approaches in other contexts, faster convergence rates are attained than is possible in the fully functional case.

Keywords: Diffusion, Drift, Infill asymptotics, Kernel density, Local time, Martingale, Nonparametric estimation, Semimartingale, Semiparametric estimation, Stochastic differential equation.

JEL Classification: C14, C22

*PRELIMINARY and INCOMPLETE VERSION. Prepared for the Cowles Foundation Conference Series: New Developments in Time Series Econometrics, Yale University, October 23-24 1999. Comments are welcome and may be sent to federico.bandi@gsb.uchicago.edu.

1 Introduction

The estimation of continuous-time models, such as those described by potentially nonlinear stochastic differential equations, has been intensively studied in recent research. Stanton (1998) provides a recent concise survey and discussion of applications in finance. In the last few years, this literature has shown a tendency to turn to fully functional procedures to identify and estimate the two functions that describe the solution to the stochastic differential equation of interest, that is the drift and diffusion functions [c.f. Jiang and Knight (1997), Stanton (1998) and Bandi and Phillips (1998) – hereafter BP, *inter alia*]. The motivation for this focus is clear. By not imposing a specific parametric structure, fully functional methods reduce the extent of potential misspecifications. Unfortunately, they do so at the expense of slower convergence rates and the potential of greater estimation error over their parametric counterparts. Yet, the informational content of accurately implemented functional methods can be put to work as a useful descriptive tool to understand more about the underlying dynamics from a general perspective and to investigate more effective procedures for parametric inference.

This paper seeks to design an estimation methodology that exploits the generality of functional methods while improving on their convergence properties. Also, we wish to utilize any available information about possible parametrizations for the two functions of interest. A natural way to proceed is to define a semiparametric estimation procedure that matches functional estimates to their parametric counterparts. In order to do so, we specify a parametric class for the underlying diffusion process and estimate the parameters of interest by minimizing two criteria which can be readily interpreted as the integrated squared differences between kernel estimates of the drift and diffusion functions and their corresponding parametric expressions.

The nonparametric estimates we use here are simplified versions of those in BP (1998). As discussed by the authors in previous work [BP (1998)], drift and diffusion functions can be identified separately using functional analogues of the true theoretical functions. Only minimal requirements need to be placed on the data generating mechanism for this approach to be justified. In particular, we do not require the existence of a time-invariant marginal data density, so stationarity is not needed.

The present work develops an asymptotic theory for the new semiparametric estimates. The limit theory relies on *infill* and *long span* asymptotics, just as in the fully nonparametric case [c.f. BP (1998) and Bandi (1999)], and the asymptotic distributions are shown to depend on the *chronological local time* of the underlying diffusion process, that is on the time that the process spends in the vicinity of each spatial point [see Protter (1990) for a general discussion and Phillips and Park (1998) for an introduction to this concept in econometrics]. As expected, semiparametric methods entail efficiency gains with respect to fully functional procedures by virtue of their faster convergence rates. The same intuition as in the standard semiparametric regression context carries over to the continuous-time model examined in this paper [c.f. Andrews (1989)].

>From a purely technical point of view, this work merges two strands of the most recent econometrics literature, namely the estimation of nonlinear models of integrated time-series [Park and Phillips (1998a,b)] and the functional identification of diffusions under minimal assumptions on the dynamics of the underlying process [Florens-Zmirou (1993), Jacod (1997) and BP (1998)]. In effect, the 'minimum distance' type of estimation that is presented in this paper can be interpreted as extremum estimation for potentially nonstationary and nonlinear continuous-time models.

The paper proceeds as follows. Section 2 discusses the model and objects of econometric interest. Section 3 details the estimation procedure. Section 4 presents the main results and Section 5 concludes. Proofs are in Section 6.

2 The model

We consider a diffusion process $\{X_t : t \geq 0\}$ generated by

$$dX_t = \mu(X_t, \theta^{drift})dt + \sigma(X_t, \theta^{diff})dB_t, \tag{1}$$

with initial condition $X_0 = \overline{X}$ and where B_t is a standard Brownian motion defined on the filtered probability space $(\Omega, \Im^B, (\Im_t^B)_{t\geq 0}, P)$. The initial condition $\overline{X} \in L^2$ and is taken to be independent of $\{B_t : t \geq 0\}$. The parameter vectors θ^{drift} and θ^{diff} are such that $\left(\theta^{drift}, \theta^{diff}\right) = \theta \in \Theta$ where Θ is a compact subset of $\Re^M$ for a generic M. More specifically, $\theta^{drift} \in \Theta^{drift} \subset \Re^{m_1}$ and $\theta^{diff} \in \Theta^{diff} \subset \Re^{m_2}$ with $m_1 + m_2 = M$. The vectors θ^{drift} and θ^{diff} jointly define a parametric family for (1).

We now define the left-continuous filtration

$$\overline{\Im}_t := \sigma(\overline{X}) \vee \Im_t^B = \sigma(\overline{X}, B_s; 0 \leq s \leq t) \qquad 0 \leq t < \infty$$

and the collection of null sets

$$\aleph := \{N \subseteq \Omega; \exists G \in \overline{\Im}_\infty \ with \ N \subseteq G \ and \ P(G) = 0\}.$$

We create the *augmented* filtration

$$\widetilde{\Im}_t^X := \sigma(\overline{\Im}_t \cup \aleph) \qquad 0 \leq t < \infty.$$

As in BP (1998), the following conditions are used in the study of (1). They guarantee the existence and pathwise uniqueness of a nonexplosive solution to (1) that is adapted to the augmented filtration $\{\widetilde{\Im}_t^X\}$.

2.1 Assumption

(A) *$\mu(\cdot, \theta^{drift})$ and $\sigma(\cdot, \theta^{diff})$ are time-homogeneous, $\mathfrak{B}$-measurable functions on $\mathfrak{D} = (l, u)$ with $-\infty \leq l < u \leq \infty$ where $\mathfrak{B}$ is the σ-field generated by Borel sets on $\mathfrak{D}$. Both functions are at least twice continuously differentiable. Hence, they satisfy local Lipschitz and growth conditions. Thus, for every compact subset $J = [1/H, H]$ with $H > 0$ of the range of the process, there exist constants C_1 and C_2 such that, for all x and y in J,*

$$|\mu(x, \theta^{drift}) - \mu(y, \theta^{drift})| + |\sigma(x, \theta^{diff}) - \sigma(y, \theta^{diff})| \leq C_1|x - y|,$$

and

$$|\mu(x, \theta^{drift})| + |\sigma(x, \theta^{diff})| \leq C_2\{1 + |x|\}.$$

(B) *$\sigma^2(\cdot, \theta^{diff}) > 0$ on $\mathfrak{D}$.*

(C) [Feller's (1952) necessary and sufficient condition for nonexplosion]. *We define $V(\alpha, \theta)$ as*

$$\int_0^\alpha S'(y, \theta)\{\int_0^y \left[\frac{2}{S'(x, \theta)\sigma^2(x, \theta^{diff})}\right] dx\}dy$$

where $S'(x,\theta)$ is the first derivative of the natural scale measure,

$$S(\alpha,\theta)=\int_0^{\alpha}\exp\{\int_0^{y}\left[-\frac{2\mu(x,\theta^{drift})}{\sigma^2(x,\theta^{diff})}\right]dx\}dy.$$

We require $V(\alpha,\theta)$ to diverge at the boundaries of $\mathfrak{D}$, i.e.

$$\lim_{\alpha\to l^+}V(\alpha,\theta)=\lim_{\alpha\to u^-}V(\alpha,\theta)=\infty.$$

(D) *$\mu(\cdot,\theta^{drift})$ and $\sigma(\cdot,\theta^{diff})$ are at least twice continuously differentiable in θ^{drift} and θ^{diff}.*

As discussed in BP (1998), under conditions (A), (B) and (C), the stochastic differential equation has a strong solution X_t that is unique, recurrent and continuous in $t\in[0,T]$. Assumption (D) will be used in the development of our asymptotics.

The objects of econometric interest are the drift, $\mu(.,\theta^{drift})$, and the diffusion term, $\sigma^2(.,\theta^{diff})$. Their conditional moment definitions are well known. They can be interpreted as representing the 'instantaneous' conditional mean and the 'instantaneous' conditional variance of increments in the process [c.f. Karlin and Taylor (1981), for example]. More precisely, $\mu(.,\theta^{drift})$ describes the conditional expected rate of change of the process for infinitesimal time changes, whereas $\sigma^2(.,\theta^{diff})$ gives the conditional rate of change of volatility.

3 The Econometric procedure

We define a 'minimum distance' type of estimation that exploits the consistency of accurately defined functional estimators and provides estimates of the parameters of interest by 'matching' the parametric expressions to their nonparametric counterparts.

The first step consists of defining the functional estimates. We consider a simplified version of the estimators in BP (1998). Assume the data X_t is recorded discretely at $\{t=t_1,t_2,..,t_n\}$ in the time interval $[0,T]$, with $T\geq T_0>0$, where T_0 is a positive constant. Also, assume equispaced data. Hence,

$$\{X_t=X_{\Delta_{n,T}},X_{2\Delta_{n,T}},X_{3\Delta_{n,T}},...,X_{n\Delta_{n,T}}\}$$

are n observations at

$$\{t_1=\Delta_{n,T},t_2=2\Delta_{n,T},t_3=3\Delta_{n,T},...,t_n=n\Delta_{n,T}\}$$

where $\Delta_{n,T}=T/n$. The diffusion function estimator is

$$\widehat{\sigma}^2_{(n,T)}(x)=\frac{1}{\Delta_{n,T}}\frac{\sum_{i=1}^{n-1}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})[X_{(i+1)\Delta_{n,T}}-X_{i\Delta_{n,T}}]^2}{\sum_{i=1}^{n}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})}. \quad (2)$$

The drift function estimator is

$$\widehat{\mu}_{(n,T)}(x)=\frac{1}{\Delta_{n,T}}\frac{\sum_{i=1}^{n-1}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})[X_{(i+1)\Delta_{n,T}}-X_{i\Delta_{n,T}}]}{\sum_{i=1}^{n}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})}. \quad (3)$$

Remark 3.1 The estimators above are defined as rather straightforward sample analogues of the theoretical functions. BP (1998) discuss the properties of consistency and asymptotic mixed normality of more general functional analogues to the true functions than (2) and (3) above. They show that recurrence (which is implied by Assumption 2.1 above), rather than stationarity, is all that is needed to achieve identification. They derive the asymptotics as (i) the time span (T) and *(ii)* the number of data points (n) increase with *(iii)* the frequency of observations ($\frac{T}{n} \rightarrow 0$). Condition *(iii)* is necessary for the consistent estimation of continuous-time models using fully functional methods. Condition (i) is crucial only for drift estimation as the local dynamic of the process does not contain sufficient information to identify its infinitesimal first moment. By letting the time span increase to infinity, the drift can be recovered in the limit since the process continues to make repeated visits to different spatial points by virtue of recurrence.

Remark 3.2 Formulae (2) and (3) can be interpreted as estimates of Stanton's first order approximations to the infinitesimal moments of a diffusion [c.f. Stanton (1997)]. They are proven to be consistent and asymptotically mixed normal under the same conditions on the time span and the sample size that were described in the previous remark [c.f. Bandi (1999)].

Remark 3.3 More general sample analogues to the true functions of the type described in BP (1998) could be used to derive the functional estimates. We decided to employ specifications based on simple smoothing rather than on convoluted kernels [as in the most general case examined by BP (1998)] for simplicity. In effect, the use of more involved specifications would not qualitatively change the asymptotic results given here.

In finite samples the use of convoluted kernels does make a difference [c.f. Bandi and Nguyen (1999)]. In particular, we know that the choice of the optimal smoothing parameter for the drift is empirically cumbersome. Yet, the use of convoluted kernels limit the effects of potentially suboptimal choices. Extension to more general kernels can be easily derived from the apparatus discussed below.

Consider the criteria

$$Q_{n,T}^{drift} = \frac{T}{n} \sum_{i=1}^{n} \left(\widehat{\mu}_{(n,T)}(X_{i\Delta n,T}) - \mu(X_{i\Delta n,T}, \theta^{drift}) \right)^2 \tag{4}$$

and

$$Q_{n,T}^{diff} = \frac{T}{n} \sum_{i=1}^{n} \left(\widehat{\sigma}^2_{(n,T)}(X_{i\Delta n,T}) - \sigma^2(X_{i\Delta n,T}, \theta^{diff}) \right)^2 \tag{5}$$

where $\widehat{\mu}_{(n,T)}(X_{i\Delta n,T})$ and $\widehat{\sigma}^2_{(n,T)}(X_{i\Delta n,T})$ are defined in (3) and (2), respectively. Then, the semi-parametric estimates $\widehat{\theta}^{drift}_{n,T}$ and $\widehat{\theta}^{diff}_{n,T}$ are obtained as follows:

$$\begin{aligned} \widehat{\theta}^{drift}_{n,T} \quad : \quad &= \underset{\theta^{drift} \in \Theta^{drift} \subset \Theta}{arg\ min} \quad Q_{n,T}^{drift} \\ &= \underset{\theta^{drift} \in \Theta^{drift} \subset \Theta}{arg\ min} \ \frac{T}{n} \sum_{i=1}^{n} \left(\widehat{\mu}_{(n,T)}(X_{i\Delta n,T}) - \mu(X_{i\Delta n,T}, \theta^{drift}) \right)^2 \end{aligned} \tag{6}$$

and

$$\begin{aligned}\widehat{\theta}_{n,T}^{diff} &:= \underset{\theta^{diff}\in\Theta^{diff}\subset\Theta}{arg\ min}\ Q_{n,T}^{diff} \\ &= \underset{\theta^{diff}\in\Theta^{diff}\subset\Theta}{arg\ min}\ \frac{T}{n}\sum_{i=1}^{n}\left(\widehat{\sigma}^2_{(n,T)}(X_{i\Delta n,T})-\sigma^2(X_{i\Delta n,T},\theta^{diff})\right)^2. \end{aligned} \tag{7}$$

Remark 3.4 As in the fully nonparametric case, we identify the drift and diffusion parameters ($\widehat{\theta}_{n,T}^{drift}$ and $\widehat{\theta}_{n,T}^{diff}$, that is) separately. This is of particular importance when we are interested in the parametrization of a specific function in situations where the other function is treated as a nuisance parameter.

Remark 3.5 In the next section we derive the consistency and asymptotic mixed normality of $\widehat{\theta}_{n,T}^{drift}$ and $\widehat{\theta}_{n,T}^{diff}$ as $T\to\infty$, $n\to\infty$ and $\frac{T}{n}\to 0$. As in the fully nonparametric case, we show that $\widehat{\theta}_{n,T}^{diff}$ can be identified over a fixed time span ($T=\overline{T}$) provided $\frac{\overline{T}}{n}\to 0$, that is as the frequency of observations increases.

4 Limit theory

First, we consider the drift case.

Theorem 4.1 (Consistency of the drift estimates) *Assume $n\to\infty$, $T\to\infty$ and $h_{n,T}\to 0$ (as $n,T\to\infty$) such that $\frac{\Delta_{n,T}}{h_{n,T}}\to 0$ and $\frac{\overline{L}_X(T,x)}{h_{n,T}}(\Delta_{n,T})^{\alpha}=O_{a.s.}(1)$ for some $\alpha\in(0,\frac{1}{2})$, then*

$$Q_{n,T}^{drift}(\theta^{drift})\overset{a.s.}{\to}Q^{drift}(\theta^{drift},\theta_0^{drift})=\int_{-\infty}^{\infty}\left(\mu(s,\theta_0^{drift})-\mu(s,\theta^{drift})\right)^2\overline{L}_X(T,s)ds, \tag{8}$$

uniformly in θ^{drift}. Also, let $B(\theta^{drift},\varepsilon)$ denote an open ball in Θ^{drift} of radius ε around θ^{drift}. Assume that for $\forall\delta>0$ and $\forall\varepsilon>0$

$$\inf_{\theta^{drift}\notin B\left(\theta_0^{drift},\varepsilon\right)}\int_{|s|\le\delta}\left(\mu(s,\theta_0^{drift})-\mu(s,\theta^{drift})\right)^2 ds>0. \tag{9}$$

Then,

$$\widehat{\theta}_{n,T}^{drift}\overset{a.s.}{\to}\theta_0^{drift}.$$

Theorem 4.2 (The limit distribution of the drift estimates) *Given $n\to\infty$, $T\to\infty$ and $h_{n,T}\to 0$ (as $n,T\to\infty$) such that $\frac{\Delta_{n,T}}{h_{n,T}}\to 0$ and $\frac{\overline{L}_X(T,x)}{h_{n,T}}(\Delta_{n,T})^{\alpha}=O_{a.s.}(1)$ for some $\alpha\in(0,\frac{1}{2})$, then*

$$(\mathbf{\Xi}_{mu}(T))^{-1/2}\left(\widehat{\theta}_{n,T}^{drift}-\theta_0^{drift}\right)\to_d N(0,\mathbf{I}), \tag{10}$$

where

$$\mathbf{\Xi}_{mu}(T)=\mathbf{B}(T)_{mu}^{-1}\mathbf{\Omega}(T)_{mu}\mathbf{B}(T)_{mu}^{-1},$$

and

$$\mathbf{B}_{mu} = \left(\int_{-\infty}^{\infty} \frac{\partial \mu_0(a)}{\partial \theta} \frac{\partial \mu_0(a)}{\partial \theta'} \overline{L}_X(T,a) da \right),$$

$$\mathbf{\Omega}_{mu} = \left(\int_{-\infty}^{\infty} \sigma^2(a) \left(\frac{\partial \mu_0(a)}{\partial \theta} \frac{\partial \mu_0(a)}{\partial \theta'} \right) \overline{L}_X(T,a) da \right).$$

Remark 4.3 The result is consistent with what we would expect to obtain in a correctly specified standard nonlinear regression context with heteroschedastic errors [c.f. Davidson and MacKinnon (1993) for a classical treatment]. The only difference is that we replace integrals with respect to probability measures with spatial integrals [c.f. Park and Phillips (1998b)]. This is due to the generality of the approach adopted here. In particular, it is due to the robustness to deviations from stationarity.

Remark 4.4 Coherently with the fully nonparametric case, the rate of convergence is path-dependent and is driven by the rate of divergence to infinity of the chronological local time of the underlying process through the integrals $\mathbf{B}_{mu}$ and $\mathbf{\Omega}_{mu}$. By virtue of the averaging, the rate is faster than in fully functional context, i.e. $\sqrt{h_{n,T}\overline{L}_X(T,x)}$. Consistently with this result, the conditions that guarantee a.s. convergence and asymptotic mixed normality are less stringent. In particular, we do not require the bandwidth parameter $h_{n,T}$ to converge to zero slowly enough to guarantee that $h_{n,T}\overline{L}_X(T,x) \overset{a.s.}{\to} \infty$ as $\overline{L}_X(T,x) \overset{a.s.}{\to} \infty$ when $T \to \infty$, $n \to \infty$ and $\frac{T}{n} \to 0$ [c.f. Bandi (1999)]. This observation is of great importance in applied work as the optimal window width for drift estimation has proven to be difficult to define due to its dependence on the statistical properties of the underlying process through its local time factor [c.f. Bandi and Nguyen (1999)].

Remark 4.5 As usual, the limit theory clarifies the sense in which enlarging the time span ($T \to \infty$) is crucial for consistent estimation of the infinitesimal first moment of a diffusion. In effect, if we fix $T(=\overline{T})$, then $\overline{L}_X(\overline{T},x) = O_p(1)$ and $\mathbf{\Xi}_{mu}(\overline{T}) = O_p(1)$. In consequence, $\widehat{\theta}_{n,T}^{drift} \overset{p}{\nrightarrow} \theta_0^{drift}$ [c.f. formula (10) above] when T is fixed.

We now turn to the diffusion parameter estimates.

Theorem 4.6 (Consistency of the diffusion estimates) *Assume $n \to \infty$, $T \to \infty$ and $h_{n,T} \to 0$ (as $n, T \to \infty$) such that $\frac{\Delta_{n,T}}{h_{n,T}} \to 0$ and $\frac{\overline{L}_X(T,x)}{h_{n,T}}(\Delta_{n,T})^{\alpha} = O_{a.s.}(1)$ for some $\alpha \in (0, \frac{1}{2})$, then*

$$Q_{n,T}^{diff}(\theta^{diff}) \overset{a.s.}{\to} Q^{diff}(\theta^{diff}, \theta_0^{diff}) = \int_{-\infty}^{\infty} \left(\sigma^2(s, \theta_0^{diff}) - \sigma^2(s, \theta^{diff}) \right)^2 \overline{L}_X(T,s) ds. \tag{11}$$

uniformly in θ^{diff}. Also, let $B(\theta^{diff}, \varepsilon)$ denote an open ball in Θ^{diff} of radius ε. Assume that for $\forall\ \delta > 0$ and $\forall\ \varepsilon > 0$

$$\inf_{\theta^{diff} \notin B\left(\theta_0^{diff}, \varepsilon\right)} \int_{|s| \le \delta} \left(\sigma^2(s, \theta_0^{diff}) - \sigma^2(s, \theta^{diff}) \right)^2 ds > 0. \tag{12}$$

Then,

$$\widehat{\theta}_{n,T}^{diff} \overset{a.s.}{\to} \theta_0^{diff}.$$

Theorem 4.7 (The limit distribution of the diffusion estimates) *Given* $n \to \infty$*,* $T \to \infty$ *and* $h_{n,T} \to 0$ *(as* $n, T \to \infty$*) such that* $\frac{\Delta_{n,T}}{h_{n,T}} \to 0$*,* $\frac{\overline{L}_X(T,x)}{h_{n,T}}(\Delta_{n,T})^{\alpha} = O_{a.s.}(1)$ *for some* $\alpha \in (0, \frac{1}{2})$ *and* $\frac{h^3_{n,T}}{\Delta_{n,T}} \to 0$*, then*

$$\frac{1}{\sqrt{\Delta_{n,T}}} \mathbf{\Xi}^{-1/2}_{sigma}(T) \left(\widehat{\theta}^{diff}_{n,T} - \theta^{diff}_0 \right) \to_d N(0, \mathbf{I}) \tag{13}$$

where

$$\mathbf{\Xi}_{sigma}(T) = \mathbf{B}(T)^{-1}_{sigma} \mathbf{\Omega}(T)_{sigma} \mathbf{B}(T)^{-1}_{sigma}$$

and

$$\mathbf{B}_{sigma} = \left(\int_{-\infty}^{\infty} \frac{\partial \sigma^2_0(a)}{\partial \theta} \frac{\partial \sigma^2_0(a)}{\partial \theta'} \overline{L}_X(T,a) da \right)$$

$$\mathbf{\Omega}_{sigma} = \left(\int_{-\infty}^{\infty} 4\sigma^4(a) \left(\frac{\partial \sigma^2_0(a)}{\partial \theta} \frac{\partial \sigma^2_0(a)}{\partial \theta'} \right) \overline{L}_X(T,a) da \right).$$

Remark 4.8 In the light of Remark 4.3, the integrals $\mathbf{B}_{sigma}$ and $\mathbf{\Omega}_{sigma}$ can be readily interpreted as spatial analogues of the expectations that would arise from the standard nonlinear estimation of conditional expectations. The term $4\sigma^4(a)$ is generated by the quadratic nature of the nonparametric estimator of the infinitesimal second moment of a diffusion.

Remark 4.9 As in the drift case, the rate of convergence is path-dependent. Also, the semiparametric estimates entail efficiency gains with respect to their nonparametric counterparts. In effect, the functional estimates have slower pointwise convergence rates given by $\frac{\sqrt{h_{n,T}\overline{L}_X(T,x)}}{\sqrt{\Delta_{n,T}}}$.

Remark 4.10 The rate of convergence of the diffusion estimates is faster than the rate of convergence of the drift estimates. The difference is remarkable and is given by the multiplicative factor $\frac{1}{\sqrt{\Delta_{n,T}}} = \sqrt{\frac{n}{T}}$. This is a standard result that perfectly reflects the difference in the pointwise convergence rates of the nonparametric estimates (that is, $\sqrt{h_{n,T}\overline{L}_X(T,x)}$ versus $\frac{\sqrt{h_{n,T}\overline{L}_X(T,x)}}{\sqrt{\Delta_{n,T}}}$).

Remark 4.11 The condition $\frac{h^3_{n,T}}{\Delta_{n,T}} \to 0$ guarantees that the limit distribution is driven by the 'variance term' in the estimation error decomposition [see the proof of Theorem 4.7]. Should $\frac{h^3_{n,T}}{\Delta_{n,T}} \to \infty$, then the 'bias term' would dominate, but the semiparametric estimates would still converge to a Gaussian distribution. No choice of the smoothing parameter can make the 'bias term' dominate in the drift case due to the slow rate of convergence of the 'variance term'.

Remark 4.12 As usual, the diffusion parameters can be identified over a fixed time span [c.f. BP (1998)]. In this case the convergence rate ceases to be path-dependent: we experience $\sqrt{n}$-convergence for the semiparametric estimates and $\sqrt{nh_{n,T}}$-convergence for the fully nonparametric counterpart in (2) above. The gain in efficiency which is assured by the adoption of the semiparametric approach is noteworthy and is perfectly consistent with more traditional semiparametric models [c.f. Andrews (1989), for example]. To summarize, if T is fixed ($=\overline{T}$) then,

$$\sqrt{n}\left(\widehat{\theta}_{n,T}^{diff}-\theta_0^{diff}\right)\rightarrow_d MN\left(0,\frac{1}{\overline{T}}\Xi_{sigma}(\overline{T})\right).$$

5 Conclusion

This paper shows how to utilize the informational content of carefully implemented nonparametric methods in the estimation of continuous-time models of the diffusion type while improving on their generally poor convergence properties. From a practical point of view, the semiparametric procedure suggested in this work combines the simplicity of limit theories that can be interpreted as spatial counterparts of the standard asymptotics for nonlinear econometric models with the generality of methods that are robust to deviations from strong distributional assumptions, such as stationarity. Furthermore, the general estimation strategy given here provides a framework which may be particularly appealing to researchers with strong beliefs about potential parametrizations for the two functions of interest.

The next step is naturally to study a testing procedure for alternative parametric specifications based on the quadratic criteria used in this paper. Due to the broadly applicable identifying information that is embodied in the estimated functional drift and diffusion functions and the finite sample accuracy of the asymptotics of the functional estimates [c.f. Bandi and Nguyen (1998)], we believe such test procedures are likely to be attractive. They can, for instance, be expected to have better size properties and more power than testing methods that are based on density-matching procedures which rely on stationarity [c.f. Pritsker (1998)]. Research on this subject will be reported by the authors in subsequent work.

6 Notation

$\rightarrow_{a.s.}$	almost sure convergence
$\rightarrow_p$	convergence in probability
$\Rightarrow, \rightarrow_d$	weak convergence
$:=$	definitional equality
$o_p(1)$	tends to zero in probability
$O_p(1)$	bounded in probability
$o_{a.s.}(1)$	tends to zero almost surely
$O_{a.s.}(1)$	bounded almost surely
$=_d$	distributional equivalence
$\mathrm{MN}(0,V)$	mixed normal distribution with variance V
$C_k, \quad k=1,2,...$	constants

7 Proofs

Proof of theorem 4.1 Uniform strong convergence of the criterion $Q_{n,T}^{drift}(\theta^{drift})$ as in (8) derives from a straightforward application of the results in BP (1998). Note that the limit quantity

$Q^{drift}(\theta^{drift})$ is continuous in θ^{drift} by virtue of Assumption 2.1.

Then, by (9) and for every $\varepsilon > 0$, $\exists \xi > 0$

$$
\begin{aligned}
& P\left(\widehat{\theta}_{n,T}^{drift} \notin B(\theta_0^{drift}, \varepsilon)\right) \\
\leq \; & P\left(Q^{drift}(\widehat{\theta}_{n,T}^{drift}, \theta_0^{drift}) \geq \xi\right) \\
\leq \; & P\left(Q^{drift}(\widehat{\theta}_{n,T}^{drift}, \theta_0^{drift}) - Q_{n,T}^{drift}(\widehat{\theta}_{n,T}^{drift}) + Q_{n,T}^{drift}(\widehat{\theta}_{n,T}^{drift}) - Q^{drift}(\theta_0^{drift}, \theta_0^{drift}) \geq \xi\right) \\
\leq \; & P\left(Q^{drift}(\widehat{\theta}_{n,T}^{drift}, \theta_0^{drift}) - Q_{n,T}^{drift}(\widehat{\theta}_{n,T}^{drift}) + Q_{n,T}^{drift}(\theta_0^{drift}) + o_{a.s.}(1) - Q^{drift}(\theta_0^{drift}, \theta_0^{drift}) \geq \xi\right) \\
\leq \; & P\left(2 \sup_{\theta^{drift} \in \Theta^{drift}} |Q^{drift}(\theta, \theta_0^{drift}) - Q_{n,T}^{drift}(\theta)| + o_{a.s.}(1) \geq \xi\right) \\
& \overset{a.s.}{\to} 0.
\end{aligned}
$$

Proof of theorem 4.2 Write

$$
\widehat{\theta}_{n,T}^{drift} - \theta_0^{drift} = -[\ddot{Q}_{n,T}^{drift}(\theta^*)]^{-1} \dot{Q}_{n,T}^{drift}(\theta_0^{drift}),
$$

where

$$
\begin{aligned}
\theta^* & \in \left(\widehat{\theta}_{n,T}^{drift}, \theta_0^{drift}\right) \\
\dot{Q}_{n,T}^{drift}(\theta_0^{drift}) & = -\frac{T}{n}\sum_{i=1}^{n}\left(\widehat{\mu}_{n,T}(X_{i\Delta_{n,T}}) - \mu(X_{i\Delta_{n,T}}, \theta_0^{drift})\right)\frac{\partial \mu(X_{i\Delta_{n,T}}, \theta_0^{drift})}{\partial \theta} \\
\ddot{Q}_{n,T}^{drift}(\theta^*) & = \underbrace{\frac{T}{n}\sum_{i=1}^{n}\frac{\partial \mu(X_{i\Delta_{n,T}}, \theta^*)}{\partial \theta}\frac{\partial \mu(X_{i\Delta_{n,T}}, \theta^*)}{\partial \theta'}}_{\ddot{Q}_{n,T}^{drift(A)}(\theta^*)} \\
& \quad \underbrace{-\frac{T}{n}\sum_{i=1}^{n}\left(\widehat{\mu}_{n,T}(X_{i\Delta_{n,T}}) - \mu(X_{i\Delta_{n,T}}, \theta^*)\right)\frac{\partial \mu(X_{i\Delta_{n,T}}, \theta^*)}{\partial \theta \partial \theta'}}_{\ddot{Q}_{n,T}^{drift(B)}(\theta^*)}.
\end{aligned}
$$

First, we examine $\ddot{Q}_{n,T}^{drift}(\theta^*)$. Consider $\ddot{Q}_{n,T}^{drift(A)}(\theta^{drift})$. Uniformly in Θ^{drift} we obtain

$$
\begin{aligned}
\ddot{Q}_{n,T}^{drift(A)}(\theta^{drift}) & = \frac{T}{n}\sum_{i=1}^{n}\frac{\partial \mu(X_{i\Delta_{n,\overline{T}}}, \theta^{drift})}{\partial \theta}\frac{\partial \mu(X_{i\Delta_{n,\overline{T}}}, \theta^{drift})}{\partial \theta'} \\
& = \int_0^T \frac{\partial \mu(X_s, \theta^{drift})}{\partial \theta}\frac{\partial \mu(X_s, \theta^{drift})}{\partial \theta'} ds + o_{a.s.}(1) \\
& = \int_{-\infty}^{\infty} \frac{\partial \mu(a, \theta^{drift})}{\partial \theta}\frac{\partial \mu(a, \theta^{drift})}{\partial \theta'}\frac{1}{\sigma^2(a)} L_X(T,a) da + o_{a.s.}(1) \\
& = \int_{-\infty}^{\infty} \frac{\partial \mu(a, \theta^{drift})}{\partial \theta}\frac{\partial \mu(a, \theta^{drift})}{\partial \theta'}\overline{L}_X(T,a) da + o_{a.s.}(1) \\
& = \ddot{Q}^{drift(A)}(\theta^{drift}) + o_{a.s.}(1), \qquad (14)
\end{aligned}
$$

by the occupation time formula [c.f. BP (1998)]. Hence,

$$
\begin{aligned}
& |\ \ddot{Q}_{n,T}^{drift(A)}\ (\widehat{\theta}_{n,T}^{drift}) - \ddot{Q}^{drift(A)}\ (\theta_0^{drift})| \\
\leq & \ |\ \ddot{Q}_{n,T}^{drift(A)}\ (\widehat{\theta}_{n,T}^{drift}) - \ddot{Q}^{drift(A)}\ (\widehat{\theta}_{n,T}^{drift}) + \ddot{Q}^{drift(A)}\ (\widehat{\theta}_{n,T}^{drift}) - \ddot{Q}^{drift(A)}\ (\theta_0^{drift})| \\
\leq & \ |\ \ddot{Q}_{n,T}^{drift(A)}\ (\widehat{\theta}_{n,T}^{drift}) - \ddot{Q}^{drift(A)}\ (\widehat{\theta}_{n,T}^{drift})| + |\ \ddot{Q}^{drift(A)}\ (\widehat{\theta}_{n,T}^{drift}) - \ddot{Q}^{drift(A)}\ (\theta_0^{drift})| \\
\leq & \ o_{a.s.}(1) + |\ \ddot{Q}^{drift(A)}\ \left(\theta_0^{drift} + o_{a.s.}(1)\right) - \ddot{Q}^{drift(A)}\ (\theta_0^{drift})| \\
\leq & \ o_{a.s.}(1) + o_{a.s.}(1) \\
& \stackrel{a.s.}{\to} 0,
\end{aligned}
$$

by the continuity of $\ddot{Q}^{drift(A)}$ (.) and (14). Since θ^* lies on the line segment between $\widehat{\theta}_{n,T}^{drift}$ and θ_0^{drift}, then

$$\ddot{Q}_{n,T}^{drift(A)}\ (\theta^*) = \ \ddot{Q}^{drift}\ (\theta_0^{drift}) + o_{a.s.}(1).$$

Following the same steps, it is simple to prove that

$$\ddot{Q}_{n,T}^{drift(B)}\ (\theta^*) \stackrel{a.s.}{\to} 0.$$

We now analyze $\dot{Q}_{n,T}\ (\theta_0^{drift})$, writing

$$
\begin{aligned}
& -\dot{Q}_{n,T}\ (\theta_0^{drift}) \\
= & \ \frac{T}{n}\sum_{i=1}^{n}\left(\widehat{\mu}_{n,T}(X_{i\Delta_{n,T}}) - \mu(X_{i\Delta_{n,T}},\theta_0^{drift})\right)\frac{\partial\mu(X_{i\Delta_{n,T}},\theta_0^{drift})}{\partial\theta} \\
= & \ \frac{T}{n}\sum_{i=1}^{n}\left(\frac{\frac{1}{h_{n,T}}\sum_{j=1}^{n}\mathbf{K}(\frac{X_{j\Delta_{n,T}}-X_{i\Delta_{n,T}}}{h_{n,T}})\left[X_{(j+1)\Delta_{n,T}} - X_{j\Delta_{n,T}}\right]}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{j=1}^{n}\mathbf{K}(\frac{X_{j\Delta_{n,T}}-X_{i\Delta_{n,T}}}{h_{n,T}})} - \mu(X_{i\Delta_{n,T}},\theta_0^{drift})\right)\frac{\partial\mu(X_{i\Delta_{n,T}},\theta_0^{drift})}{\partial\theta}.
\end{aligned}
$$

We can write $\mu(X_{i\Delta_{n,T}},\theta_0^{drift})$ as $\mu_0(X_{i\Delta_{n,T}})$ and obtain

$$
\begin{aligned}
& -\dot{Q}_{n,T}\ (\theta_0^{drift}) \\
& = \underbrace{\frac{T}{n}\sum_{i=1}^{n}\left(\frac{\frac{1}{h_{n,T}}\sum_{j=1}^{n}\mathbf{K}(\frac{X_{j\Delta_{n,T}}-X_{i\Delta_{n,T}}}{h_{n,T}})\int_{j\Delta_{n,T}}^{(j+1)\Delta_{n,T}}\left(\mu_0(X_s) - \mu_0(X_{i\Delta_{n,T}})\right)ds}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{j=1}^{n}\mathbf{K}(\frac{X_{j\Delta_{n,T}}-X_{i\Delta_{n,T}}}{h_{n,T}})}\right)\frac{\partial\mu_0(X_{i\Delta_{n,T}})}{\partial\theta}}_{\mathbf{A}_{n,T}} \\
& + \underbrace{\frac{T}{n}\sum_{i=1}^{n}\left(\frac{\frac{1}{h_{n,T}}\sum_{j=1}^{n}\mathbf{K}(\frac{X_{j\Delta_{n,T}}-X_{i\Delta_{n,T}}}{h_{n,T}})\int_{j\Delta_{n,T}}^{(j+1)\Delta_{n,T}}\sigma_0(X_s)dB_s}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{j=1}^{n}\mathbf{K}(\frac{X_{j\Delta_{n,T}}-X_{i\Delta_{n,T}}}{h_{n,T}})}\right)\frac{\partial\mu_0(X_{i\Delta_{n,T}})}{\partial\theta}}_{\mathbf{B}_{n,T}(1)}.
\end{aligned}
$$

First, we examine the second term, $\mathbf{B}_{n,T}(1)$. Consider the martingale

$$\mathbf{B}_{n,T}(r) = \frac{T}{n}\sum_{i=1}^{[nr]}\left(\frac{\frac{1}{h_{n,T}}\sum_{j=1}^{n}\mathbf{K}(\frac{X_{j\Delta_{n,T}}-X_{i\Delta_{n,T}}}{h_{n,T}})\int_{j\Delta_{n,T}}^{(j+1)\Delta_{n,T}}\sigma_0(X_s)dB_s}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{j=1}^{n}\mathbf{K}(\frac{X_{j\Delta_{n,T}}-X_{i\Delta_{n,T}}}{h_{n,T}})}\right)\frac{\partial\mu_0(X_{i\Delta_{n,T}})}{\partial\theta}.$$

As $n \to \infty$, $T \to \infty$ and $h_{n,T} \to 0$ such that $\frac{\Delta_{n,T}}{h_{n,T}} \to 0$ and $\frac{\overline{L}_X(T,x)}{h_{n,T}}(\Delta_{n,T})^{\alpha} = O_{a.s.}(1)$ for some $\alpha \in (0, \frac{1}{2})$ [c.f. BP (1998)], its quadratic variation process can be written as

$$\begin{aligned}
&[\mathbf{B}_{n,T}]_r \\
= \;& \left(\frac{T}{n}\right)^2\sum_{i=1}^{[nr]}\sum_{k=1}^{[nr]}\left(\frac{\left(\frac{1}{h_{n,T}}\right)^2\sum_{j=1}^{n}\mathbf{K}(\frac{X_{j\Delta_{n,T}}-X_{i\Delta_{n,T}}}{h_{n,T}})\mathbf{K}(\frac{X_{j\Delta_{n,T}}-X_{k\Delta_{n,T}}}{h_{n,T}})\int_{j\Delta_{n,T}}^{(j+1)\Delta_{n,T}}\sigma_0^2(X_s)ds}{\left(\frac{\Delta_{n,T}}{h_{n,T}}\sum_{j=1}^{n}\mathbf{K}(\frac{X_{j\Delta_{n,T}}-X_{k\Delta_{n,T}}}{h_{n,T}})\right)\left(\frac{\Delta_{n,T}}{h_{n,T}}\sum_{j=1}^{n}\mathbf{K}(\frac{X_{j\Delta_{n,T}}-X_{i\Delta_{n,T}}}{h_{n,T}})\right)}\right)\times \\
&\times\left(\frac{\partial\mu_0(X_{i\Delta_{n,T}})}{\partial\theta}\frac{\partial\mu_0(X_{k\Delta_{n,T}})}{\partial\theta'}\right) \\
= \;& \left(\frac{T}{n}\right)^2\sum_{i=1}^{[nr]}\sum_{k=1}^{[nr]}\left(\sum_{j=1}^{n}\frac{\left(\frac{1}{h_{n,T}}\right)^2\mathbf{K}(\frac{X_{j\Delta_{n,T}}-X_{i\Delta_{n,T}}}{h_{n,T}})\mathbf{K}(\frac{X_{j\Delta_{n,T}}-X_{k\Delta_{n,T}}}{h_{n,T}})\sigma_0^2(X_{j\Delta_{n,T}}+o_{a.s.}(1))\Delta_{n,T}}{\left(\frac{\Delta_{n,T}}{h_{n,T}}\sum_{j=1}^{n}\mathbf{K}(\frac{X_{j\Delta_{n,T}}-X_{k\Delta_{n,T}}}{h_{n,T}})\right)\left(\frac{\Delta_{n,T}}{h_{n,T}}\sum_{j=1}^{n}\mathbf{K}(\frac{X_{j\Delta_{n,T}}-X_{i\Delta_{n,T}}}{h_{n,T}})\right)}\right)+o_{a.s.}(1) \\
= \;& \int_0^{rT}\int_0^{rT}\left(\int_0^T\frac{\left(\frac{1}{h_{n,T}}\right)^2\mathbf{K}(\frac{X_u-X_a}{h_{n,T}})\mathbf{K}(\frac{X_u-X_b}{h_{n,T}})\sigma_0^2(X_u)du}{\left(\frac{1}{h_{n,T}}\int_0^T\mathbf{K}(\frac{X_u-X_a}{h_{n,T}})du\right)\left(\frac{1}{h_{n,T}}\int_0^T\mathbf{K}(\frac{X_u-X_b}{h_{n,T}})du\right)}\right)\left(\frac{\partial\mu_0(X_a)}{\partial\theta}\frac{\partial\mu_0(X_b)}{\partial\theta'}\right)dadb+o_{a.s.}(1) \\
= \;& \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty}\frac{\left(\frac{1}{h_{n,T}}\right)^2\mathbf{K}(\frac{u-a}{h_{n,T}})\mathbf{K}(\frac{u-b}{h_{n,T}})\sigma_0^2(u)\overline{L}_X(T,u)du}{\left(\frac{1}{h_{n,T}}\int_{-\infty}^{\infty}\mathbf{K}(\frac{u-a}{h_{n,T}})\overline{L}_X(T,u)du\right)\left(\frac{1}{h_{n,T}}\int_{-\infty}^{\infty}\mathbf{K}(\frac{u-b}{h_{n,T}})\overline{L}_X(T,u)du\right)}\right)\times \\
&\times\left(\frac{\partial\mu_0(a)}{\partial\theta}\frac{\partial\mu_0(b)}{\partial\theta'}\right)\overline{L}_X(rT,a)\overline{L}_X(rT,b)dadb+o_{a.s.}(1).
\end{aligned}$$

Thus, setting

$$\frac{u-a}{h_{n,T}} = z,$$

we can write

$$\begin{aligned}
&\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty}\frac{\frac{1}{h_{n,T}}\mathbf{K}(z)\mathbf{K}(\frac{a+h_{n,T}z-b}{h_{n,T}})\sigma_0^2(a+h_{n,T}z)\overline{L}_X(T,a+h_{n,T}z)dz}{\left(\int_{-\infty}^{\infty}\mathbf{K}(z)\overline{L}_X(T,a+h_{n,T}z)dz\right)\left(\frac{1}{h_{n,T}}\int_{-\infty}^{\infty}\mathbf{K}(\frac{a+h_{n,T}z-b}{h_{n,T}})\overline{L}_X(T,a+h_{n,T}z)dz\right)}\right)\times \\
&\times\left(\frac{\partial\mu_0(a)}{\partial\theta}\frac{\partial\mu_0(b)}{\partial\theta'}\right)\overline{L}_X(rT,a)\overline{L}_X(rT,b)dadb+o_{a.s.}(1).
\end{aligned}$$

Further, setting

$$\frac{a-b}{h_{n,T}} = k,$$

we obtain

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty}\frac{\mathbf{K}(z)\mathbf{K}(z+k)\sigma_0^2(a+h_{n,T}z)\overline{L}_X(T,a+h_{n,T}z)dz}{\left(\int_{-\infty}^{\infty}\mathbf{K}(z)\overline{L}_X(T,a+h_{n,T}z)dz\right)\left(\int_{-\infty}^{\infty}\mathbf{K}(z+k)\overline{L}_X(T,a+h_{n,T}z)dz\right)}\right)\times$$
$$\times\left(\frac{\partial\mu_0(a)}{\partial\theta}\frac{\partial\mu_0(a-kh_{n,T})}{\partial\theta'}\right)\overline{L}_X(rT,a)\overline{L}_X(rT,a-kh_{n,T})dadk+o_{a.s.}(1)$$
$$\stackrel{a.s.}{\rightarrow}\int_{-\infty}^{\infty}\sigma_0^2(a)\left(\frac{\partial\mu_0(a)}{\partial\theta}\frac{\partial\mu_0(a)}{\partial\theta'}\right)\frac{\left(\overline{L}_X(rT,a)\right)^2}{\overline{L}_X(T,a)}da.$$

Then, by Brownian motion embedding [c.f. Revuz and Yor (1994)]

$$\mathbf{B}_{n,T}(1)\rightarrow_d N\left(0,\left(\int_{-\infty}^{\infty}\sigma^2(a)\left(\frac{\partial\mu_0(a)}{\partial\theta}\frac{\partial\mu_0(a)}{\partial\theta'}\right)\overline{L}_X(T,a)da\right)\right).$$

Next, we examine $\mathbf{A}_{n,T}$.

$$\begin{aligned}
&\mathbf{A}_{n,T}\\
=\ &\frac{T}{n}\sum_{i=1}^{n}\left(\frac{\frac{1}{h_{n,T}}\sum_{j=1}^{n}\mathbf{K}(\frac{X_{j\Delta_{n,T}}-X_{i\Delta_{n,T}}}{h_{n,T}})\int_{j\Delta_{n,T}}^{(j+1)\Delta_{n,T}}\left(\mu_0(X_s)-\mu_0(X_{i\Delta_{n,T}})\right)ds}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{j=1}^{n}\mathbf{K}(\frac{X_{j\Delta_{n,T}}-X_{i\Delta_{n,T}}}{h_{n,T}})}\right)\frac{\partial\mu_0(X_{i\Delta_{n,T}})}{\partial\theta}\\
=\ &\underbrace{\frac{T}{n}\sum_{i=1}^{n}\left(\frac{\frac{1}{h_{n,T}}\sum_{j=1}^{n}\mathbf{K}(\frac{X_{j\Delta_{n,T}}-X_{i\Delta_{n,T}}}{h_{n,T}})\left(\mu_0(X^*)-\mu_0(X_{j\Delta_{n,T}})\right)\Delta_{n,T}}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{j=1}^{n}\mathbf{K}(\frac{X_{j\Delta_{n,T}}-X_{i\Delta_{n,T}}}{h_{n,T}})}\right)\frac{\partial\mu_0(X_{i\Delta_{n,T}})}{\partial\theta}}_{\mathbf{A}_{n,T}^1}\\
&+\underbrace{\frac{T}{n}\sum_{i=1}^{n}\left(\frac{\frac{1}{h_{n,T}}\sum_{j=1}^{n}\mathbf{K}(\frac{X_{j\Delta_{n,T}}-X_{i\Delta_{n,T}}}{h_{n,T}})\left(\mu_0(X_{j\Delta_{n,T}})-\mu_0(X_{i\Delta_{n,T}})\right)\Delta_{n,T}}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{j=1}^{n}\mathbf{K}(\frac{X_{j\Delta_{n,T}}-X_{i\Delta_{n,T}}}{h_{n,T}})}\right)\frac{\partial\mu_0(X_{i\Delta_{n,T}})}{\partial\theta}}_{\mathbf{A}_{n,T}^2},
\end{aligned}$$

where $X^*\in(X_{(j+1)\Delta_{n,T}},X_{j\Delta_{n,T}})$ by the mean value theorem. $\mathbf{A}_{n,T}^1$ can be bounded as follows

$$\mathbf{A}_{n,T}^1\leq const.\kappa_{n,T}\left(\frac{T}{n}\sum_{i=1}^{n}\frac{\partial\mu_0(X_{i\Delta_{n,T}})}{\partial\theta}\right)$$

where $\kappa_{n,T}=\max_{i\leq n}\sup_{j\Delta_{n,T}\leq s\leq(j+1)\Delta_{n,T}}|X_s-X_{j\Delta_{n,T}}|$. We know that $\kappa_{n,T}=O_{a.s.}\left({\Delta_{n,T}}^{1/2}\right)$. Then the bound becomes,

$$const.O_{a.s.}\left({\Delta_{n,T}}^{1/2}\right)\left(\left(\int_{-\infty}^{\infty}\frac{\partial\mu_0(a)}{\partial\theta}\overline{L}_X(T,a)da\right)+o_{a.s.}(1)\right).$$

Now consider term $\mathbf{A}_{n,T}^2$.

$$\mathbf{A}_{n,T}^2$$

$$= \frac{T}{n}\sum_{i=1}^{n}\left(\frac{\frac{1}{h_{n,T}}\sum_{j=1}^{n}\mathbf{K}(\frac{X_{j\Delta_{n,T}}-X_{i\Delta_{n,T}}}{h_{n,T}})\left(\mu_0(X_{j\Delta_{n,T}})-\mu_0(X_{i\Delta_{n,T}})\right)\Delta_{n,T}}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{j=1}^{n}\mathbf{K}(\frac{X_{j\Delta_{n,T}}-X_{i\Delta_{n,\overline{T}}}}{h_{n,T}})}\right)\frac{\partial\mu_0(X_{i\Delta_{n,T}})}{\partial\theta}$$

$$= \frac{T}{n}\sum_{i=1}^{n}\left(\frac{\frac{1}{h_{n,T}}\sum_{j=1}^{n}\mathbf{K}(\frac{X_{j\Delta_{n,T}}-X_{i\Delta_{n,T}}}{h_{n,T}})\left(\mu_0^{'}(x_{ij}^{*})\left(X_{j\Delta_{n,T}}-X_{i\Delta_{n,T}}\right)\right)\Delta_{n,T}}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{j=1}^{n}\mathbf{K}(\frac{X_{j\Delta_{n,T}}-X_{i\Delta_{n,\overline{T}}}}{h_{n,T}})}\right)\frac{\partial\mu_0(X_{i\Delta_{n,T}})}{\partial\theta},$$

where $x^{*}=f(X_{j\Delta_{n,T}},X_{i\Delta_{n,T}})\in[X_{j\Delta_{n,T}},X_{i\Delta_{n,T}}]$. Hence,

$$\begin{aligned}
&\mathbf{A}_{n,T}^{2}\\
&= \int_0^T\frac{\frac{1}{h_{n,T}}\int_0^T\mathbf{K}(\frac{X_u-X_s}{h_{n,T}})\mu_0^{'}\left(f(X_u,X_s)\right)(X_u-X_s)}{\frac{1}{h_{n,T}}\int_0^T\mathbf{K}(\frac{X_u-X_s}{h_{n,T}})du}\frac{\partial\mu_0(X_s)}{\partial\theta}duds+o_{a.s}(1)\\
&= \int_{-\infty}^{\infty}\frac{\frac{1}{h_{n,T}}\int_{-\infty}^{\infty}\mathbf{K}(\frac{u-s}{h_{n,T}})\mu_0^{'}\left(f(s,u)\right)(u-s)}{\frac{1}{h_{n,T}}\int_{-\infty}^{\infty}\mathbf{K}(\frac{u-s}{h_{n,T}})\overline{L}(T,u)du}\frac{\partial\mu_0(s)}{\partial\theta}\overline{L}_X(T,u)\overline{L}_X(T,s)duds+o_{a.s}(1).
\end{aligned}$$

In consequence,

$$\begin{aligned}
&\frac{1}{h_{n,T}}\mathbf{A}_{n,T}^{2}\\
&= \int_{-\infty}^{\infty}\frac{\frac{1}{h_{n,T}}\int_{-\infty}^{\infty}\mathbf{K}(\frac{u-s}{h_{n,T}})\mu_0^{'}\left(f(s,u)\right)\left(\frac{u-s}{h_{n,T}}\right)}{\frac{1}{h_{n,T}}\int_{-\infty}^{\infty}\mathbf{K}(\frac{u-s}{h_{n,T}})\overline{L}(T,u)du}\frac{\partial\mu_0(s)}{\partial\theta}\overline{L}_X(T,u)\overline{L}_X(T,s)duds+o_{a.s}(1)\\
&= \int_{-\infty}^{\infty}\frac{\int_{-\infty}^{\infty}c\mathbf{K}(c)\mu_0^{'}\left(f(s,s+h_{n,T}c)\right)}{\int_{-\infty}^{\infty}\mathbf{K}(c)\overline{L}(T,s+h_{n,T}c)dc}\frac{\partial\mu_0(s)}{\partial\theta}\overline{L}_X(T,s+h_{n,T}c)\overline{L}_X(T,s)dsdc+o_{a.s}(1)\\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}c\mathbf{K}(c)\frac{\mu_0^{'}(s)}{\sigma_0^2(s)}\frac{\partial\mu_0(s)}{\partial\theta}L_X(T,s+h_{n,T}c)dsdc\\
&-\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}c\mathbf{K}(c)\frac{\mu_0^{'}(s)}{\sigma_0^2(s)}\frac{\partial\mu_0(s)}{\partial\theta}L_X(T,s)dsdc\\
&+o_{a.s}(1).
\end{aligned}$$

Then,

$$\frac{1}{h_{n,T}}\mathbf{A}_{n,T}^{2}=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}c\mathbf{K}(c)\frac{\mu_0^{'}(s)}{\sigma_0^2(s)}\frac{\partial\mu_0(s)}{\partial\theta}\left(L_X(T,s+h_{n,T}c)-L_X(T,s)\right)dsdc.$$

In consequence,

$$\frac{1}{h_{n,T}^{3/2}}\mathbf{A}_{n,T}^{2}=2\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}c\mathbf{K}(c)\frac{\mu_0^{'}(s)}{\sigma_0^2(s)}\frac{\partial\mu_0(s)}{\partial\theta}\frac{1}{2\sqrt{h_{n,T}}}\left(L_X(T,s+h_{n,T}c)-L_X(T,s)\right)dsdc.$$

By a simple application of the results in BP (1998) we can write,

$$\frac{1}{h_{n,T}^{3/2}}\mathbf{A}_{n,T}^{2}\to_d 2\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(c\mathbf{K}(c)\left(\frac{\mu_0^{'}(s)}{\sigma_0^2(s)}\right)\frac{\partial\mu_0(s)}{\partial\theta}\right)\mathfrak{B}\left(L_X(T,s),c\right)dsdc,$$

where $\mathfrak{B}(.,.)$ is a standard Brownian sheet. Hence,

$$
\begin{aligned}
& \left(\widehat{\theta}_{n,T}^{drift} - \theta_0^{drift}\right) \\
= & \left(\left(\int_{-\infty}^{\infty} \frac{\partial \mu_0(a)}{\partial \theta} \frac{\partial \mu_0(a)}{\partial \theta^{'}} \overline{L}_X(T,a) da\right) + o_{a.s.}(1)\right)^{-1} \left[\mathbf{A}_{n,T}^1 + \mathbf{A}_{n,T}^2 + \mathbf{B}_{n,T}(1)\right] \\
\rightarrow & _d \left(\int_{-\infty}^{\infty} \frac{\partial \mu_0(a)}{\partial \theta} \frac{\partial \mu_0(a)}{\partial \theta^{'}} \overline{L}_X(T,a) da\right)^{-1} MN\left(0, \left(\int_{-\infty}^{\infty} \sigma^2(a) \left(\frac{\partial \mu_0(a)}{\partial \theta} \frac{\partial \mu_0(a)}{\partial \theta^{'}}\right) \overline{L}_X(T,a) da\right)\right) \\
& \stackrel{d}{=} N\left(0, \boldsymbol{\Xi}_{mu}(T)\right).
\end{aligned}
$$

This, in turn, implies

$$
\boldsymbol{\Xi}_{mu}^{-1/2}(T)\left(\widehat{\theta}_{n,T}^{drift} - \theta_0^{drift}\right) \rightarrow_d N(0, \mathbf{I}),
$$

where

$$
\boldsymbol{\Xi}_{mu}(T) = \mathbf{B}(T)_{mu}^{-1} \boldsymbol{\Omega}(T)_{mu} \mathbf{B}(T)_{mu}^{-1},
$$

$$
\mathbf{B}_{mu} = \left(\int_{-\infty}^{\infty} \frac{\partial \mu_0(a)}{\partial \theta} \frac{\partial \mu_0(a)}{\partial \theta^{'}} \overline{L}_X(T,a) da\right)
$$

and

$$
\boldsymbol{\Omega}_{mu} = \left(\int_{-\infty}^{\infty} \sigma^2(a) \left(\frac{\partial \mu_0(a)}{\partial \theta} \frac{\partial \mu_0(a)}{\partial \theta^{'}}\right) \overline{L}_X(T,a) da\right).
$$

This proves the stated result.

Proof of theorem 4.6 We can follow the same steps as in the proof of Theorem 4.1.

Proof of theorem 4.7 As in the proof of Theorem 4.2, write

$$
\widehat{\theta}_{n,T}^{diff} - \theta_0^{diff} = -[\ddot{Q}_{n,T}^{diff}(\theta^*)]^{-1} \dot{Q}_{n,T}^{diff}(\theta_0^{diff}),
$$

where

$$
\begin{aligned}
\theta^* & \in \left(\widehat{\theta}_{n,T}^{diff}, \theta_0^{diff}\right) \\
\dot{Q}_{n,T}^{diff}(\theta_0^{diff}) & = -\frac{T}{n} \sum_{i=1}^{n} \left(\widehat{\sigma}_{n,T}^2(X_{i\Delta_{n,T}}) - \sigma^2(X_{i\Delta_{n,T}}, \theta_0^{diff})\right) \frac{\partial \sigma^2(X_{i\Delta_{n,T}}, \theta_0^{diff})}{\partial \theta} \\
\ddot{Q}_{n,T}^{diff}(\theta^*) & = \underbrace{\frac{T}{n} \sum_{i=1}^{n} \frac{\partial \sigma^2(X_{i\Delta_{n,T}}, \theta^*)}{\partial \theta} \frac{\partial \sigma^2(X_{i\Delta_{n,T}}, \theta^*)}{\partial \theta^{'}}}_{\ddot{Q}_{n,T}^{diff(A)}(\theta^*)} \\
& \quad \underbrace{- \frac{T}{n} \sum_{i=1}^{n} \left(\widehat{\sigma}_{n,T}^2(X_{i\Delta_{n,T}}) - \sigma^2(X_{i\Delta_{n,T}}, \theta^*)\right) \frac{\partial \sigma^2(X_{i\Delta_{n,T}}, \theta^*)}{\partial \theta \partial \theta^{'}}}_{\ddot{Q}_{n,T}^{diff(B)}(\theta^*)}.
\end{aligned}
$$

First, we examine $\ddot{Q}_{n,T}^{diff}(\theta^*)$. Consider $\ddot{Q}_{n,T}^{diff(A)}(\theta^{diff})$. Uniformly in Θ^{diff} we obtain

$$\begin{aligned}
\ddot{Q}_{n,T}^{diff(A)}(\theta^{diff}) &= \frac{T}{n}\sum_{i=1}^{n}\frac{\partial\sigma^2(X_{i\Delta_{n,\overline{T}}},\theta^{diff})}{\partial\theta}\frac{\partial\sigma^2(X_{i\Delta_{n,\overline{T}}},\theta^{diff})}{\partial\theta'} \\
&= \int_0^T\frac{\partial\sigma^2(X_s,\theta^{diff})}{\partial\theta}\frac{\partial\sigma^2(X_s,\theta^{diff})}{\partial\theta'}ds+o_{a.s.}(1) \\
&= \int_{-\infty}^{\infty}\frac{\partial\sigma^2(a,\theta^{diff})}{\partial\theta}\frac{\partial\sigma^2(a,\theta^{diff})}{\partial\theta'}\frac{1}{\sigma^2(a)}L_X(T,a)da+o_{a.s.}(1) \\
&= \int_{-\infty}^{\infty}\frac{\partial\mu(a,\theta^{diff})}{\partial\theta}\frac{\partial\mu(a,\theta^{diff})}{\partial\theta'}\overline{L}_X(T,a)da+o_{a.s.}(1) \\
&= \ddot{Q}^{diff(A)}(\theta^{diff})+o_{a.s.}(1), \qquad (15)
\end{aligned}$$

by the occupation time formula [c.f. BP (1998)]. Then, using the continuity of $\ddot{Q}^{diff}(.)$, the a.s. consistency of $\widehat{\theta}_{n,T}^{diff}$ and result (15) as in the proof of Theorem 4.2, we can write

$$\ddot{Q}_{n,T}^{diff(A)}(\theta^*) = \ddot{Q}^{diff(A)}(\theta_0^{diff})+o_{a.s.}(1),$$

since θ^* lies on the line segment connecting $\widehat{\theta}_{n,T}^{diff}$ and θ_0^{diff}. Further,

$$\ddot{Q}_{n,T}^{diff(B)}(\theta^*) = \ddot{Q}^{diff(B)}(\theta_0^{diff})+o_{a.s.}(1)\overset{a.s.}{\to}0.$$

Now, consider $\dot{Q}_{n,T}^{diff}(\theta_0^{diff})$.

$$\begin{aligned}
&-\dot{Q}_{n,T}^{diff}(\theta_0^{diff}) \\
&= \frac{T}{n}\sum_{i=1}^{n}\left(\widehat{\sigma}^2(X_{i\Delta_{n,T}})-\sigma^2(X_{i\Delta_{n,T}},\theta_0^{diff})\right)\frac{\partial\sigma^2(X_{i\Delta_{n,T}},\theta_0^{diff})}{\partial\theta} \\
&= \frac{T}{n}\sum_{i=1}^{n}\left(\frac{\frac{1}{h_{n,T}}\sum_{j=1}^{n}\mathbf{K}(\frac{X_{j\Delta_{n,T}}-X_{i\Delta_{n,T}}}{h_{n,T}})\left(X_{(j+1)\Delta_{n,T}}-X_{j\Delta_{n,T}}\right)^2}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{j=1}^{n}\mathbf{K}(\frac{X_{j\Delta_{n,T}}-X_{i\Delta_{n,T}}}{h_{n,T}})}-\sigma^2(X_{i\Delta_{n,T}},\theta_0^{diff})\right)\frac{\partial\sigma^2(X_{i\Delta_{n,T}},\theta_0^{diff})}{\partial\theta}.
\end{aligned}$$

Then, writing $\sigma^2(X_{i\Delta_{n,\overline{T}}},\theta_0^{diff})=\sigma_0^2(X_{i\Delta_{n,\overline{T}}})$, we obtain

$$\begin{aligned}
&-\dot{Q}_{n,T}^{diff}(\theta_0^{diff}) \\
&= \underbrace{\frac{T}{n}\sum_{i=1}^{n}\left(\frac{\frac{1}{h_{n,T}}\sum_{j=1}^{n}\mathbf{K}(\frac{X_{j\Delta_{n,T}}-X_{i\Delta_{n,T}}}{h_{n,T}})\int_{j\Delta_{n,T}}^{(j+1)\Delta_{n,T}}2\left(X_s-X_{j\Delta_{n,T}}\right)\mu_0(X_s)ds}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{j=1}^{n}\mathbf{K}(\frac{X_{j\Delta_{n,T}}-X_{i\Delta_{n,T}}}{h_{n,T}})}\right)\frac{\partial\sigma_0^2(X_{i\Delta_{n,T}})}{\partial\theta}}_{\mathbf{A}_{n,T}} \\
&+\underbrace{\frac{T}{n}\sum_{i=1}^{n}\left(\frac{\frac{1}{h_{n,T}}\sum_{j=1}^{n}\mathbf{K}(\frac{X_{j\Delta_{n,T}}-X_{i\Delta_{n,T}}}{h_{n,T}})\int_{j\Delta_{n,T}}^{(j+1)\Delta_{n,T}}2\left(X_s-X_{j\Delta_{n,T}}\right)\sigma_0(X_s)dB_s}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{j=1}^{n}\mathbf{K}(\frac{X_{j\Delta_{n,T}}-X_{i\Delta_{n,T}}}{h_{n,T}})}\right)\frac{\partial\sigma_0^2(X_{i\Delta_{n,T}})}{\partial\theta}}_{\mathbf{B}_{n,T}(1)}
\end{aligned}$$

$$+\underbrace{\frac{T}{n}\sum_{i=1}^{n}\left(\frac{\frac{1}{h_{n,T}}\sum_{j=1}^{n}\mathbf{K}(\frac{X_{j\Delta_{n,T}}-X_{i\Delta_{n,T}}}{h_{n,T}})\int_{j\Delta_{n,T}}^{(j+1)\Delta_{n,T}}\left(\sigma_0^2(X_s)-\sigma_0^2(X_{i\Delta_{n,T}})\right)ds}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{j=1}^{n}\mathbf{K}(\frac{X_{j\Delta_{n,T}}-X_{i\Delta_{n,T}}}{h_{n,T}})}\right)\frac{\partial\sigma_0^2(X_{i\Delta_{n,T}})}{\partial\theta}}_{\mathbf{C}_{n,T}}.$$

Fisrt, we examine the second term $\mathbf{B}_{n,T}(1)$. Consider the martingale

$$\begin{aligned}&\sqrt{\frac{1}{\Delta_{n,T}}}\mathbf{B}_{n,T}(r)\\ =\ &\sqrt{\frac{T}{n}}\sum_{i=1}^{[nr]}\left(\frac{\frac{1}{h_{n,T}}\sum_{j=1}^{n}\mathbf{K}(\frac{X_{j\Delta_{n,T}}-X_{i\Delta_{n,T}}}{h_{n,T}})\int_{j\Delta_{n,T}}^{(j+1)\Delta_{n,T}}2\left(X_s-X_{j\Delta_{n,T}}\right)\sigma_0(X_s)dB_s}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{j=1}^{n}\mathbf{K}(\frac{X_{j\Delta_{n,T}}-X_{i\Delta_{n,T}}}{h_{n,T}})}\right)\frac{\partial\sigma_0^2(X_{i\Delta_{n,T}})}{\partial\theta}.\end{aligned}$$

Then, as $n\to\infty$, $T\to\infty$ and $h_{n,T}\to 0$ such that $\frac{\Delta_{n,T}}{h_{n,T}}\to 0$ and $\frac{\overline{L}_X(T,x)}{h_{n,T}}(\Delta_{n,T})^{\alpha}=O_{a.s.}(1)$ for some $\alpha\in(0,\frac{1}{2})$ [c.f. BP (1998)], we can write the quadratic variation process as

$$\begin{aligned}&[\mathbf{B}_{n,T}]_r\\ =\ &\frac{T}{n}\sum_{i=1}^{[nr]}\sum_{k=1}^{[nr]}\left(\frac{\left(\frac{1}{h_{n,T}}\right)^2\sum_{j=1}^{n}\mathbf{K}(\frac{X_{j\Delta_{n,T}}-X_{i\Delta_{n,T}}}{h_{n,T}})\mathbf{K}(\frac{X_{j\Delta_{n,T}}-X_{k\Delta_{n,T}}}{h_{n,T}})\int_{j\Delta_{n,T}}^{(j+1)\Delta_{n,T}}2\left(X_s-X_{j\Delta_{n,T}}\right)\sigma_0^2(X_s)ds}{\left(\frac{\Delta_{n,T}}{h_{n,T}}\sum_{j=1}^{n}\mathbf{K}(\frac{X_{j\Delta_{n,T}}-X_{k\Delta_{n,T}}}{h_{n,T}})\right)\left(\frac{\Delta_{n,T}}{h_{n,T}}\sum_{j=1}^{n}\mathbf{K}(\frac{X_{j\Delta_{n,T}}-X_{i\Delta_{n,T}}}{h_{n,T}})\right)}\right)\times\\ &\times\left(\frac{\partial\sigma_0^2(X_{i\Delta_{n,T}})}{\partial\theta}\frac{\partial\sigma_0^2(X_{k\Delta_{n,T}})}{\partial\theta'}\right)\\ =\ &\frac{T}{n}\sum_{i=1}^{[nr]}\sum_{k=1}^{[nr]}\left(\sum_{j=1}^{n}\frac{4\left(\frac{1}{h_{n,T}}\right)^2\mathbf{K}(\frac{X_{j\Delta_{n,T}}-X_{i\Delta_{n,\overline{T}}}}{h_{n,T}})\mathbf{K}(\frac{X_{j\Delta_{n,T}}-X_{k\Delta_{n,\overline{T}}}}{h_{n,T}})\sigma_0^4\left(X_{j\Delta_{n,T}}+o_{a.s.}(1)\right)(\Delta_{n,T})^2}{\left(\frac{\Delta_{n,T}}{h_{n,T}}\sum_{j=1}^{n}\mathbf{K}(\frac{X_{j\Delta_{n,T}}-X_{k\Delta_{n,T}}}{h_{n,T}})\right)\left(\frac{\Delta_{n,T}}{h_{n,T}}\sum_{j=1}^{n}\mathbf{K}(\frac{X_{j\Delta_{n,T}}-X_{i\Delta_{n,T}}}{h_{n,T}})\right)}\right)\times\\ &\times\left(\frac{\partial\sigma_0^2(X_{i\Delta_{n,T}})}{\partial\theta}\frac{\partial\sigma_0^2(X_{k\Delta_{n,T}})}{\partial\theta'}\right)\\ =\ &\int_0^{rT}\int_0^{rT}\left(\int_0^T\frac{4\left(\frac{1}{h_{n,T}}\right)^2\mathbf{K}(\frac{X_u-X_a}{h_{n,T}})\mathbf{K}(\frac{X_u-X_b}{h_{n,T}})\sigma_0^4(X_u)du}{\left(\frac{1}{h_{n,T}}\int_0^T\mathbf{K}(\frac{X_u-X_a}{h_{n,T}})du\right)\left(\frac{1}{h_{n,T}}\int_0^T\mathbf{K}(\frac{X_u-X_b}{h_{n,T}})du\right)}\right)\left(\frac{\partial\sigma_0^2(X_a)}{\partial\theta}\frac{\partial\sigma_0^2(X_b)}{\partial\theta'}\right)dadb\\ &+o_{a.s.}(1)\\ =\ &\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty}\frac{4\left(\frac{1}{h_{n,T}}\right)^2\mathbf{K}(\frac{u-a}{h_{n,T}})\mathbf{K}(\frac{u-b}{h_{n,T}})\sigma_0^4(u)\overline{L}_X(T,u)du}{\left(\frac{1}{h_{n,T}}\int_{-\infty}^{\infty}\mathbf{K}(\frac{u-a}{h_{n,T}})\overline{L}_X(T,u)du\right)\left(\frac{1}{h_{n,T}}\int_{-\infty}^{\infty}\mathbf{K}(\frac{u-b}{h_{n,T}})\overline{L}_X(T,u)du\right)}\right)\times\\ &\times\left(\frac{\partial\sigma_0^2(a)}{\partial\theta}\frac{\partial\sigma_0^2(b)}{\partial\theta'}\right)\overline{L}_X(rT,a)\overline{L}_X(rT,b)dadb+o_{a.s.}(1).\end{aligned}$$

Setting

$$\frac{u-a}{h_{n,T}}=z,$$

we can write

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty}\frac{4\frac{1}{h_{n,T}}\mathbf{K}(z)\mathbf{K}(\frac{a+h_{n,T}z-b}{h_{n,T}})\sigma_0^4(a+h_{n,T}z)\overline{L}_X(T,a+h_{n,T}z)dz}{\left(\int_{-\infty}^{\infty}\mathbf{K}(z)\overline{L}_X(T,a+h_{n,T}z)dz\right)\left(\frac{1}{h_{n,T}}\int_{-\infty}^{\infty}\mathbf{K}(\frac{a+h_{n,T}z-b}{h_{n,T}})\overline{L}_X(T,a+h_{n,T}z)dz\right)}\right)\times$$
$$\times\left(\frac{\partial\sigma_0^2(a)}{\partial\theta}\frac{\partial\sigma_0^2(b)}{\partial\theta'}\right)\overline{L}_X(rT,a)\overline{L}_X(rT,b)dadb+o_{a.s.}(1).$$

Further, setting

$$\frac{a-b}{h_{n,T}}=k,$$

we obtain

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty}\frac{4\mathbf{K}(z)\mathbf{K}(z+k)\sigma_0^4(a+h_{n,T}z)\overline{L}_X(T,a+h_{n,T}z)dz}{\left(\int_{-\infty}^{\infty}\mathbf{K}(z)\overline{L}_X(T,a+h_{n,T}z)dz\right)\left(\int_{-\infty}^{\infty}\mathbf{K}(z+k)\overline{L}_X(T,a+h_{n,T}z)dz\right)}\right)\times$$
$$\times\left(\frac{\partial\sigma_0^2(a)}{\partial\theta}\frac{\partial\sigma_0^2(a-kh_{n,T})}{\partial\theta'}\right)\overline{L}_X(rT,a)\overline{L}_X(rT,a-kh_{n,T})dadk+o_{a.s.}(1)$$
$$\overset{a.s.}{\to}\int_{-\infty}^{\infty}4\sigma_0^4(a)\left(\frac{\partial\sigma_0^2(a)}{\partial\theta}\frac{\partial\sigma_0^2(a)}{\partial\theta'}\right)\frac{\left(\overline{L}_X(rT,a)\right)^2}{\overline{L}_X(T,a)}da.$$

Then, as in the proof of Theorem 4.2,

$$\frac{1}{\sqrt{\Delta_{n,T}}}\mathbf{B}_{n,T}(1)\to_d MN\left(0,\left(\int_{-\infty}^{\infty}4\sigma_0^4(a)\left(\frac{\partial\sigma_0^2(a)}{\partial\theta}\frac{\partial\sigma_0^2(a)}{\partial\theta'}\right)\overline{L}_X(T,a)da\right)\right).$$

Now examine $\mathbf{C}_{n,T}$.

$$\begin{aligned}
&\mathbf{C}_{n,T}\\
=\ &\frac{T}{n}\sum_{i=1}^{n}\left(\frac{\frac{1}{h_{n,T}}\sum_{j=1}^{n}\mathbf{K}(\frac{X_{j\Delta_{n,T}}-X_{i\Delta_{n,T}}}{h_{n,T}})\int_{j\Delta_{n,T}}^{(j+1)\Delta_{n,T}}\left(\sigma_0^2(X_s)-\sigma_0^2(X_{i\Delta_{n,T}})\right)ds}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{j=1}^{n}\mathbf{K}(\frac{X_{j\Delta_{n,T}}-X_{i\Delta_{n,T}}}{h_{n,T}})}\right)\frac{\partial\sigma_0^2(X_{i\Delta_{n,T}})}{\partial\theta}\\
=\ &\underbrace{\frac{T}{n}\sum_{i=1}^{n}\left(\frac{\frac{1}{h_{n,T}}\sum_{j=1}^{n}\mathbf{K}(\frac{X_{j\Delta_{n,T}}-X_{i\Delta_{n,T}}}{h_{n,T}})\left(\sigma_0^2(X^*)-\sigma_0^2(X_{j\Delta_{n,T}})\right)\Delta_{n,T}}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{j=1}^{n}\mathbf{K}(\frac{X_{j\Delta_{n,T}}-X_{i\Delta_{n,T}}}{h_{n,T}})}\right)\frac{\partial\sigma_0^2(X_{i\Delta_{n,T}})}{\partial\theta}}_{\mathbf{C}_{n,T}^1}\\
&+\underbrace{\frac{T}{n}\sum_{i=1}^{n}\left(\frac{\frac{1}{h_{n,T}}\sum_{j=1}^{n}\mathbf{K}(\frac{X_{j\Delta_{n,T}}-X_{i\Delta_{n,T}}}{h_{n,T}})\left(\sigma_0^2(X_{j\Delta_{n,T}})-\sigma_0^2(X_{i\Delta_{n,T}})\right)\Delta_{n,T}}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{j=1}^{n}\mathbf{K}(\frac{X_{j\Delta_{n,T}}-X_{i\Delta_{n,T}}}{h_{n,T}})}\right)\frac{\partial\sigma_0^2(X_{i\Delta_{n,T}})}{\partial\theta}}_{\mathbf{C}_{n,T}^2},
\end{aligned}$$

where $X^*\in(X_{(j+1)\Delta_{n,T}},X_{j\Delta_{n,T}})$ by the mean value theorem. First, we analyse term $\mathbf{C}_{n,T}^2$.

$$\mathbf{C}^2_{n,T} = \frac{T}{n}\sum_{i=1}^{n}\left(\frac{\frac{1}{h_{n,T}}\sum_{j=1}^{n}\mathbf{K}(\frac{X_{j\Delta_{n,T}}-X_{i\Delta_{n,T}}}{h_{n,T}})\left(\sigma^2_0(X_{j\Delta_{n,T}})-\sigma^2_0(X_{i\Delta_{n,T}})\right)\Delta_{n,T}}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{j=1}^{n}\mathbf{K}(\frac{X_{j\Delta_{n,T}}-X_{i\Delta_{n,T}}}{h_{n,T}})}\right)\frac{\partial\sigma^2_0(X_{i\Delta_{n,T}})}{\partial\theta}$$

Following the same steps as for $\mathbf{A}^2_{n,T}$ in the proof of Theorem 4.2 we can prove that

$$\frac{1}{h^{3/2}_{n,T}}\mathbf{C}^2_{n,T} \to_d 4\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} c\mathbf{K}(c)\left(\frac{\sigma'_0(s)}{\sigma_0(s)}\right)\frac{\partial\sigma^2_0(s)}{\partial\theta}\mathfrak{B}\left(L_X(T,s),c\right)dsdc,$$

where $\mathfrak{B}(.,.)$ is a standard Brownian sheet. As for $\mathbf{A}_{n,T}$ and $\mathbf{C}^1_{n,T}$, it is easy to see that $\mathbf{A}_{n,T} = o_{a.s.}(\mathbf{B}_{n,T})$ and $\mathbf{C}^1_{n,T} = o_{a.s.}(\mathbf{C}^2_{n,T})$. Then,

$$\begin{aligned}
& \widehat{\theta}^{diff}_{n,T} - \theta^{diff}_0 \\
= & -[\ddot{Q}^{diff}_{n,T}(\theta^*)]^{-1}\dot{Q}^{diff}_{n,T}(\theta^{diff}_0) \\
= & \left(\left(\int_{-\infty}^{\infty}\frac{\partial\sigma^2_0(a)}{\partial\theta}\frac{\partial\sigma^2_0(a)}{\partial\theta'}\overline{L}_X(T,a)da\right)+o_{a.s.}(1)\right)^{-1}[\mathbf{B}_{n,T}+\mathbf{A}_{n,T}+\mathbf{C}^1_{n,T}+\mathbf{C}^2_{n,T}] \\
= & \left(\left(\int_{-\infty}^{\infty}\frac{\partial\sigma^2_0(a)}{\partial\theta}\frac{\partial\sigma^2_0(a)}{\partial\theta'}\overline{L}_X(T,a)da\right)+o_{a.s.}(1)\right)^{-1}[\mathbf{B}_{n,T}+o_{a.s.}(\mathbf{B}_{n,T})+o_{a.s.}(\mathbf{C}^2_{n,T})+\mathbf{C}^2_{n,T}].
\end{aligned}$$

If $\frac{h^3_{n,T}}{\Delta_{n,T}} \to 0$, then

$$\begin{aligned}
& \sqrt{\frac{1}{\Delta_{n,T}}}\left(\widehat{\theta}^{diff}_{n,T}-\theta^{diff}_0\right) \\
& \to_d \left(\int_{-\infty}^{\infty}\frac{\partial\sigma^2_0(a)}{\partial\theta}\frac{\partial\sigma^2_0(a)}{\partial\theta'}\overline{L}_X(T,a)da\right)^{-1} MN\left(0,\left(\int_{-\infty}^{\infty}4\sigma^4(a)\left(\frac{\partial\sigma^2_0(a)}{\partial\theta}\frac{\partial\sigma^2_0(a)}{\partial\theta'}\right)\overline{L}_X(T,a)da\right)\right) \\
& \stackrel{d}{=} N\left(0,\mathbf{\Xi}_{sigma}(T)\right).
\end{aligned}$$

In consequence,

$$\frac{1}{\sqrt{\Delta_{n,T}}}\mathbf{\Xi}^{-1/2}_{sigma}(T)\left(\widehat{\theta}^{diff}_{n,T}-\theta^{diff}_0\right) \to_d N(0,\mathbf{I}),$$

where

$$\mathbf{\Xi}_{sigma}(T) = \mathbf{B}(T)^{-1}_{sigma}\mathbf{\Omega}(T)_{sigma}\mathbf{B}(T)^{-1}_{sigma},$$

$$\mathbf{B}_{sigma} = \left(\int_{-\infty}^{\infty}\frac{\partial\sigma^2_0(a)}{\partial\theta}\frac{\partial\sigma^2_0(a)}{\partial\theta'}\overline{L}_X(T,a)da\right),$$

and

$$\mathbf{\Omega}_{sigma} = \left(\int_{-\infty}^{\infty}4\sigma^4(a)\left(\frac{\partial\sigma^2_0(a)}{\partial\theta}\frac{\partial\sigma^2_0(a)}{\partial\theta'}\right)\overline{L}_X(T,a)da\right).$$

This concludes the proof for the diffusion estimator.

References

[1] Aït-Sahalia, Y., 1996b. Testing continuous-time models of the spot interest rate. The Review of Financial Studies 2, 385-426.

[2] Andrews, D.W.K., 1989. Asymptotics for Semiparametric Econometric Models: III. Testing and Examples. C.F.D.P. No. 910, Yale University.

[3] Bandi, F., 1999. Stanton's (1997) approximations to the first and second moment of a diffusion: statistical properties and implications for pricing. Unpublished working paper, Yale University.

[4] Bandi, F., Nguyen T., 1999. Fully nonparametric estimators for diffusions: a small sample analysis. Unpublished working paper, Yale University.

[5] Bandi, F., Phillips P.C.B., 1998. Econometric estimation of diffusion models. Unpublished working paper, Yale University.

[6] Baxter, M., Rennie, A., 1996. Financial Calculus. An Introduction to Derivative Pricing. Cambridge University Press, Cambridge.

[7] Davidson, R., MacKinnon, J., 1993. Estimation and Inference in Econometrics. Oxford University Press, New York.

[8] Duffie, D., 1992. Dynamic Asset Pricing Theory. Princeton University Press.

[9] Florens-Zmirou, D., 1993. On estimating the diffusion coefficient from discrete observations. Journal of Applied Probability 30, 790-804.

[10] Härdle, W., 1990. Applied Nonparametric Regression. Cambridge University Press, Cambridge.

[11] Jacod, J., 1997. Nonparametric kernel estimation of the diffusion coefficient of a diffusion. Prépublication No. 405 du Laboratoire de l'Université Paris VI.

[12] Jiang, G.J., Knight, J., 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.

[13] Karatzas, I., Shreve, S.E., 1988. Brownian Motion and Stochastic Calculus. Springer-Verlag, New York.

[14] Karlin, S., Taylor, 1981. A Second Course in Stochastic Processes. Academic Press, New York.

[15] Park, J., Phillips P.C.B., 1999. Asymptotics for nonlinear transformations of integrated time series. Econometric Theory, **15**, 269-298.

[16] Park, J., Phillips P.C.B., 1998b. Nonlinear regressions with integrated time series. Unpublished working paper, Yale University.

[17] Phillips, P.C.B., Park, J., 1997. Nonstationary density estimation and kernel autoregression. Unpublished working paper, Yale University.

[18] Pritsker, M., 1998. Nonparametric density estimation and tests of continuous-time interest rate models. The Review of Financial Studies 3.

[19] Protter, P., 1990. Stochastic Integration and Differential Equations. Springer-Verlag, New York.

[20] Revuz, D., Yor, M., 1994. Continuous Martingales and Brownian Motion. Second Edition, Springer-Verlag, New York.

[21] Stanton, R., 1998. A nonparametric model of term strucure dynamics and the market price of interest rate risk. Journal of Finance 5, 1973-2002.