A SIMPLE APPROACH TO THE PARAMETRIC ESTIMATION OF POTENTIALLY NONSTATIONARY DIFFUSIONS

BY

FEDERICO M. BANDI and PETER C. B. PHILLIPS

COWLES FOUNDATION PAPER NO. 1205

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

2007

http://cowles.econ.yale.edu/

Available online at www.sciencedirect.com
ScienceDirect
ELSEVIER
Journal of Econometrics 137 (2007) 354–395
JOURNAL OF Econometrics
www.elsevier.com/locate/jeconom

A simple approach to the parametric estimation of potentially nonstationary diffusions ☆

Federico M. Bandi[a,*], Peter C.B. Phillips[b]

[a]Graduate School of Business, University of Chicago, 5807 South Woodlawn Avenue, Chicago, IL 60637, USA
[b]Cowles Foundation for Research in Economics, Yale University, USA

Received 9 October 2002; accepted 13 June 2005
Available online 2 May 2006

Abstract

A simple and robust approach is proposed for the parametric estimation of scalar homogeneous stochastic differential equations. We specify a parametric class of diffusions and estimate the parameters of interest by minimizing criteria based on the integrated squared difference between kernel estimates of the drift and diffusion functions and their parametric counterparts. The procedure does not require simulations or approximations to the true transition density and has the simplicity of standard nonlinear least-squares methods in discrete time. A complete asymptotic theory for the parametric estimates is developed. The limit theory relies on infill and long span asymptotics and is robust to deviations from stationarity, requiring only recurrence.

JEL classification: C14; C22

Keywords: Diffusion; Drift; Local time; Parametric estimation; Stochastic differential equation

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.

☆A preliminary version of this paper entitled "Accelerated asymptotics for diffusion model estimation" was written for the Cowles Foundation Conference "New Developments in Time Series Econometrics", Yale University, October 23–24, 1999.

*Corresponding author. Tel.: +1 773 834 4352; fax: +1 773 702 0458.
E-mail address: federico.bandi@gsb.uchicago.edu (F.M. Bandi).

0304-4076/$ - see front matter
doi:10.1016/j.jeconom.2005.06.033

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, namely the drift and diffusion functions (the interested reader is referred to the review by Bandi and Phillips, 2002, hereafter BP, and the references therein). 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 inferior efficiency 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 a simple parametric estimation method that matches parametric estimates of the drift and diffusion functions to their functional counterparts. In order to do so, we specify a parametric class for the underlying diffusion process and estimate the drift and diffusion parameters *individually* by minimizing two criteria which can be readily interpreted as the integrated squared differences between functional estimates of drift and diffusion and their corresponding parametric expressions. The first-stage nonparametric estimates are defined as straightforward sample analogs to the theoretical functions. Drift and diffusion function are known to have conditional moment representations. Hence, the nonparametric estimates are empirical analogs to conditional moments written as weighted averages. The weights are constructed using conventional kernels (Bandi and Phillips, 2003).

The limit theory relies on infill (i.e., increasingly frequent observations over time) and long span asymptotics (i.e., increasing span of data). Both features are crucial to derive the consistency of the first-stage nonparametric estimates and, as consequence, of the final parameter estimates under recurrence. Recurrence is the identifying assumption used in this paper. It guarantees return of the continuous sample path of the scalar diffusion process to sets of nonzero Lebesgue measure in its range an infinite number of times over time. Being the infinitesimal moments defined pointwise, the return of the path of the process to neighborhoods of each spatial level appears to be an important property to exploit for the purpose of their identification. More precisely, the infill assumption allows us to approximate the continuous sample path of the underlying process with its discrete counterpart while replicating the infinitesimal features of the conditional moments of interest by virtue of sample analogs. The long span assumption permits us to make use of the dynamic properties of the underlying Markov process for the sake of the consistent estimation of drift and diffusion through repeated visits to each spatial set, as implied by recurrence.

Recurrence is known to be a milder assumption than stationarity and mixing (see, e.g., Meyn and Tweedie, 1993). Recurrent processes do not have to possess a time-invariant probability measure. They are called null recurrent in this case. Positive recurrent processes are recurrent processes that are endowed with a stationary density to which they converge in the limit. Stationary processes are positive recurrent processes that either have reached the time-invariant stationary density or are started at it. The validity of the limit theory in this paper only requires recurrence. Even though our theory could (and will) be specialized to the positive recurrent and stationary case, in general potential users do not have to make assumptions about the stationarity properties of the process when estimating individual infinitesimal moments. Consistency of the drift (diffusion) parameter estimates is preserved under misspecification of the diffusion (drift) function in the recurrent class. Furthermore, while it is true that the dynamic features of the underlying process shape the asymptotic distributions in general, we show that all the relevant information about such features is embodied in estimable random objects that define the variances of asymptotically normal variates. Hence, from the sole point of view of statistical inference, the limiting distributions do not depend on whether the process is stationary or not, being defined in terms of random norming. Such invariance is a valuable feature for applied work.

Some additional observations are in order. Starting with the fundamental work of Gouriéroux et al. (1993) and Gallant and Tauchen (1996), a variety of simulation-based methods have been recently introduced to consistently estimate parametric models for diffusions. For example, Brandt and Santa-Clara (2002), Durham and Gallant (2002), Elerian et al. (2001), and Eraker (2001), among others, suggest simulation-based procedures for maximum likelihood estimation. Somewhat different is the approach in Aït-Sahalia (2002) who recommends approximations to the true, generally unknown, transition density of the discretely sampled process for the purpose of consistent likelihood estimation. Carrasco et al. (2002), Chacko and Viceira (2003), Jiang and Knight (2002), and Singleton (2001) suggest characteristic function-based generalized method of moment (GMM) estimation. GMM-based estimation is also discussed in Conley et al. (1997), Duffie and Glynn (2004), and Hansen and Scheinkman (1995), inter alia. While some of these techniques permit to achieve the same efficiency that (the generally infeasible) maximum likelihood estimation would guarantee,[1] they do so at the cost of some computational burden. In addition, most of these methods explicitly trade off robustness for efficiency.

The parametric procedure that we discuss in this paper has two main features. The first feature is computational *simplicity*: the methodology only requires straightforward estimation of nonparametric functionals à la Nadaraya–Watson type in the first stage and implementation of a minimization routine similar to conventional nonlinear least-squares in the second stage. The second feature is *robustness*. Specifically, the statistical assumptions that are used for consistency are minimal and the information contained in the nonparametric estimates of drift and diffusion is fully exploited for the purpose of parametric inference. As such, our method can be employed as a preliminary descriptive tool and be regarded as complementary rather than alternative to some existing methods.

Furthermore, the ‘‘minimum distance’’ type of estimation that is discussed in this work might be interpreted as extremum estimation for potentially nonstationary and nonlinear continuous-time models of the diffusion type. Minimum distance methods for robust estimation have a long history in statistics (the interested reader is referred to Chiang, 1956; Ferguson, 1958; Koul, 1992; and the review papers in Maddala and Rao, 1997) and have been recently applied to potentially nonlinear, but strictly stationary, diffusion processes by Aït-Sahalia (1996). Altissimo and Mele (2003) have recently extended the procedure in Aït-Sahalia (1996) to estimate multivariate models with unobservables through simulation methods. Aït-Sahalia estimates nonparametrically the stationary density of the process and, given a parametric class for drift and diffusion, designs an estimation method that matches the nonparametric density function of the process to its uniquely specified parametric counterpart. Specifically, matching is obtained through minimization of the mean-squared difference between the nonparametric estimate of the density function of the process and its parametric counterpart.

[1]The true conditional distribution of the discretely sample data is known in closed-form only for few processes (see, e.g., Lo, 1988).

There are several differences between our approach and the methodology in Aït-Sahalia (1996). First, we do not employ the informational content of the nonparametric density. As pointed out earlier, the adopted parametric model might not imply the existence of a time-invariant measure and can be null recurrent. Second, in our framework, parametric inference on the second infinitesimal moment does not depend on inference on the first infinitesimal moment. In other words, when interested in the identification of the diffusion function, as is often the case in practise, the econometrician does not have to estimate the first infinitesimal moment or specify a parametric class for it. Interestingly, we show that this is true even if the two moments imply explosion (or attraction) of the underlying diffusion and finite returns (rather than infinite returns, as implied by recurrence) to sets of nonzero Lebesgue measure. Hence, the consistency of the diffusion parameters, as well as the feasibility of their asymptotic distribution, are not affected by potential misspecifications of the drift function. In addition, the process can be transient. As far as the drift parameters are concerned, only their asymptotic covariance depends on the true infinitesimal second moment. However, the drift parameters may be consistently estimated even when the diffusion function is misspecified provided the underlying process is in the recurrent class.

The above-mentioned properties are achieved through the use of increasingly frequent data points in the limit as well as increasing spans of data. The appropriateness of this twofold limit theory is an empirical issue which depends on the application. Nonetheless, it is known to be a realistic approximation in fields, such as finance, where data sets are often characterized by a large number of observations sampled at relatively high frequencies over long spans of time. The simulation studies of Bandi and Nguyen (1999) and Jiang and Knight (1999) show that daily data, for example, are good approximations to very frequent observations for estimators relying on very frequent observations. Long spans of daily data are commonplace in finance. Higher than daily frequencies are also now available in finance, albeit over generally shorter time spans. However, the use of very high-frequency (intradaily, for example) observations poses microstructure-related issues (see Bandi and Russell, 2005 for a review of recent contributions on this topic). Dealing with these issues is beyond the scope of the present paper.

The paper proceeds as follows. Section 2 presents the model and the objects of econometric interest. Section 3 details the estimation procedure. Section 4 lays out the limiting results. In Section 5 we specialize our general theory to the Brownian motion and positive recurrent case, as well as to the stationary case. Section 6 discusses covariance matrix estimation. Section 7 focuses on efficiency issues. Section 8 concludes and discusses extensions. Appendix A provides proofs and technicalities. A glossary of notation is in Appendix B.

2. The model

We consider a filtered complete probability space $(\Omega, \Im, (\Im_t)_{t\geqslant 0}, P)$ on which is defined the continuous adapted process

$$X_t = X_0 + \int_0^t \mu(X_s, \theta^\mu)\,\mathrm{d}s + \int_0^t \sigma(X_s, \theta^\sigma)\,\mathrm{d}B_s, \tag{1}$$

where $\{B_t : t\geqslant 0\}$ is a standard Brownian motion. The initial condition X_0 is square integrable and is taken to be independent of $\{B_t : t\geqslant 0\}$. The probability space satisfies the "usual hypotheses" (Protter, 1995), namely (i) $\Im_0$ contains all the null sets of $\Im$ and (ii) $(\Im_t)_{t\geqslant 0}$ is right continuous, i.e., $\Im_t = \bigcap_{u>t}\Im_u\ \forall t$. The parameter vectors θ^μ and θ^σ are such that $(\theta^\mu, \theta^\sigma) = \theta \in \Theta$, where Θ is an open and bounded subset of $\Re^M$ for a generic M. More specifically, $\theta^{\mu} \in \Theta^{\mu} \subset \Re^{m_1}$ and $\theta^{\sigma} \in \Theta^{\sigma} \subset \Re^{m_2}$ with $m_1 + m_2 = M$. The vectors θ^{μ} and θ^{σ} jointly define a parametric family for the process in Eq. (1). Since we will be dealing with extremum estimation procedures, it is convenient to denote the true values of these parameters by θ_0^{μ} and θ_0^{σ}.

As in BP (2003), the following conditions are used in the study of the continuous process in Eq. (1). In what follows the symbol $\mathfrak{D}$ denotes the admissible range of X_t.

Assumption 1. (i) $\mu(.,\theta^{\mu})$ and $\sigma(.,\theta^{\sigma})$ are time-homogeneous, $\mathfrak{B}$-measurable functions on $\mathfrak{D} = (l,u)$ with $-\infty \leqslant l < u \leqslant \infty$, 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 of the range of the process, there exist constants C_1^J and C_2^J such that, for all x and y in J,

$$|\mu(x,\theta^{\mu}) - \mu(y,\theta^{\mu})| + |\sigma(x,\theta^{\sigma}) - \sigma(y,\theta^{\sigma})| \leqslant C_1^J |x-y|$$

and

$$|\mu(x,\theta^{\mu})| + |\sigma(x,\theta^{\sigma})| \leqslant C_2^J \{1+|x|\}.$$

(ii) $\sigma^2(.,\theta^{\sigma}) > 0$ on $\mathfrak{D}$.

(iii) We define $S(\alpha,\theta)$, the natural scale function, as

$$S(\alpha,\theta) = \int_c^{\alpha} \exp\left\{ \int_c^{y} \left[-\frac{2\mu(x,\theta^{\mu})}{\sigma^2(x,\theta^{\sigma})} \right] \mathrm{d}x \right\} \mathrm{d}y, \tag{2}$$

where c is a generic fixed number belonging to $\mathfrak{D}$. We require $S(\alpha,\theta)$ to satisfy

$$\lim_{\alpha \to l} S(\alpha,\theta) = -\infty$$

and

$$\lim_{\alpha \to u} S(\alpha,\theta) = \infty.$$

(iv) $\mu(x,\theta^{\mu})$ and $\sigma(x,\theta^{\sigma})$ are at least twice continuously differentiable in θ^{μ} and θ^{σ} for all $x \in \mathfrak{D}$.

Under Conditions (i)–(iii), the adapted process in Eq. (1) is recurrent (see, e.g., Karatzas and Shreve, 1991). Condition (iv) will be used in the development of our asymptotics. If, in addition to Conditions (i)–(iii), we have

$$\overline{m} = \int_{\mathfrak{D}} m(a,\theta)\,\mathrm{d}a < \infty,$$

where $m(.,\theta)$ is the so-called speed function defined as

$$m(.,\theta) = \frac{2}{\sigma^2(.,\theta^{\sigma}) S'(.,\theta)},$$

with $S'(.,\theta)$ being the first derivative of the scale function in Eq. (2), then the process is positive recurrent and possesses a time-invariant probability measure $f(.,\theta) = m(.,\theta)/\overline{m}$ according to which it is distributed, at least in the limit. As mentioned, our theory also applies to processes for which Conditions (i)–(iii) are satisfied and $\overline{m} = \infty$. Such processes are nonstationary. They are typically called null recurrent. Brownian motion is an example of null recurrent diffusion. Nonetheless, the class of null recurrent diffusion processes is substantially broader than Brownian motion and is known to include highly nonlinear processes (see, e.g., BP, 2002).

As discussed in the Introduction, if interest centers on the identification of the second infinitesimal moment, recurrence can be further relaxed. In fact, this moment can be estimated consistently under transience, that is, in situations where the process of interest is not guaranteed to visit every level in its admissible range an infinite number of times over time with probability one, as implied by our Assumption 1(iii). We will come back to this observation (see Remark 11).

The objects of econometric interest in this paper are the drift, $\mu(., \theta^{\mu})$, and the diffusion term, $\sigma^2(., \theta^{\sigma})$. The conditional moment interpretations of these objects are well known, representing the ''instantaneous'' conditional mean and the ''instantaneous'' conditional variance of increments in the process (see, e.g., Karlin and Taylor, 1981). More precisely, $\mu(., \theta^{\mu})$ describes the conditional expected rate of change of the process for infinitesimal time changes, whereas $\sigma^2(., \theta^{\sigma})$ gives the conditional rate of change of volatility, for infinitesimal variations in time.

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 the estimators in BP (2003) in their single smoothing versions. Assume the data X_t is recorded discretely at $\{t = t_1, t_2, \ldots, t_n\}$ in the time interval $(0, T]$, where T is a positive constant. Also, assume equispaced data. Hence,

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

are n observations at

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

where $\Delta_{n,T} = T/n$. The drift estimator is defined as

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

The diffusion estimator is defined as

$$\widehat{\sigma}^2_{(n,T)}(.) = \frac{1}{\Delta_{n,T}} \frac{\sum_{j=1}^{n-1} \mathbf{K}\left(\frac{X_{j\Delta_{n,T}} - .}{h_{n,T}}\right)[X_{(j+1)\Delta_{n,T}} - X_{j\Delta_{n,T}}]^2}{\sum_{j=1}^{n} \mathbf{K}\left(\frac{X_{j\Delta_{n,T}} - .}{h_{n,T}}\right)}. \tag{4}$$

The function $\mathbf{K}(.)$ that appears in Eqs. (3) and (4) is a conventional kernel whose properties are listed below.

Assumption 2. The kernel $\mathbf{K}(.)$ is a continuously differentiable, symmetric and non-negative function whose derivative $\mathbf{K}'(.)$ is absolutely integrable and for which

$$\int_{-\infty}^{\infty}\mathbf{K}(s)\,\mathrm{d}s = 1, \quad \int_{-\infty}^{\infty}\mathbf{K}^2(s)\,\mathrm{d}s < \infty, \quad \sup_s \mathbf{K}(s) < C_3,$$

and

$$\mathbf{K}_2 = \int_{-\infty}^{\infty} s^2\mathbf{K}(s)\,\mathrm{d}s < \infty.$$

Remark 1. The estimators in Eqs. (3) and (4) are straightforward sample analogs to the theoretical functions. BP (2003) discuss their properties of consistency and asymptotic normality. They show that recurrence, which is implied by Assumption 1, rather than positive recurrence or stationarity, is all that is needed to achieve identification. BP (2003) derive the asymptotics as the time span (T) and the number of data points (n) increase with the frequency of observations ($\Delta_{n,T} = T/n \to 0$). Increasing the data frequency over time is crucial for the consistent estimation of continuous-time models using fully functional methods under general assumptions on the statistical evolution of the underlying process and equispaced data. By letting the time span increase to infinity, the drift and diffusion function can be recovered in the limit since the process continues to make repeated visits to all spatial points in its range by virtue of recurrence. However, enlarging the time span is necessary only for consistent drift estimation. The local dynamics of the process contain sufficient information to identify consistently the infinitesimal second moment.

In other words, recurrence suffices for the pointwise estimation of diffusion processes since it is all that one needs to imply infinite returns to each spatial level x with probability one. When we combine the recurrence property with differences between adjacent observations $X_{j\Delta_{n,T}}, X_{(j+1)\Delta_{n,T}}$ going to zero as $\Delta_{n,T} \to 0$, it is intuitive to understand why $\widehat{\mu}_{(n,T)}(x)$ and $\widehat{\sigma}^2_{(n,T)}(x)$ represent consistent estimates of the infinitesimal first and second moments for all $x \in \mathfrak{D}$ (BP, 2002 contains further discussions).

Remark 2. More general sample analogs to the true functions of the convoluted type described in BP (2003) could be used instead to derive the functional estimates. Here we employ specifications based on simple smoothing rather than on convoluted kernels, as in the most general case examined by BP (2003), for simplicity in the proofs.

The use of more involved specifications is known to potentially improve the asymptotic mean-squared error of the pointwise functional estimates and be beneficial in a finite sample (see, e.g., 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 limits the effects of potentially suboptimal choices. Extensions to convoluted kernels can be easily derived from the apparatus presented below. BP (2003) discuss bandwidth selection.

We now turn to parametric estimation. Consider a subset of $\overline{n} \leqslant n$ observations over a *fixed* time span $\overline{T} \leqslant T$. Assume the observations are equispaced with distance between adjacent data points given by $\Delta_{\overline{n},\overline{T}} = \overline{T}/\overline{n}$. Let $\widehat{\boldsymbol{\mu}}$ be the column vector of nonparametric drift estimates at the $\overline{n}$ data points $X_{i\Delta_{\overline{n},\overline{T}}}$ with $i = 1, \ldots, \overline{n}$, i.e., $\widehat{\boldsymbol{\mu}} = (\widehat{\mu}_{(n,T)}(X_{\Delta_{\overline{n},\overline{T}}}), \ldots, \widehat{\mu}_{(n,T)}(X_{\overline{n}\Delta_{\overline{n},\overline{T}}}))'$. Let $\boldsymbol{\mu}(\theta^\mu)$ be the column vector of the parametric drift specifications at the same $\overline{n}$ data points, i.e., $\boldsymbol{\mu}(\theta^\mu) = (\mu(X_{\Delta_{\overline{n},\overline{T}}}, \theta^\mu), \ldots, \mu(X_{\overline{n}\Delta_{\overline{n},\overline{T}}}, \theta^\mu))'$. Assume $\widehat{\boldsymbol{\sigma}}^2$ and $\boldsymbol{\sigma}^2(\theta^\sigma)$ are defined analogously. Consider the criteria

$$Q^{\mu}_{n,\overline{n},T} = \frac{\overline{T}}{\overline{n}} ||\widehat{\boldsymbol{\mu}} - \boldsymbol{\mu}(\theta^{\mu})||^2 = \frac{\overline{T}}{\overline{n}} \sum_{i=1}^{\overline{n}} (\widehat{\mu}_{(n,T)}(X_{i\Delta_{\overline{n},\overline{T}}}) - \mu(X_{i\Delta_{\overline{n},\overline{T}}}, \theta^{\mu}))^2 \tag{5}$$

and

$$Q^{\sigma}_{n,\overline{n},T} = \frac{\overline{T}}{\overline{n}} ||\widehat{\boldsymbol{\sigma}}^2 - \boldsymbol{\sigma}^2(\theta^{\sigma})||^2 = \frac{\overline{T}}{\overline{n}} \sum_{i=1}^{\overline{n}} (\widehat{\sigma}^2_{(n,T)}(X_{i\Delta_{\overline{n},\overline{T}}}) - \sigma^2(X_{i\Delta_{\overline{n},\overline{T}}}, \theta^{\sigma}))^2, \tag{6}$$

where $\widehat{\mu}_{(n,T)}(.)$ and $\widehat{\sigma}^2_{(n,T)}(.)$ are defined in Eqs. (3) and (4), respectively. Eqs. (5) and (6) can be interpreted as the integrated mean-squared differences between the kernel estimates and their corresponding parametric specifications.

The kernel estimates are defined over an enlarging time span T, whereas the criteria are defined over a fixed time span $\overline{T} \leqslant T$. In both cases, we assume that the distance between observations goes to zero asymptotically, namely $\Delta_{\overline{n},\overline{T}} \to 0$ and $\Delta_{n,T} \to 0$. Our sampling scheme can be easily understood with an example.[2] Assume $T = \sqrt{n}$, for instance, but a different increasing function of n could be adopted. Then, the observations in the full sample are equispaced at times $\{1/\sqrt{n}, 2/\sqrt{n}, \ldots, 1, 1 + 1/\sqrt{n}, \ldots, \sqrt{n}\}$ since $\Delta_{n,T} = T/n = 1/\sqrt{n}$. We can now split the sample into two parts, namely observations in $(0, \overline{T}]$ and observations in $(\overline{T}, T]$. Assume, without loss of generality, that $\overline{T} = 1$. Also, in agreement with our previous notation, assume that there are $\overline{n}$ equispaced observations in the first part of the sample. Then, $1/\overline{n} = 1/\sqrt{n}$. This implies that the number of observations in the first part of the sample, which is defined over a fixed time span$\overline{T} = 1$, grows with $\sqrt{n}$, whereas the number of observations in the second part of the sample grows with n. Given this discussion, one should really write T_n and $\overline{n}_n$ to make the dependence of T and $\overline{n}$ on n explicit. We choose to simply write T and $\overline{n}$ for conciseness in the formulae.

From a theoretical standpoint, fixing the time span over which the criteria are defined is a convenient way to discuss consistency issues, as in Theorems 1 and 3, without having to deal with a possibly unbalanced criterion function. The intuition is as follows. As we show in Theorems 1 and 3, the criteria depend on a random quantity, i.e., local time, which diverges to infinity almost surely in the case of recurrent processes. In order for the criteria to be bounded in probability, local time would have to be defined over a fixed observation span. This is what our sampling scheme accomplishes. Alternatively, one could let $\overline{T}$ go off to infinity just like T, but local time would have to be standardized appropriately for the criteria to be bounded in probability. The standardization would have to be process specific and, as such, would defeat the goal of the present paper.[3] Having made this point, we should stress that it is relatively straightforward to obtain weak convergence results even when $\overline{T} \to \infty$ (see, e.g., Remarks 12 and 13).

From an applied standpoint, fixing the time span $\overline{T}$ over which the criteria are constructed while defining the kernel estimates over an enlarging time span T is immaterial. It simply implies that the entire sample (i.e., data between 0 and T) is used to define the kernel estimates, whereas the first part of the sample (i.e., data between 0 and $\overline{T}$

[2]We thank an anonymous referee for suggesting this example.

[3]Even given a complete parametric model that fully specifies drift and diffusion in the recurrent class, the relevant standardization would be known only in few specific cases (see Section 5). In general, however, one might be simply interested in either the drift or the diffusion function. In this case, one might wish to avoid imposing unnecessary structure on the other infinitesimal moment.

with $\overline{T} \leqslant T$) is used to define the criteria. But, of course, the first part of the sample can be chosen to be large (i.e., $\overline{T}$ can be chosen to be approximately equal to T, if not equal to T).

To summarize, in the sequel the notation $T \to \infty$ will refer to the situation where the kernel estimates are defined over an enlarging span of time. The criteria in Eqs. (5) and (6) will always be defined over a *fixed* time span $\overline{T} \leqslant T$ unless otherwise noted (cf., Remarks 12 and 13). In all cases n, the number of equispaced observations between 0 and T, and $\bar{n}$, the number of equispaced observations between 0 and $\overline{T}$, will be assumed to diverge to infinity with $\Delta_{n,T} = T/n$ and $\Delta_{\bar{n},\overline{T}} = \overline{T}/\bar{n}$ going to zero.

Specifically, we will use the notation $\underset{n,\bar{n},T\to\infty}{\overset{\text{p}}{\to}}$ and $\underset{n,\bar{n},T\to\infty}{\Rightarrow}$ for consistency and weak convergence results obtained as the time span T over which the kernel estimates are defined increases while the time span $\overline{T}$ over which the criteria are defined are fixed. We will use the notation $\underset{n\to\infty}{\overset{\text{p}}{\to}}$ and $\underset{n\to\infty}{\Rightarrow}$ for consistency and weak convergence results obtained as both the time span T over which the kernel estimates are defined and the time span $\overline{T}$ over which the criteria are defined are fixed (in this case we will also assume that $\overline{T} = T =$ constant and $n = \bar{n}$). Finally, we will use the notation $\underset{n,\overline{T}\to\infty}{\Rightarrow}$ to define weak convergence results obtained as both the time span T over which the kernel estimates are defined and the time span $\overline{T}$ over which the criteria are defined increase asymptotically (in this case, again, we will assume that $\overline{T} = T$and $n = \bar{n}$).

The parametric estimates $\widehat{\theta}^{\mu}_{n,\bar{n},T}$ and $\widehat{\theta}^{\sigma}_{n,\bar{n},T}$ are obtained as follows:

$$\widehat{\theta}^{\mu}_{n,\bar{n},T} := \underset{\theta^{\mu}\in\Theta^{\mu}\subset\Theta}{\arg\min}\, Q^{\mu}_{n,\bar{n},T} = \underset{\theta^{\mu}\in\Theta^{\mu}\subset\Theta}{\arg\min} \frac{\overline{T}}{\bar{n}} ||\widehat{\boldsymbol{\mu}} - \boldsymbol{\mu}(\theta^{\mu})||^2 \tag{7}$$

and

$$\widehat{\theta}^{\sigma}_{n,\bar{n},T} := \underset{\theta^{\sigma}\in\Theta^{\sigma}\subset\Theta}{\arg\min}\, Q^{\sigma}_{n,\bar{n},T} = \underset{\theta^{\sigma}\in\Theta^{\sigma}\subset\Theta}{\arg\min} \frac{\overline{T}}{\bar{n}} ||\widehat{\boldsymbol{\sigma}}^2 - \boldsymbol{\sigma}^2(\theta^{\sigma})||^2. \tag{8}$$

Remark 3. As in the fully nonparametric case discussed in BP (2003), we identify the drift and diffusion parameters separately. This is of particular importance when one is interested in the parametrization of a specific function in situations where the other function is treated as a nuisance parameter. On the other hand, the drift and the diffusion function can have parameters in common. If this is the case, one should entertain the possibility of achieving efficiency gains by accounting for this commonality. We discuss the case of common elements in Section 7.

4. Limit theory

We start with the drift case. In what follows, we use the notation $\mu(a, \theta^{\mu}_0)$ and $\mu_0(a)$ interchangeably. Equivalently, we use interchangeably the notation $\sigma^2(a, \theta^{\sigma}_0)$ and $\sigma^2_0(a)$. These notations are convenient and should cause no confusion.

Theorem 1 (*Consistency of the drift parameter estimates*). *Assume* $n, \bar{n} \to \infty$, $T \to \infty$, *and* $h_{n,T} \to 0$ (*as* $n, T \to \infty$) *so that* $(\overline{L}_X(T,x)/h_{n,T})(\Delta_{n,T}\log(1/\Delta_{n,T}))^{1/2} = \mathrm{o}_{\text{a.s.}}(1)$ *and* $\overline{L}_X(T,x)h_{n,T} \overset{\text{a.s.}}{\to} \infty\ \forall x \in \mathfrak{D}$, *then*

$$Q^{\mu}_{n,\bar{n},T}(\theta^{\mu}) \underset{n,\bar{n},T\to\infty}{\overset{\text{p}}{\to}} Q^{\mu}(\theta^{\mu}, \theta_0) = \int_{\mathfrak{D}} (\mu(a, \theta^{\mu}_0) - \mu(a, \theta^{\mu}))^2 \overline{L}_X(\overline{T}, a)\,\mathrm{d}a \tag{9}$$

uniformly in θ^{μ}, *where* $\overline{L}_X(\overline{T},a)$ *is the chronological local time of the underlying diffusion process at* $\overline{T}$ *and* a, *i.e.*, *the nondecreasing* (*in* $\overline{T}$) *random process which satisfies*

$$\overline{L}_X(\overline{T},a)=\lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\frac{1}{\sigma_0^2(a)}\int_0^{\overline{T}}\mathbf{1}_{[a,a+\varepsilon)}(X_s)\sigma_0^2(X_s)\,\mathrm{d}s,$$

with probability one. *Now*, *let* $B(\theta^{\mu},\varepsilon)$ *denote an open ball of radius* ε *around* θ^{μ} *in* Θ^{μ}. *Assume that* $\forall\varepsilon>0$

$$\inf_{\theta^{\mu}\notin B(\theta_0^{\mu},\varepsilon)}\int_{\mathfrak{D}}(\mu(a,\theta_0^{\mu})-\mu(a,\theta^{\mu}))^2\overline{L}_X(\overline{T},a)\,\mathrm{d}a>0\quad\text{a.s.}\tag{10}$$

Then,

$$\widehat{\theta}^{\mu}_{n,\overline{n},T}\underset{n,\overline{n},T\to\infty}{\overset{\mathrm{p}}{\to}}\theta_0^{\mu}.$$

Theorem 2 (*The limit distribution of the drift parameter estimates*). *Given* $n,\overline{n}\to\infty$, $T\to\infty$, *and* $h_{n,T}\to 0$ (*as* $n,T\to\infty$) *such that* $(\overline{L}_X(T,x)/h_{n,T})(\Delta_{n,T}\log(1/\Delta_{n,T}))^{1/2}=\mathrm{o}_{\mathrm{a.s.}}(1)$, $\Xi_{\mu}^{-1}(T)h_{n,T}^4\overset{\mathrm{a.s.}}{\to}0$, *and* $\overline{L}_X(T,x)h_{n,T}\overset{\mathrm{a.s.}}{\to}\infty$ $\forall x\in\mathfrak{D}$, *then*

$$(\widehat{\Xi}_{\mu}(T))^{-1/2}(\widehat{\theta}^{\mu}_{n,\overline{n},T}-\theta_0^{\mu})\underset{n,\overline{n},T\to\infty}{\Rightarrow}\mathbf{N}(0,\mathbf{I}_{m_1}),\tag{11}$$

where $\widehat{\Xi}_{\mu}(T)$ *is a consistent estimate of* $\Xi_{\mu}(T)$ *as defined by*

$$\Xi_{\mu}(T)=B(\overline{T})_{\mu}^{-1}V(T)_{\mu}B(\overline{T})_{\mu}^{-1},\tag{12}$$

with

$$B(\overline{T})_{\mu}=\left(\int_{\mathfrak{D}}\frac{\partial\mu_0(a)}{\partial\theta^{\mu}}\frac{\partial\mu_0(a)}{\partial\theta^{\mu'}}\overline{L}_X(\overline{T},a)\,\mathrm{d}a\right),$$

$$V(T)_{\mu}=\left(\int_{\mathfrak{D}}\sigma_0^2(a)\left(\frac{\partial\mu_0(a)}{\partial\theta^{\mu}}\frac{\partial\mu_0(a)}{\partial\theta^{\mu'}}\right)\frac{(\overline{L}_X(\overline{T},a))^2}{\overline{L}_X(T,a)}\,\mathrm{d}a\right),$$

where $\overline{L}_X(.,a)$ *is the chronological local time of the underlying diffusion process at* a. *If* $\Xi_{\mu}^{-1}(T)h_{n,T}^4=\mathrm{O}_{\mathrm{a.s.}}(1)$, *then*

$$(\widehat{\Xi}_{\mu}(T))^{-1/2}(\widehat{\theta}^{\mu}_{n,\overline{n},T}-\theta_0^{\mu}-\Gamma^{\mu})\underset{n,\overline{n},T\to\infty}{\Rightarrow}\mathbf{N}(0,\mathbf{I}_{m_1}),$$

where

$$\Gamma^{\mu}=h_{n,T}^2B(\overline{T})_{\mu}^{-1}\int_{\mathfrak{D}}\mathbf{K}_2\left(\frac{\partial\mu_0(a)}{\partial a}\frac{\frac{\partial m(a)}{\partial a}}{m(a)}+\frac{1}{2}\frac{\partial^2\mu_0(a)}{\partial a}\right)\frac{\partial\mu_0(a)}{\partial\theta^{\mu}}\overline{L}_X(\overline{T},a)\,\mathrm{d}a,$$

$m(.)$ *is the speed function of the underlying diffusion and* $\mathbf{K}_2=\int_{-\infty}^{\infty}c^2\mathbf{K}(c)\,\mathrm{d}c$.

Remark 4. Both the chronological local time $\overline{L}_X(T,x)$, i.e., the random amount of time that the diffusion spends in the local neighborhood of the generic spatial point x, and the speed function of the process of interest play a role in the definition of our asymptotics. This is a by-product of the generality of our assumptions.

As opposed to the time-invariant probability density that emerges from stationary estimation procedures, both quantities are known to be well-defined for stationary as well as for nonstationary diffusion processes, while having a close connection to the stationary density $f(x)$ should positive recurrence, or strict stationary, be satisfied. In fact,

$$\frac{\overline{L}_X(T,x)}{T} \xrightarrow{\text{p}} f(x) = \frac{m(x)}{\overline{m}} \tag{13}$$

$\forall x \in \mathfrak{D}$ as $T \to \infty$ when the process is positive recurrent ($\overline{m} < \infty$). Theorem 6.3, in Bosq (1998, p. 150) contains an even stronger (with probability one) consistency result in the case of strictly stationary processes. For positive recurrent processes, the result can be derived from an application of the Darling–Kac theorem, for example (see, e.g., Darling and Kac, 1957; Bandi and Moloche, 2001 for a recent use of the theorem).

Eq. (13) says that the standardized local time of a positive recurrent diffusion process converges to its stationary density. Additionally, $\overline{L}_X(T,x)$ diverges linearly with T. If $\overline{m} = \infty$ and the process is null recurrent, then local time diverges at a speed slower than T. In general, the local time of a recurrent process diverges to infinity with T almost surely since (i) the process visits every level in its range an infinity number of times as the time span increases indefinitely and (ii) local time measures data density. As shown, the divergence properties of the local time factor affect the convergence properties of the drift parameter estimates (cf., Eq. (11)). A similar result applies to the diffusion case that we discuss below.

Remark 5. For a smoothing sequence converging to zero at a fast enough rate as to eliminate the asymptotic bias term Γ^{μ} (i.e., so that $\Xi_{\mu}^{-1}(T)h_{n,T}^{4} \xrightarrow{\text{a.s.}} 0$), the weak convergence result in Eq. (11) is consistent with what we would expect to obtain in a correctly specified standard nonlinear regression context with heteroskedastic errors (see, e.g., Davidson and MacKinnon, 1993 for a classical treatment). The only difference is that we replace integrals with respect to probability measures with spatial integrals, i.e., integrals defined with respect to local time (see, e.g., Park and Phillips, 1999, 2001 for discussions in the context of unit-root models for discrete time series).

Remark 6. Coherently with the fully nonparametric case discussed elsewhere (BP, 2003), the rate of convergence is path-dependent and is driven by the rate of divergence to infinity of the local time factor through the spatial integral V_{μ}. By virtue of the averaging, this rate is generally faster than in the fully functional context where it is known to be equal to $\sqrt{h_{n,T}\,\overline{L}_X(T,x)}$.

Remark 7. 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 $\overline{L}_X(\overline{T},.)$ is bounded in probability and does not diverge to infinity with probability one. Consequently, the matrix $\widehat{\Xi}_{\mu}(T) = \widehat{\Xi}_{\mu}(\overline{T}) = \Xi_{\mu}(\overline{T}) + \text{o}_{\text{p}}(1)$ is also bounded in probability. Hence, $\widehat{\theta}^{\mu}_{n(=\overline{n}),\overline{T}} \underset{n\to\infty}{\overset{\text{p}}{\nrightarrow}} \theta^{\mu}_{0}$ when T is fixed (cf., Eq. (11)). Thus, even though we define the criterion over a fixed span of data $\overline{T}$, the drift kernel estimates ought to be defined over an enlarging span of observations to obtain consistency of the drift parameter estimates. This result mirrors the analogous result in the fully functional case where it was shown that, contrary to the diffusion function, the drift term cannot be estimated over a fixed observation span (see, e.g., BP, 2003).

We now turn to the diffusion parameter estimates.

Theorem 3 (*Consistency of the diffusion parameter estimates*). *Assume $n, \overline{n} \to \infty$, $T \to \infty$, and $h_{n,T} \to 0$ (as $n, T \to \infty$) such that $(\overline{L}_X(T,x)/h_{n,T})(\Delta_{n,T}\log(1/\Delta_{n,T}))^{1/2} = \text{o}_{\text{a.s.}}(1)$*

$\forall x \in \mathfrak{D}$, *then*

$$Q^{\sigma}_{n,\bar{n},T}(\theta^{\sigma}) \underset{n,\bar{n},T\to\infty}{\overset{\mathrm{p}}{\to}} Q^{\sigma}(\theta^{\sigma},\theta_0) = \int_{\mathfrak{D}} (\sigma^2(a,\theta_0^{\sigma}) - \sigma^2(a,\theta^{\sigma}))^2 \overline{L}_X(\overline{T},a)\,\mathrm{d}a \tag{14}$$

uniformly in θ^{σ}, *where* $\overline{L}_X(\overline{T},a)$ *is the chronological local time of the underlying diffusion process at* $\overline{T}$ *and* a, *i.e., the nondecreasing (in* $\overline{T}$*) random process which satisfies*

$$\overline{L}_X(\overline{T},a) = \lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\frac{1}{\sigma_0^2(a)}\int_0^{\overline{T}} \mathbf{1}_{[a,a+\varepsilon)}(X_s)\sigma_0^2(X_s)\,\mathrm{d}s,$$

with probability one. Now, let $B(\theta^{\sigma},\varepsilon)$ *denote an open ball of radius* ε *around* θ^{σ} *in* Θ^{σ}. *Assume that* $\forall\varepsilon > 0$

$$\inf_{\theta^{\sigma}\notin B(\theta_0^{\sigma},\varepsilon)} \int_{\mathfrak{D}} (\sigma^2(a,\theta_0^{\sigma}) - \sigma^2(a,\theta^{\sigma}))^2 \overline{L}_X(\overline{T},a)\,\mathrm{d}a > 0 \quad \text{a.s.} \tag{15}$$

Then,

$$\widehat{\theta}^{\sigma}_{n,\bar{n},T} \underset{n,\bar{n},T\to\infty}{\overset{\mathrm{p}}{\to}} \theta_0^{\sigma}.$$

Theorem 4 (*The limit distribution of the diffusion parameter estimates*). *Given* $n,\bar{n}\to\infty$, $T\to\infty$, *and* $h_{n,T}\to 0$ *(as* $n,T\to\infty$*) such that* $(\overline{L}_X(T,x)/h_{n,T})(\Delta_{n,T}\log(1/\Delta_{n,T}))^{1/2} = o_{\mathrm{a.s.}}(1)$ $\forall x\in\mathfrak{D}$ *and* $\Xi_{\sigma}^{-1}(T)h^4_{n,T}/\Delta_{n,T} \overset{\mathrm{a.s.}}{\to} 0$, *then*

$$\frac{1}{\sqrt{\Delta_{n,T}}}(\widehat{\Xi}_{\sigma}(T))^{-1/2}(\widehat{\theta}^{\sigma}_{n,\bar{n},T} - \theta_0^{\sigma}) \underset{n,\bar{n},T\to\infty}{\Rightarrow} \mathbf{N}(0,\mathbf{I}_{m_2}), \tag{16}$$

where $\widehat{\Xi}_{\sigma}(T)$ *is a consistent estimate of* $\Xi_{\sigma}(T)$ *as defined by*

$$\Xi_{\sigma}(T) = B(\overline{T})_{\sigma}^{-1}V(T)_{\sigma}B(\overline{T})_{\sigma}^{-1} \tag{17}$$

with

$$B(\overline{T})_{\sigma} = \left(\int_{\mathfrak{D}} \frac{\partial\sigma_0^2(a)}{\partial\theta^{\sigma}}\frac{\partial\sigma_0^2(a)}{\partial\theta^{\sigma'}}\overline{L}_X(\overline{T},a)\,\mathrm{d}a\right),$$

$$V(T)_{\sigma} = \left(\int_{\mathfrak{D}} 2\sigma_0^4(a)\left(\frac{\partial\sigma_0^2(a)}{\partial\theta^{\sigma}}\frac{\partial\sigma_0^2(a)}{\partial\theta^{\sigma'}}\right)\frac{(\overline{L}_X(\overline{T},a))^2}{\overline{L}_X(T,a)}\,\mathrm{d}a\right),$$

where $\overline{L}_X(.,a)$ *is the chronological local time of the underlying diffusion process at* a. *If* $\Xi_{\sigma}^{-1}(T)h^4_{n,T}/\Delta_{n,T} = O_{\mathrm{a.s.}}(1)$, *then*

$$\frac{1}{\sqrt{\Delta_{n,T}}}(\widehat{\Xi}_{\sigma}(T))^{-1/2}(\widehat{\theta}^{\sigma}_{n,\bar{n},T} - \theta_0^{\sigma} - \Gamma^{\sigma}) \underset{n,\bar{n},T\to\infty}{\Rightarrow} \mathbf{N}(0,\mathbf{I}_{m_2}),$$

where

$$\Gamma^{\sigma} = h^2_{n,T}B(\overline{T})_{\sigma}^{-1}\int_{\mathfrak{D}} \mathbf{K}_2\left(\frac{\partial\sigma_0^2(a)}{\partial a}\frac{\frac{\partial m(a)}{\partial a}}{m(a)} + \frac{1}{2}\frac{\partial^2\sigma_0^2(a)}{\partial a}\right)\frac{\partial\sigma_0^2(a)}{\partial\theta^{\sigma}}\overline{L}_X(\overline{T},a)\,\mathrm{d}a,$$

$m(.)$ *is the speed function of the underlying diffusion and* $\mathbf{K}_2 = \int_{-\infty}^{\infty} c^2\mathbf{K}(c)\,\mathrm{d}c$.

Remark 8. In light of Remark 5, the integrals B_{σ} and V_{σ} can be interpreted as spatial analogs of the integrals with respect to probability measures that would arise from the standard nonlinear estimation of conditional expectations in discrete time. The term $2\sigma_0^4(a)$ is due to the quadratic nature of the nonparametric estimator of the infinitesimal second moment.

Remark 9. As in the drift case, the rate of convergence is path-dependent being driven by a local time factor. Also, the parametric estimates entail efficiency gains with respect to their nonparametric counterparts. In fact, the functional estimates have generally slower pointwise convergence rates given by $\sqrt{h_{n,T}\overline{L}_X(T,x)}/\sqrt{\Delta_{n,T}}$ (BP, 2003).

Remark 10. The rate of convergence of the diffusion estimates is faster than the rate of convergence of the drift estimates. The difference is given by the multiplicative factor $1/\sqrt{\Delta_{n,T}} = \sqrt{n/T}$ and is consistent with corresponding results in the fully functional case.

We now consider the case where the diffusion parameters are estimated by defining *both* the kernel estimates and the relevant criterion over a fixed observation span. In other words, we assume that $T = \overline{T}$ and is fixed. The symbol **MN** in Theorem 5 denotes a mixed normal distribution.

Theorem 5 (*The limit distribution of the diffusion parameter estimates with T fixed*). *Given* $n(=\overline{n}) \to \infty$ *and* $h_{n,\overline{T}} \to 0$ *(as* $n \to \infty$*) so that* $(1/h_{n,\overline{T}})(\Delta_{n,\overline{T}}\log(1/\Delta_{n,\overline{T}}))^{1/2} = \mathrm{o}(1)$ *and* $h^4_{n,\overline{T}}/\Delta_{n,\overline{T}} \to 0$, *then*

$$\frac{1}{\sqrt{\Delta_{n,\overline{T}}}}(\widehat{\theta}^{\sigma}_{n,\overline{T}} - \theta^{\sigma}_0) \underset{n\to\infty}{\Rightarrow} \mathbf{MN}(0, \Xi_\sigma(\overline{T}))$$

with

$$\Xi_\sigma(\overline{T}) = B(\overline{T})^{-1}_\sigma V(\overline{T})_\sigma B(\overline{T})^{-1}_\sigma \tag{18}$$

and

$$B(\overline{T})_\sigma = \left(\int_{\mathfrak{D}} \frac{\partial\sigma_0^2(a)}{\partial\theta^\sigma}\frac{\partial\sigma_0^2(a)}{\partial\theta^{\sigma'}}\overline{L}_X(\overline{T},a)\,\mathrm{d}a\right),$$

$$V(\overline{T})_\sigma = \left(\int_{\mathfrak{D}} 2\sigma_0^4(a)\left(\frac{\partial\sigma_0^2(a)}{\partial\theta^\sigma}\frac{\partial\sigma_0^2(a)}{\partial\theta^{\sigma'}}\right)\overline{L}_X(\overline{T},a)\,\mathrm{d}a\right),$$

where $\overline{L}_X(.,a)$ *is the chronological local time of the underlying diffusion process at* a.

Remark 11. The diffusion parameters can be identified over a fixed time span. Hence, recurrence is not necessary to identify the second infinitesimal moment and the process can be transient. In this case, the convergence rate ceases to be path-dependent. We experience $\sqrt{n}$-convergence for the parametric estimates (since $1/\sqrt{\Delta_{n,\overline{T}}} = \sqrt{n}/\sqrt{\overline{T}}$ and $\overline{T}$ is fixed) and $\sqrt{nh_{n,\overline{T}}}$-convergence for the nonparametric estimates in Eq. (4) above (see, e.g., BP, 2003). The gain in efficiency which is guaranteed by the adoption of the parametric approach in this paper is noteworthy and coherent with more traditional semiparametric models in discrete time (see, e.g., Andrews, 1989).

5. Some special cases: Brownian motion and stationary processes

Since the rate of convergence of the estimates is influenced by the rate of divergence to infinity of the chronological local time factor, it is worth analyzing the cases for which such a rate is known in closed-form, namely Brownian motion and the wide class of positive recurrent and stationary processes.

It is important to point out again that consistent estimation of either infinitesimal moment of interest does not require a complete parametrization of the underlying process. Hence, potential users do not have to take an a priori stand on the stationarity properties of the process in general. This is an important aspect of our methodology. Furthermore, the dynamic features of the process affect the limiting distributions only through estimable random objects that characterize the variance of asymptotically normal variates. While null recurrent processes are expected to converge at a slower pace than positive recurrent and strictly stationary processes due to the slower divergence rates of the corresponding local time factors (cf., Remark 4), the convergence rates are embodied in random variance–covariance matrices (in Eqs. (12) and (17)) which can be estimated from the data as we discuss in Section 6. Consistent estimation of the variance–covariance matrices only requires recurrence.

In what follows we explicitly discuss the convergence rates of the parametric estimates in the two cases that were mentioned above: Brownian motion and positive recurrent (as well as strictly stationary) processes. The results in this section are mainly of a theoretical interest but can also be of help to the user should stationarity of the underlying process be known, for instance. We start with Brownian motion.

5.1. Brownian motion

Assume the data are generated from a Brownian motion $\widetilde{B} = \sigma B_t$ with local variance σ^2. We parametrize the diffusion process as

$$X_t = X_0 + \int_0^t \mu \, \mathrm{d}s + \int_0^t \sigma \, \mathrm{d}B_s$$

and minimize the criteria in Eqs. (5) and (6). It follows that

$$\widehat{\theta}^{\mu}_{n,\bar{n},T} = \frac{\overline{T}}{\bar{n}} \sum_{i=1}^{\bar{n}} \widehat{\mu}_{(n,T)}(X_{i\Delta_{\bar{n},\overline{T}}}) = \widehat{\mu}_{n,\bar{n},T}$$

and

$$\widehat{\theta}^{\sigma}_{n,\bar{n},T} = \sqrt{\frac{\overline{T}}{\bar{n}} \sum_{i=1}^{\bar{n}} \widehat{\sigma}^2_{(n,T)}(X_{i\Delta_{\bar{n},\overline{T}}})} = \widehat{\sigma}_{n,\bar{n},T}.$$

The limit theories can be expressed in closed-form since the rate of divergence to infinity of the Brownian local time is known. In particular, $\mathfrak{D} = (-\infty, \infty)$ and

$$\begin{aligned} B_{\mu} &= \left(\int_{-\infty}^{\infty} \frac{\partial \mu_0(a)}{\partial \theta^{\mu}} \frac{\partial \mu_0(a)}{\partial \theta^{\mu'}} \overline{L}_X(\overline{T}, a) \, \mathrm{d}a \right) \\ &= \int_{-\infty}^{\infty} \overline{L}_{\widetilde{B}}(\overline{T}, a) \, \mathrm{d}a \end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{\sigma^2}[\widetilde{B}]_{\overline{T}} \\
&= \int_{-\infty}^{\infty} \frac{1}{\sigma}\overline{T}^{1/2} L_B\left(1, \frac{1}{\overline{T}^{1/2}}\frac{a}{\sigma}\right) \mathrm{d}a \\
&= \int_{-\infty}^{\infty} \overline{T} L_B(1,x)\,\mathrm{d}x \\
&= \overline{T}[B]_1 \\
&= \overline{T},
\end{aligned}
$$

also

$$
\begin{aligned}
V_\mu &= \left(\int_{-\infty}^{\infty} \sigma_0^2(a)\left(\frac{\partial\mu_0(a)}{\partial\theta^\mu}\frac{\partial\mu_0(a)}{\partial\theta^{\mu'}}\right)\frac{(\overline{L}_X(\overline{T},a))^2}{\overline{L}_X(T,a)}\,\mathrm{d}a\right) \\
&= \int_{-\infty}^{\infty} \sigma^2 \frac{\frac{1}{\sigma^2}(\overline{T}^{1/2}L_B(1,\frac{1}{\overline{T}^{1/2}}\frac{a}{\sigma}))^2}{\frac{1}{\sigma}T^{1/2}L_B(1,\frac{1}{T^{1/2}}\frac{a}{\sigma})}\,\mathrm{d}a \\
&= \int_{-\infty}^{\infty} \sigma^2 \frac{1}{\sigma}\frac{\overline{T}^{1/2}}{\overline{T}^{1/2}}\frac{(\overline{T}^{1/2}L_B(1,\frac{1}{\overline{T}^{1/2}}\frac{a}{\sigma}))^2}{T^{1/2}L_B(1,\frac{1}{T^{1/2}}\frac{\overline{T}^{1/2}}{\overline{T}^{1/2}}\frac{a}{\sigma})}\,\mathrm{d}a \\
&= \int_{-\infty}^{\infty} \sigma^2\overline{T}^{1/2}\frac{(\overline{T}^{1/2}L_B(1,x))^2}{T^{1/2}L_B(1,\frac{\overline{T}^{1/2}}{T^{1/2}}x)}\,\mathrm{d}x \\
&= \frac{1}{T^{1/2}L_B(1,0+\mathrm{o}(1))}\int_{-\infty}^{\infty}\sigma^2\overline{T}^{3/2}(L_B(1,x))^2\,\mathrm{d}x.
\end{aligned}
$$

Then,

$$
T^{1/4}\left(\int_{-\infty}^{\infty}\sigma^2\left(\frac{(L_B(1,x))^2}{\overline{T}^{1/2}L_B(1,0)}\right)\mathrm{d}x + \mathrm{o_p}(1)\right)^{-1/2}(\widehat{\theta}^\mu_{n,\bar{n},T}-\mu_0)\underset{n,\bar{n},T\to\infty}{\Rightarrow}\mathbf{N}(0,1)
$$

with $\mu_0 = 0$. The rate of convergence, $T^{1/4}$, is faster than in the fully nonparametric case, where it is known to be $T^{1/4}h_{n,T}^{1/2}$ (BP, 2003).

We now turn to diffusion estimation. Write

$$
\begin{aligned}
B_\sigma &= \left(\int_{-\infty}^{\infty}\frac{\partial\sigma_0^2(a)}{\partial\theta^\sigma}\frac{\partial\sigma_0^2(a)}{\partial\theta^{\sigma'}}\overline{L}_X(\overline{T},a)\,\mathrm{d}a\right) \\
&= \int_{-\infty}^{\infty}4\sigma^2\overline{L}_{\widetilde{B}}(\overline{T},a)\,\mathrm{d}a \\
&= 4\sigma^2\overline{T}
\end{aligned}
$$

and

$$
\begin{aligned}
V_\sigma &= \left(\int_{-\infty}^{\infty}2\sigma_0^4(a)\left(\frac{\partial\sigma_0^2(a)}{\partial\theta^\sigma}\frac{\partial\sigma_0^2(a)}{\partial\theta^{\sigma'}}\right)\frac{(\overline{L}_X(\overline{T},a))^2}{\overline{L}_X(T,a)}\,\mathrm{d}a\right) \\
&= \int_{-\infty}^{\infty}2\sigma^4(4\sigma^2)\left(\frac{(\overline{L}_{\widetilde{B}}(\overline{T},a))^2}{\overline{L}_{\widetilde{B}}(T,a)}\right)\mathrm{d}a
\end{aligned}
$$

$$= \frac{1}{T^{1/2} L_B(1, 0 + \mathrm{o}(1))} \int_{-\infty}^{\infty} 2\sigma^4 (4\sigma^2) \overline{T}^{3/2} (L_B(1, x))^2 \, \mathrm{d}x.$$

In consequence,

$$\frac{T^{1/4}}{\sqrt{\Delta_{n,T}}} \left(\int_{-\infty}^{\infty} \frac{\sigma^2}{2} \left(\frac{(L_B(1, x))^2}{\overline{T}^{1/2} L_B(1, 0)} \right) \mathrm{d}x + \mathrm{o_p}(1) \right) (\widehat{\theta}^{\sigma}_{n,\overline{n},T} - \sigma) \underset{n,\overline{n},T \to \infty}{\Rightarrow} \mathbf{N}(0, 1).$$

As in the previous case, the rate of convergence that would emerge from purely functional estimation is slower and equals $(T^{1/4} \sqrt{\Delta_{n,T}}) h_{n,T}^{1/2}$.

Remark 12. It appears that we can increase further the rate of convergence by working with criteria defined over an enlarging time span $\overline{T} = T(\to \infty)$ implying $n = \overline{n}$. In this case,

$$\sqrt{T}(\widehat{\theta}^{\mu}_{n,\overline{T}} - \mu_0) \underset{n,\overline{T} \to \infty}{\Rightarrow} \mathbf{N}(0, \sigma^2),$$

with $\mu_0 = 0$, and

$$\sqrt{n}(\widehat{\theta}^{\sigma}_{n,\overline{T}} - \sigma) \underset{n,\overline{T} \to \infty}{\Rightarrow} \mathbf{N}(0, \tfrac{1}{2}\sigma^2).$$

5.2. Positive recurrent and stationary processes

Since local time converges to the stationary density of the process $f(.)$ when standardized by T (cf., Remark 4), in the drift case we obtain

$$\sqrt{T}(\Xi_{\mu} + \mathrm{o_p}(1))^{-1/2} (\widehat{\theta}^{\mu}_{n,\overline{n},T} - \theta^{\mu}_0) \underset{n,\overline{n},T \to \infty}{\Rightarrow} \mathbf{N}(0, \mathbf{I}_{m_1}),$$

where

$$\Xi_{\mu} = B(\overline{T})^{-1}_{\mu} V_{\mu} B(\overline{T})^{-1}_{\mu}$$

and

$$B(\overline{T})_{\mu} = \left(\int_{\mathfrak{D}} \frac{\partial \mu_0(a)}{\partial \theta^{\mu}} \frac{\partial \mu_0(a)}{\partial \theta^{\mu'}} \overline{L}_X(\overline{T}, a) \, \mathrm{d}a \right),$$

$$V_{\mu} = \left(\int_{\mathfrak{D}} \sigma_0^2(a) \left(\frac{\partial \mu_0(a)}{\partial \theta^{\mu}} \frac{\partial \mu_0(a)}{\partial \theta^{\mu'}} \right) \frac{(\overline{L}_X(\overline{T}, a))^2}{f(a)} \, \mathrm{d}a \right).$$

In agreement with the Brownian motion case, the rate of convergence, $\sqrt{T}$, is faster than in the fully nonparametric case where it was shown to be $\sqrt{h_{n,T} T}$ (BP, 2003). As for the diffusion case, we can write

$$\frac{\sqrt{T}}{\sqrt{\Delta_{n,T}}} (\Xi_{\sigma} + \mathrm{o_p}(1))^{-1/2} (\widehat{\theta}^{\sigma}_{n,\overline{n},T} - \theta^{\sigma}_0)$$
$$= \sqrt{n} (\Xi_{\sigma} + \mathrm{o_p}(1))^{-1/2} (\widehat{\theta}^{\sigma}_{n,\overline{n},T} - \theta^{\sigma}_0) \underset{n,\overline{n},T \to \infty}{\Rightarrow} \mathbf{N}(0, \mathbf{I}_{m_2}),$$

where

$$\Xi_{\sigma} = B(\overline{T})^{-1}_{\sigma} V_{\sigma} B(\overline{T})^{-1}_{\sigma}$$

and

$$B(\overline{T})_\sigma = \left(\int_{\mathfrak{D}} \frac{\partial\sigma_0^2(a)}{\partial\theta^\sigma}\frac{\partial\sigma_0^2(a)}{\partial\theta^{\sigma'}}\overline{L}_X(\overline{T},a)\,\mathrm{d}a\right),$$

$$V_\sigma = \left(\int_{\mathfrak{D}} 2\sigma_0^4(a)\left(\frac{\partial\sigma_0^2(a)}{\partial\theta^\sigma}\frac{\partial\sigma_0^2(a)}{\partial\theta^{\sigma'}}\right)\frac{(\overline{L}_X(\overline{T},a))^2}{f(a)}\,\mathrm{d}a\right).$$

Again, the diffusion estimates converge at a faster speed, $\sqrt{n}$, than in the fully functional case, $\sqrt{nh_{n,T}}$.

Remark 13. We can now define the criteria over an enlarging time span $\overline{T} = T(\to\infty)$ with $\overline{n} = n$. Contrary to the Brownian motion case, no additional improvement in the convergence rates is obtained over the situation illustrated above. Nonetheless, the asymptotic variances have a more familiar look. In fact,

$$\sqrt{T}(\Xi_\mu + \mathrm{o_p}(1))^{-1/2}(\widehat{\theta}^\mu_{n,\overline{T}} - \theta^\mu_0) \underset{n,\overline{T}\to\infty}{\Rightarrow} \mathbf{N}(0, \mathbf{I}_{m_1}),$$

where

$$\Xi_\mu = B_\mu^{-1}V_\mu B_\mu^{-1},$$

and

$$B_\mu = \left(\int_{\mathfrak{D}} \frac{\partial\mu_0(a)}{\partial\theta^\mu}\frac{\partial\mu_0(a)}{\partial\theta^{\mu'}}f(a)\,\mathrm{d}a\right) = \mathbf{E}\left(\frac{\partial\mu_0(X)}{\partial\theta^\mu}\frac{\partial\mu_0(X)}{\partial\theta^{\mu'}}\right),$$

$$V_\mu = \left(\int_{\mathfrak{D}} \sigma_0^2(a)\left(\frac{\partial\mu_0(a)}{\partial\theta^\mu}\frac{\partial\mu_0(a)}{\partial\theta^{\mu'}}\right)f(a)\,\mathrm{d}a\right) = \mathbf{E}\left(\sigma_0^2(X)\frac{\partial\mu_0(X)}{\partial\theta^\mu}\frac{\partial\mu_0(X)}{\partial\theta^{\mu'}}\right).$$

Additionally,

$$\sqrt{n}(\Xi_\sigma + \mathrm{o_p}(1))^{-1/2}(\widehat{\theta}^\sigma_{n,\overline{T}} - \theta^\sigma_0) \underset{n,\overline{T}\to\infty}{\Rightarrow} \mathbf{N}(0, \mathbf{I}_{m_2}),$$

where

$$\Xi_\sigma = B_\sigma^{-1}V_\sigma B_\sigma^{-1}$$

and

$$B_\sigma = \left(\int_{\mathfrak{D}} \frac{\partial\sigma_0^2(a)}{\partial\theta^\sigma}\frac{\partial\sigma_0^2(a)}{\partial\theta^{\sigma'}}f(a)\,\mathrm{d}a\right) = \mathbf{E}\left(\frac{\partial\sigma_0^2(X)}{\partial\theta^\sigma}\frac{\partial\sigma_0^2(X)}{\partial\theta^{\sigma'}}\right),$$

$$V_\sigma = \left(\int_{\mathfrak{D}} 2\sigma_0^4(a)\left(\frac{\partial\sigma_0^2(a)}{\partial\theta^\sigma}\frac{\partial\sigma_0^2(a)}{\partial\theta^{\sigma'}}\right)f(a)\,\mathrm{d}a\right) = \mathbf{E}\left(2\sigma_0^4(X)\frac{\partial\sigma_0^2(X)}{\partial\theta^\sigma}\frac{\partial\sigma_0^2(X)}{\partial\theta^{\sigma'}}\right).$$

6. Covariance matrix estimation

We now discuss estimation of the covariance matrices in Theorems 2 and 4. We only focus on the first infinitesimal moment. The results readily extend to the diffusion case with obvious modifications. From Theorem 2, write the asymptotic covariance as

$$acov(\widehat{\theta}^{\mu}_{n,\bar{n},T}) = \Xi_{\mu}(\theta_0) = (B_{\mu}(\theta_0))^{-1}(V_{\mu}(\theta_0))(B_{\mu}(\theta_0))^{-1}$$

with

$$B_{\mu}(\theta_0) = \left(\int_{\mathfrak{D}} \frac{\partial\mu(a,\theta_0^{\mu})}{\partial\theta^{\mu}}\frac{\partial\mu(a,\theta_0^{\mu})}{\partial\theta^{\mu'}}\overline{L}_X(\overline{T},a)\,\mathrm{d}a\right)$$

and

$$V_{\mu}(\theta_0) = \left(\int_{\mathfrak{D}} \sigma^2(a,\theta_0^{\sigma})\left(\frac{\partial\mu(a,\theta_0^{\mu})}{\partial\theta^{\mu}}\frac{\partial\mu(a,\theta_0^{\mu})}{\partial\theta^{\mu'}}\right)\frac{(\overline{L}_X(\overline{T},a))^2}{\overline{L}_X(T,a)}\,\mathrm{d}a\right).$$

It is straightforward to show (see the proof of Theorem 4) that

$$\begin{aligned}
&\widehat{B}_{\mu}(\theta^{\mu})_{\bar{n},\overline{T}}\\
&= \Delta_{\bar{n},\overline{T}}\sum_{i=1}^{\bar{n}}\frac{\partial\mu(X_{i\Delta_{\bar{n},\overline{T}}},\theta^{\mu})}{\partial\theta^{\mu}}\frac{\partial\mu(X_{i\Delta_{\bar{n},\overline{T}}},\theta^{\mu})}{\partial\theta^{\mu'}}\\
&\underset{\bar{n}\to\infty}{\overset{\text{a.s.}}{\to}} \int_{\mathfrak{D}}\frac{\partial\mu(a,\theta^{\mu})}{\partial\theta^{\mu}}\frac{\partial\mu(a,\theta^{\mu})}{\partial\theta^{\mu'}}\overline{L}_X(\overline{T},a)\,\mathrm{d}a\\
&= B_{\mu}(\theta)
\end{aligned}$$

and

$$\begin{aligned}
&\widehat{V}_{\mu}(\theta)_{n,\bar{n},T}\\
&= \frac{h_{n,T}}{h_{\bar{n},\overline{T}}}\frac{(\Delta_{\bar{n},\overline{T}})^2}{\Delta_{n,T}}\sum_{i=1}^{\bar{n}}\sigma^2(X_{i\Delta_{\bar{n},\overline{T}}},\theta^{\sigma})\frac{\partial\mu(X_{i\Delta_{\bar{n},\overline{T}}},\theta^{\mu})}{\partial\theta^{\mu}}\frac{\partial\mu(X_{i\Delta_{\bar{n},\overline{T}}},\theta^{\mu})}{\partial\theta^{\mu'}}\frac{\sum_{j=1}^{\bar{n}}\mathbf{K}\left(\frac{X_{j\Delta_{\bar{n},\overline{T}}}-X_{i\Delta_{\bar{n},\overline{T}}}}{h_{\bar{n},\overline{T}}}\right)}{\sum_{j=1}^{n}\mathbf{K}\left(\frac{X_{j\Delta_{n,T}}-X_{i\Delta_{\bar{n},\overline{T}}}}{h_{n,T}}\right)}\\
&\underset{n,\bar{n},T\to\infty}{\overset{\text{a.s.}}{\to}} \int_{\mathfrak{D}}\sigma^2(a,\theta^{\sigma})\left(\frac{\partial\mu(a,\theta^{\mu})}{\partial\theta^{\mu}}\frac{\partial\mu(a,\theta^{\mu})}{\partial\theta^{\mu'}}\right)\frac{(\overline{L}_X(\overline{T},a))^2}{\overline{L}_X(T,a)}\,\mathrm{d}a = V_{\mu}(\theta)
\end{aligned}$$

uniformly in θ. We combine this result with the continuity of $\partial\mu(.,\theta^{\mu})/\partial\theta^{\mu}$ and $\sigma^2(.,\theta^{\sigma})$ at θ_0^{μ} and θ_0^{σ} (cf., Assumption 1) and the consistency of $\widehat{\theta}^{\mu}_{n,\bar{n},T}$ and $\widehat{\theta}^{\sigma}_{n,\bar{n},T}$ (from Theorems 1 and 3) to yield

$$\widehat{B}_{\mu}(\widehat{\theta}^{\mu}_{n,\bar{n},T})_{\bar{n},\overline{T}} \underset{n,\bar{n},T\to\infty}{\overset{\text{p}}{\to}} B_{\mu}(\theta_0)$$

and

$$\widehat{V}_{\mu}(\widehat{\theta}_{n,\bar{n},T})_{n,\bar{n},T} \underset{n,\bar{n},T\to\infty}{\overset{\text{p}}{\to}} V_{\mu}(\theta_0).$$

The proof follows standard arguments in extremum estimation (see the proof of Theorem 2 for a similar derivation). In consequence,

$$\begin{aligned}
\widehat{\Xi}_{\mu}(\widehat{\theta}^{\mu}_{n,\bar{n},T}) &= (\widehat{B}_{\mu}(\widehat{\theta}^{\mu}_{n,\bar{n},T})_{\bar{n},\overline{T}})^{-1}(\widehat{V}_{\mu}(\widehat{\theta}_{n,\bar{n},T})_{n,\bar{n},T})(\widehat{B}_{\mu}(\widehat{\theta}^{\mu}_{n,\bar{n},T})_{\bar{n},\overline{T}})^{-1}\\
&\underset{n,\bar{n},T\to\infty}{\overset{\text{p}}{\to}} (B_{\mu}(\theta_0))^{-1}V_{\mu}(\theta_0)(B_{\mu}(\theta_0))^{-1} = \Xi_{\mu}(\theta_0).
\end{aligned}$$

Defining the criterion over an enlarging time span as in Remarks 12 and 13, we obtain

$$\begin{aligned}
&\widehat{B}_\mu(\widehat{\theta}^\mu_{n,\overline{T}})_{n,T} \\
&= \Delta_{n,T} \sum_{i=1}^{n} \frac{\partial\mu(X_{i\Delta_{n,T}}, \widehat{\theta}^\mu_{n,\overline{T}})}{\partial\theta^\mu} \frac{\partial\mu(X_{i\Delta_{n,T}}, \widehat{\theta}^\mu_{n,\overline{T}})}{\partial\theta^{\mu'}} \\
&\underset{n(=\overline{n}),\overline{T}(=T)\to\infty}{\overset{p}{\to}} \int_{\mathfrak{D}} \frac{\partial\mu(a,\theta^\mu_0)}{\partial\theta^\mu} \frac{\partial\mu(a,\theta^\mu_0)}{\partial\theta^{\mu'}} \overline{L}_X(T,a)\,\mathrm{d}a \\
&= B_\mu(\theta_0)
\end{aligned} \tag{19}$$

and

$$\begin{aligned}
&\widehat{V}_\mu(\widehat{\theta}_{n,\overline{T}})_{n,\overline{T}} \\
&= \Delta_{n,T} \sum_{i=1}^{n} \sigma^2(X_{i\Delta_{n,T}}, \widehat{\theta}^\sigma_{n,\overline{T}}) \frac{\partial\mu(X_{i\Delta_{n,T}}, \widehat{\theta}^\mu_{n,\overline{T}})}{\partial\theta^\sigma} \frac{\partial\mu(X_{i\Delta_{n,T}}, \widehat{\theta}^\mu_{n,\overline{T}})}{\partial\theta^{\sigma'}} \\
&\underset{n(=\overline{n}),\overline{T}(=T)\to\infty}{\overset{p}{\to}} \int_{\mathfrak{D}} \sigma^2(a,\theta^\sigma_0) \left(\frac{\partial\mu(a,\theta^\mu_0)}{\partial\theta^\sigma} \frac{\partial\mu(a,\theta^\mu_0)}{\partial\theta^{\sigma'}} \right) \overline{L}_X(T,a)\,\mathrm{d}a \\
&= V_\mu(\theta_0).
\end{aligned} \tag{20}$$

This discussion further clarifies the analogy between the methods developed here and more standard nonlinear estimation problems. As conventional in correctly specified nonlinear regression models with heterogeneous errors, the asymptotic covariance matrix can be consistently estimated using sample averages involving the outer-product of the gradient of the conditional expectation calculated at the estimated parameter vector.

In sum, the methods proposed here can be viewed as nonlinear least-squares in continuous time. The main difference between the standard approach in discrete time and the approach in this paper is that preliminary kernel estimates of drift and diffusion function must be obtained. Normality of the resulting estimates can be fruitfully used for inference. As always, the asymptotic covariance matrices can be estimated by virtue of sample analogs.

7. Efficiency issues

7.1. Presence of cross-restrictions between drift and diffusion function

Standard econometric theory suggests that if the first and second moment have elements in common (namely if $\Theta^\mu \cap \Theta^\sigma \neq 0$ in our case), one should consider taking an optimally defined convex combination of the estimated common parameters for the purpose of minimizing their asymptotic variance and increase efficiency. In general, though, the drift and diffusion parameter converge at different rates (cf., Theorems 2 and 4). In this sense, our problem is nonstandard. In the limit, in fact, a linear combination of drift and diffusion parameters would have an asymptotic distribution that is dominated by the terms that converge at the slowest pace, namely the drift parameters. Thus, should the drift and diffusion have parameters in common, we recommend recovering the parameters of interest from the diffusion estimates. Not only are these estimates consistent over a relatively short time span (as indicated by Theorem 5), but they also converge at a faster speed than the corresponding drift estimates.

7.2. Weighted least-squares in continuous time

We can push the analogy between our methods and conventional least-squares procedures with heteroskedastic errors a step forward. Specifically, given the form of the asymptotic variances, one can employ generalized or weighted least-squares methods to increase efficiency.

Consider estimation of the diffusion function over a fixed time span $\overline{T}$ as in Theorem 5. Let $\Psi^{\sigma}_{\overline{n},\overline{T}}$ be a diagonal matrix of size $\overline{n}\times\overline{n}$ (or, equivalently in this case, of size $n\times n$) with diagonal elements given by $2\widehat{\sigma}^4_{(n,T)}(X_{\Delta_{\overline{n},\overline{T}}}),\ldots,2\widehat{\sigma}^4_{(n,T)}(X_{\overline{n}\Delta_{\overline{n},\overline{T}}})$. Write now the criterion

$$\widehat{\theta}^{\sigma,\mathrm{GLS}}_{n,\overline{T}} := \arg\min_{\theta^{\sigma}\in\Theta^{\sigma}\subset\Theta} \frac{\overline{T}}{\overline{n}}||(\Psi^{\sigma}_{\overline{n},\overline{T}})^{-1/2}(\widehat{\boldsymbol{\sigma}}^2-\boldsymbol{\sigma}^2(\theta^{\sigma}))||^2. \tag{21}$$

The following corollary to Theorem 5 readily derives.

Corollary to Theorem 5. *Given $n(=\overline{n})\to\infty$ and $h_{n,\overline{T}}\to 0$ (as $n\to\infty$) so that*

$$\frac{1}{h_{n,\overline{T}}}(\Delta_{n,\overline{T}}\log(1/\Delta_{n,\overline{T}}))^{1/2}=\mathrm{o}(1)$$

and $h^4_{n,\overline{T}}/\Delta_{n,\overline{T}}\to 0$, then

$$\frac{1}{\sqrt{\Delta_{n,\overline{T}}}}(\widehat{\theta}^{\sigma,\mathrm{GLS}}_{n,\overline{T}}-\theta^{\sigma}_0)\underset{n\to\infty}{\Rightarrow}\mathbf{MN}(0,\Xi^{\mathrm{GLS}}_{\sigma}(\overline{T}))$$

with

$$\Xi^{\mathrm{GLS}}_{\sigma}(\overline{T})=\left(\int_{\mathfrak{D}}\frac{1}{2\sigma_0^4(a)}\frac{\partial\sigma_0^2(a)}{\partial\theta^{\sigma}}\frac{\partial\sigma_0^2(a)}{\partial\theta^{\sigma'}}\overline{L}_X(\overline{T},a)\,\mathrm{d}a\right)^{-1},$$

where $\overline{L}_X(.,a)$ is the chronological local time of the underlying diffusion process at a.

Both $\Xi_{\sigma}(\overline{T})$ in Eq. (18) and $\Xi^{\mathrm{GLS}}_{\sigma}(\overline{T})$ can be converted from spatial integrals to integrals over time by virtue of the occupation time formula. $\Xi^{\mathrm{GLS}}_{\sigma}(\overline{T})$, for example, can be expressed as follows:

$$\left(\int_0^{\overline{T}}\frac{1}{2\sigma_0^4(X_s)}\frac{\partial\sigma_0^2(X_s)}{\partial\theta^{\sigma}}\frac{\partial\sigma_0^2(X_s)}{\partial\theta^{\sigma'}}\,\mathrm{d}s\right)^{-1}.$$

Conventional geometry in $L^2[0,\overline{T}]$, therefore, reveals[4] that the random matrix $\Xi_{\sigma}(\overline{T})-\Xi^{\mathrm{GLS}}_{\sigma}(\overline{T})$ is positive semi-definite with probability one. Hence, the weighting guarantees efficiency gains.

Generalized or weighted least-squares are expected to be beneficial even in the case where the kernel estimates are defined over an enlarging time span while the criteria are defined over a fixed span of observations as in Theorems 1–4. However, due to the path-dependency of the rates of convergence in this case, the results are, at least theoretically, less clean than in the case

[4]By the Cauchy–Schwarz inequality, in the scalar function case we have $\left(\int fh\right)^2\leqslant\int f^2\int h^2$. Setting $f=g/\sigma$ and $h=\sigma g$ this leads to $\left(\int g^2\right)^2\leqslant\int(g^2/\sigma^2)\int g^2\sigma^2$ so that $\left(\int g^2/\sigma^2\right)^{-1}\leqslant\left(\int g^2\right)^{-1}\left(\int g^2\sigma^2\right)\left(\int g^2\right)^{-1}$. In a similar way, in the vector function case when $\int hh'$ is positive definite we have $\begin{vmatrix}\int ff' & \int fh'\\ \int hf' & \int hh'\end{vmatrix}=|\int hh'|\left|\int ff'-\int fh\left(\int hh'\right)^{-1}\int hf'\right|\geqslant 0$. Setting $f=g/\sigma$ and $h=\sigma g$ as before, this leads to $\int(1/\sigma^2)gg'-\int gg'\left(\int\sigma^2gg'\right)^{-1}\int gg'\geqslant 0$ or $\left(\int(1/\sigma^2)gg'\right)^{-1}\leqslant\left(\int gg'\right)^{-1}\left(\int\sigma^2gg'\right)\left(\int gg'\right)^{-1}$, as required.

of Theorem 5. Let $\Psi^{\mu}_{\overline{n},\overline{T}}$ be a diagonal matrix of size $\overline{n} \times \overline{n}$ $(= n \times n)$ with diagonal elements given by $\widehat{\sigma}^2_{(n,T)}(X_{\Delta_{\overline{n},\overline{T}}}), \ldots, \widehat{\sigma}^2_{(n,T)}(X_{\overline{n}\Delta_{\overline{n},\overline{T}}})$. Assume the criteria in Eqs. (5) and (6)) are weighted by $\Psi^{\mu}_{\overline{n},\overline{T}}$ and $\Psi^{\sigma}_{\overline{n},\overline{T}}$, respectively, as in the case of Eq. (21). Hence, the limiting covariance matrices in Theorems 2 and 4 can be represented as follows:

$$\Xi^{\mathrm{GLS}}_{\mu} = B^{\mathrm{GLS}}(\overline{T})^{-1}_{\mu} V^{\mathrm{GLS}}_{\mu}(T) B^{\mathrm{GLS}}(\overline{T})^{-1}_{\mu}$$

with

$$B^{\mathrm{GLS}}(\overline{T})_{\mu} = \left(\int_{\mathfrak{D}} \frac{1}{\sigma_0^2(a)} \frac{\partial \mu_0(a)}{\partial \theta^{\mu}} \frac{\partial \mu_0(a)}{\partial \theta^{\mu'}} \overline{L}_X(\overline{T}, a)\, \mathrm{d}a \right),$$

$$V^{\mathrm{GLS}}(T)_{\mu} = \left(\int_{\mathfrak{D}} \frac{1}{\sigma_0^2(a)} \left(\frac{\partial \mu_0(a)}{\partial \theta^{\mu}} \frac{\partial \mu_0(a)}{\partial \theta^{\mu'}} \right) \frac{(\overline{L}_X(\overline{T}, a))^2}{\overline{L}_X(T, a)}\, \mathrm{d}a \right),$$

and

$$\Xi^{\mathrm{GLS}}_{\sigma}(T) = B^{\mathrm{GLS}}(\overline{T})^{-1}_{\sigma} V^{\mathrm{GLS}}(T)_{\sigma} B^{\mathrm{GLS}}(\overline{T})^{-1}_{\sigma}$$

with

$$B^{\mathrm{GLS}}(\overline{T})_{\sigma} = \left(\int_{\mathfrak{D}} \frac{1}{2\sigma_0^4(a)} \frac{\partial \sigma_0^2(a)}{\partial \theta^{\sigma}} \frac{\partial \sigma_0^2(a)}{\partial \theta^{\sigma'}} \overline{L}_X(\overline{T}, a)\, \mathrm{d}a \right),$$

$$V^{\mathrm{GLS}}(T)_{\sigma} = \left(\int_{\mathfrak{D}} \frac{1}{2\sigma_0^4(a)} \left(\frac{\partial \sigma_0^2(a)}{\partial \theta^{\sigma}} \frac{\partial \sigma_0^2(a)}{\partial \theta^{\sigma'}} \right) \frac{(\overline{L}_X(\overline{T}, a))^2}{\overline{L}_X(T, a)}\, \mathrm{d}a \right).$$

Since potential users will typically choose $\overline{T}$ close to T, if not equal to it (see the discussion in Section 3), then

$$\Xi^{\mathrm{GLS}}_{\mu} \underset{T \approx \overline{T}}{\approx} \left(\int_{\mathfrak{D}} \frac{1}{\sigma_0^2(a)} \frac{\partial \mu_0(a)}{\partial \theta^{\mu}} \frac{\partial \mu_0(a)}{\partial \theta^{\mu'}} \overline{L}_X(\overline{T}, a)\, \mathrm{d}a \right)^{-1} \tag{22}$$

and

$$\Xi^{\mathrm{GLS}}_{\sigma}(T) \underset{T \approx \overline{T}}{\approx} \left(\int_{\mathfrak{D}} \frac{1}{2\sigma_0^4(a)} \frac{\partial \sigma_0^2(a)}{\partial \theta^{\sigma}} \frac{\partial \sigma_0^2(a)}{\partial \theta^{\sigma'}} \overline{L}_X(\overline{T}, a)\, \mathrm{d}a \right)^{-1}. \tag{23}$$

The expressions in Eqs. (22) and (23) confirm the benefit of weighted least-squares for the case where the kernel estimates are defined over expanding spans of observations.

8. Conclusions and extensions

This paper discusses a methodology that utilizes the informational content of nonparametric methods in the parametric estimation of continuous-time models of the diffusion type while improving on their generally poor convergence properties.

The technique presented here allows us to estimate the parameters of the infinitesimal moments of potentially nonlinear stochastic differential equations in situations where the transition density of the discretely sampled process is unknown, as is typically the case in practice. Our procedure does not require simulations, or approximations to the true transition density, and has the simplicity of standard nonlinear least-squares methods in discrete time.

The method combines the appeal of limit theories that can be interpreted as spatial counterparts of the standard asymptotics for nonlinear econometric models with the generality of procedures that are robust to deviations from strong distributional assumptions, such as positive recurrence or strict stationarity. In both the stationary and nonstationary (but recurrent) cases the limiting distributions are normal with limiting variance–covariance matrices that can be readily estimated from the data. Several extensions can be considered.

(1) *Parametric estimation of multivariate diffusions and jump-diffusion processes*—Given the nature of our criteria, both extensions would require preliminary consistent estimates of the corresponding infinitesimal moments under recurrence. However, these moments can be evaluated as in recent work by Bandi and Moloche (2001) in the case of multivariate diffusions and Bandi and Nguyen (2003) in the case of jump-diffusion processes. In particular, in Bandi and Moloche (2001) it was shown that the absence of a notion of local time for multivariate semimartingales does not represent an impediment when deriving a fully nonparametric theory of inference for functionals of multidimensional diffusions. Similarly, the absence of a notion of local time is not expected to hamper parametric estimation by virtue of (weighted) least-squares methods as in this paper.

(2) *Specification tests for possibly nonstationary diffusions*—A testing procedure for alternative parametric specifications for diffusions based on our quadratic criteria can be provided. Designing specification tests for diffusions is a vibrant area of recent research. Aït-Sahalia (1996) provides a specification test for parametric drift and diffusion function based on the stationary density of the process. Corradi and White (1999) focus on the infinitesimal second moment but dispense with the assumption of stationarity. Hong and Li (2003) discuss specification tests for both the drift and the diffusion function of a stationary diffusion process relying on the informational content of the process' transition density. Empirical distribution function-based tests for stationary scalar and multivariate diffusion processes are discussed in Corradi and Swanson (2005). In order to fix ideas in our framework, consider the drift case. Assume one wishes to test the hypothesis $\mathrm{H}_0 : \mu_0(x) = \mu(x, \theta^{\mu})$ against $\mathrm{H}_1 : \mu_0(x) \neq \mu(x, \theta^{\mu})$. Provided a consistent (under the null) parametric estimate of θ^{μ}, $\widetilde{\theta}_n^{\mu}$ say, is obtained and the distribution of $Q^{\mu}_{n,\bar{n},T}(\widetilde{\theta}_n^{\mu})$ is derived under the null, intuitions and methods typically employed in discrete time can be put to work to construct a consistent test. Interestingly, while the drift parameter estimates discussed in this paper are natural candidates for $\widetilde{\theta}_n^{\mu}$, alternative estimates, eventually obtained by virtue of one of the existing consistent methods for diffusions, such as those cited in the Introduction, can be employed. In consequence, a testing method relying on $Q^{\mu}_{n,\bar{n},T}$ or $Q^{\sigma}_{n,\bar{n},T}$ might be regarded as a specification test for a chosen parametric model versus a consistent functional alternative. This procedure would be in the tradition of more conventional semiparametric tests of parametric specifications for marginal densities as in Bickel and Rosenblatt (1973), Fan (1994), Rosenblatt (1975), and, more recently, Aït-Sahalia (1996) in the context of diffusion estimation. 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 (Bandi and Nguyen, 1999), such a testing methodology is likely to be attractive. It can, for instance, be expected to have better size properties and more power than testing methods for potentially nonlinear continuous-time processes based on density-matching methods relying on stationarity (Pritsker, 1998). Research on this subject is under way and will be reported in later work.

Acknowledgments

We are grateful to an anonymous associate editor and two anonymous referees for their valuable comments. We thank Shin Kanaya for suggesting two technical clarifications. We thank Robert de Jong, Ron Gallant, Eric Renault, and seminar participants at the NBER Working Group on empirical methods in macroeconomics and finance (Summer Institute 2000), the 2000 Econometrics Society World Conference in Seattle, the University of Chicago and the Université de Montréal for useful discussions. Bandi thanks the Graduate School of Business of the University of Chicago for financial support. Phillips thanks the NSF for support under Grant nos. SBR 97-30295, SES 00-92509 and SES 04-142254.

Appendix A. Proofs

Proof of Theorem 1. In the proof of Theorems 1–4 we assume that $h_{n,T} \to 0$ and

$$\frac{\overline{L}_X(T,x)}{h_{n,T}}(\Delta_{n,T}\log(1/\Delta_{n,T}))^{1/2} = \mathrm{o}_{\text{a.s.}}(1) \quad \forall x \in \mathfrak{D}$$

given $n, T \to \infty$ with $\Delta_{n,T} = T/n \to 0$. First, we prove uniform convergence of the criterion $Q^{\mu}_{n,\bar{n},T}(\theta^{\mu})$ as in Eq. (9). Write

$$\begin{aligned}
&Q^{\mu}_{n,\bar{n},T}(\theta^{\mu}) \\
&= \frac{\overline{T}}{\bar{n}}\sum_{i=1}^{\bar{n}}(\widehat{\mu}_{(n,T)}(X_{i\Delta_{\bar{n},\overline{T}}}) - \mu(X_{i\Delta_{\bar{n},\overline{T}}},\theta^{\mu}))^2 \\
&= \frac{\overline{T}}{\bar{n}}\sum_{i=1}^{\bar{n}}\left[\frac{1}{\Delta_{n,T}}\frac{\sum_{j=1}^{n-1}\mathbf{K}\left(\frac{X_{j\Delta_{n,T}}-X_{i\Delta_{\bar{n},\overline{T}}}}{h_{n,T}}\right)[X_{(j+1)\Delta_{n,T}} - X_{j\Delta_{n,T}}]}{\sum_{j=1}^{n}\mathbf{K}\left(\frac{X_{j\Delta_{n,T}}-X_{i\Delta_{\bar{n},\overline{T}}}}{h_{n,T}}\right)} - \mu(X_{i\Delta_{\bar{n},\overline{T}}},\theta^{\mu})\right]^2 \\
&= \underbrace{\frac{\overline{T}}{\bar{n}}\sum_{i=1}^{\bar{n}}\left[\frac{1}{\Delta_{n,T}}\frac{\sum_{j=1}^{n-1}\mathbf{K}\left(\frac{X_{j\Delta_{n,T}}-X_{i\Delta_{\bar{n},\overline{T}}}}{h_{n,T}}\right)[X_{(j+1)\Delta_{n,T}} - X_{j\Delta_{n,T}}]}{\sum_{j=1}^{n}\mathbf{K}\left(\frac{X_{j\Delta_{n,T}}-X_{i\Delta_{\bar{n},\overline{T}}}}{h_{n,T}}\right)}\right]^2}_{\mathbf{a}_{n,\bar{n},T}} + \underbrace{\frac{\overline{T}}{\bar{n}}\sum_{i=1}^{\bar{n}}\mu^2(X_{i\Delta_{\bar{n},\overline{T}}},\theta^{\mu})}_{\mathbf{b}_{n,\bar{n},T}} \\
&\quad \underbrace{-2\frac{\overline{T}}{\bar{n}}\sum_{i=1}^{\bar{n}}\mu(X_{i\Delta_{\bar{n},\overline{T}}},\theta^{\mu})\left(\frac{1}{\Delta_{n,T}}\frac{\sum_{j=1}^{n-1}\mathbf{K}\left(\frac{X_{j\Delta_{n,T}}-X_{i\Delta_{\bar{n},\overline{T}}}}{h_{n,T}}\right)[X_{(j+1)\Delta_{n,T}} - X_{j\Delta_{n,T}}]}{\sum_{j=1}^{n}\mathbf{K}\left(\frac{X_{j\Delta_{n,T}}-X_{i\Delta_{\bar{n},\overline{T}}}}{h_{n,T}}\right)}\right)}_{\mathbf{c}_{n,\bar{n},T}}.
\end{aligned}$$

Using the modulus of continuity of a diffusion as in Florens-Zmirou (1993, p. 797), as well as the occupation time formula for continuous semimartingales (Revuz and Yor, 1994, Corollary 1.6, p. 15), we can readily show that

$$\begin{aligned}\mathbf{b}_{n,\bar{n},T} &= \frac{\overline{T}}{\bar{n}}\sum_{i=1}^{\bar{n}}\mu^2(X_{i\Delta_{\bar{n},\overline{T}}},\theta^\mu)\\ &= \int_0^{\overline{T}}\mu^2(X_s,\theta^\mu)\,\mathrm{d}s + \mathrm{o}_{\text{a.s.}}(1)\\ &= \int_{\mathfrak{D}}\mu^2(a,\theta^\mu)\overline{L}_X(\overline{T},a)\,\mathrm{d}a + \mathrm{o}_{\text{a.s.}}(1)\end{aligned}$$

(see, e.g., BP, 2003). Furthermore, we can write

$$\begin{aligned}\mathbf{c}_{n,\bar{n},T} = &\underbrace{-2\frac{\overline{T}}{\bar{n}}\sum_{i=1}^{\bar{n}}\mu(X_{i\Delta_{\bar{n},\overline{T}}},\theta^\mu)\left(\frac{1}{\Delta_{n,T}}\frac{\sum_{j=1}^{n-1}\mathbf{K}\left(\frac{X_{j\Delta_{n,T}}-X_{i\Delta_{\bar{n},\overline{T}}}}{h_{n,T}}\right)\int_{j\Delta_{n,T}}^{(j+1)\Delta_{n,T}}\mu_0(X_s)\,\mathrm{d}s}{\sum_{j=1}^{n}\mathbf{K}\left(\frac{X_{j\Delta_{n,T}}-X_{i\Delta_{\bar{n},\overline{T}}}}{h_{n,T}}\right)}\right)}_{\mathbf{c1}_{n,\bar{n},T}}\\ &\underbrace{-2\frac{\overline{T}}{\bar{n}}\sum_{i=1}^{\bar{n}}\mu(X_{i\Delta_{\bar{n},\overline{T}}},\theta^\mu)\left(\frac{1}{\Delta_{n,T}}\frac{\sum_{j=1}^{n-1}\mathbf{K}\left(\frac{X_{j\Delta_{n,T}}-X_{i\Delta_{\bar{n},\overline{T}}}}{h_{n,T}}\right)\int_{j\Delta_{n,T}}^{(j+1)\Delta_{n,T}}\sigma_0(X_s)\,\mathrm{d}B_s}{\sum_{j=1}^{n}\mathbf{K}\left(\frac{X_{j\Delta_{n,T}}-X_{i\Delta_{\bar{n},\overline{T}}}}{h_{n,T}}\right)}\right)}_{\mathbf{c2}_{n,\bar{n},T}}.\end{aligned}$$

It follows that

$$\begin{aligned}\mathbf{c1}_{n,\bar{n},T} &= -2\frac{\overline{T}}{\bar{n}}\sum_{i=1}^{\bar{n}}\mu(X_{i\Delta_{\bar{n},\overline{T}}},\theta^\mu)\left(\frac{\sum_{j=1}^{n-1}\mathbf{K}\left(\frac{X_{j\Delta_{n,T}}-X_{i\Delta_{\bar{n},\overline{T}}}}{h_{n,T}}\right)\mu_0(X_{j\Delta_{n,T}}+\mathrm{o}_{\text{a.s.}}(1))}{\sum_{j=1}^{n}\mathbf{K}\left(\frac{X_{j\Delta_{n,T}}-X_{i\Delta_{\bar{n},\overline{T}}}}{h_{n,T}}\right)}\right)\\ &= -2\int_0^{\overline{T}}\mu(X_s,\theta^\mu)\left(\frac{\frac{1}{h_{n,T}}\int_0^T\mathbf{K}\left(\frac{X_u-X_s}{h_{n,T}}\right)\mu(X_u+\mathrm{o}_{\text{a.s.}}(1))\,\mathrm{d}u}{\frac{1}{h_{n,T}}\int_0^T\mathbf{K}\left(\frac{X_a-X_s}{h_{n,T}}\right)\mathrm{d}a}\right)\mathrm{d}s+\mathrm{o}_{\text{a.s.}}(1)\\ &= -2\int_{-\infty}^{\infty}\mu(s,\theta^\mu)\left(\frac{\frac{1}{h_{n,T}}\int_{-\infty}^{\infty}\mathbf{K}\left(\frac{u-s}{h_{n,T}}\right)\mu_0(u+\mathrm{o}_{\text{a.s.}}(1))\overline{L}_X(T,u)\,\mathrm{d}u}{\frac{1}{h_{n,T}}\int_{-\infty}^{\infty}\mathbf{K}\left(\frac{a-s}{h_{n,T}}\right)\overline{L}_X(T,a)\,\mathrm{d}a}\right)\overline{L}_X(\overline{T},s)\,\mathrm{d}s+\mathrm{o}_{\text{a.s.}}(1)\\ &= -2\int_{-\infty}^{\infty}\mu(s,\theta^\mu)\left(\frac{\int_{-\infty}^{\infty}\mathbf{K}(c)\mu_0(s+h_{n,T}c)\overline{L}_X(T,s+h_{n,T}c)\,\mathrm{d}c}{\int_{-\infty}^{\infty}\mathbf{K}(e)\overline{L}_X(T,s+h_{n,T}e)\,\mathrm{d}e}\right)\overline{L}_X(\overline{T},s)\,\mathrm{d}s+\mathrm{o}_{\text{a.s.}}(1)\\ &\underset{n,\bar{n},T\to\infty}{\overset{\text{a.s.}}{\to}} -2\int_{\mathfrak{D}}\mu(a,\theta^\mu)\mu_0(a)\overline{L}_X(\overline{T},a)\,\mathrm{d}a,\end{aligned}$$

given Assumption 2, namely $\int_{-\infty}^{\infty}\mathbf{K}(s)\,\mathrm{d}s = 1$. As in the proof of Theorem 2, one can also show that

$$\mathbf{c2}_{n,\bar{n},T}\underset{n,\bar{n},T\to\infty}{\overset{\text{p}}{\to}}0,$$

and

$$\mathbf{\Phi}^{-1/2}(T)\mathbf{c2}_{n,\bar{n},T} = \mathrm{O}_{\text{p}}(1),$$

with

$$\mathbf{\Phi}(T) = \int_{\mathfrak{D}} \left(4\sigma^2(a, \theta_0^{\mu})\mu^2(a, \theta^{\mu}) \frac{(\overline{L}_X(\overline{T}, a))^2}{\overline{L}_X(T, a)} \right) \mathrm{d}a.$$

We now examine the quadratic term $\mathbf{a}_{n,\overline{n},T}$. Write

$$\frac{\overline{T}}{\overline{n}} \sum_{i=1}^{\overline{n}} \left[\frac{1}{\Delta_{n,T}} \frac{\sum_{j=1}^{n-1} \mathbf{K}\left(\frac{X_{j\Delta_{n,T}} - X_{i\Delta_{\overline{n},\overline{T}}}}{h_{n,T}}\right)[X_{(j+1)\Delta_{n,T}} - X_{j\Delta_{n,T}}]}{\sum_{j=1}^{n} \mathbf{K}\left(\frac{X_{j\Delta_{n,T}} - X_{i\Delta_{\overline{n},\overline{T}}}}{h_{n,T}}\right)} \right]^2$$

$$= \underbrace{\frac{\overline{T}}{\overline{n}} \sum_{i=1}^{\overline{n}} \left[\frac{1}{\Delta_{n,T}} \frac{\sum_{j=1}^{n-1} \mathbf{K}\left(\frac{X_{j\Delta_{n,T}} - X_{i\Delta_{\overline{n},\overline{T}}}}{h_{n,T}}\right) \int_{j\Delta_{n,T}}^{(j+1)\Delta_{n,T}} \mu_0(X_s)\,\mathrm{d}s}{\sum_{j=1}^{n} \mathbf{K}\left(\frac{X_{j\Delta_{n,T}} - X_{i\Delta_{\overline{n},\overline{T}}}}{h_{n,T}}\right)} \right]^2}_{\mathbf{a1}_{n,\overline{n},T}}$$

$$+ \underbrace{\frac{\overline{T}}{\overline{n}} \sum_{i=1}^{\overline{n}} \left[\frac{1}{\Delta_{n,T}} \frac{\sum_{j=1}^{n-1} \mathbf{K}\left(\frac{X_{j\Delta_{n,T}} - X_{i\Delta_{\overline{n},\overline{T}}}}{h_{n,T}}\right) \int_{j\Delta_{n,T}}^{(j+1)\Delta_{n,T}} \sigma_0(X_s)\,\mathrm{d}B_s}{\sum_{j=1}^{n} \mathbf{K}\left(\frac{X_{j\Delta_{n,T}} - X_{i\Delta_{\overline{n},\overline{T}}}}{h_{n,T}}\right)} \right]^2}_{\mathbf{a2}_{n,\overline{n},T}}$$

$$+ \underbrace{\sum_{i=1}^{\overline{n}} \left(\frac{\frac{2\overline{T}}{\overline{n}} \left(\frac{1}{\Delta_{n,T}}\right)^2 \sum_{k,j=1}^{n-1} \mathbf{K}\left(\frac{X_{j\Delta_{n,T}} - X_{i\Delta_{\overline{n},\overline{T}}}}{h_{n,T}}\right) \mathbf{K}\left(\frac{X_{k\Delta_{n,T}} - X_{i\Delta_{\overline{n},\overline{T}}}}{h_{n,T}}\right) \int_{k\Delta_{n,T}}^{(k+1)\Delta_{n,T}} \mu_0(X_s)\,\mathrm{d}s \int_{j\Delta_{n,T}}^{(j+1)\Delta_{n,T}} \sigma_0(X_s)\,\mathrm{d}B_s}{\sum_{j=1}^{n} \mathbf{K}\left(\frac{X_{j\Delta_{n,T}} - X_{i\Delta_{\overline{n},\overline{T}}}}{h_{n,T}}\right) \sum_{k=1}^{n} \mathbf{K}\left(\frac{X_{k\Delta_{n,T}} - X_{i\Delta_{\overline{n},\overline{T}}}}{h_{n,T}}\right)} \right)}_{\mathbf{a3}_{n,\overline{n},T}}.$$

We start with the first term, namely $\mathbf{a1}_{n,\overline{n},T}$. Following the same steps leading to the asymptotic expression of term $\mathbf{c1}_{n,\overline{n},T}$ above we deduce that

$$\mathbf{a1}_{n,\overline{n},T} \underset{n,\overline{n},T \to \infty}{\overset{\text{a.s.}}{\to}} \int_{\mathfrak{D}} \mu^2(a, \theta_0^{\mu}) \overline{L}_X(\overline{T}, a)\,\mathrm{d}a.$$

We now examine term $\mathbf{a2}_{n,\overline{n},T}$. Write

$$\mathbf{a2}_{n,\overline{n},T} = \frac{\overline{T}}{\overline{n}} \sum_{i=1}^{\overline{n}} \frac{1}{\left(\frac{\Delta_{n,T}}{h_{n,T}} \sum_{j=1}^{n} \mathbf{K}\left(\frac{X_{j\Delta_{n,T}} - X_{i\Delta_{\overline{n},\overline{T}}}}{h_{n,T}}\right) \right)^2} \left[\frac{1}{h_{n,T}} \sum_{j=1}^{n-1} \mathbf{K}\left(\frac{X_{j\Delta_{n,T}} - X_{i\Delta_{\overline{n},\overline{T}}}}{h_{n,T}}\right) \int_{j\Delta_{n,T}}^{(j+1)\Delta_{n,T}} \sigma_0(X_s)\,\mathrm{d}B_s \right]^2$$

$$= \frac{\overline{T}}{\overline{n}} \sum_{i=1}^{\overline{n}} \frac{1}{\left(\frac{\Delta_{n,T}}{h_{n,T}} \sum_{j=1}^{n} \mathbf{K}\left(\frac{X_{j\Delta_{n,T}} - X_{i\Delta_{\overline{n},\overline{T}}}}{h_{n,T}}\right) \right)^2} (M_{n,T}^i(1))^2,$$

where

$$M_{n,T}^i(r) = \frac{1}{h_{n,T}} \sum_{j=1}^{J-1} \mathbf{K}\left(\frac{X_{j\Delta_{n,T}} - X_{i\Delta_{\overline{n},\overline{T}}}}{h_{n,T}}\right) \mathbf{1}_{\{rn\Delta_{n,T} \geqslant i\Delta_{\overline{n}}\overline{T}\}} \int_{j\Delta_{n,T}}^{(j+1)\Delta_{n,T}} \sigma_0(X_s)\,\mathrm{d}B_s$$

$$+\frac{1}{h_{n,T}}\mathbf{K}\left(\frac{X_{J\Delta_{n,T}}-X_{i\Delta_{\bar{n},\bar{T}}}}{h_{n,T}}\right)\mathbf{1}_{\{rn\Delta_{n,T}\geqslant i\Delta_{\bar{n}}\bar{T}\}}\int_{J\Delta_{n,T}}^{rn\Delta_{n,T}}\sigma_0(X_s)\,\mathrm{d}B_s$$

for $J=[rn]$ with $[x]$ denoting, as conventional, the largest integral that is less than or equal to x. $M^i_{n,T}(r)$ is a martingale. Since we focus on the case $r=1$, here (and in similar arguments below) we drop the indicator function for notational convenience. Now notice that

$$\mathrm{d}M^i_{n,T}(r)=\frac{1}{h_{n,T}}\mathbf{K}\left(\frac{X_{J\Delta_{n,T}}-X_{i\Delta_{\bar{n},\bar{T}}}}{h_{n,T}}\right)\sigma_0(X_{rn\Delta_{n,T}})\,\mathrm{d}B_{rn\Delta_{n,T}},$$

and

$$\begin{aligned}\mathrm{d}[M^i_{n,T},M^k_{n,T}]_r &= \mathrm{d}M^i_{n,T}(r)\,\mathrm{d}M^k_{n,T}(r)\\ &= \left[\frac{n\Delta_{n,T}}{h^2_{n,T}}\mathbf{K}\left(\frac{X_{J\Delta_{n,T}}-X_{i\Delta_{\bar{n},\bar{T}}}}{h_{n,T}}\right)\mathbf{K}\left(\frac{X_{J\Delta_{n,T}}-X_{k\Delta_{\bar{n},\bar{T}}}}{h_{n,T}}\right)\sigma_0^2(X_{rn\Delta_{n,T}})\right]\mathrm{d}r.\end{aligned}$$

Then,

$$\begin{aligned}[M^i_{n,T},M^k_{n,T}]_r &= \int_0^r \mathrm{d}[M^i_{n,T},M^k_{n,T}]_s\\ &= \int_0^r\left[\frac{n\Delta_{n,T}}{h^2_{n,T}}\mathbf{K}\left(\frac{X_{[sn]\Delta_{n,T}}-X_{i\Delta_{\bar{n},\bar{T}}}}{h_{n,T}}\right)\mathbf{K}\left(\frac{X_{[sn]\Delta_{n,T}}-X_{k\Delta_{\bar{n},\bar{T}}}}{h_{n,T}}\right)\sigma_0^2(X_{sn\Delta_{n,T}})\right]\mathrm{d}s\\ &= \frac{1}{h^2_{n,T}}\int_0^{rT}\left[\mathbf{K}\left(\frac{X_u-X_{i\Delta_{\bar{n},\bar{T}}}}{h_{n,T}}\right)\mathbf{K}\left(\frac{X_u-X_{k\Delta_{\bar{n},\bar{T}}}}{h_{n,T}}\right)\sigma_0^2(X_u)\right]\mathrm{d}u+\mathrm{o_p}(1).\end{aligned}\tag{24}$$

Furthermore,

$$\begin{aligned}M^i_{n,T}(r) &= \int_0^r \mathrm{d}M^i_n(s)\\ &= \int_0^r\frac{1}{h_{n,T}}\mathbf{K}\left(\frac{X_{[sn]\Delta_{n,T}}-X_{i\Delta_{\bar{n},\bar{T}}}}{h_{n,T}}\right)\sigma_0(X_{sn\Delta_{n,T}})\,\mathrm{d}B_{sn\Delta_{n,T}},\end{aligned}$$

and

$$M^i_{n,T}(r)M^k_{n,T}(r)=\int_0^r M^i_{n,T}(s)\,dM^i_{n,T}(s)+\int_0^r M^k_{n,T}(s)\,dM^i_{n,T}(s)+[M^i_{n,T},M^k_{n,T}]_r\quad\forall r\in[0,1],\tag{25}$$

(see, e.g., Chung and Williams, 1990). Hence,

$$\begin{aligned}\mathbf{a2}_{n,\bar{n},T/r} &= \frac{\bar{T}}{\bar{n}}\sum_{i=1}^{\bar{n}}\frac{1}{\left(\frac{\Delta_{n,T}}{h_{n,T}}\sum_{j=1}^n\mathbf{K}\left(\frac{X_{j\Delta_{n,T}}-X_{i\Delta_{\bar{n},\bar{T}}}}{h_{n,T}}\right)\right)^2}(M^i_{n,T}(r))^2\\ &= \underbrace{\frac{\bar{T}}{\bar{n}}\sum_{i=1}^{\bar{n}}\frac{1}{\left(\frac{\Delta_{n,T}}{h_{n,T}}\sum_{j=1}^n\mathbf{K}\left(\frac{X_{j\Delta_{n,T}}-X_{i\Delta_{\bar{n},\bar{T}}}}{h_{n,T}}\right)\right)^2}[M^i_{n,T}]_r}_{\overline{\mathbf{a2}}_{n,\bar{n},T/r}}\end{aligned}$$

$$\underbrace{+2\frac{\overline{T}}{\overline{n}}\sum_{i=1}^{\overline{n}}\frac{1}{\left(\frac{\Delta_{n,T}}{h_{n,T}}\sum_{j=1}^{n}\mathbf{K}\left(\frac{X_{j\Delta_{n,T}}-X_{i\Delta_{\overline{n},\overline{T}}}}{h_{n,T}}\right)\right)^2}\left(\int_0^r M_{n,T}^i(s)\,\mathrm{d}M_{n,T}^i(s)\right)}_{\overline{\overline{\mathbf{a2}}}_{n,\overline{n},T/r}}. \tag{26}$$

Using Eq. (24) and proceeding as for term $\mathbf{b}_{n,\overline{n},T}$ and term $\mathbf{c1}_{n,\overline{n},T}$ above, the quantity $\overline{\mathbf{a2}}_{n,\overline{n},T/r}$ can be readily evaluated at $r=1$ and represented as follows:

$$\begin{aligned}
&\overline{(\mathbf{a2})}_{n,\overline{n},T/1}\\
&=\frac{\overline{T}}{\overline{n}}\sum_{i=1}^{\overline{n}}\frac{\frac{1}{h_{n,T}^2}\int_0^T\mathbf{K}^2\left(\frac{X_u-X_{i\Delta_{\overline{n},\overline{T}}}}{h_{n,T}}\right)\sigma_0^2(X_u)\,\mathrm{d}u}{\left(\frac{\Delta_{n,T}}{h_{n,T}}\sum_{j=1}^{n}\mathbf{K}\left(\frac{X_{j\Delta_{n,T}}-X_{i\Delta_{\overline{n},\overline{T}}}}{h_{n,T}}\right)\right)^2}+\mathrm{o_p}(1)\\
&=\frac{1}{h_{n,T}}\int_0^{\overline{T}}\frac{\frac{1}{h_{n,T}}\int_0^T\mathbf{K}^2\left(\frac{X_u-X_s}{h_{n,T}}\right)\sigma_0^2(X_u)\,\mathrm{d}u}{\left(\frac{1}{h_{n,T}}\int_0^T\mathbf{K}\left(\frac{X_a-X_s}{h_{n,T}}\right)\mathrm{d}a\right)^2}\,\mathrm{d}s+\mathrm{o_p}(1)\\
&=\frac{1}{h_{n,T}}\int_{-\infty}^{\infty}\frac{\frac{1}{h_{n,T}}\int_{-\infty}^{\infty}\mathbf{K}^2\left(\frac{u-s}{h_{n,T}}\right)\sigma_0^2(u)\overline{L}_X(T,u)\,\mathrm{d}u}{\left(\frac{1}{h_{n,T}}\int_{-\infty}^{\infty}\mathbf{K}\left(\frac{a-s}{h_{n,T}}\right)\overline{L}_X(T,a)\,\mathrm{d}a\right)^2}\overline{L}_X(\overline{T},s)\,\mathrm{d}s+\mathrm{o_p}(1)\\
&=\frac{1}{h_{n,T}}\mathbf{K}_2\left(\int_{\mathfrak{D}}\sigma_0^2(a)\frac{\overline{L}_X(\overline{T},a)}{\overline{L}_X(T,a)}\,\mathrm{d}a\right)+\mathrm{o_p}(1),
\end{aligned}$$

where $\mathbf{K}_2=\int_{-\infty}^{\infty}\mathbf{K}^2(s)\,\mathrm{d}s<\infty$, by Assumption 2. In consequence,

$$\begin{aligned}
&\frac{\overline{T}}{\overline{n}}\sum_{i=1}^{\overline{n}}\frac{1}{\left(\sum_{j=1}^{n}\mathbf{K}\left(\frac{X_{j\Delta_{n,T}}-X_{i\Delta_{\overline{n},\overline{T}}}}{h_{n,T}}\right)\right)^2}[M_{n,T}^i]_1\\
&=\frac{1}{h_{n,T}}\mathbf{K}_2\left(\int_{\mathfrak{D}}\sigma_0^2(a)\frac{\overline{L}_X(\overline{T},a)}{\overline{L}_X(T,a)}\,\mathrm{d}a\right)+\mathrm{o_p}(1)\underset{n,\overline{n},T\to\infty}{\overset{\mathrm{p}}{\to}}0
\end{aligned}$$

if $h_{n,T}\overline{L}_X(T,x)\overset{\mathrm{a.s.}}{\to}\infty\ \forall x\in\mathfrak{D}$, as stated in our assumptions. We now analyze the second component of term $\mathbf{a2}_{n,\overline{n},T/r}$, namely $\overline{\overline{\mathbf{a2}}}_{n,\overline{n},T/r}$. Given $\{X_{i\Delta_{\overline{n},\overline{T}}}:i=1,\ldots,\overline{n}\}$, $\overline{\overline{\mathbf{a2}}}_{n,\overline{n},T/r}$ constitutes a weighted sum of continuous martingales evaluated at $r\in[0,1]$. By virtue of Eq. (25) and noting that $\int_0^r[M_{n,T}^i,M_{n,T}^k]_s\,\mathrm{d}[M_{n,T}^i,M_{n,T}^k]_s=[M_{n,T}^i,M_{n,T}^k]_r^2/2$, the variation process of $\overline{\overline{\mathbf{a2}}}_{n,\overline{n},T/r}$at $r=1$ can be expressed as follows. Write

$$\begin{aligned}
&[\overline{\overline{\mathbf{a2}}}_{n,\overline{n},T}]_{r=1}\\
&=4\left(\frac{\overline{T}}{\overline{n}}\right)^2\sum_{i=1}^{\overline{n}}\sum_{k=1}^{\overline{n}}\frac{\int_0^1 M_{n,T}^i(s)M_{n,T}^k(s)\,\mathrm{d}[M_{n,T}^i,M_{n,T}^k]_s}{\left(\frac{\Delta_{n,T}}{h_{n,T}}\sum_{j=1}^{n}\mathbf{K}\left(\frac{X_{j\Delta_{n,T}}-X_{i\Delta_{\overline{n},\overline{T}}}}{h_{n,T}}\right)\right)^2\left(\frac{\Delta_{n,T}}{h_{n,T}}\sum_{j=1}^{n}\mathbf{K}\left(\frac{X_{j\Delta_{n,T}}-X_{k\Delta_{\overline{n},\overline{T}}}}{h_{n,T}}\right)\right)^2}\\
&=2\left(\frac{\overline{T}}{\overline{n}}\right)^2\sum_{i=1}^{\overline{n}}\sum_{k=1}^{\overline{n}}\frac{[M_{n,T}^i,M_{n,T}^k]_1^2}{\left(\frac{\Delta_{n,T}}{h_{n,T}}\sum_{j=1}^{n}\mathbf{K}\left(\frac{X_{j\Delta_{n,T}}-X_{i\Delta_{\overline{n},\overline{T}}}}{h_{n,T}}\right)\right)^2\left(\frac{\Delta_{n,T}}{h_{n,T}}\sum_{j=1}^{n}\mathbf{K}\left(\frac{X_{j\Delta_{n,T}}-X_{k\Delta_{\overline{n},\overline{T}}}}{h_{n,T}}\right)\right)^2}+\mathrm{o_p}(1)
\end{aligned}$$

$$
\begin{aligned}
&= 2\left(\frac{\overline{T}}{\overline{n}}\right)^2 \sum_{i=1}^{\overline{n}}\sum_{k=1}^{\overline{n}} \frac{\left(\frac{1}{h_{n,T}^2}\int_0^T \mathbf{K}\left(\frac{X_u - X_{i\Delta_{\overline{n},\overline{T}}}}{h_{n,T}}\right)\mathbf{K}\left(\frac{X_u - X_{k\Delta_{\overline{n},\overline{T}}}}{h_{n,T}}\right)\sigma_0^2(X_u)\,\mathrm{d}u\right)^2}{\left(\frac{\Delta_{n,T}}{h_{n,T}}\sum_{j=1}^n \mathbf{K}\left(\frac{X_{j\Delta_{n,T}} - X_{i\Delta_{\overline{n},\overline{T}}}}{h_{n,T}}\right)\right)^2\left(\frac{\Delta_{n,T}}{h_{n,T}}\sum_{j=1}^n \mathbf{K}\left(\frac{X_{j\Delta_{n,T}} - X_{k\Delta_{\overline{n},\overline{T}}}}{h_{n,T}}\right)\right)^2} + \mathrm{o_p}(1)\\
&= 2\int_0^{\overline{T}}\int_0^{\overline{T}} \frac{\left(\int_0^T \frac{1}{h_{n,T}^2}\mathbf{K}\left(\frac{X_u - X_a}{h_{n,T}}\right)\mathbf{K}\left(\frac{X_u - X_b}{h_{n,T}}\right)\sigma_0^2(X_u)\,\mathrm{d}u\right)^2}{\left(\frac{1}{h_{n,T}}\int_0^T \mathbf{K}\left(\frac{X_f - X_a}{h_{n,T}}\right)\mathrm{d}f\right)^2\left(\frac{1}{h_{n,T}}\int_0^T \mathbf{K}\left(\frac{X_e - X_b}{h_{n,T}}\right)\mathrm{d}e\right)^2}\,\mathrm{d}a\,\mathrm{d}b + \mathrm{o_p}(1)\\
&= 2\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{\left(\int_{-\infty}^{\infty}\frac{1}{h_{n,T}^2}\mathbf{K}\left(\frac{u-a}{h_{n,T}}\right)\mathbf{K}\left(\frac{u-b}{h_{n,T}}\right)\sigma_0^2(u)\overline{L}_X(T,u)\,\mathrm{d}u\right)^2}{\left(\frac{1}{h_{n,T}}\int_{-\infty}^{\infty}\mathbf{K}\left(\frac{f-a}{h_{n,T}}\right)\overline{L}_X(T,f)\,\mathrm{d}f\right)^2\left(\frac{1}{h_{n,T}}\int_{-\infty}^{\infty}\mathbf{K}\left(\frac{e-b}{h_{n,T}}\right)\overline{L}_X(T,e)\,\mathrm{d}e\right)^2}\overline{L}_X(\overline{T},a)\overline{L}_X(\overline{T},b)\,\mathrm{d}a\,\mathrm{d}b + \mathrm{o_p}(1)\\
&= 2\frac{1}{h_{n,T}}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{\frac{1}{h_{n,T}}\left(\int_{-\infty}^{\infty}\frac{1}{h_{n,T}}\mathbf{K}\left(\frac{u-a}{h_{n,T}}\right)\mathbf{K}\left(\frac{u-b}{h_{n,T}}\right)\sigma_0^2(u)\overline{L}_X(T,u)\,\mathrm{d}u\right)^2\overline{L}_X(\overline{T},a)\overline{L}_X(\overline{T},b)\,\mathrm{d}a\,\mathrm{d}b}{\left(\frac{1}{h_{n,T}}\int_{-\infty}^{\infty}\mathbf{K}\left(\frac{f-a}{h_{n,T}}\right)\overline{L}_X(T,f)\,\mathrm{d}f\right)^2\left(\frac{1}{h_{n,T}}\int_{-\infty}^{\infty}\mathbf{K}\left(\frac{e-b}{h_{n,T}}\right)\overline{L}_X(T,e)\,\mathrm{d}e\right)^2} + \mathrm{o_p}(1)\\
&= 2\frac{1}{h_{n,T}}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{\frac{1}{h_{n,T}}\left(\int_{-\infty}^{\infty}\mathbf{K}(c)\mathbf{K}\left(\frac{a-b}{h_{n,T}}+c\right)\sigma_0^2(a+ch_{n,T})\overline{L}_X(T,a+ch_{n,T})\,\mathrm{d}c\right)^2\overline{L}_X(\overline{T},a)\overline{L}_X(\overline{T},b)\,\mathrm{d}a\,\mathrm{d}b}{(\overline{L}_X(T,a))^2(\overline{L}_X(T,b))^2} + \mathrm{o_p}(1)\\
&= 2\frac{1}{h_{n,T}}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{\left(\int_{-\infty}^{\infty}\mathbf{K}(c)\mathbf{K}(k+c)\sigma_0^2(a)\overline{L}_X(T,a)\,\mathrm{d}c\right)^2\overline{L}_X(\overline{T},a)\overline{L}_X(\overline{T},a-h_{n,T}k)\,\mathrm{d}a\,\mathrm{d}k}{(\overline{L}_X(T,a))^2(\overline{L}_X(T,a-h_{n,T}k))^2} + \mathrm{o_p}(1)\\
&= 2\frac{1}{h_{n,T}}\left(\int_{\mathfrak{D}}\frac{\sigma_0^4(a)\overline{L}_X^2(\overline{T},a)}{(\overline{L}_X(T,a))^2}\,\mathrm{d}a\right)\left(\int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty}\mathbf{K}(c)\mathbf{K}(k+c)\,\mathrm{d}c\right)^2\mathrm{d}k\right) + \mathrm{o_p}(1).
\end{aligned}
$$

Now fix $T = \overline{T}$. By virtue of conventional arguments (see, e.g., Revuz and Yor, 1994, Theorem 1.9, p. 175, and Theorem 2.3, p. 496) each martingale $\int_0^r M^i_{n,\overline{T}}(s)\,\mathrm{d}M^i_{n,\overline{T}}(s)$ can be embedded in a Brownian motion with quadratic variation process given by $\left[\int_0 M^i_{n,\overline{T}}(s)\,\mathrm{d}M^i_{n,\overline{T}}(s)\right]_r$. Let T increase. Thus,

$$
\left[\int_0 M^i_{n,T}(s)\,\mathrm{d}M^i_{n,T}(s)\right]_r^{-1/2}\int_0^r M^i_{n,T}(s)\,\mathrm{d}M^i_{n,T}(s) \underset{n,\overline{n},T\to\infty}{\Rightarrow} \mathbf{N}(0,1) \quad \forall i = 1,\ldots,\overline{n}.
$$

Similarly, when standardized by its vanishing variation process at $r = 1$, the linear combination $\overline{\overline{\mathbf{a2}}}_{n,\overline{n},T/r=1}$ is normally distributed in the limit. In fact,

$$
[\overline{\overline{\mathbf{a2}}}_{n,\overline{n},T}]_{r=1}^{-1/2}\overline{\overline{\mathbf{a2}}}_{n,\overline{n},T/r=1} = \mathrm{O_p}(1),
$$

where

$$
\begin{aligned}
[\overline{\overline{\mathbf{a2}}}_{n,\overline{n},T}]_{r=1} &= \frac{2}{h_{n,T}}\left(\int_{\mathfrak{D}}\frac{\sigma_0^4(a)\overline{L}_X^2(\overline{T},a)}{(\overline{L}_X(T,a))^2}\,\mathrm{d}a\right)\left(\int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty}\mathbf{K}(c)\mathbf{K}(k+c)\,\mathrm{d}c\right)^2\mathrm{d}k\right) + \mathrm{o_p}(1)\\
&\underset{n,\overline{n},T\to\infty}{\overset{\mathrm{p}}{\to}} 0.
\end{aligned}
$$

This, in turn, implies that

$$2\frac{\overline{T}}{\overline{n}}\sum_{i=1}^{\overline{n}}\frac{1}{\left(\frac{\Delta_{n,T}}{h_{n,T}}\sum_{j=1}^{n}\mathbf{K}\left(\frac{X_{j\Delta_{n,T}}-X_{i\Delta_{\overline{n},T}}}{h_{n,T}}\right)\right)^2}\left(\int_0^1 M_{n,T}^i(s)\,\mathrm{d}M_{n,T}^i(s)\right)\underset{n,\overline{n},T\to\infty}{\overset{\mathrm{p}}{\to}}0.$$

Similar steps allow us to show that the term $\mathbf{a3}_{n,\overline{n},T}\underset{n,\overline{n},T\to\infty}{\overset{\mathrm{p}}{\to}}0$. This proves pointwise weak convergence of $Q^{\mu}_{n,\overline{n},T}(\theta^{\mu})$ to $Q^{\mu}(\theta^{\mu},\theta_0)$. We now prove uniform convergence. Define $Z_{n,\overline{n},T}(\theta,\theta_0)=Q^{\mu}_{n,\overline{n},T}(\theta)-Q^{\mu}(\theta,\theta_0)$. Using our regularity conditions from Assumptions 1 and 2, it readily follows that $\forall\varepsilon>0\ \exists\gamma>0$ such that

$$\overline{\lim_{n,\overline{n},T\to\infty}}P\left(\sup_{\theta\in\Theta^{\mu}}\sup_{\theta^*\in B(\theta,\gamma)}(|Z_{n,\overline{n},T}(\theta^*,\theta_0)-Z_{n,\overline{n},T}(\theta,\theta_0)|>\varepsilon)\right)<\varepsilon, \tag{27}$$

where $B(\theta,\gamma)$ is an open ball of radius γ centered at θ. The expression in Eq. (27) is a stochastic equicontinuity condition. By virtue of the boundedness of Θ, let $\{B(\theta_{\bar{j}},\gamma):\bar{j}=1,\ldots,\overline{J}\}$ be a finite cover of $\Theta^{\mu}\subset\Theta$ so that $\bigcup_{\bar{j}=1}^{\overline{J}}B(\theta_{\bar{j}},\gamma)\supseteq\Theta^{\mu}$. We wish to show that $\exists\widetilde{T},\widetilde{n}$, and $\widetilde{\overline{n}}$ so that for $T>\widetilde{T}$, $n>\widetilde{n}$, and $\overline{n}>\widetilde{\overline{n}}$ there exists an arbitrarily small $\varepsilon>0$ such that

$$P\left(\sup_{\theta\in\Theta^{\mu}}|Q^{\mu}_{n,\overline{n},T}(\theta)-Q^{\mu}(\theta,\theta_0)|>2\varepsilon\right)<\varepsilon. \tag{28}$$

Write

$$\begin{aligned}
&\overline{\lim_{n,\overline{n},T\to\infty}}P\left(\sup_{\theta\in\Theta^{\mu}}|Z_{n,\overline{n},T}(\theta,\theta_0)|>2\varepsilon\right)\\
&\leqslant\overline{\lim_{n,\overline{n},T\to\infty}}P\left(\max_{1\leqslant\bar{j}\leqslant\overline{J}}\sup_{\theta\in B(\theta_{\bar{j}},\gamma)}|Z_{n,\overline{n},T}(\theta,\theta_0)|>2\varepsilon\right)\\
&\leqslant\overline{\lim_{n,\overline{n},T\to\infty}}P\left(\max_{1\leqslant\bar{j}\leqslant\overline{J}}\sup_{\theta^*\in B(\theta_{\bar{j}},\gamma)}|Z_{n,\overline{n},T}(\theta^*,\theta_0)-Z_{n,\overline{n},T}(\theta_{\bar{j}},\theta_0)+Z_{n,\overline{n},T}(\theta_{\bar{j}},\theta_0)|>2\varepsilon\right)\\
&\leqslant\overline{\lim_{n,\overline{n},T\to\infty}}P\left(\sup_{\theta\in\Theta^{\mu}}\sup_{\theta^*\in B(\theta,\gamma)}|Z_{n,\overline{n},T}(\theta^*,\theta_0)-Z_{n,\overline{n},T}(\theta,\theta_0)|>\varepsilon\right)\\
&\quad+\overline{\lim_{n,\overline{n},T\to\infty}}P\left(\max_{1\leqslant\bar{j}\leqslant\overline{J}}|Z_{n,\overline{n},T}(\theta_{\bar{j}},\theta_0)|>\varepsilon\right)\\
&<\varepsilon,
\end{aligned}$$

where the last inequality follows from (i) the condition in Eq. (27) and (ii) pointwise weak convergence of $Q^{\mu}_{n,\overline{n},T}(\theta)$ to $Q^{\mu}(\theta,\theta_0)$, as shown earlier. Hence, uniform convergence of the criterion function holds. This result proves the first part of the theorem. We now discuss consistency. For every $\varepsilon>0$, $\exists\xi>0$ such that

$$P(\widehat{\theta}^{\mu}_{n,\overline{n},T}\notin B(\theta^{\mu}_0,\varepsilon))$$

$$\leqslant P(Q^{\mu}(\widehat{\theta}^{\mu}_{n,\bar{n},T},\theta_0) - Q^{\mu}(\theta^{\mu}_0,\theta_0) \geqslant \xi)$$
$$\leqslant P(Q^{\mu}(\widehat{\theta}^{\mu}_{n,\bar{n},T},\theta_0) - Q^{\mu}_{n,\bar{n},T}(\widehat{\theta}^{\mu}_{n,\bar{n},T}) + Q^{\mu}_{n,\bar{n},T}(\widehat{\theta}^{\mu}_{n,\bar{n},T}) - Q^{\mu}(\theta^{\mu}_0,\theta_0) \geqslant \xi)$$
$$\leqslant P(Q^{\mu}(\widehat{\theta}^{\mu}_{n,\bar{n},T},\theta_0) - Q^{\mu}_{n,\bar{n},T}(\widehat{\theta}^{\mu}_{n,\bar{n},T}) + Q^{\mu}_{n,\bar{n},T}(\theta^{\mu}_0) + o_p(1) - Q^{\mu}(\theta^{\mu}_0,\theta_0) \geqslant \xi)$$
$$\leqslant P\left(2\sup_{\theta\in\Theta^{\mu}} |Q^{\mu}(\theta,\theta^{\mu}_0) - Q^{\mu}_{n,\bar{n},T}(\theta)| + o_p(1) \geqslant \xi\right) \underset{n,\bar{n},T\to\infty}{\to} 0,$$

where the first inequality follows from the identification condition implied by Eq. (10), the third inequality derives from the fact that $\widehat{\theta}^{\mu}_{n,\bar{n},T}$ is defined to satisfy

$$\widehat{\theta}^{\mu}_{n,\bar{n},T} \in \Theta^{\mu} \subset \Theta \quad \text{and} \quad Q^{\mu}_{n,\bar{n},T}(\widehat{\theta}^{\mu}_{n,\bar{n},T}) \leqslant \min_{\theta\in\Theta^{\mu}\subset\Theta} Q^{\mu}_{n,\bar{n},T}(\theta) + o_p(1)$$

and the final result is implied by uniform convergence of the criterion. This result proves the second part of the theorem. □

Proof of Theorem 2. Using the mean-value theorem, write

$$\widehat{\theta}^{\mu}_{n,\bar{n},T} - \theta^{\mu}_0 = -[\ddot{Q}^{\mu}_{n,\bar{n},T}(\theta^*_{n,\bar{n},T})]^{-1}\dot{Q}^{\mu}_{n,\bar{n},T}(\theta^{\mu}_0),$$

where

$$\theta^*_{n,\bar{n},T} \in (\widehat{\theta}^{\mu}_{n,\bar{n},T},\theta^{\mu}_0),$$
$$-\dot{Q}^{\mu}_{n,\bar{n},T}(\theta^{\mu}_0) = \frac{\overline{T}}{\bar{n}}\sum_{i=1}^{\bar{n}} (\widehat{\mu}_{(n,T)}(X_{i\Delta_{\bar{n},\overline{T}}}) - \mu(X_{i\Delta_{\bar{n},\overline{T}}},\theta^{\mu}_0))\frac{\partial\mu(X_{i\Delta_{\bar{n},\overline{T}}},\theta^{\mu}_0)}{\partial\theta^{\mu}},$$
$$\ddot{Q}^{\mu}_{n,\bar{n},T}(\theta^*_{n,\bar{n},T}) = \underbrace{\frac{\overline{T}}{\bar{n}}\sum_{i=1}^{\bar{n}} \frac{\partial\mu(X_{i\Delta_{\bar{n},\overline{T}}},\theta^*_{n,\bar{n},T})}{\partial\theta^{\mu}}\frac{\partial\mu(X_{i\Delta_{\bar{n},\overline{T}}},\theta^*_{n,\bar{n},T})}{\partial\theta^{\mu'}}}_{\ddot{Q}^{\mu(A)}_{n,\bar{n},T}(\theta^*_{n,\bar{n},T})}$$
$$\underbrace{-\frac{\overline{T}}{\bar{n}}\sum_{i=1}^{\bar{n}}(\widehat{\mu}_{(n,T)}(X_{i\Delta_{\bar{n},\overline{T}}}) - \mu(X_{i\Delta_{n,\overline{T}}},\theta^*_{n,\bar{n},T}))\frac{\partial\mu(X_{i\Delta_{n,\overline{T}}},\theta^*_{n,\bar{n},T})}{\partial\theta^{\mu}\,\partial\theta^{\mu'}}}_{\ddot{Q}^{\mu(B)}_{n,\bar{n},T}(\theta^*_{n,\bar{n},T})}.$$

Notice that $\theta^*_{n,\bar{n},T} \underset{n,\bar{n},T\to\infty}{\overset{p}{\to}} \theta^{\mu}_0$ since $\widehat{\theta}^{\mu}_{n,\bar{n},T} \underset{n,\bar{n},T\to\infty}{\overset{p}{\to}} \theta^{\mu}_0$ and $\theta^*_{n,\bar{n},T}$ lies on the line segment connecting $\widehat{\theta}^{\mu}_{n,\bar{n},T}$ and θ^{μ}_0. First, we examine $\ddot{Q}^{\mu(A)}_{n,\bar{n},T}(\theta^*_{n,\bar{n},T})$. Consider $\ddot{Q}^{\mu(A)}_{n,\bar{n},T}(\theta^{\mu})$.Using previous methods, we obtain

$$\ddot{Q}^{\mu(A)}_{n,\bar{n},T}(\theta^{\mu}) = \frac{\overline{T}}{\bar{n}}\sum_{i=1}^{\bar{n}} \frac{\partial\mu(X_{i\Delta_{\bar{n},\overline{T}}},\theta^{\mu})}{\partial\theta^{\mu}}\frac{\partial\mu(X_{i\Delta_{\bar{n},\overline{T}}},\theta^{\mu})}{\partial\theta^{\mu'}}$$
$$= \int_0^{\overline{T}} \frac{\partial\mu(X_s,\theta^{\mu})}{\partial\theta^{\mu}}\frac{\partial\mu(X_s,\theta^{\mu})}{\partial\theta^{\mu'}}\,ds + o_{a.s.}(1)$$

$$= \int_{\mathfrak{D}} \frac{\partial\mu(a,\theta^{\mu})}{\partial\theta^{\mu}} \frac{\partial\mu(a,\theta^{\mu})}{\partial\theta^{\mu'}} \frac{1}{\sigma_0^2(a)} L_X(\overline{T},a)\,\mathrm{d}a + \mathrm{o}_{\mathrm{a.s.}}(1)$$
$$= \int_{\mathfrak{D}} \frac{\partial\mu(a,\theta^{\mu})}{\partial\theta^{\mu}} \frac{\partial\mu(a,\theta^{\mu})}{\partial\theta^{\mu'}} \overline{L}_X(\overline{T},a)\,\mathrm{d}a + \mathrm{o}_{\mathrm{a.s.}}(1)$$
$$= \ddot{Q}^{\mu(A)}(\theta^{\mu},\theta_0) + \mathrm{o}_{\mathrm{a.s.}}(1). \tag{29}$$

Now notice that, given pointwise strong convergence as implied by Eq. (29) and the continuity of $\ddot{Q}^{\mu(A)}(.,\theta_0)$ from Assumption 1, the result

$$\sup_{\theta\in\Theta^{\mu}} |\ddot{Q}^{\mu(A)}_{n,\bar{n},T}(\theta) - \ddot{Q}^{\mu(A)}(\theta,\theta_0)| \underset{n,\bar{n},T\to\infty}{\overset{\mathrm{p}}{\to}} 0 \tag{30}$$

can be proved by using the same methods that were discussed in the proof of Theorem 1 to obtain Eq. (28). Hence,

$$|\ddot{Q}^{\mu(A)}_{n,\bar{n},T}(\widehat{\theta}^{\mu}_{n,\bar{n},T}) - \ddot{Q}^{\mu(A)}(\theta^{\mu}_0,\theta_0)|$$
$$\leqslant |\ddot{Q}^{\mu(A)}_{n,\bar{n},T}(\widehat{\theta}^{\mu}_{n,\bar{n},T}) - \ddot{Q}^{\mu(A)}(\widehat{\theta}^{\mu}_{n,\bar{n},T},\theta_0) + \ddot{Q}^{\mu(A)}(\widehat{\theta}^{\mu}_{n,\bar{n},T},\theta_0) - \ddot{Q}^{\mu(A)}(\theta^{\mu}_0,\theta_0)|$$
$$\leqslant |\ddot{Q}^{\mu(A)}_{n,\bar{n},T}(\widehat{\theta}^{\mu}_{n,\bar{n},T}) - \ddot{Q}^{\mu(A)}(\widehat{\theta}^{\mu}_{n,\bar{n},T},\theta_0)| + |\ddot{Q}^{\mu(A)}(\widehat{\theta}^{\mu}_{n,\bar{n},T},\theta_0) - \ddot{Q}^{\mu(A)}(\theta^{\mu}_0,\theta_0)|$$
$$\leqslant \mathrm{o_p}(1) + \mathrm{o_p}(1) \underset{n,\bar{n},T\to\infty}{\overset{\mathrm{p}}{\to}} 0,$$

where the second inequality holds by the triangle inequality and the final result follows from uniform convergence of $\ddot{Q}^{\mu(A)}_{n,\bar{n},T}(\theta)$ to $\ddot{Q}^{\mu(A)}(\theta,\theta_0)$ over Θ^{μ} (cf., Eq. (30)), continuity of $\ddot{Q}^{\mu(A)}(.,\theta_0)$, and consistency of $\widehat{\theta}^{\mu}_{n,\bar{n},T}$. Since $\theta^{*}_{n,\bar{n},T}$ lies on the line segment connecting $\widehat{\theta}^{\mu}_{n,\bar{n},T}$ and θ^{μ}_0, then

$$\ddot{Q}^{\mu(A)}_{n,\bar{n},T}(\theta^{*}_{n,\bar{n},T}) = \ddot{Q}^{\mu(A)}(\theta^{\mu}_0,\theta_0) + \mathrm{o_p}(1).$$

We now analyze $-\dot{Q}^{\mu}_{n,\bar{n},T}(\theta^{\mu}_0)$, writing

$$-\dot{Q}^{\mu}_{n,\bar{n},T}(\theta^{\mu}_0) = \frac{\overline{T}}{\bar{n}} \sum_{i=1}^{\bar{n}} (\widehat{\mu}_{(n,T)}(X_{i\Delta_{\bar{n},\overline{T}}}) - \mu(X_{i\Delta_{\bar{n},\overline{T}}},\theta^{\mu}_0)) \frac{\partial\mu(X_{i\Delta_{\bar{n},\overline{T}}},\theta^{\mu}_0)}{\partial\theta^{\mu}}$$
$$= \frac{\overline{T}}{\bar{n}} \sum_{i=1}^{\bar{n}} \left(\frac{\frac{1}{h_{n,T}} \sum_{j=1}^{n-1} \mathbf{K}\left(\frac{X_{j\Delta_{n,T}} - X_{i\Delta_{\bar{n},\overline{T}}}}{h_{n,T}}\right) [X_{(j+1)\Delta_{n,T}} - X_{j\Delta_{n,T}}]}{\frac{\Delta_{n,T}}{h_{n,T}} \sum_{j=1}^{n} \mathbf{K}\left(\frac{X_{j\Delta_{n,T}} - X_{i\Delta_{\bar{n},\overline{T}}}}{h_{n,T}}\right)} - \mu(X_{i\Delta_{\bar{n},\overline{T}}},\theta^{\mu}_0) \right) \frac{\partial\mu(X_{i\Delta_{n,\overline{T}}},\theta^{\mu}_0)}{\partial\theta^{\mu}}.$$

For simplicity, and given that there is no ambiguity with the notation, we express $\mu(X_{i\Delta_{\bar{n},\overline{T}}},\theta^{\mu}_0)$ as $\mu_0(X_{i\Delta_{\bar{n},\overline{T}}})$ and obtain

$$-\dot{Q}^{\mu}_{n,\bar{n},T}(\theta^{\mu}_0)$$

$$= \underbrace{\frac{\overline{T}}{\overline{n}}\sum_{i=1}^{\overline{n}}\left(\frac{\frac{1}{h_{n,\overline{T}}}\sum_{j=1}^{n-1}\mathbf{K}\left(\frac{X_{j\Delta_{n,T}}-X_{i\Delta_{\overline{n},\overline{T}}}}{h_{n,T}}\right)\int_{j\Delta_{n,T}}^{(j+1)\Delta_{n,T}}(\mu_0(X_s)-\mu_0(X_{i\Delta_{\overline{n},\overline{T}}}))\,\mathrm{d}s}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{j=1}^{n}\mathbf{K}\left(\frac{X_{j\Delta_{n,T}}-X_{i\Delta_{\overline{n},\overline{T}}}}{h_{n,T}}\right)}\right)\frac{\partial\mu_0(X_{i\Delta_{\overline{n},\overline{T}}})}{\partial\theta^{\mu}}}_{\mathbf{A}_{n,\overline{n},T}}$$

$$+ \underbrace{\frac{\overline{T}}{\overline{n}}\sum_{i=1}^{\overline{n}}\left(\frac{\frac{1}{h_{n,T}}\sum_{j=1}^{n-1}\mathbf{K}\left(\frac{X_{j\Delta_{n,T}}-X_{i\Delta_{\overline{n},\overline{T}}}}{h_{n,T}}\right)\int_{j\Delta_{n,T}}^{(j+1)\Delta_{n,T}}\sigma_0(X_s)\,\mathrm{d}B_s}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{j=1}^{n}\mathbf{K}\left(\frac{X_{j\Delta_{n,T}}-X_{i\Delta_{\overline{n},\overline{T}}}}{h_{n,T}}\right)}\right)\frac{\partial\mu_0(X_{i\Delta_{\overline{n},\overline{T}}})}{\partial\theta^{\mu}}}_{\mathbf{B}_{n,\overline{n},T}(1)}.$$

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

$$\mathbf{B}_{n,\overline{n},T}(r) = \sum_{j=1}^{[nr]-1} w(X_{j\Delta_{n,T}})\int_{j\Delta_{n,T}}^{(j+1)\Delta_{n,T}}\sigma_0(X_s)\,\mathrm{d}B_s$$

with

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

Also, write

$$\overline{\mathbf{B}}_{n,\overline{n},T}(r) = (\ddot{Q}^{\mu(A)}(\theta_0^{\mu},\theta_0)+\mathrm{o_p}(1))^{-1}\mathbf{B}_{n,\overline{n},T}(r) = \sum_{j=1}^{[nr]-1}\overline{w}(X_{j\Delta_{n,T}})\int_{j\Delta_{n,T}}^{(j+1)\Delta_{n,T}}\sigma_0(X_s)\,\mathrm{d}B_s,$$

where

$$\overline{w}(X_{j\Delta_{n,T}}) = (\ddot{Q}^{\mu(A)}(\theta_0^{\mu},\theta_0)+\mathrm{o_p}(1))^{-1}w(X_{j\Delta_{n,T}}).$$

The quantity $\overline{\mathbf{B}}_{n,\overline{n},T}(r)$ is a weighted sum of Brownian integrals whose quadratic variation can be characterized as

$$[\overline{\mathbf{B}}_{n,\overline{n},T}]_r \underset{n,\overline{n},T\to\infty}{\overset{\mathrm{p}}{\to}} (\ddot{Q}^{\mu(A)}(\theta_0^{\mu},\theta_0))^{-1}\left(\int_{\mathfrak{D}}\sigma_0^2(a)\left(\frac{\partial\mu_0(a)}{\partial\theta^{\mu}}\frac{\partial\mu_0(a)}{\partial\theta^{\mu'}}\right)\frac{(\overline{L}_X(\overline{T},a))^2}{\overline{L}_X^2(T,a)}\overline{L}_X(rT,a)\,\mathrm{d}a\right) \times(\ddot{Q}^{\mu(A)}(\theta_0^{\mu},\theta_0))^{-1}.$$

Now consider $\overline{\overline{\mathbf{B}}}_{n,\overline{n},T}(r) = ([\overline{\mathbf{B}}_{n,\overline{n},T}]_1)^{-1/2}\overline{\mathbf{B}}_{n,\overline{n},T}(r)$. Hence,

$$[\overline{\overline{\mathbf{B}}}_{n,\overline{n},T}(r)]_r = ([\overline{\mathbf{B}}_{n,\overline{n},T}]_1)^{-1/2}[\overline{\mathbf{B}}_{n,\overline{n},T}]_r([\overline{\mathbf{B}}_{n,\overline{n},T}]_1)^{-1/2}.$$

For $r=1$, given the asymptotic orthogonality between its elements, the vector $\overline{\overline{\mathbf{B}}}_{n,\overline{n},T}(r)$ can be embedded in a vector Brownian motion with quadratic variation $[\overline{\overline{\mathbf{B}}}_{n,\overline{n},T}(r)]_{r=1} = \mathbf{I}_{m_1}$ (see, e.g., Revuz and Yor, 1994, Corollary 2.4, p. 497). Then,

$$([\overline{\mathbf{B}}_{n,\overline{n},T}]_1)^{-1/2}\overline{\mathbf{B}}_{n,\overline{n},T}(1) \underset{n,\overline{n},T\to\infty}{\Rightarrow} \mathbf{N}(0,\mathbf{I}_{m_1}). \tag{31}$$

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

$$\mathbf{A}_{n,\overline{n},T} = \underbrace{\frac{\overline{T}}{\overline{n}}\sum_{i=1}^{\overline{n}}\left(\frac{\frac{1}{h_{n,T}}\sum_{j=1}^{n-1}\mathbf{K}\left(\frac{X_{j\Delta_{n,T}}-X_{i\Delta_{\overline{n},\overline{T}}}}{h_{n,T}}\right)(\mu_0(X^*)-\mu_0(X_{j\Delta_{n,T}}))\Delta_{n,T}}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{j=1}^{n}\mathbf{K}\left(\frac{X_{j\Delta_{n,T}}-X_{i\Delta_{\overline{n},\overline{T}}}}{h_{n,T}}\right)}\right)\frac{\partial\mu_0(X_{i\Delta_{\overline{n},\overline{T}}})}{\partial\theta^{\mu}}}_{\mathbf{A}^1_{n,\overline{n},T}} + \underbrace{\frac{\overline{T}}{\overline{n}}\sum_{i=1}^{\overline{n}}\left(\frac{\frac{1}{h_{n,T}}\sum_{j=1}^{n-1}\mathbf{K}\left(\frac{X_{j\Delta_{n,T}}-X_{i\Delta_{\overline{n},\overline{T}}}}{h_{n,T}}\right)(\mu_0(X_{j\Delta_{n,T}})-\mu_0(X_{i\Delta_{\overline{n},\overline{T}}}))\Delta_{n,T}}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{j=1}^{n}\mathbf{K}\left(\frac{X_{j\Delta_{n,T}}-X_{i\Delta_{\overline{n},\overline{T}}}}{h_{n,T}}\right)}\right)\frac{\partial\mu_0(X_{i\Delta_{\overline{n},\overline{T}}})}{\partial\theta^{\mu}}}_{\mathbf{A}^2_{n,\overline{n},T}},$$

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

$$\mathbf{A}^1_{n,\overline{n},T} \leqslant C_4\kappa_{n,T}\left(\frac{\overline{T}}{\overline{n}}\sum_{i=1}^{\overline{n}}\frac{\partial\mu_0(X_{i\Delta_{\overline{n},\overline{T}}})}{\partial\theta^{\mu}}\right),$$

where $\kappa_{n,T} = \max_{j\leqslant n}\sup_{j\Delta_{n,T}\leqslant s\leqslant (j+1)\Delta_{n,T}}|X_s - X_{j\Delta_{n,T}}|$. We know that

$$\kappa_{n,T} = \mathrm{O}_{\mathrm{a.s.}}(\Delta_{n,T}^{1/2}\log(1/\Delta_{n,T})).$$

Then, the bound becomes

$$C_4\mathrm{O}_{\mathrm{a.s.}}(\Delta_{n,T}^{1/2}\log(1/\Delta_{n,T}))\left(\left(\int_{\mathfrak{D}}\frac{\partial\mu_0(a)}{\partial\theta^{\mu}}\overline{L}_X(\overline{T},a)\,\mathrm{d}a\right)+\mathrm{o}_{\mathrm{a.s.}}(1)\right).$$

Now consider the term $\mathbf{A}^2_{n,\overline{n},T}$. Using the Quotient Limit Theorem (Revuz and Yor, 1994, Theorem 3.12, p. 408) as in Bandi and Moloche (2001), we obtain

$$\mathbf{A}^2_{n,\overline{n},T} = \frac{\overline{T}}{\overline{n}}\sum_{i=1}^{\overline{n}}\left(\frac{\frac{1}{h_{n,T}}\sum_{j=1}^{n-1}\mathbf{K}\left(\frac{X_{j\Delta_{n,T}}-X_{i\Delta_{\overline{n},\overline{T}}}}{h_{n,T}}\right)(\mu_0(X_{j\Delta_{n,T}})-\mu_0(X_{i\Delta_{\overline{n},\overline{T}}}))\Delta_{n,T}}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{j=1}^{n}\mathbf{K}\left(\frac{X_{j\Delta_{n,T}}-X_{i\Delta_{\overline{n},\overline{T}}}}{h_{n,T}}\right)}\right)\frac{\partial\mu_0(X_{i\Delta_{\overline{n},\overline{T}}})}{\partial\theta^{\mu}}$$
$$= h^2_{n,T}\int_{\mathfrak{D}}\mathbf{K}_2\left(\frac{\partial\mu_0(a)}{\partial a}\frac{\frac{\partial m(a)}{\partial a}}{m(a)}+\frac{1}{2}\frac{\partial^2\mu_0(a)}{\partial a}\right)\frac{\partial\mu_0(a)}{\partial\theta^{\mu}}\overline{L}_X(\overline{T},a)\,\mathrm{d}a+\mathrm{o}(h^2_{n,T}),$$

where $m(.)$ is the speed function of the process and $\mathbf{K}_2 = \int_{-\infty}^{\infty}c^2\mathbf{K}(c)\,\mathrm{d}c<\infty$. Following similar steps as those leading to Eq. (31), we can show that

$$\ddot{Q}^{\mu(B)}_{n,\overline{n},T}(\theta^*_{n,\overline{n},T}) \underset{n,\overline{n},T\to\infty}{\overset{\mathrm{p}}{\to}} 0.$$

Then,

$$\Xi_{\mu}^{-1/2}(T)(\widehat{\theta}^{\mu}_{n,\overline{n},T}-\theta^{\mu}_0)$$

$$= \Xi_\mu^{-1/2}(T)\left(\int_{\mathfrak{D}} \frac{\partial\mu_0(a)}{\partial\theta^\mu}\frac{\partial\mu_0(a)}{\partial\theta^{\mu'}}\overline{L}_X(\overline{T},a)\,\mathrm{d}a + \mathrm{o_p}(1)\right)^{-1}[\mathbf{A}^1_{n,\overline{n},T} + \mathbf{A}^2_{n,\overline{n},T} + \mathbf{B}_{n,\overline{n},T}(1)]$$
$$\overset{\mathrm{d}}{=} \Xi_\mu^{-1/2}(T)\left(\int_{\mathfrak{D}} \frac{\partial\mu_0(a)}{\partial\theta^\mu}\frac{\partial\mu_0(a)}{\partial\theta^{\mu'}}\overline{L}_X(\overline{T},a)\,\mathrm{d}a + \mathrm{o_p}(1)\right)^{-1}[\mathrm{O_p}(h^2_{n,T}) + \mathbf{B}_{n,\overline{n},T}(1)]$$
$$\underset{n,\overline{n},T\to\infty}{\Rightarrow} \mathbf{N}(0,\mathbf{I}_{m_1}),$$

where

$$\Xi_\mu(T) = B(\overline{T})^{-1}_\mu V(T)_\mu B(\overline{T})^{-1}_\mu,$$

$$B_\mu = \left(\int_{\mathfrak{D}} \frac{\partial\mu_0(a)}{\partial\theta^\mu}\frac{\partial\mu_0(a)}{\partial\theta^{\mu'}}\overline{L}_X(\overline{T},a)\,\mathrm{d}a\right),$$

and

$$V_\mu = \left(\int_{\mathfrak{D}} \sigma_0^2(a)\left(\frac{\partial\mu_0(a)}{\partial\theta^\mu}\frac{\partial\mu_0(a)}{\partial\theta^{\mu'}}\right)\frac{(\overline{L}_X(\overline{T},a))^2}{\overline{L}_X(T,a)}\,\mathrm{d}a\right)$$

provided $h^4_{n,T}\Xi_\mu^{-1}(T) \overset{\mathrm{a.s.}}{\to} 0$. If $h^4_{n,T}\Xi_\mu^{-1}(T) = \mathrm{O_{a.s.}}(1)$, thus

$$\Xi_\mu^{-1/2}(T)(\widehat{\theta}^\mu_{n,\overline{n},T} - \theta^\mu_0 - \Gamma^\mu) \underset{n,\overline{n},T\to\infty}{\Rightarrow} \mathbf{N}(0,\mathbf{I}_{m_1}),$$

where

$$\Gamma^\mu = (h^2_{n,T})B_\mu^{-1}\int_{\mathfrak{D}} \mathbf{K}_2\left(\frac{\partial\mu_0(a)}{\partial a}\frac{\frac{\partial m(a)}{\partial a}}{m(a)} + \frac{1}{2}\frac{\partial^2\mu_0(a)}{\partial a}\right)\frac{\partial\mu_0(a)}{\partial\theta^\mu}\overline{L}_X(\overline{T},a)\,\mathrm{d}a$$

with $\mathbf{K}_2 = \int_{-\infty}^{\infty} c^2\mathbf{K}(c)\,\mathrm{d}c$. This proves the stated result. □

Proof of Theorem 3. We can follow similar steps as for the proof of Theorem 1. □

Proof of Theorem 4. Write

$$\widehat{\theta}^\sigma_{n,\overline{n},T} - \theta^\sigma_0 = -[\ddot{Q}^\sigma_{n,\overline{n},T}(\theta^*_{n,\overline{n},T})]^{-1}\dot{Q}^\sigma_{n,\overline{n},T}(\theta^\sigma_0),$$

where

$$\theta^*_{n,\overline{n},T} \in (\widehat{\theta}^\sigma_{n,\overline{n},T}, \theta^\sigma_0),$$
$$-\dot{Q}^\sigma_{n,\overline{n},T}(\theta^\sigma_0) = \frac{\overline{T}}{\overline{n}}\sum_{i=1}^{\overline{n}}(\widehat{\sigma}^2_{(n,T)}(X_{i\Delta_{\overline{n},\overline{T}}}) - \sigma^2(X_{i\Delta_{\overline{n},\overline{T}}},\theta^\sigma_0))\frac{\partial\sigma^2(X_{i\Delta_{\overline{n},\overline{T}}},\theta^\sigma_0)}{\partial\theta^\sigma},$$
$$\ddot{Q}^\sigma_{n,\overline{n},T}(\theta^*_{n,\overline{n},T}) = \underbrace{\frac{\overline{T}}{\overline{n}}\sum_{i=1}^{\overline{n}}\frac{\partial\sigma^2(X_{i\Delta_{\overline{n},\overline{T}}},\theta^*_{n,\overline{n},T})}{\partial\theta^\sigma}\frac{\partial\sigma^2(X_{i\Delta_{\overline{n},\overline{T}}},\theta^*_{n,\overline{n},T})}{\partial\theta^{\sigma'}}}_{\ddot{Q}^{\sigma(A)}_{n,\overline{n},T}(\theta^*_{n,\overline{n},T})}$$
$$\underbrace{-\frac{\overline{T}}{n}\sum_{i=1}^{\overline{n}}(\widehat{\sigma}^2_{(n,T)}(X_{i\Delta_{\overline{n},\overline{T}}}) - \sigma^2(X_{i\Delta_{\overline{n},\overline{T}}},\theta^*_{n,\overline{n},T}))\frac{\partial\sigma^2(X_{i\Delta_{\overline{n},\overline{T}}},\theta^*_{n,\overline{n},T})}{\partial\theta^\sigma\,\partial\theta^{\sigma'}}}_{\ddot{Q}^{\sigma(B)}_{n,\overline{n},T}(\theta^*_{n,\overline{n},T})}.$$

First, we examine $\ddot{Q}^{\sigma}_{n,\overline{n},T}(\theta^{*}_{n,\overline{n},T})$. Consider $\ddot{Q}^{\sigma(A)}_{n,\overline{n},T}(\theta^{\sigma})$. We obtain

$$
\begin{aligned}
\ddot{Q}^{\sigma(A)}_{n,\overline{n},T}(\theta^{\sigma}) &= \frac{\overline{T}}{\overline{n}}\sum_{i=1}^{n}\frac{\partial\sigma^2(X_{i\Delta_{\overline{n},\overline{T}}},\theta^{\sigma})}{\partial\theta^{\sigma}}\frac{\partial\sigma^2(X_{i\Delta_{\overline{n},\overline{T}}},\theta^{\sigma})}{\partial\theta^{\sigma'}}\\
&= \int_0^{\overline{T}}\frac{\partial\sigma^2(X_s,\theta^{\sigma})}{\partial\theta^{\sigma}}\frac{\partial\sigma^2(X_s,\theta^{\sigma})}{\partial\theta^{\sigma'}}\,\mathrm{d}s + \mathrm{o}_{\mathrm{a.s.}}(1)\\
&= \int_{\mathfrak{D}}\frac{\partial\sigma^2(a,\theta^{\sigma})}{\partial\theta^{\sigma}}\frac{\partial\sigma^2(a,\theta^{\sigma})}{\partial\theta^{\sigma'}}\frac{1}{\sigma_0^2(a)}L_X(\overline{T},a)\,\mathrm{d}a + \mathrm{o}_{\mathrm{a.s.}}(1)\\
&= \int_{\mathfrak{D}}\frac{\partial\sigma^2(a,\theta^{\sigma})}{\partial\theta^{\sigma}}\frac{\partial\sigma^2(a,\theta^{\sigma})}{\partial\theta^{\sigma'}}\overline{L}_X(\overline{T},a)\,\mathrm{d}a + \mathrm{o}_{\mathrm{a.s.}}(1)\\
&= \ddot{Q}^{\sigma(A)}(\theta^{\sigma},\theta_0) + \mathrm{o}_{\mathrm{a.s.}}(1), \qquad (32)
\end{aligned}
$$

using the continuity of the underlying semimartingale as in previous proofs and the occupation time formula. Uniform strong convergence over Θ^{σ} can be shown following the same steps leading to Eq. (28) in the proof of Theorem 1. Hence,

$$
\sup_{\theta\in\Theta^{\sigma}}|\ddot{Q}^{\sigma(A)}_{n,\overline{n},T}(\theta) - \ddot{Q}^{\sigma(A)}(\theta,\theta_0)| \underset{n,\overline{n},T\to\infty}{\overset{\mathrm{p}}{\to}} 0. \qquad (33)
$$

Then, using the continuity of $\ddot{Q}^{\sigma(A)}(.,\theta_0)$, the consistency of $\widehat{\theta}^{\sigma}_{n,\overline{n},T}$, and the result in Eq. (33) as in the proof of Theorem 2, we can readily obtain

$$
\ddot{Q}^{\sigma(A)}_{n,\overline{n},T}(\theta^{*}_{n,\overline{n},T}) = \ddot{Q}^{\sigma(A)}(\theta^{\sigma}_0,\theta_0) + \mathrm{o}_{\mathrm{p}}(1),
$$

since, as earlier, $\theta^{*}_{n,\overline{n},T}$ lies on the line segment connecting $\widehat{\theta}^{\sigma}_{n,\overline{n},T}$ and θ^{σ}_0. Furthermore,

$$
\ddot{Q}^{\sigma(B)}_{n,\overline{n},T}(\theta^{*}_{n,\overline{n},T}) = \ddot{Q}^{\sigma(B)}(\theta^{\sigma}_0,\theta_0) + \mathrm{o}_{\mathrm{p}}(1) = \mathrm{o}_{\mathrm{p}}(1).
$$

Now, consider $-\dot{Q}^{\sigma}_{n,\overline{n},T}(\theta^{\sigma}_0)$. Write

$$
\begin{aligned}
&-\dot{Q}^{\sigma}_{n,\overline{n},T}(\theta^{\sigma}_0)\\
&\quad= \frac{\overline{T}}{\overline{n}}\sum_{i=1}^{\overline{n}}(\widehat{\sigma}^2_{(n,T)}(X_{i\Delta_{\overline{n},\overline{T}}}) - \sigma^2(X_{i\Delta_{\overline{n},\overline{T}}},\theta^{\sigma}_0))\frac{\partial\sigma^2(X_{i\Delta_{\overline{n},\overline{T}}},\theta^{\sigma}_0)}{\partial\theta^{\sigma}}\\
&\quad= \frac{\overline{T}}{\overline{n}}\sum_{i=1}^{\overline{n}}\left(\frac{\frac{1}{h_{n,T}}\sum_{j=1}^{n-1}\mathbf{K}\left(\frac{X_{j\Delta_{n,T}}-X_{i\Delta_{\overline{n},\overline{T}}}}{h_{n,T}}\right)(X_{(j+1)\Delta_{n,T}} - X_{j\Delta_{n,T}})^2}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{j=1}^{n}\mathbf{K}\left(\frac{X_{j\Delta_{n,T}}-X_{i\Delta_{\overline{n},\overline{T}}}}{h_{n,T}}\right)} - \sigma^2(X_{i\Delta_{\overline{n},\overline{T}}},\theta^{\sigma}_0)\right)\\
&\qquad\times\frac{\partial\sigma^2(X_{i\Delta_{\overline{n},\overline{T}}},\theta^{\sigma}_0)}{\partial\theta^{\sigma}}.
\end{aligned}
$$

Using the notation $\sigma^2(X_{i\Delta_{n,\overline{T}}}, \theta_0^\sigma) = \sigma_0^2(X_{i\Delta_{n,\overline{T}}})$, we obtain

$$
\begin{aligned}
&-\dot{Q}^{\sigma}_{n,\bar{n},T}(\theta_0^\sigma)\\
&= \underbrace{\frac{\overline{T}}{\bar{n}}\sum_{i=1}^{\bar{n}}\left(\frac{\frac{1}{h_{n,T}}\sum_{j=1}^{n-1}\mathbf{K}\left(\frac{X_{j\Delta_{n,T}}-X_{i\Delta_{\bar{n},\overline{T}}}}{h_{n,T}}\right)\int_{j\Delta_{n,T}}^{(j+1)\Delta_{n,T}} 2(X_s - X_{j\Delta_{n,T}})\mu_0(X_s)\,\mathrm{d}s}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{j=1}^{n}\mathbf{K}\left(\frac{X_{j\Delta_{n,T}}-X_{i\Delta_{\bar{n},\overline{T}}}}{h_{n,T}}\right)}\right)\frac{\partial\sigma_0^2(X_{i\Delta_{\bar{n},\overline{T}}})}{\partial\theta^\sigma}}_{\mathbf{A}_{n,\bar{n},T}}\\
&\quad+\underbrace{\frac{\overline{T}}{\bar{n}}\sum_{i=1}^{\bar{n}}\left(\frac{\frac{1}{h_{n,T}}\sum_{j=1}^{n-1}\mathbf{K}\left(\frac{X_{j\Delta_{n,T}}-X_{i\Delta_{\bar{n},\overline{T}}}}{h_{n,T}}\right)\int_{j\Delta_{n,T}}^{(j+1)\Delta_{n,T}} 2\left(X_s - X_{j\Delta_{n,T}}\right)\sigma_0(X_s)\,\mathrm{d}B_s}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{j=1}^{n}\mathbf{K}\left(\frac{X_{j\Delta_{n,T}}-X_{i\Delta_{\bar{n},\overline{T}}}}{h_{n,T}}\right)}\right)\frac{\partial\sigma_0^2(X_{i\Delta_{\bar{n},\overline{T}}})}{\partial\theta^\sigma}}_{\mathbf{B}_{n,\bar{n},T}(1)}\\
&\quad+\underbrace{\frac{\overline{T}}{\bar{n}}\sum_{i=1}^{\bar{n}}\left(\frac{\frac{1}{h_{n,T}}\sum_{j=1}^{n-1}\mathbf{K}\left(\frac{X_{j\Delta_{n,T}}-X_{i\Delta_{\bar{n},\overline{T}}}}{h_{n,T}}\right)\int_{j\Delta_{n,T}}^{(j+1)\Delta_{n,T}} (\sigma_0^2(X_s) - \sigma_0^2(X_{i\Delta_{n,\overline{T}}}))\,\mathrm{d}s}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{j=1}^{n}\mathbf{K}\left(\frac{X_{j\Delta_{n,T}}-X_{i\Delta_{\bar{n},\overline{T}}}}{h_{n,T}}\right)}\right)\frac{\partial\sigma_0^2(X_{i\Delta_{\bar{n},\overline{T}}})}{\partial\theta^\sigma}}_{\mathbf{C}_{n,\bar{n},T}}.
\end{aligned}
$$

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

$$
\sqrt{\frac{1}{\Delta_{n,T}}}\mathbf{B}_{n,\bar{n},T}(r) = \sqrt{\frac{1}{\Delta_{n,T}}}\sum_{j=1}^{[nr]-1} w(X_{j\Delta_{n,T}})\int_{j\Delta_{n,T}}^{(j+1)\Delta_{n,T}} 2(X_s - X_{j\Delta_{n,T}})\sigma_0(X_s)\,\mathrm{d}B_s,
$$

where

$$
w(X_{j\Delta_{n,T}}) = \frac{\overline{T}}{\bar{n}}\sum_{i=1}^{\bar{n}}\frac{\frac{1}{h_{n,T}}\mathbf{K}\left(\frac{X_{j\Delta_{n,T}}-X_{i\Delta_{\bar{n},\overline{T}}}}{h_{n,T}}\right)\frac{\partial\sigma_0^2(X_{i\Delta_{\bar{n},\overline{T}}})}{\partial\theta^\sigma}}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{j=1}^{n}\mathbf{K}\left(\frac{X_{j\Delta_{n,T}}-X_{i\Delta_{\bar{n},\overline{T}}}}{h_{n,T}}\right)}. \tag{34}
$$

As in the case of the corresponding term in the proof of Theorem 2, $\sqrt{(1/\Delta_{n,T})}\mathbf{B}_{n,\bar{n},T}(r)$ is a weighted sum of Brownian integrals whose quadratic variation can be expressed as

$$
\left[\sqrt{\frac{1}{\Delta_{n,T}}}\mathbf{B}_{n,\bar{n},T}\right]_r = \frac{1}{\Delta_{n,T}}\sum_{j=1}^{[nr]-1} w^2(X_{j\Delta_{n,T}})\left[2\int M_s\mathrm{d}M_s\right]_{j\Delta_{n,T},(j+1)\Delta_{n,T}} + \mathrm{o_p}(1) \tag{35}
$$

with $M_s = \int_{j\Delta_{n,T}}^{s}\sigma_0(X_u)\,\mathrm{d}B_u$. For simplicity, in Eq. (35) we abuse notation by writing $w^2(.)$ even though $w^2(.)$ is an m_1-vector. The simplification should cause no confusion. Hence,

$$
\begin{aligned}
\left[\sqrt{\frac{1}{\Delta_{n,T}}}\mathbf{B}_{n,\bar{n},T}\right]_r &= \frac{4}{\Delta_{n,T}}\sum_{j=1}^{[nr]-1} w^2(X_{j\Delta_{n,T}})\left[\int M_s\,\mathrm{d}M_s\right]_{j\Delta_{n,T},(j+1)\Delta_{n,T}} + \mathrm{o_p}(1)\\
&= \frac{4}{\Delta_{n,T}}\sum_{j=1}^{[nr]-1} w^2(X_{j\Delta_{n,T}})\int_{j\Delta_{n,T}}^{(j+1)\Delta_{n,T}} M_s^2\,\mathrm{d}[M]_s + \mathrm{o_p}(1)
\end{aligned}
$$

$$= \frac{4}{\Delta_{n,T}} \sum_{j=1}^{[nr]-1} w^2(X_{j\Delta_{n,T}}) \int_{j\Delta_{n,T}}^{(j+1)\Delta_{n,T}} [M]_s \, \mathrm{d}[M]_s + \mathrm{o_p}(1)$$

$$= \frac{4}{\Delta_{n,T}} \sum_{j=1}^{[nr]-1} w^2(X_{j\Delta_{n,T}}) \left(\frac{[M]^2}{2} \Big|_{j\Delta_{n,T}}^{(j+1)\Delta_{n,T}} \right) + \mathrm{o_p}(1)$$

$$= \frac{2}{\Delta_{n,T}} \sum_{j=1}^{[nr]-1} w^2(X_{j\Delta_{n,T}}) \left(\int_{j\Delta_{n,T}}^{(j+1)\Delta_{n,T}} \sigma_0^2(X_s) \, \mathrm{d}s \right)^2 + \mathrm{o_p}(1)$$

$$= \int_{\mathfrak{D}} 2\sigma_0^4(a) \left(\frac{\partial \sigma_0^2(a)}{\partial \theta^{\sigma}} \frac{\partial \sigma_0^2(a)}{\partial \theta^{\sigma'}} \right) \frac{\left(\overline{L}_X(\overline{T}, a)\right)^2}{\overline{L}_X^2(T, a)} \overline{L}_X(rT, a) \, \mathrm{d}a + \mathrm{o_p}(1).$$

By using the same steps leading to Eq. (31) in the proof of Theorem 2, we can show that

$$\sqrt{\frac{1}{\Delta_{n,T}}} ([\overline{\mathbf{B}}_{n,\bar{n},T}]_1)^{-1/2} \overline{\mathbf{B}}_{n,\bar{n},T}(1) \underset{n,\bar{n},T \to \infty}{\Rightarrow} \mathbf{N}(0, \mathbf{I}_{m_2}),$$

where

$$[\overline{\mathbf{B}}_{n,\bar{n},T}]_1 = (\ddot{Q}^{\sigma(A)}(\theta_0^{\sigma}, \theta_0))^{-1} \left(\int_{\mathfrak{D}} 2\sigma_0^4(a) \left(\frac{\partial \sigma_0^2(a)}{\partial \theta^{\sigma}} \frac{\partial \sigma_0^2(a)}{\partial \theta^{\sigma'}} \right) \frac{(\overline{L}_X(\overline{T}, a))^2}{\overline{L}_X(T, a)} \, \mathrm{d}a \right) (\ddot{Q}^{\sigma(A)}(\theta_0^{\sigma}, \theta_0))^{-1}$$

and

$$\overline{\mathbf{B}}_{n,\bar{n},T}(1) = (\ddot{Q}^{\sigma(A)}(\theta_0^{\sigma}, \theta_0))^{-1} \mathbf{B}_{n,\bar{n},T}(r).$$

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

$$\mathbf{C}_{n,\bar{n},T} = \underbrace{\frac{\overline{T}}{\bar{n}} \sum_{i=1}^{\bar{n}} \left(\frac{\frac{1}{h_{n,T}} \sum_{j=1}^{n} \mathbf{K}\left(\frac{X_{j\Delta_{n,T}} - X_{i\Delta_{\bar{n},\overline{T}}}}{h_{n,T}} \right) (\sigma_0^2(X^*) - \sigma_0^2(X_{j\Delta_{n,T}})) \Delta_{n,T}}{\frac{\Delta_{n,T}}{h_{n,T}} \sum_{j=1}^{n} \mathbf{K}\left(\frac{X_{j\Delta_{n,T}} - X_{i\Delta_{\bar{n},\overline{T}}}}{h_{n,T}} \right)} \right) \frac{\partial \sigma_0^2(X_{i\Delta_{\bar{n},\overline{T}}})}{\partial \theta^{\sigma}}}_{\mathbf{C}^1_{n,\bar{n},T}}$$

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

where $X^* \in (X_{(j+1)\Delta_{n,T}}, X_{j\Delta_{n,T}})$ by the mean-value theorem, as earlier. Analogously to $\mathbf{A}^2_{n,\bar{n},T}$ in the proof of Theorem 2, we can show that

$$\mathbf{C}^2_{n,\bar{n},T} = h^2_{n,T} \int_{\mathfrak{D}} \mathbf{K}_2 \left(\frac{\partial \sigma_0^2(a)}{\partial a} \frac{\frac{\partial m(a)}{\partial a}}{m(a)} + \frac{1}{2} \frac{\partial^2 \sigma_0^2(a)}{\partial a} \right) \frac{\partial \sigma_0^2(a)}{\partial \theta^{\sigma}} \overline{L}_X(\overline{T}, a) \, \mathrm{d}a + \mathrm{o}(h^2_{n,T}),$$

where $\mathbf{K}_2 = \int_{-\infty}^{\infty} c^2 \mathbf{K}(c)\,\mathrm{d}c$, by virtue of the Quotient Limit Theorem. As for $\mathbf{A}_{n,\bar{n},T}$ and $\mathbf{C}^1_{n,\bar{n},T}$, it is immediate to see that $\mathbf{A}_{n,\bar{n},T} = \mathrm{o_p}(\mathbf{B}_{n,\bar{n},T}(1))$ and $\mathbf{C}^1_{n,\bar{n},T} = \mathrm{o_p}(\mathbf{C}^2_{n,\bar{n},T})$. Then,

$$\begin{aligned}
&\frac{1}{\sqrt{\Delta_{n,T}}} \Xi_\sigma^{-1/2}(T)(\widehat{\theta}^\sigma_{n,\bar{n},T} - \theta^\sigma_0) \\
&= -\frac{1}{\sqrt{\Delta_{n,T}}} \Xi_\sigma^{-1/2}(T)[\ddot{Q}^\sigma_{n,\bar{n},T}(\theta^*)]^{-1} \dot{Q}^\sigma_{n,\bar{n},T}(\theta^\sigma_0) \\
&= \frac{1}{\sqrt{\Delta_{n,T}}} \Xi_\sigma^{-1/2}(T)\left(\left(\int_{\mathfrak{D}} \frac{\partial\sigma_0^2(a)}{\partial\theta^\sigma}\frac{\partial\sigma_0^2(a)}{\partial\theta^{\sigma'}} \bar{L}_X(\overline{T},a)\,\mathrm{d}a\right) + \mathrm{o_p}(1)\right)^{-1} \\
&\quad \times[\mathbf{B}_{n,\bar{n},T}(1) + \mathbf{A}_{n,\bar{n},T} + \mathbf{C}_{n,\bar{n},T}] \\
&\overset{\mathrm{d}}{=} \frac{1}{\sqrt{\Delta_{n,T}}} \Xi_\sigma^{-1/2}(T)\left(\left(\int_{\mathfrak{D}} \frac{\partial\sigma_0^2(a)}{\partial\theta^\sigma}\frac{\partial\sigma_0^2(a)}{\partial\theta^{\sigma'}} \bar{L}_X(\overline{T},a)\,\mathrm{d}a\right) + \mathrm{o_p}(1)\right)^{-1} [\mathrm{O_p}(h^2_{n,T}) + \mathbf{B}_{n,\bar{n},T}(1)].
\end{aligned}$$

If $h^4_{n,T}\Xi_\sigma^{-1}(T)/\Delta_{n,T} \overset{\text{a.s.}}{\to} 0$, then

$$\frac{1}{\sqrt{\Delta_{n,T}}} \Xi_\sigma^{-1/2}(T)(\widehat{\theta}^\sigma_{n,\bar{n},T} - \theta^\sigma_0) \underset{n,\bar{n},T\to\infty}{\Rightarrow} \mathbf{N}(0, \mathbf{I}_{m_2}),$$

where

$$\Xi_\sigma(T) = B(\overline{T})_\sigma^{-1} V(T)_\sigma B(\overline{T})_\sigma^{-1},$$

$$B_\sigma = \left(\int_{\mathfrak{D}} \frac{\partial\sigma_0^2(a)}{\partial\theta^\sigma}\frac{\partial\sigma_0^2(a)}{\partial\theta^{\sigma'}} \bar{L}_X(\overline{T},a)\,\mathrm{d}a\right),$$

and

$$V_\sigma = \left(\int_{\mathfrak{D}} 2\sigma_0^4(a)\left(\frac{\partial\sigma_0^2(a)}{\partial\theta^\sigma}\frac{\partial\sigma_0^2(a)}{\partial\theta^{\sigma'}}\right) \frac{(\bar{L}_X(\overline{T},a))^2}{\bar{L}_X(T,a)}\,\mathrm{d}a\right).$$

If $h^4_{n,T}\Xi_\sigma^{-1}(T)/\Delta_{n,T} = \mathrm{O_{a.s.}}(1)$, then

$$\frac{1}{\sqrt{\Delta_{n,T}}} \Xi_\sigma^{-1/2}(T)(\widehat{\theta}^\sigma_{n,\bar{n},T} - \theta^\sigma_0 - \Gamma^\sigma) \underset{n,\bar{n},T\to\infty}{\Rightarrow} \mathbf{N}(0, \mathbf{I}_{m_2}),$$

where

$$\Gamma^\sigma = h^2_{n,T} B_\sigma^{-1} \int_{\mathfrak{D}} \mathbf{K}_2 \left(\frac{\partial\sigma_0^2(a)}{\partial a}\frac{\frac{\partial m(a)}{\partial a}}{m(a)} + \frac{1}{2}\frac{\partial^2\sigma_0^2(a)}{\partial a}\right)\frac{\partial\sigma_0^2(a)}{\partial\theta^\sigma}\bar{L}_X(\overline{T},a)\,\mathrm{d}a$$

with $\mathbf{K}_2 = \int_{-\infty}^{\infty} c^2\mathbf{K}(c)\,\mathrm{d}c$. This concludes the proof of Theorem 4. □

Proof of Theorem 5. The proof largely follows the proof of Theorem 4. We simply need to show mixed normality of the limiting distribution when performing the asymptotics over a fixed time interval $\overline{T}$ (with $n = \bar{n}$). Consider

$$\sqrt{\frac{1}{\Delta_{n,\overline{T}}}}\mathbf{B}_{n,\overline{T}}(r) = \sqrt{\frac{1}{\Delta_{n,\overline{T}}}} \sum_{j=1}^{[nr]-1} w(X_{j\Delta_{n,\overline{T}}}) \int_{j\Delta_{n,\overline{T}}}^{(j+1)\Delta_{n,\overline{T}}} 2(X_s - X_{j\Delta_{n,\overline{T}}})\sigma_0(X_s)\,\mathrm{d}B_s$$

with $w(X_{j\Delta_{n,\overline{T}}})$ as defined in Eq. (34). By using standard embedding arguments (see, e.g., Revuz and Yor, 1994, Theorem 2.3, p. 496), it is simple to show that

$$\sqrt{\frac{1}{\Delta_{n,\overline{T}}}}\mathbf{B}_{n,\overline{T}}(r) \underset{n\to\infty}{\Rightarrow} W\left(\int_{\mathfrak{D}} 2\sigma_0^4(a)\left(\frac{\partial\sigma_0^2(a)}{\partial\theta^{\sigma}}\frac{\partial\sigma_0^2(a)}{\partial\theta^{\sigma'}}\right)\overline{L}_X(r\overline{T},a)\,\mathrm{d}a\right),$$

where W denotes Brownian motion. We now need to prove that W is independent of the asymptotic variance or, equivalently, of $\overline{L}_X(r\overline{T},.)$, for mixed normality to hold. To this extent, we evaluate the limiting covariation process between $\sqrt{(1/\Delta_{n,\overline{T}})}\mathbf{B}_{n,\overline{T}}(r)$ and

$$X_{r\overline{T}} = X_{r\overline{T}} - X_0 + X_0 = \underbrace{\sum_{j=1}^{[nr]-1}(X_{(j+1)\Delta_{n,\overline{T}}} - X_{j\Delta_{n,\overline{T}}})}_{\alpha_{n,\overline{T}}} + \underbrace{(X_{\Delta_{n,\overline{T}}} - X_0) + X_0}_{\beta_{n,\overline{T}}}.$$

Consider

$$\begin{aligned}
&\left[\sqrt{\frac{1}{\Delta_{n,\overline{T}}}}\mathbf{B}_{n,\overline{T}},\alpha_{n,\overline{T}}\right]_r \\
&\quad= \sqrt{\frac{1}{\Delta_{n,\overline{T}}}}\sum_{j=1}^{[nr]-1} w(X_{j\Delta_{n,\overline{T}}})\int_{j\Delta_{n,\overline{T}}}^{(j+1)\Delta_{n,\overline{T}}} 2(X_s - X_{j\Delta_{n,\overline{T}}})\sigma_0^2(X_s)\,\mathrm{d}s + \mathrm{o_p}(1) \\
&\quad= \sqrt{\frac{n}{\overline{T}}}\sum_{j=1}^{[nr]-1} w(X_{j\Delta_{n,\overline{T}}})\int_{j\Delta_{n,\overline{T}}}^{(j+1)\Delta_{n,\overline{T}}} 2\left(\int_{j\Delta_{n,\overline{T}}}^{s}\sigma_0(X_s)\,\mathrm{d}B_s\right)\sigma_0^2(X_s)\,\mathrm{d}s + \mathrm{o_p}(1) \\
&\quad= \sqrt{\frac{n}{\overline{T}}}\sum_{j=1}^{[nr]-1} w(X_{j\Delta_{n,\overline{T}}})2\left(\int_{j\Delta_{n,\overline{T}}}^{(j+1)\Delta_{n,\overline{T}}}\sigma_0^2(X_s)\,\mathrm{d}s\right)\int_{j\Delta_{n,\overline{T}}}^{(j+1)\Delta_{n,\overline{T}}}\sigma_0(X_s)\,\mathrm{d}B_s \\
&\qquad+ \sqrt{\frac{n}{\overline{T}}}\sum_{j=1}^{[nr]-1} w(X_{j\Delta_{n,\overline{T}}})2\left(\int_{j\Delta_{n,\overline{T}}}^{(j+1)\Delta_{n,\overline{T}}}\left(\int_{j\Delta_{n,\overline{T}}}^{s}\sigma_0^2(X_s)\,\mathrm{d}s\right)\sigma_0(X_s)\,\mathrm{d}B_s\right) + \mathrm{o_p}(1) \\
&\quad= qv^1_{n,\overline{T}} + qv^2_{n,\overline{T}} + \mathrm{o_p}(1),
\end{aligned}$$

where the first and the second asymptotic approximations derive from the asymptotic vanishing rate of the diffusion's finite variation component and the penultimate line follows from integration by parts. The term $qv^2_{n,\overline{T}}$ has a variation which could be expressed as

$$\begin{aligned}
&[qv^2_{n,\overline{T}}]_r \\
&\quad\leqslant \frac{n}{\overline{T}}\left(\max_{1\leqslant j\leqslant [nr]-1}\int_{j\Delta_{n,\overline{T}}}^{(j+1)\Delta_{n,\overline{T}}}\sigma_0^2(X_s)\,\mathrm{d}s\right)^2\sum_{j=1}^{[nr]-1} w^2(X_{j\Delta_{n,\overline{T}}})4\left(\int_{j\Delta_{n,\overline{T}}}^{(j+1)\Delta_{n,\overline{T}}}\sigma_0^2(X_s)\,\mathrm{d}s\right) \\
&\quad= \mathrm{O_p}\left(\frac{1}{n}\right)\sum_{j=1}^{[nr]-1} w^2(X_{j\Delta_{n,\overline{T}}})4\left(\int_{j\Delta_{n,\overline{T}}}^{(j+1)\Delta_{n,\overline{T}}}\sigma_0^2(X_s)\,\mathrm{d}s\right) \underset{n\to\infty}{\overset{\mathrm{p}}{\to}} 0.
\end{aligned}$$

(As earlier in the proof of Theorem 4 we abuse notation slightly by squaring the vector weight w.) Hence, $qv^2_{n,\overline{T}} = o_p(1)$. As for $qv^1_{n,T}$, this term is trivially $o_p(1)$ since $\sum_{j=1}^{[nr]-1} w(X_{j\Delta_{n,\overline{T}}}) \int_{j\Delta_{n,\overline{T}}}^{(j+1)\Delta_{n,\overline{T}}} \sigma_0(X_s)\,dB_s$ is bounded in probability and $\sqrt{(n/\overline{T})} \int_{j\Delta_{n,\overline{T}}}^{(j+1)\Delta_{n,\overline{T}}} \sigma_0^2(X_s)\,ds = o_p(1)$ uniformly in j. Finally, the covariation process between $\sqrt{(1/\Delta_{n,\overline{T}})}\mathbf{B}_{n,\overline{T}}$ and $\beta_{n,\overline{T}}$ is zero since X_0 is independent of the Brownian path and the Brownian increments are independent of each other. This proves the stated result. □

Appendix B. Notation

$\xrightarrow{\text{a.s.}}$	almost sure convergence
$\xrightarrow{\text{p}}$	convergence in probability
$\Rightarrow, \xrightarrow{\text{d}}$	weak convergence
$\underset{n,\bar{n},T\to\infty}{\to}, \underset{n,\bar{n},T\to\infty}{\Rightarrow}$	convergence, weak convergence with $\bar{T}$ fixed and $T \to \infty$
$\underset{n\to\infty}{\to}, \underset{n\to\infty}{\Rightarrow}$	convergence, weak convergence with $\bar{T} = T$ fixed
$\underset{n,\overline{T}\to\infty}{\Rightarrow}$	weak convergence with $\bar{T} = T \to \infty$
$:=$	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
$\overset{d}{=}$	distributional equivalence
$MN(0, V)$	mixed normal distribution with variance V
$C_k, \quad k = 1, 2, \ldots$	constants
$[X]_t$	quadratic variation of X at t

References

Aït-Sahalia, Y., 1996. Testing continuous-time models of the spot interest rate. Review of Financial Studies 2, 385–426.

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

Altissimo, F., Mele, A., 2003. Simulated nonparametric estimation of continuous-time models of asset prices and returns. Working paper.

Andrews, D.W.K., 1989. Asymptotics for semiparametric econometric models: III. Testing and examples. Cowles Foundation working paper No. 910, Yale University.

Bandi, F.M., Moloche G., 2001. On the functional estimation of multivariate diffusions. Working paper.

Bandi, F.M., Nguyen T., 1999. Fully nonparametric estimators for diffusions: a small sample analysis. Working paper.

Bandi, F.M., Nguyen, T., 2003. On the functional estimation of jump-diffusion processes. Journal of Econometrics 116, 293–328.

Bandi, F.M., Phillips, P.C.B., 2002. Nonstationary continuous-time models. in: Aït-Sahalia Y., Hansen, L.P. (Eds.), Handbook of Financial Econometrics, forthcoming.
Bandi, F.M., Phillips, P.C.B., 2003. Fully nonparametric estimation of scalar diffusion models. Econometrica 71, 241–283.
Bandi, F.M., Russell, J.R., 2005. Volatility. In: Birge J.R., Linetsky V. (Eds.), Handbook of Financial Engineering, forthcoming.
Bickel, P.J., Rosenblatt, M., 1973. On some global measures of the deviations of density function estimates. Annals of Statistics 1, 1071–1095.
Bosq, D., 1998. Nonparametric Statistic for Stochastic Processes. Springer, New York.
Brandt, M.W., Santa-Clara, P., 2002. Simulated likelihood estimation of diffusions with an application to exchange rate dynamics in incomplete markets. Journal of Financial Economics 63, 161–210.
Carrasco, M., Chernov, M., Florens J.P., Ghysels, E., 2002. Estimation of jump-diffusions with a continuum of moment conditions. Working paper.
Chacko, G., Viceira, L., 2003. Spectral GMM estimation of continuous-time processes. Journal of Econometrics 116, 259–292.
Chiang, G.L., 1956. On regular best asymptotically normal estimates. Annals of Mathematical Statistics 27, 336–351.
Chung, K.L., Williams, R.J., 1990. Introduction to Stochastic Integration. Birkhäuser, Boston.
Conley, T., Hansen, L.P., Luttmer, E., Scheinkman, J., 1997. Short-term interest rates as subordinated diffusions. Review of Financial Studies 10, 525–577.
Corradi, V., Swanson, N.R., 2005. Bootstrap specification tests for diffusion processes. Journal of Econometrics 124, 117–148.
Corradi, V., White, H., 1999. Specification tests for the variance of a diffusion. Journal of Time Series Analysis 20, 253–270.
Darling, D.A., Kac, M., 1957. On occupation times for Markoff processes. Transactions of the American Mathematical Society 84, 444–458.
Davidson, R., MacKinnon, J.G., 1993. Estimation and Inference in Econometrics. Oxford University Press, New York.
Duffie, D., Glynn, P., 2004. Estimation of continuous-time Markov processes sampled at random time intervals. Econometrica 72, 1773–1808.
Durham, G., Gallant, R., 2002. Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes. Journal of Business and Economic Statistics 20, 279–316.
Elerian, O., Chib, S., Shephard, N., 2001. Likelihood inference for discretely observed nonlinear diffusions. Econometrica 69, 959–1012.
Eraker, B., 2001. MCMC analysis of diffusion models with application to finance. Journal of Business and Economic Statistics 19, 177–191.
Fan, Y., 1994. Testing the goodness of fit of a parametric density function by kernel methods. Econometric Theory 10, 316–356.
Ferguson, T.S., 1958. A method of generating best asymptotically normal estimates with application to the estimation of bacterial densities. Annals of Mathematical Statistics 29, 1046–1062.
Florens-Zmirou, D., 1993. On estimating the diffusion coefficient from discrete observations. Journal of Applied Probability 30, 790–804.
Gallant, A.R., Tauchen, G., 1996. Which moments to match? Econometric Theory 12, 657–681.
Gouriéroux, C., Monfort, A., Renault, E., 1993. Indirect inference. Journal of Applied Econometrics 8S, 85–118.
Hansen, L.P., Scheinkman, J., 1995. Back to the future: generating moment implications for continuous-time Markov processes. Econometrica 63, 767–804.
Hong, Y., Li, H., 2003. Nonparametric specification testing for continuous-time models with application to spot interest rates. Review of Financial Studies 18, 37–84.
Jiang, G.J., Knight, J., 1999. Finite sample comparison of alternative estimators of Itô diffusion processes: a Monte-Carlo study. Journal of Computational Finance 2, 5–38.
Jiang, G.J., Knight, J., 2002. Estimation of continuous-time processes via the empirical characteristic function. Journal of Business and Economic Statistics, 198–212.
Karatzas, I., Shreve, S.E., 1991. Brownian Motion and Stochastic Calculus. Springer, New York.
Karlin, S., Taylor, H.M., 1981. A Second Course in Stochastic Processes. Academic Press, New York.
Koul, H.L., 1992. Weighted Empiricals and Linear Models. Institute of Mathematical Statistics, Hayward, CA.

Lo, A.W., 1988. Maximum likelihood estimation of generalized Ito processes with discretely sampled data. Econometric Theory 4, 231–247.

Maddala, G.S., Rao, C.R. (Eds.), 1997. Handbook of Statistics 15: Robust Inference. North-Holland, Amsterdam.

Meyn, S.P., Tweedie, R.L., 1993. Markov Chains and Stochastic Stability. Springer, London.

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

Park, J., Phillips, P.C.B., 2001. Nonlinear regressions with integrated time series. Econometrica 69, 117–161.

Pritsker, M., 1998. Nonparametric density estimation and tests of continuous-time interest rate models. Review of Financial Studies 11, 449–487.

Protter, P., 1995. Stochastic Integration and Differential Equations: A New Approach. Springer, Berlin.

Revuz, D., Yor, M., 1994. Continuous Martingales and Brownian Motion. Springer, Berlin.

Rosenblatt, M., 1975. Quadratic measure of deviation of two-dimensional density estimates and a test of independence. Annals of Statistics 3, 1–14.

Singleton, K., 2001. Estimation of affine pricing models using the empirical characteristic function. Journal of Econometrics 102, 111–141.