CrossMark

Econometric Theory, **34**, 2018, 896–946.
doi:10.1017/S0266466617000305

ON THE FUNCTIONAL ESTIMATION OF MULTIVARIATE DIFFUSION PROCESSES

FEDERICO M. BANDI
Johns Hopkins University and Edhec-Risk Institute
GUILLERMO MOLOCHE
fermat.ai

We propose a nonparametric estimation theory for the occupation density, the drift vector, and the diffusion matrix of multivariate diffusion processes. The estimators are sample analogues to infinitesimal conditional expectations constructed as Nadaraya-Watson kernel averages. Mild assumptions are imposed on the statistical properties of the multivariate system to obtain limiting results. Harris recurrence is all that we require to show consistency and asymptotic (mixed) normality of the proposed functional estimators. The identification method and asymptotic theory apply to both stationary and nonstationary multivariate diffusion processes of the recurrent type.

1. INTRODUCTION

The estimation of stochastic differential equations, often conducted in conjunction with the study of valuation models for derivative securities, has drawn substantial attention in recent years. Parametric, semiparametric, and nonparametric estimation methods for scalar diffusion processes are now well established under a variety of assumptions on the statistical properties of the underlying continuous-time series (see, e.g., the review papers by Aït-Sahalia, Hansen, and Scheinkman, 2010, Bandi and Phillips, 2010, Cai and Hong, 2003, Fan, 2005, Gallant and Tauchen, 2010, and Johannes and Polson, 2010, for discussions).

Increasing the dimensionality of the system poses substantial complications to the econometrics of continuous-time models. While existing methods allow us to deal rather efficiently with involved *parametric* specifications for multivariate diffusions (see, e.g., Gallant and Tauchen, 2010, and the references therein), less progress has been made in the *nonparametric* estimation of multidimensional specifications for continuous-time models of the diffusion type. Such a

We are especially grateful to Valentina Corradi for helpful discussions. We thank Xiaohong Chen, the Editor Peter C.B. Phillips, and four anonymous referees for their very useful comments. Seminar participants at various institutions and conferences have also provided suggestions for which we are thankful. Bandi acknowledges financial support from the IBM Corporation Faculty Research Fund at Chicago Booth, University of Chicago, and from Carey Business School, Johns Hopkins University. Address correspondence to Federico Bandi, Johns Hopkins Carey Business School, 100 International Drive, Baltimore, MD 21202, USA; e-mail: fbandi1@jhu.edu.

© Cambridge University Press 2017

https://doi.org/10.1017/S0266466617000305 Published online by Cambridge University Press

development appears to be important in view of the robustness to potential misspecifications offered by fully functional estimation methods as well as their descriptive power and usefulness in building more accurate parametric models.

As is well known, the dynamic evolution of a multivariate diffusion process depends on the form of its drift vector and diffusion matrix. Importantly, the drift vector and the diffusion matrix have infinitesimal (first and second) conditional moment definitions which lend themselves to nonparametric kernel estimation. Sample analogues to the infinitesimal first and second moment can therefore be constructed by employing, among other methods, classical Nadaraya-Watson kernel estimators. To this extent, this article derives the limiting properties of Nadaraya-Watson kernel estimators of the drift vector and diffusion matrix of a multivariate diffusion process under Harris recurrence. Harris recurrence is known to be a milder assumption than stationarity and mixing (see, e.g., Meyn and Tweedie, 1993). Intuitively, it solely requires the continuous trajectory of the process to visit sets of nonzero Lebesgue measure in its admissible range an infinite number of times over time. Thus, it represents a mild condition for local identification, as we show formally below. Harris recurrent processes may be strictly stationary, stationary in the limit (ergodic or positive recurrent) or nonstationary (null recurrent).

The asymptotic behavior of the drift and diffusion estimators is examined as the observation frequency increases (infill asymptotics) and as the time span lengthens (long span asymptotics) *jointly*. We prove consistency of the functional estimates and convergence to mixtures of normal laws, where the mixing variates depend on a random object which drives the convergence rates of the functional estimates and whose divergence properties depend on the specific process being considered. Such a random object is, in general, *not* a chronological local time as in the scalar diffusion case examined elsewhere (Bandi and Phillips, 2003, BP henceforth, and Aït-Sahalia and Park, 2016). Nevertheless, it can be interpreted as an estimate of the density of the occupation time measure of the underlying process, where the latter represents the amount of time spent by the process in a certain spatial set of nonzero Lebesgue measure. Unlike the concept of local time, the notion of occupation density extends to multivariate frameworks (see, e.g., Geman and Horowitz, 1980). Of course, when the dimensionality of the problem collapses to one, the random object driving the rates of convergence of the nonparametric estimates is, indeed, a chronological local time.

From a technical standpoint, the nonexistence of a notion of local time for multivariate continuous semimartingales represents a considerable theoretical difficulty to overcome when studying nonparametric kernel estimation for multidimensional diffusion processes under potential nonstationarities. This article provides a solution to this problem while offering additional insights about the scalar diffusion case which, as discussed, may be viewed as a subcase of our more general theory of inference.

Relaxing stationarity for identification is theoretically and empirically important. We show that the dimensionality of the problem has a two-fold effect on the

https://doi.org/10.1017/S0266466617000305 Published online by Cambridge University Press

rate of convergence of the functional estimates when the system is nonstationary. The first effect is coherent with the conventional "curse of dimensionality" resulting from the estimation of conditional expectations in multivariate discrete-time frameworks under stationarity. The second effect operates through the random quantity which drives the rates of convergence of the functional estimates and, as emphasized earlier, may be interpreted as an estimate of the density of the occupation measure of the process. This quantity inherits the divergence rate of the process' occupation measure. The occupation measure is known to diverge linearly with the time span when examining positive Harris recurrent Markov processes of any dimension. However, the dimensionality of the system affects negatively its divergence rate in the null recurrent (nonstationary) case, thereby delivering an additional curse of dimensionality which is a genuine by-product of the mildness of our assumptions. We call this double effect "double curse of dimensionality." We emphasize that the double curse of dimensionality is not specific to the use of functional methods for multivariate, potentially nonstationary, diffusion processes. We expect the same effect to arise from the functional estimation of multivariate, nonstationary processes in discrete time should the adopted asymptotic theory allow for general recurrent dynamics.

Research on the use of recurrence as an identifying assumption appears warranted in discrete time as well. It is, as in the continuous-time case, still in its infancy. Fundamental progress in discrete time, both in the context of autoregressive processes—as in this article—and in the context of cointegrating models—has been made by, e.g., Guerre (2004), Karlsen and Tjøstheim (2001), Karlsen, Myklebust, and Tjøstheim (2007), Moloche (2004b), and Schienle (2008). We refer the reader to Wang and Phillips (2009a, 2009b), and the references therein, for a promising alternative approach based on Skorohod embedding and nonlinear transformations of the embedded process. Wang (2015) provides a rich account of existing results and new perspectives.

Previous work on the functional estimation of multivariate diffusion processes has largely focused on the diffusion matrix. Brugiére (1991) extends the nonparametric estimator of the second infinitesimal moment suggested by Florens-Zmirou (1993) in the scalar case to a system of diffusions and provides a proof of consistency in probability (see, also, Genon-Catalot and Jacod, 1993). In follow-up work, Brugiére (1993) derives the limiting distribution of his estimator and shows asymptotic normality. Importantly, the methodology in Brugiére (1991,1993) does not rest on stationarity. However, his limiting results are derived using increasing frequencies over a fixed span of data. Hence, the methods cannot be extended to drift estimation since identification of the drift necessitates an asymptotically enlarging data span. Boudoukh, Downing, Richardson, Stanton, and Whitelaw (2007) generalize the univariate procedure in Stanton (1997) to propose nonparametric kernel analogues to drift and diffusion matrices for multidimensional diffusions. The asymptotic properties of their Nadaraya-Watson-style estimators are not discussed, thereby rendering statistical inference difficult to implement and interpret in their framework.

https://doi.org/10.1017/S0266466617000305 Published online by Cambridge University Press

In sum, when dealing with multidimensional diffusion processes, a complete theory of inference for traditional Nadaraya-Watson kernel estimates of drift and diffusion function under discreteness of the observation sampling has yet to be established, to the best of our knowledge, even in the stationary case. We contribute to the literature by studying nonparametric identification of the occupation density, drift, and diffusion of stationary and nonstationary multidimensional diffusion processes in the recurrent class.

The article is organized as follows. Section 2 introduces the model and the nonparametric estimates. Section 3 discusses Harris recurrence for multidimensional diffusion processes. Section 4 reports important preliminaries about Harris recurrent processes. These results are used in the development and interpretation of our limit theory. In Section 5 we discuss our asymptotic findings. Section 6 concludes. Proofs and technical details are in the Appendix. In what follows, the symbols $\Rightarrow$, $\xrightarrow{a.s.}$, $\xrightarrow{p}$ and $\overset{d}{=}$ stand for weak convergence, convergence with probability one, convergence in probability, and distributional equivalence, respectively. For two real-valued positive sequences a_T and b_T, the notation $a_T \sim b_T$ means that $a_T = O(b_T)$. When applied to a generic matrix A the operator *vec* stacks the columns of A. With a squared matrix A, we use the notation $A^{1/2}$ to signify $E\Lambda^{1/2}E^{-1}$, where E contains columns of eigenvectors and Λ is a diagonal matrix of eigenvalues. The symbols $\otimes$ and $\mathbf{1}_B$ denote the Kronecker product and the indicator function of the set B, respectively.

2. DESCRIPTION OF THE MODEL AND ESTIMATORS

Consider the probability space $(\Omega, \mathcal{F}, P)$, the filtration of sub-σ-fields $\mathcal{F}_t$, and the continuous adapted process $\{X_t, \mathcal{F}_t; 0 \le t < \infty\}$ with

$$X_t = X_0 + \int_0^t \boldsymbol{\mu}(X_s)\, ds + \int_0^t \boldsymbol{\sigma}(X_s)\, \mathrm{d}\mathbf{B}_s, \tag{1}$$

where X_0 is a given initial condition, $\mathbf{B} = \{B_t, \mathcal{F}_t; 0 \le t < \infty\}$ is an m-dimensional standard Brownian motion, $\boldsymbol{\mu}(.) = \left\{\mu_j(.)\right\}_{1 \le j \le d}$ is a $d \times 1$ Borel measurable drift vector and $\boldsymbol{\sigma}(.) = \left\{\sigma_{jk}(.)\right\}_{\substack{1 \le j \le d \\ 1 \le k \le m}}$ is a $d \times m$ Borel measurable matrix. Assume X_0 is independent of $\mathbf{B}$ and X_t takes values in $I \subseteq \mathcal{R}^d$. Each coordinate X_t^j of the process can be written as

$$X_t^j = X_0^j + \int_0^t \mu_j(X_s)\, ds + \sum_{k=1}^{m} \int_0^t \sigma_{jk}(X_s)\, \mathrm{d}B_s^k, \qquad 0 \le t < \infty, 1 \le j \le d. \tag{2}$$

Define the $d \times d$ symmetric and nonnegative (diffusion) matrix $\mathbf{a}(x) = \boldsymbol{\sigma}(x)\boldsymbol{\sigma}(x)^{\mathsf{T}}$ with generic element $a_{jp}(x) = \sum_{k=1}^{m} \sigma_{jk}(x)\sigma_{pk}(x)$ $1 \le j \le d, 1 \le p \le d, \forall x \in I \subseteq \mathcal{R}^d$. Write the expectation (probability) conditional on $X_0 = x$, where $x = (x^1, x^2, \ldots, x^d)$ is a d-dimensional initial condition, as E^x (P^x). The drift vector $\boldsymbol{\mu}(.)$ and the diffusion matrix $\mathbf{a}(.)$ have classical representations in terms of

https://doi.org/10.1017/S0266466617000305 Published online by Cambridge University Press

infinitesimal conditional moments, i.e.,

$$\mathrm{E}^x\left[X_t^j - x^j\right] = t\mu_j(x) + o(t),$$
$$\mathrm{E}^x\left[\left(X_t^j - x^j\right)\left(X_t^p - x^p\right)\right] = ta_{jp}(x) + o(t),$$

as $t \downarrow 0$ (see, e.g., Karatzas and Shreve, 1991, page 282).

Now, assume the process $\{X_t : t \geq 0\}$ is sampled at equispaced times $\{t = t_1, t_2, \ldots, t_n\}$ in the interval $[0, T]$. It readily follows that $\{X_t = X_{\Delta_{n,T}}, X_{2\Delta_{n,T}}, X_{3\Delta_{n,T}}, \ldots, X_{n\Delta_{n,T}}\}$ are n observations on the process X_t 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 $\Delta_{n,T} = T/n$.

We estimate the drift vector $\boldsymbol{\mu}(x)$ and the diffusion matrix $\mathbf{a}(x)$ $\forall x \in I \subseteq \mathcal{R}^d$ by employing

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

and

$$\widehat{\mathbf{a}}_{n,T}(x) = \frac{1}{\Delta_{n,T}} \frac{\sum_{i=1}^{n-1} \mathbf{K}_{h_{n,T}}\left(X_{i\Delta_{n,T}} - x\right)\left(X_{(i+1)\Delta_{n,T}} - X_{i\Delta_{n,T}}\right)\left(X_{(i+1)\Delta_{n,T}} - X_{i\Delta_{n,T}}\right)^{\mathsf{T}}}{\sum_{i=1}^{n} \mathbf{K}_{h_{n,T}}\left(X_{i\Delta_{n,T}} - x\right)}, \tag{4}$$

where $\mathbf{K}_h(X_{i\Delta_{n,T}} - x) = \frac{1}{h^d}\prod_{j=1}^{d} K\left(\frac{X^j_{i\Delta_{n,T}} - x^j}{h}\right)$ is a product kernel function and $h_{n,T}$ is a bandwidth sequence. The properties of the kernel $K(.)$ are spelled out in Assumption 2 below.

The estimators in equations (3) and (4) belong to the general class of Nadaraya-Watson kernel estimators (see, e.g., Pagan and Ullah, 1999). They are multidimensional counterparts of those discussed in BP (2003). In light of the absence of a notion of local time and corresponding theory, our analysis of the more compelling, from an empirical and theoretical standpoint, multivariate case poses technical complications which, as said, were absent in the scalar case.

Our limiting results will yield consistency and asymptotic mixed normality of the estimates in an asymptotic design which lets the time span increase without bound ($T \to \infty$) with a distance between observations going to zero ($\Delta_{n,T} = T/n \to 0$). The former assumption (long span asymptotics) is necessary for drift estimation. The latter (infill asymptotics) is key to approximate the continuous trajectory of the process with a sample of discretely sampled observations while replicating the infinitesimal properties of the relevant moments. This assumption is important for nonparametric identification in the absence of a time-invariant stationary density. All limits are *joint* limits over $T, n \to \infty$ with $\Delta_{n,T} = T/n \to 0$.

We view our asymptotic design as a realistic approximation in fields, such as finance, in which data sets comprise observations sampled at relatively high frequencies over sufficiently long spans of time. To this extent, several simulation

https://doi.org/10.1017/S0266466617000305 Published online by Cambridge University Press

studies (see, e.g., Jiang and Knight, 1999) have shown that daily data, for instance, are sensible approximations to frequent observations for nonparametric estimators relying on frequent observations. Higher than daily frequencies are also now available (in finance, for instance), albeit over generally shorter time spans. The use of intradaily data, however, poses microstructure-related issues which are beyond the scope of the present article.

3. HARRIS RECURRENCE

This section introduces conditions on the underlying continuous-time process. In particular, we emphasize *recurrence*, a property such that the path of the process of interest returns to open sets in its range an infinite number of times over time, regardless of the initial condition. Specifically, let A be a measurable set of $I \subseteq \mathcal{R}^d$ and define $\tau_{\overline{A}} = \inf\left\{t \geq 0 : X_t \in \overline{A}\right\}$, i.e., the first hitting time of the closure $\overline{A}$ of the set A. The process X_t is null Harris recurrent if $P^x[\tau_{\overline{A}} < \infty] = 1$ for every $x \in I \backslash \overline{A}$. The process X_t is positive Harris recurrent if $\mathrm{E}^x[\tau_{\overline{A}}] < \infty$ for every $x \in I \backslash \overline{A}$.

We now turn to formal assumptions. Under Assumption 1(b), the process is Harris (null or positive) recurrent. Assumption 1(c) implies positive recurrence (ergodicity) and simply strengthens Assumption 1(b). While Assumption 1(c) is not necessary, it will be interesting to specialize our results to the more familiar case of positive recurrent or strictly stationary processes.

Assumption 1.

(a) $\boldsymbol{\mu}(.)$ and $\boldsymbol{\sigma}(.)$ are time-homogeneous, $\mathcal{B}$-measurable functions on $I \subseteq \mathcal{R}^d$, where $\mathcal{B}$ is the σ-field generated by Borel sets on I. They are such that a strong solution to the d-dimensional process $\{X_t : t \geq 0\}$ in equation (1) exists and is unique up to null sets.

(b) (Recurrence). Assume that, for every open and bounded set $A \subset I$,

$$\min_{x \in \overline{A}} a_{jj}(x) > 0$$

for some $1 \leq j \leq d$. Define the second-order elliptic operator

$$\mathcal{L}\varphi(.) = \sum_{j=1}^{d} \mu_j(.) \frac{\partial \varphi(.)}{\partial x^j} + \frac{1}{2} \sum_{j=1}^{d} \sum_{p=1}^{d} a_{jp}(.) \frac{\partial^2 \varphi(.)}{\partial x^j \partial x^p}.$$

There is a function $\varphi(.) : \mathcal{R}^d \backslash \{0\} \to \mathcal{R}$ of class C^2 (i.e., twice continuously differentiable) in the domain of the operator that satisfies

$$\mathcal{L}\varphi(.) \leq 0 \qquad \text{on } \mathcal{R}^d \backslash \{0\}$$

and is such that $\Psi(r) := \min_{\|x\|=r} \varphi(x)$ is strictly increasing with $\lim_{r \to \infty} \Psi(r) = \infty$ (c.f. Karatzas and Shreve, 1991, Exercise 7.13, part (i), page 371).

(c) (Positive recurrence). There is a function $\varphi(.) : \mathcal{R}^d \backslash \{0\} \to \mathcal{R}$ of class C^2 in the domain of the operator that satisfies

https://doi.org/10.1017/S0266466617000305 Published online by Cambridge University Press

$$\mathcal{L}\varphi(.) \leq -1 \qquad \text{on } \mathcal{R}^d \backslash \{0\}$$

and is such that $\Psi(r) := \min_{\|x\|=r} \varphi(x)$ is strictly increasing with $\lim_{r\to\infty} \Psi(r) = \infty$ (c.f. Karatzas and Shreve, 1991, Exercise 7.13, part (ii), page 371).

Assume $X^{(x)}$ is the unique strong solution of equation (1) with initial condition $X_0^{(x)} = x \in I \subseteq \mathcal{R}^d$, then the measure ϕ is invariant for equation (1) if, and only if,

$$\phi(A) = \int_I P\left(X_t^{(x)} \in A\right) \phi(dx) \quad \forall A \in \mathcal{B}(I)$$

for every $0 \leq t < \infty$ (see, e.g., Karatzas and Shreve, 1991, Exercise 6.18, page 362). Harris recurrence is a sufficient condition for the existence of a σ-finite invariant measure. This measure is unique up to multiplication by a constant. If the invariant measure can be normalized to a probability measure (see equation (6) below), then we say that the process is positive Harris recurrent as implied by Assumption 1(c) above. Otherwise, the process is null Harris recurrent (and nonstationary).

For illustration, consider the scalar case ($d = 1$) with range $I = (l, u)$ such that $-\infty \leq l < u \leq \infty$. The "speed measure," i.e.,

$$m(\mathrm{d}x) = \frac{2\mathrm{d}x}{v'(x)\sigma^2(x)}, \qquad \forall x \in I \subseteq \mathcal{R}, \tag{5}$$

where $v'(x)$ is the first derivative of the "scale function"

$$v(x) = \int_c^x \exp\left\{-2\int_c^\xi \frac{\mu(\varepsilon)}{\sigma^2(\varepsilon)}\mathrm{d}\varepsilon\right\}\mathrm{d}\xi, \qquad \forall x \in I \subseteq \mathcal{R}$$

is the unique invariant measure for some $c \in I$. Under positive recurrence, the process admits a time-invariant probability measure, to which it converges, and the normalized speed measure, i.e., $m(\mathrm{d}x)/m(I) = p(\mathrm{d}x)$, is the time-invariant probability measure of X, namely

$$\lim_{t\to\infty} P^x(X_t < y) = \frac{m((l,y))}{m(I)} \qquad \forall x, y \in I \subseteq \mathcal{R}, \tag{6}$$

c.f., Karatzas and Shreve (1991, Exercise 5.40, page 353).

Scalar Brownian motion and Brownian motion on the plane are classical examples of univariate and bivariate null Harris recurrent diffusion processes. In higher dimensions Brownian motion is not recurrent. Hence, while Harris recurrence might not apply to certain highly dimensional nonstationary systems, it does of course apply to all strictly stationary or ergodic systems regardless of their linearity properties and dimensionality. The well-known class of multivariate *affine* diffusions $\left(\text{linear drift vector } \boldsymbol{\mu}(x) \text{ and linear diffusion matrix } \mathbf{a}(x)\right)$ is, trivially, positive recurrent under standard assumptions.

https://doi.org/10.1017/S0266466617000305 Published online by Cambridge University Press

With the exception of Brownian motion in one and two dimensions, examples of null recurrent scalar and multivariate diffusion processes are, admittedly, conspicuously absent from the econometric literature. Similarly, the existing discrete-time literature has a tendency to solely present examples based on unit-root processes. An interesting exception for our purposes is Chen and Hansen (2002). They provide an example of a multidimensional diffusion displaying various recurrence properties depending on the relation between the dimensionality of the system and the parameters of the invariant measure and diffusion function. Specifically, assume $\phi(\mathrm{d}x) = c_1(1+\|x\|^2)^{-\vartheta}\mathrm{d}x$ and $\mathbf{a}(x) = c_2(1+\|x\|^2)^{\theta}\mathbf{I}_d$. Then, the d-dimensional diffusion is null Harris recurrent if $\frac{d}{2} \geq \vartheta \geq \theta - 1 + \frac{d}{2} \geq -1 + \frac{d}{2}$. It is positive Harris recurrent if $\vartheta > \frac{d}{2}$ and $\vartheta \geq \theta - 1 + \frac{d}{2}$. This is a case of volatility-induced stationarity (Conley, Hansen, Luttmer, and Scheinkman, 1997).

As discussed, Harris recurrence is a weaker assumption than stationarity and mixing (see, e.g., Meyn and Tweedie, 1993). Intuitively, point-wise identification requires return of the sample path of the process to local neighborhoods. This is precisely what Harris recurrence yields.

4. PRELIMINARIES ABOUT HARRIS RECURRENT PROCESSES

We now present two theorems which are central—in their various forms—to the analysis of Harris recurrent processes. Both theorems apply to Harris recurrent continuous-time Markov processes potentially more general than multivariate diffusion processes.

THEOREM 1 (The ratio-limit theorem). *Consider the continuous-time Markov process X_t defined on the filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, P)$. Assume X_t is Harris recurrent with invariant measure ϕ. Then, for any Borel measurable pair of functions $f(.)$ and $g(.)$ which are integrable with respect to ϕ, the ratio of the functionals $\int_0^T f(X_s)\mathrm{d}s$ and $\int_0^T g(X_s)\mathrm{d}s$ is so that*

$$P^x\left(\lim_{T\to\infty} \frac{\int_0^T f(X_s)\mathrm{d}s}{\int_0^T g(X_s)\mathrm{d}s} = \frac{<\phi, f>}{<\phi, g>}\right) = 1, \tag{7}$$

provided $<\phi, g> = \int g(x)\phi(\mathrm{d}x) > 0$.

THEOREM 2 (The Darling-Kac theorem). *Consider the continuous-time Markov process X_t defined on the filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, P)$. Assume X_t is Harris recurrent with a stationary transition density and invariant measure ϕ. For a given nonnegative function $f(.)$ such that $0 < \int f(x)\phi(\mathrm{d}x) < \infty$, we have*

$$v(T) = U(T)T^{\alpha} \sim \mathrm{E}^x\left[\int_0^{\infty} \exp\left\{-\frac{t}{T}\right\} f(X_t)\mathrm{d}t\right] \qquad 0 \leq \alpha \leq 1 \tag{8}$$

so that $U(T)$ is slowly varying as $T \to \infty$.[1] *It follows that*

https://doi.org/10.1017/S0266466617000305 Published online by Cambridge University Press

$$\lim_{T\to\infty} P^x\left(\frac{1}{C_X v(T)}\int_0^T f(X_s)\mathrm{d}s < u\right) = G_\alpha(u), \tag{9}$$

where C_X is a process-specific constant, and $G_\alpha(u)$ is the cumulative distribution function of a Mittag-Leffler random variable g_α, i.e., a random variable with moments

$$\mathrm{E}\left(\mathrm{g}_\alpha^r\right) = \frac{r!}{\Gamma(1+r\alpha)}, \qquad r \geq 0,$$

where $\Gamma(.)$ is the gamma function.

Remark 1. Theorem 1 can be interpreted as an ergodic theorem for potentially nonstationary continuous-time Markov processes. Integrals with respect to the invariant measure replace standard integrals with respect to the process' time-invariant probability density. The interested reader is referred to Azéma, Kaplan-Duflo, and Revuz (1966) for additional details.

Remark 2. Theorem 2 provides a weak convergence result for additive functionals of potentially nonstationary continuous-time Markov processes. For $\alpha = 0$, the Mittag-Leffler density becomes the exponential density with parameter 1 and $G_0(u) = 1 - e^{-u}$ with $u \geq 0$. In fact, $\mathrm{E}\left(\mathrm{g}_0^r\right) = \frac{r!}{\Gamma(1)} = r!$ defines its moments. For $\alpha = \frac{1}{2}$, the Mittag-Leffler density corresponds to the truncated (to positive draws) normal density with sigma $\sqrt{2}$ and $G_{\frac{1}{2}}(u) = \frac{1}{\sqrt{\pi}}\int_0^u e^{-y^2/4}\mathrm{d}y$ with $u \geq 0$. In this case, $\mathrm{E}\left(\mathrm{g}_{1/2}^r\right) = \frac{r!}{\Gamma\left(1+\frac{1}{2}r\right)}$. Given the properties of the gamma function, the first two moments, for instance, are $\mathrm{E}\left(\mathrm{g}_{1/2}^1\right) = \frac{1!}{\Gamma\left(\frac{3}{2}\right)} = \frac{2}{\sqrt{\pi}}$ and $\mathrm{E}\left(\mathrm{g}_{1/2}^2\right) = \frac{2!}{\Gamma(2)} = 2$, thereby yielding a variance equal to $2\left(1-\frac{2}{\pi}\right)$.

Remark 3. If $f(.)$ is the indicator function applied to the generic set A, then the additive functional $\int_0^T f(X_s)\mathrm{d}s$ defines the occupation time of the set A, i.e.,

$$\mathbb{L}_X(T,A) = \int_0^T \mathbf{1}_{\{X_s\in A\}}\mathrm{d}s.$$

The Darling-Kac theorem provides the asymptotic distribution of the normalized occupation times of (possibly multivariate) Harris recurrent Markov processes over measurable sets. Under Harris (positive or null) recurrence, $\mathbb{L}_X(T,A) \xrightarrow{p} \infty$ $\forall A \subset I$ since the number of visits made to each set A by these processes diverges to infinity as the time span enlarges. The probability rate of divergence of the occupation time $\mathbb{L}_X(T,.)$ is given by the features of the underlying process through the function $v(.)$. This rate, and the corresponding limiting distribution, are known in closed-form only for a few processes. In the scalar Brownian motion case, $\alpha = \frac{1}{2}$ and $U(.) = 1$ (i.e., $v(T) = \sqrt{T}$) yielding

$$\lim_{T\to\infty} P^x\left(\frac{1}{C_X\sqrt{T}}\int_0^T \mathbf{1}_{\{X_s\in A\}}\mathrm{d}s < u\right) = G_{\frac{1}{2}}(u) = \frac{1}{\sqrt{\pi}}\int_0^u e^{-y^2/4}\mathrm{d}y, \qquad u \geq 0.^2 \tag{10}$$

https://doi.org/10.1017/S0266466617000305 Published online by Cambridge University Press

In the planar Brownian motion case, $\alpha = 0$ and $U(.) = \log(.)$ (i.e., $v(T) = \log T$) yielding,

$$\lim_{T\to\infty} P^x\left(\frac{1}{C_X \log T}\int_0^T \mathbf{1}_{\{X_s\in A\}}\mathrm{d}s < u\right) = G_0(u) = 1 - e^{-u}, \qquad u \geq 0. \tag{11}$$

The weak convergence results in equations (10) and (11) are versions of the Kallianpur-Robbins Theorem (Kallianpur and Robbins, 1953).

Remark 4. If the process X_t is positive Harris recurrent or strictly stationary, then we obtain the degenerate case $\alpha = 1$ and $U(.) = 1$ (i.e., $v(T) = T$) yielding a weak ergodic theorem:

$$\frac{1}{T}\int_0^T f(X_s)\mathrm{d}s \xrightarrow{p} \int_{-\infty}^{\infty} f(x)p(\mathrm{d}x), \tag{12}$$

where $m(\mathrm{d}x)/m(I) = p(\mathrm{d}x)$ is the time-invariant probability measure of X_t. This result can be strengthened to almost-sure convergence using methods analogous to those leading to Theorem 1. Equation (12) shows that additive functionals of positive Harris recurrent (and, of course, strictly stationary) processes increase like T (see, e.g., Revuz and Yor, 1998, page 409). While positive recurrent processes have occupation times that increase linearly with T regardless of their dimension, the dimensionality of the system affects, in general, the divergence rates of the occupation times of a null recurrent process, as shown in the previous remark. This observation will be important to understand the convergence properties of the proposed functional estimates.

Remark 5. The limit theory for nonstationary continuous-time Harris recurrent processes depends on occupation times which, when appropriately rescaled by the unknown function $v(T)$, tend to a Mittag-Leffler random variable, in general. In agreement with this result, we will show that the martingale terms of the estimation error decompositions of the estimators in equations (3) and (4) converge (stably) to a mixture Gaussian distribution with a random Mittag-Leffler variance. In practice, the scaling, which depends on the Mittag-Leffler parameter α—and, hence, on the recurrence properties of the process—is unknown. The form of the estimators is, however, such that one can, instead, use a data-driven scaling. Since this data-driven scaling converges to the same Mittag-Leffler density, the martingale terms of the estimation error decompositions, and the limiting distributions of the estimators, will be shown to follow a classical normal distribution.

5. ASYMPTOTICS

Before discussing our limiting results, we impose more structure on the model. We begin, however, with assumptions on the kernel function $\mathbf{K}(.)$ appearing in the definition of the estimators in equations (3) and (4).

https://doi.org/10.1017/S0266466617000305 Published online by Cambridge University Press

Assumption 2. The function $\mathbf{K}(x)$ is a product kernel function $\Pi_{j=1}^{d} K(x^j)$. $K(.)$ is a bounded, continuous, and symmetric density defined on a compact set S. The derivatives of the density $K(.)$ are bounded.

Assumption 3.

1. Let $\lim_{T\to\infty} \frac{1}{v(T)} \int_0^T \int_0^T \|\widetilde{g}_{s,t}(a,b)\|_r \mathrm{d}s\mathrm{d}t < \infty$, with $\widetilde{g}_{s,t}(a,b) = \widetilde{p}_{s,t}(a,b) - \widetilde{p}_s(a)\widetilde{p}_t(b)$, where $\widetilde{p}_{s,t}(.,.)$ is the joint density for values of the process at times s and t and $\widetilde{p}_s(.)$ is the marginal density for values of the process at time s, with some $r \in [1,\infty]$, $\forall a,b$.[3]
2. $v(T)h_{n,T}^{2d/r} \to \infty$.
3. $h_{n,T}^{d-2} v(T) \to \infty$.

Assumption 3.1 is a regularity condition controlling dependence in the process sample path. It amounts, effectively, to condition $A'(r)$ in Bosq (1998, page 105). If $r = \infty$, the bandwidth requirement in Assumption 3.2 is trivially met since $v(T) \to \infty$. This is the case if, for example, the process is ergodic ($\widetilde{g}_{s,t} = \widetilde{g}_{|t-s|} = \widetilde{g}_u$) and Castellana and Leadbetter's condition (Castellana and Leadbetter, 1986) is satisfied: $|\widetilde{g}_u(a,b)| \leq \mathrm{M}(u) \in L_1(0,\infty)$ for all a and b. In fact, if $r = \infty$ and the process is ergodic,

$$\lim_{T\to\infty} \frac{1}{T} \int_0^T \int_0^T \|\widetilde{g}_{s,t}(a,b)\|_r \mathrm{d}s\mathrm{d}t = \lim_{T\to\infty} \int_0^T \left(1 - \frac{u}{T}\right) \|\widetilde{g}_u(a,b)\|_\infty \mathrm{d}u$$
$$= \int_0^\infty \|\widetilde{g}_u(a,b)\|_\infty \mathrm{d}u < \infty.$$

It is known that the Castellana and Leadbetter's condition is satisfied by a very large class of ergodic diffusions (Leblanc, 1995, Kutoyants, 1995, and Veretennikov, 1999, among others). In general, the requirement in Assumption 3.2 (i.e., $v(T)h_{n,T}^{2d/r} \to \infty$) links bandwidth choice to the "regularity" of the process' sample path. The more regular the path ($r \to \infty$), i.e., the more locally informative the path, the less binding the bandwidth condition (since, when $r \to \infty$, $v(T)h_{n,T}^{2d/r} \to v(T)$). For $r = 2$, we obtain the "optimal" rate $v(T)h_{n,T}^{d} \to \infty$ (see, e.g., Bosq, 1998, Theorem 4.2, page 94). For $r = 1$, we have $v(T)h_{n,T}^{2d} \to \infty$ (see, e.g., Bosq, 1998, Theorem 4.1, page 91). The "dimensionality" condition in Assumption 3.3 (i.e., $h_{n,T}^{d-2} v(T) \to \infty$) is never binding when $d = 1$ or $d = 2$.

Assumption 4.

1. $\boldsymbol{\mu}(.)$ and $\boldsymbol{\sigma}(.)$ are bounded, with bounded derivatives of any order, and such that a strong Harris recurrent solution to the system exists. The uniform Hörmander condition is met (see, e.g., Kusuoka and Strook, 1985).
2. Regular variation at 0 of resolvents of the diffusion—as in equation (8)—is satisfied.
3. The invariant measure $\phi(\mathrm{d}x)$ admits a continuous Lebesgue density $\widetilde{\phi}(x)$ that is stricly positive everywhere ($\widetilde{\phi}(x) = \frac{\phi(\mathrm{d}x)}{\mathrm{d}x}$).[4]

https://doi.org/10.1017/S0266466617000305 Published online by Cambridge University Press

Since the inferential problem is nonparametric, the boundedness assumption on the coefficients (Assumption 4.1) is not restrictive. The available data are, of course, clearly bounded.[5] Assumption 4.2 allows us to express the *deterministic equivalent*(s) of the process as being proportional to $v(T) = U(T)T^{\alpha}$ and employ limiting results for additive functionals of Harris recurrent processes as in Darling and Kac (1957). Assumptions 4.1, 4.2, and 4.3 imply Assumption 3.1 in Löcherbach and Loukianova (LL, henceforth) (2008). As in LL (2008), we will find it convenient in the proof of Lemma A.1 (see Appendix) to use Nummelin splitting for multidimensional continuous-time processes, a technique which LL (2008) introduced.

As pointed out earlier, all limiting results are *joint* over $T, n \to \infty$ with $\Delta_{n,T} = T/n \to 0$. We start with the convergence properties of the averaged kernel function.

THEOREM 3. *Let Assumptions 1, 2, 3, and 4 be satisfied. Define*

$$\widehat{L}_{n,T}(x) = \Delta_{n,T} \sum_{i=1}^{n} \mathbf{K}_{h_{n,T}}(X_{i\Delta_{n,T}} - x). \tag{13}$$

Let $n, T \to \infty$ with $\Delta_{n,T} \to 0$. Assume $h_{n,T}$ is a deterministic, vanishing sequence so that $\frac{\Delta_{n,T}}{h_{n,T}^2} \to 0$ jointly with $n, T \to \infty$. Then,

$$\frac{\widehat{L}_{n,T}(x)}{v(T)} \Rightarrow C_X \widetilde{\phi}\,(x)\, \mathrm{g}_{\alpha}, \quad \forall x \in I \subseteq \mathcal{R}^d. \tag{14}$$

Remark 6 ($d = 1$)**.** In the univariate case, for which a notion of local time is available, a version of Theorem 3 was derived to link the divergence properties of the averaged kernel function to those of the local time of the process for an increasing time span and vanishing distances between observations (BP, 2003). This result was particularly important since, in the case of scalar recurrent diffusions, the growth rate of $\widehat{L}_{n,T}(x)$ was shown to affect the rate of convergence of the drift and diffusion function estimators (BP, 2003). A couple of examples are in order. If X is Brownian motion (i.e., in the $\frac{1}{2}$-null recurrent situation), then $\alpha = \frac{1}{2}$ (c.f., Remark 3 above) and

$$\widehat{L}_{n,T}(x) = O_p(\sqrt{T}). \tag{15}$$

If X is positive recurrent (or strictly stationary) as in Remark 4, then $\alpha = 1$ and

$$\widehat{L}_{n,T}(x) = O_p(T). \tag{16}$$

Analogous findings were previously discussed by BP (2003) using different arguments. Phillips (2001), Moloche (2004a), and Park (2006) provide interesting discussions of the asymptotic properties of the expected local time and its estimates.

https://doi.org/10.1017/S0266466617000305 Published online by Cambridge University Press

Remark 7 ($d > 1$)**.** In the more general multivariate case, local time is not defined but $\widehat{L}_{n,T}(x)$ is shown, in Theorem 3, to converge to a random process whose rate of divergence to infinity is driven by a deterministic function of time which is regularly varying at infinity. In particular, under suitable bandwidth conditions, the averaged kernel function $\widehat{L}_{n,T}(x)$ has divergence properties which mimic those of the additive functionals of the underlying process (from Theorem 2 above). Similarly to the scalar case presented elsewhere (BP, 2003), this result will prove important when discussing the convergence properties of the functional estimates of the drift vector and diffusion matrix since their rates of convergence will depend on the rate of divergence of $\widehat{L}_{n,T}(x)$.

Remark 8 (The strictly stationary *or* ergodic case)**.** Theorem 3 readily implies that, if the process is positive Harris recurrent or strictly stationary, then

$$\frac{1}{T}\widehat{L}_{n,T}(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{K}_{h_{n,T}}\left(X_{i\Delta_{n,T}} - x\right) \xrightarrow{p} \widetilde{p}(x) \tag{17}$$

with $\widetilde{p}(x) = \frac{p(\mathrm{d}x)}{\mathrm{d}x}$, provided $\frac{\Delta_{n,T}}{h_{n,T}^2} \to 0$ as $h_{n,T} \to 0$ with $n, T \to \infty$ so that $\Delta_{n,T} \to 0$. As expected, if the underlying process is endowed with a time-invariant probability measure, then the standardized (by T) averaged kernel function represents a well-defined density estimator for every dimension. We refer the reader to Castellana and Leadbetter (1986, Theorems 5.1 and 5.2) for a discussion of the strictly stationary ($r = \infty$) case in the presence of a continuous record of observations. Our emphasis is on discretely sampled (with time distance $\Delta_{n,T}$) observations on the process' continuous sample path.

Remark 9. Theorem 3 uses a condition on the vanishing rate of the bandwidth as a function of the divergence rate of the occupation density $\left(v(T)h_{n,T}^{2d/r} \to \infty\right)$ which has not been made completely explicit by the existing literature (see, e.g., the review by Bandi and Phillips, 2010). This condition may often not be binding. It is, in fact, not binding in the ergodic case when $r = \infty$ (c.f., the discussion below Assumption 3).

Remark 10. The discretization condition (namely, $\frac{\Delta_{n,T}}{h_{n,T}^2} \to 0$) is milder than the corresponding condition imposed by, e.g., BP (2003) and Xu (2009), among others, for local time estimation (effectively, $\frac{\Delta_{n,T}\log(1/\Delta_{n,T})\widehat{L}_{n,T}^2(x)}{h_{n,T}^2} \xrightarrow{a.s.} 0$).

5.1. Estimating the Drift Vector

Theorem 4 discusses the consistency and limiting distribution of the drift vector estimator in equation (3).

THEOREM 4 (The asymptotic distribution of the drift vector estimator). *Let Assumptions 1, 2, 3, and 4 be satisfied. Assume the deterministic sequence $h_{n,T}$ is such that $h_{n,T} \to 0$ with $n, T \to \infty$ and $\Delta_{n,T} \to 0$ jointly. If $h_{n,T}^d \widehat{L}_{n,T}(x) \xrightarrow{p} \infty$,*

https://doi.org/10.1017/S0266466617000305 Published online by Cambridge University Press

$h_{n,T}^{d+4}\widehat{L}_{n,T}(x) \xrightarrow{p} C$, and $\frac{\Delta_{n,T}\sqrt{h_{n,T}^{d}\widehat{L}_{n,T}(x)}}{h_{n,T}^{2}} \xrightarrow{p} 0$, where C is a constant, then

$$\sqrt{h_{n,T}^{d}\widehat{L}_{n,T}(x)}\left(\widehat{\boldsymbol{\mu}}_{n,T}(x)-\boldsymbol{\mu}(x)-\boldsymbol{\Gamma}^{\mu}(x)\right)$$

$$\Rightarrow (\mathbf{a}(x))^{1/2}\mathbf{N}\left(\mathbf{0},\left(\int_{\mathrm{S}}K^{2}(u)\,\mathrm{d}u\right)^{d}\mathbf{I}_{d}\right),\quad \forall x\in I\subseteq\mathcal{R}^{d}, \tag{18}$$

with

$$\boldsymbol{\Gamma}^{\mu}(x)=(bias_{1},bias_{2},\ldots,bias_{d})^{\intercal}(x),$$

$$bias_{g}(x)=h_{n,T}^{2}\left(\int_{\mathrm{S}}s^{2}K(s)\mathrm{d}s\right)\left(\sum_{j=1}^{d}\frac{\partial\mu_{g}(x)}{\partial x^{j}}\frac{\frac{\partial\widetilde{\phi}(x)}{\partial x^{j}}}{\widetilde{\phi}(x)}+\frac{1}{2}\sum_{j=1}^{d}\frac{\partial^{2}\mu_{g}(x)}{\partial x^{j}\partial x^{j}}\right),\quad \forall g=1,\ldots,d.$$

Remark 11 (Occupation density and consistency)**.** The quantity $\widehat{L}_{n,T}(x)$ plays the same role here that is played by the number of data points in the more standard estimation of conditional expectations in the stationary, discrete time, context. What matters, in general, is not the speed at which T diverges to infinity but rather the speed at which the occupation time of a set diverges to infinity. The proof of the theorem in the Appendix makes clear that an important assumption for consistency is $h_{n,T}^{d}\widehat{L}_{n,T}(x)\xrightarrow{p}\infty$ (along with the requirement $\frac{\Delta_{n,T}}{h_{n,T}^{2}}\to 0$ for a vanishing discretization error.)

Remark 12 (Bandwidth choice)**.** Theorem 3 implies that $\widehat{L}_{n,T}(x)=O_{p}(v(T))$ under suitable bandwidth conditions. Under these conditions, the bandwidth requirements in Theorem 4 can be restated as follows: $h_{n,T}^{d}v(T)\to\infty$, $h_{n,T}^{d+4}v(T)\to C$ and $\frac{\Delta_{n,T}\sqrt{h_{n,T}^{d}v(T)}}{h_{n,T}^{2}}\to 0$. The asymptotic mean squared error (AMSE) of the drift vector estimator is, therefore, of order $O(h_{n,T}^{4})+O_{p}\left(\frac{1}{v(T)h_{n,T}^{d}}\right)$. Thus, the optimal drift bandwidth in the mean-squared-error sense should be proportional to $v(T)^{-\frac{1}{d+4}}$, where d is—again—the number of equations in the system. The smoothing sequence may be set equal to

$$h_{n,T}^{dr}:=c^{dr}\left(\frac{1}{\log v(T)}\right)v(T)^{-\frac{1}{d+4}}, \tag{19}$$

where c^{dr} is a constant. In principle, this choice would allow one to eliminate the influence of the bias term from the limiting distribution of the drift estimates and obtain centering at zero, while achieving a close-to-optimal speed of convergence.

Four related issues, however, arise. First, choosing c^{dr} is, as in more typical nonparametric problems, cumbersome. Second, $v(T)$ is process specific and unknown, in general. Third, one could replace $v(T)$ with $\widehat{L}_{n,T}(x)$ in equation (19). However, $\widehat{L}_{n,T}(x)$ depends on a bandwidth sequence as well. In fact, in principle,

https://doi.org/10.1017/S0266466617000305 Published online by Cambridge University Press

it depends on the same bandwidth sequence. Thus, the choice variable $h_{n,T}^{dr}$ would appear on both sides of the equation leading to a circularity which may be bypassed in the classical stationary framework (in which the optimal bandwidth is proportional to $T^{-\frac{1}{d+4}}$) but is hard to avoid in nonstationary environments (in which $v(T)$ is generally unknown). Bandi, Corradi, and Wilhelm (2016) discuss this issue in discrete time. Fourth, the selected bandwidth also has to satisfy the condition for a vanishing discretization, namely $\frac{\Delta_{n,T}\sqrt{h_{n,T}^d \widehat{L}_{n,T}(x)}}{h_{n,T}^2} \xrightarrow{p} 0$. Any effective bandwidth selection criterion for general continuous-time models ought to account for these observations.

The issue of bandwidth selection for recurrent processes is, therefore, particularly delicate. Guerre (2004) provides an enlightening discussion in discrete time.

Remark 13 (The strictly stationary *or* ergodic case)**.** In the positive Harris recurrent case, under suitable bandwidth conditions laid out earlier, $\frac{1}{T}\widehat{L}_{n,T}(x) \xrightarrow{p} \widetilde{p}(x)$ and

$$\frac{\frac{\partial \widetilde{\phi}(x)}{\partial x^j}}{\widetilde{\phi}(x)} = \frac{\frac{\partial \widetilde{\phi}(x)}{\partial x^j}/\widetilde{\phi}(I)}{\widetilde{\phi}(x)/\widetilde{\phi}(I)} = \frac{\frac{\partial \widetilde{p}(x)}{\partial x^j}}{\widetilde{p}(x)}$$

since the invariant measure is integrable, i.e., $\widetilde{\phi}(I) < \infty$ (see the comments in Section 3). In consequence, if $\frac{\Delta_{n,T}\sqrt{h_{n,T}^d T}}{h_{n,T}^2} \to 0$ and $h_{n,T} = O\left(T^{-\frac{1}{d+4}}\right)$ as $h_{n,T} \to 0$ with $n, T \to \infty$ and $\Delta_{n,T} \to 0$, then

$$\sqrt{T h_{n,T}^d}\left(\widehat{\mu}_g(x) - \mu_g(x) - bias_g(x)\right) \Rightarrow \mathbf{N}\left(0, \left(\int_S K^2(u)\,du\right)^d \frac{\sum_{k=1}^m \sigma_{gk}^2(x)}{\widetilde{p}(x)}\right), \quad \forall x \in I \subseteq \mathcal{R}^d, \forall g = 1, \ldots, d, \tag{20}$$

and

$$bias_g(x) = h_{n,T}^2 \left(\int_S s^2 K(s) ds\right)\left(\sum_{j=1}^d \frac{\partial \mu_g(x)}{\partial x^j}\frac{\frac{\partial \widetilde{p}(x)}{\partial x^j}}{\widetilde{p}(x)} + \frac{1}{2}\sum_{j=1}^d \frac{\partial^2 \mu_g(x)}{\partial x^j \partial x^j}\right).$$

An analogous corollary can, of course, be obtained for $h_{n,T} = o\left(T^{-\frac{1}{d+4}}\right)$. In this case, the asymptotic expression would be identical with the sole exception of the absence of the bias term.

Remark 14 (The *two* curses of dimensionality)**.** Consistent with more conventional models in discrete time, an increase in the dimensionality of the system leads to a decrease in the rate of convergence of the nonparametric estimates. Contrary to standard problems, though, this effect operates through two channels, i.e., a *deterministic* effect which depends on d and a "*stochastic*" effect which

https://doi.org/10.1017/S0266466617000305 Published online by Cambridge University Press

depends on the speed of divergence to infinity of the quantity $\widehat{L}_{n,T}(x)$, i.e., $v(T)$. The first effect is standard. As d increases, the optimal $h_{n,T}$ should converge to zero at a slower rate. In the positive Harris recurrent case, the rate at which the bandwidth has to be adjusted as the dimension increases for a given sample size n and time span T depends on powers of d. We recall, in fact, that the AMSE-optimal bandwidth is $h_{n,T} = O\left(T^{-\frac{1}{d+4}}\right)$. In other words, when increasing the dimensionality of the problem, the kernel window width must be made wider to offset the sparser density of the data points. The second effect is novel. Null Harris recurrent processes induce divergence rates for the quantity $\widehat{L}_{n,T}(x)$ which are inversely related to the dimensionality of the system (see Theorem 3). Importantly, while these rates cannot be quantified in general, practical implementation of our procedures does not require their evaluation a-priori, as we discuss further in Remark 15. The scalar and planar Brownian motion cases are notable exceptions for which the rates can be computed in closed form (c.f., Remark 3).

To summarize, under positive Harris recurrence (or strict stationarity), the curse of dimensionality operates through only one channel, i.e., the dimension d, since $\widehat{L}_{n,T}(x)$ diverges at the constant rate T independently of the number of equations in the system (c.f. Remark 8). In the general (possibly nonstationary) case, the optimal bandwidth should account for the conventional curse of dimensionality given by d as well as for a second curse of dimensionality caused by the expected, smaller values of $\widehat{L}_{n,T}(x)$. The optimal (in an AMSE sense) bandwidth in Remark 12, i.e., $h_{n,T} \sim v(T)^{-\frac{1}{d+4}}$, accounts for both effects.

We expect this novel notion of "double curse of dimensionality" to carry over to functional estimation procedures for multivariate (possibly nonstationary) discrete-time processes. In other words, such an effect is truly a by-product of our minimal assumptions and does not appear to be specific to the study of continuous-time processes.

Remark 15. The asymptotic distribution in equation (18) depends on quantities which may, in principle, be estimated from the data, i.e., $\widehat{L}_{n,T}(x)$ and $\mathbf{a}(.)$ (c.f. Theorem 3 and, below, Theorem 5). Statistical inference on the drift vector does not require any conjectures about the dynamic features of the underlying process, such as stationarity, aside from recurrence. While traditional asymptotic theory is derived based on explicit assumptions of either stationarity or nonstationarity (often of the unit-root or $\frac{1}{2}$-null recurrent type), which are imposed before inference begins, our weak convergence results reflect the mildness of recurrence as an identifying assumption. Inference hinges on random norming in the context of asymptotic normal distributions and can be implemented in the simple framework of mixed normal models with estimable random variances. As discussed earlier, from a theoretical standpoint, the rates of convergence are, indeed, affected by both the stationarity features of the underlying process and the dimensionality of the system through the constant α which drives the rate of divergence of $\widehat{L}_{n,T}(x)$ (c.f. Theorems 4 and 5, below). Importantly, however, such a constant does not

https://doi.org/10.1017/S0266466617000305 Published online by Cambridge University Press

have—in principle—to be identified empirically for meaningful statistical inference to be conducted.

Remark 16. Again, the condition on vanishing discretization for weak convergence is a bit milder than in the existing literature. For the case $d = 1$, the subject of previous work, it is $\frac{\Delta_{n,T}\sqrt{\widehat{L}_{n,T}(x)}}{h_{n,T}^{3/2}} \xrightarrow{p} 0$, rather than $\frac{\Delta_{n,T}\log(1/\Delta_{n,T})\widehat{L}_{n,T}^2(x)}{h_{n,T}^2} \xrightarrow{a.s.} 0$.

It will be apparent that similar arguments to those in the previous remarks apply to the estimation of the diffusion matrix to which we now turn. Below, for conciseness, we only place emphasis on those aspects that are specific to diffusion matrix evaluation.

5.2. Estimating the Diffusion Matrix

Theorem 5 discusses the consistency and limiting distribution of the diffusion matrix estimator in equation (4).

THEOREM 5 (The limiting distribution of the diffusion matrix estimator). *Let Assumptions 1, 2, 3, and 4 be satisfied. Assume the deterministic sequence $h_{n,T}$ is such that $h_{n,T} \to 0$ with $n, T \to \infty$ and $\Delta_{n,T} \to 0$. Define $P_D = (D^\intercal D)^{-1} D^\intercal$, where D is the standard duplication matrix, i.e., the unique $d^2 \times (d(d+1))/2$ matrix which, given a positive definite $d \times d$ matrix* $\mathbf{u}$, *yields*

$$vech\mathbf{u} = P_D vec\mathbf{u} = \begin{bmatrix} u_{1,1} \\ u_{2,1} \\ \cdots \\ u_{d,d} \end{bmatrix}.$$

Let $\frac{h_{n,T}^d \widehat{L}_{n,T}(x)}{\Delta_{n,T}} \xrightarrow{p} \infty$, $\frac{h_{n,T}^{d+4} \widehat{L}_{n,T}(x)}{\Delta_{n,T}} \xrightarrow{p} C$, *where* C *is a constant, and* $h_{n,T}^{d-4} \widehat{L}_{n,T}(x) \Delta_{n,T} \xrightarrow{p} 0$. *Then,*

$$\sqrt{\frac{h_{n,T}^d \widehat{L}_{n,T}(x)}{\Delta_{n,T}}} \left(vech\widehat{\mathbf{a}}_{n,T}(x) - vech\mathbf{a}(x) - \mathbf{\Gamma}^{\sigma^2}(x) \right)$$
$$\Rightarrow (\mathbf{\Xi}(x))^{1/2} \mathbf{N}\left(\mathbf{0}, \left(\int_S K^2(s) ds \right)^d \mathbf{I}_{\frac{d(d+1)}{2}} \right), \quad \forall x \in I \subseteq \mathcal{R}^d, \tag{21}$$

where

$$\mathbf{\Xi}(x) = P_D \left(2\mathbf{a}(x) \otimes \mathbf{a}(x) \right) P_D^\intercal,$$

$$\mathbf{\Gamma}^{\sigma^2}(x) = (bias_{1,1}, bias_{2,1}, \ldots, bias_{d,d})^\intercal(x),$$

and

https://doi.org/10.1017/S0266466617000305 Published online by Cambridge University Press

$$bias_{g,q}(x) = h_{n,T}^2 \left(\int_S s^2 K(s) \mathrm{d}s \right) \times \left(\sum_{j=1}^{d} \frac{\partial a_{g,q}(x)}{\partial x^j} \frac{\frac{\partial \widetilde{\phi}(x)}{\partial x^j}}{\widetilde{\phi}(x)} + \frac{1}{2} \sum_{j=1}^{d} \frac{\partial^2 a_{g,q}(x)}{\partial x^j \partial x^j} \right), \quad for\ g,q = (1,1),\ldots,(d,d).$$

Remark 17 (The fixed T case)**.** Contrary to drift estimation, Theorem 5 clarifies that the diffusion matrix can be consistently estimated over a fixed span of data $T = \overline{T}$ (c.f., Florens-Zmirou, 1993, in the scalar case and Brugiére, 1991 and 1993, in the multivariate case). Hence, if interest is placed on volatility estimation, the process can be transient for identification. Should the focus be on both the diffusion matrix and the drift vector (namely, on the full system's dynamics, as in this article), then recurrence represents, as discussed, a rather mild condition for the identification of the full system.

Remark 18 (The bivariate case)**.** It is now worth being explicit about the form of the limiting variance in equation (21). For clarity, we consider the bivariate case, i.e., $d = 2$. Write

$$\begin{aligned}
\Xi(x) &= P_D \left(2\mathbf{a}(x) \otimes \mathbf{a}(x)\right) P_D^{\intercal} \\
&= 2 \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \\
&\quad \times \begin{bmatrix} \left(\sum \sigma_{1k}^2\right)^2 & \left(\sum \sigma_{1k}^2\right)\left(\sum \sigma_{1k}\sigma_{2k}\right) & \left(\sum \sigma_{1k}^2\right)\left(\sum \sigma_{1k}\sigma_{2k}\right) & \left(\sum \sigma_{1k}\sigma_{2k}\right)^2 \\ \left(\sum \sigma_{1k}^2\right)\left(\sum \sigma_{1k}\sigma_{2k}\right) & \left(\sum \sigma_{1k}^2\right)\left(\sum \sigma_{2k}^2\right) & \left(\sum \sigma_{1k}\sigma_{2k}\right)^2 & \left(\sum \sigma_{2k}^2\right)\left(\sum \sigma_{1k}\sigma_{2k}\right) \\ \left(\sum \sigma_{1k}^2\right)\left(\sum \sigma_{1k}\sigma_{2k}\right) & \left(\sum \sigma_{1k}\sigma_{2k}\right)^2 & \left(\sum \sigma_{1k}^2\right)\left(\sum \sigma_{2k}^2\right) & \left(\sum \sigma_{2k}^2\right)\left(\sum \sigma_{1k}\sigma_{2k}\right) \\ \left(\sum \sigma_{1k}\sigma_{2k}\right)^2 & \left(\sum \sigma_{2k}^2\right)\left(\sum \sigma_{1k}\sigma_{2k}\right) & \left(\sum \sigma_{2k}^2\right)\left(\sum \sigma_{1k}\sigma_{2k}\right) & \left(\sum \sigma_{2k}^2\right)^2 \end{bmatrix} \\
&\quad \times \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & 1 \end{bmatrix} \\
&= \begin{bmatrix} 2\left(\sum \sigma_{1k}^2\right)^2 & 2\left(\sum \sigma_{1k}^2\right)\left(\sum \sigma_{1k}\sigma_{2k}\right) & 2\left(\sum \sigma_{1k}\sigma_{2k}\right)^2 \\ 2\left(\sum \sigma_{1k}^2\right)\left(\sum \sigma_{1k}\sigma_{2k}\right) & \left(\sum \sigma_{1k}\sigma_{2k}\right)^2 + \left(\sum \sigma_{2k}^2\right)\left(\sum \sigma_{1k}^2\right) & 2\left(\sum \sigma_{2k}^2\right)\left(\sum \sigma_{1k}\sigma_{2k}\right) \\ 2\left(\sum \sigma_{1k}\sigma_{2k}\right)^2 & 2\left(\sum \sigma_{2k}^2\right)\left(\sum \sigma_{1k}\sigma_{2k}\right) & 2\left(\sum \sigma_{2k}^2\right)^2 \end{bmatrix},
\end{aligned}$$

where $\sum = \sum_{k=1}^{m}$. We now compute the asymptotic variance of

$$vec\widehat{\mathbf{a}}_{n,T}(x) - vec\mathbf{a}(x).$$

Notice that

$$Dvech\mathbf{a}(x) = vec\mathbf{a}(x).$$

Then, the limiting variance of $vec\widehat{\mathbf{a}}_{n,T}(x)$ can be written as

$$\begin{aligned}
D\Xi(x)D^{\intercal} &= DP_D \left(2\mathbf{a}(x) \otimes \mathbf{a}(x)\right) P_D^{\intercal} D^{\intercal} \\
&= 2D(D^{\intercal}D)^{-1} D^{\intercal} \left(\mathbf{a}(x) \otimes \mathbf{a}(x)\right) D(D^{\intercal}D)^{-1} D^{\intercal}
\end{aligned}$$

https://doi.org/10.1017/S0266466617000305 Published online by Cambridge University Press

$$= 2\overline{P}_D\left(\mathbf{a}(x)\otimes\mathbf{a}(x)\right)\overline{P}_D$$
$$= 2\overline{P}_D\left(\mathbf{a}(x)\otimes\mathbf{a}(x)\right)$$

since $\overline{P}_D\left(\mathbf{a}(x)\otimes\mathbf{a}(x)\right)D = \left(\mathbf{a}(x)\otimes\mathbf{a}(x)\right)D$, where $\overline{P}_D$ is the $d^2\times d^2$ matrix that projects $\mathcal{R}^{d^2}$ orthogonally onto $R(D)$, i.e., the range space of D. In particular,

$$\overline{P}_D = \frac{1}{2}\left(\mathbf{I}_{d^2} + \mathbf{G}_{d^2}\right),$$

where $\mathbf{I}_{d^2}$ is a $d^2\times d^2$ identity matrix and $\mathbf{G}_{d^2}$ is such that

$$\mathbf{G}_{d^2} = \sum_i\sum_j U_{ji}\otimes U_{ij},$$

where the $d\times d$ matrix U_{ji} has 1 in the (j,i) position and 0 elsewhere. In the $d=2$ case, for example, we obtain

$$\mathbf{G}_{d^2} = \begin{bmatrix} 1 & 0 & 0 & 0\\ 0 & 0 & 1 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & 0 & 1 \end{bmatrix}$$

and

$$\overline{P}_D = \begin{bmatrix} 1 & 0 & 0 & 0\\ 0 & \frac{1}{2} & \frac{1}{2} & 0\\ 0 & \frac{1}{2} & \frac{1}{2} & 0\\ 0 & 0 & 0 & 1 \end{bmatrix},$$

which implies

$$2\overline{P}_D\left(\mathbf{a}(x)\otimes\mathbf{a}(x)\right)$$
$$= \begin{bmatrix} 2\left(\sum\sigma_{1k}^2\right)^2 & 2\left(\sum\sigma_{1k}^2\right)\left(\sum\sigma_{1k}\sigma_{2k}\right) & 2\left(\sum\sigma_{1k}^2\right)\left(\sum\sigma_{1k}\sigma_{2k}\right) & 2\left(\sum\sigma_{1k}\sigma_{2k}\right)^2\\ 2\left(\sum\sigma_{1k}^2\right)\left(\sum\sigma_{1k}\sigma_{2k}\right) & \left(\sum\sigma_{1k}^2\right)\left(\sum\sigma_{2k}^2\right)+\left(\sum\sigma_{1k}\sigma_{2k}\right)^2 & \left(\sum\sigma_{1k}^2\right)\left(\sum\sigma_{2k}^2\right)+\left(\sum\sigma_{1k}\sigma_{2k}\right)^2 & 2\left(\sum\sigma_{2k}^2\right)\left(\sum\sigma_{1k}\sigma_{2k}\right)\\ 2\left(\sum\sigma_{1k}^2\right)\left(\sum\sigma_{1k}\sigma_{2k}\right) & \left(\sum\sigma_{1k}^2\right)\left(\sum\sigma_{2k}^2\right)+\left(\sum\sigma_{1k}\sigma_{2k}\right)^2 & \left(\sum\sigma_{1k}^2\right)\left(\sum\sigma_{2k}^2\right)+\left(\sum\sigma_{1k}\sigma_{2k}\right)^2 & 2\left(\sum\sigma_{2k}^2\right)\left(\sum\sigma_{1k}\sigma_{2k}\right)\\ 2\left(\sum\sigma_{1k}\sigma_{2k}\right)^2 & 2\left(\sum\sigma_{2k}^2\right)\left(\sum\sigma_{1k}\sigma_{2k}\right) & 2\left(\sum\sigma_{2k}^2\right)\left(\sum\sigma_{1k}\sigma_{2k}\right) & 2\left(\sum\sigma_{2k}^2\right)^2 \end{bmatrix}.$$

Remark 19 (Bandwidth choice)**.** A "nearly" optimal selection rule for the bandwidth in the diffusion case is given by

$$h_{n,T}^{dif} := c^{dif}\left(\frac{1}{\log\left(v(T)/\Delta_{n,T}\right)}\right)\left(v(T)/\Delta_{n,T}\right)^{-\frac{1}{d+4}}, \tag{22}$$

where c^{dif} is a constant. Coherently with our previous discussion in the drift case (c.f., Remark 12), this choice would allow us to eliminate the influence of the bias term from the limiting distribution of the diffusion estimates and obtain

https://doi.org/10.1017/S0266466617000305 Published online by Cambridge University Press

centering about zero, while achieving a close-to-optimal speed of convergence. Barring issues with the selection of c^{dif}, this choice is—as earlier in the drift case—infeasible since $v(T)$ is generally unknown. Replacing $v(T)$ with $\widehat{L}_{n,T}(x)$ would make the problem circular (in light of the dependence of $\widehat{L}_{n,T}(x)$ on the same bandwidth $h_{n,T}^{dif}$). It would also not account for the bandwidth condition induced by discretization. As discussed in Remark 12, both issues need to be accommodated.

Being the diffusion function estimable over a fixed span of observations (c.f., Remark 17), there is, however, relatively less scope for local adaptation of the bandwidth sequence than in the drift case examined earlier. An approximate (optimal) rule to select the bandwidth for a slowly diverging T would, in fact, be

$$h_{n,T}^{dif} := c^{dif} \left(\frac{1}{\log n} \right) n^{-\frac{1}{d+4}},$$

which is standard in nonparametric statistics and does not require evaluation of the process' occupation density.

6. CONCLUSION

This article studies kernel methods for multivariate diffusion processes. We provide an estimation theory which is easily interpretable based on traditional results in nonparametric analysis for multidimensional discrete-time series but has the additional advantage of robustness to deviations from strong distributional assumptions, such as stationarity. Harris recurrence is the identifying assumption used in the present work to show consistency and asymptotic (mixed) normality of the functional estimates of drift vector and diffusion matrix. First, this assumption is known to be milder than stationarity and mixing and might prove useful to study multivariate (discrete-time or continuous-time) processes whose stationarity can neither be guaranteed nor ruled out a priori. Second, even in the stationary case, functional methods which do not rely on the information contained in the process' stationary density may be useful when the stationary density can hardly be identified reliably, as is the case for persistent time series. In both cases, our asymptotic theory appears to provide an intuitive assessment of statistical uncertainty by (inversely) relating the size of the confidence intervals to the occupation time of the underlying empirical process. Similarly, the "double curse of dimensionality" may be viewed as a theoretical representation of the risks of empirical work conducted in the context of multidimensional, as well as highly persistent, processes.

We introduce the methods in the context of classical Nadaraya-Watson kernel estimators. While these estimators are arguably the most widely used in applied work, they can be improved upon. Coherently with the more classical analysis of stationary discrete-time series, the methods may be extended to a variety of multidimensional nonparametric procedures like local linear and polynomial fitting, among others (see, e.g., Fan, 1992, and Masry, 1996a, 1996b, and, for interesting

https://doi.org/10.1017/S0266466617000305 Published online by Cambridge University Press

work in the scalar diffusion case, Fan and Zhang, 2003, Moloche, 2004a, and Aït-Sahalia and Park, 2016).

Even though the focus of this article is on nonparametric estimation, the procedures we discuss might be used to evaluate parametric models. For instance, parametric specifications for multivariate diffusions may be tested by utilizing criteria which compare functional estimates of drift and diffusion matrix to their parametric counterparts. The integrated squared error employed by Bickel and Rosenblatt (1975), Fan (1994), and Aït-Sahalia (1996), among others, is a possible criterion. Such a comparison might be conducted independently for drift and diffusion since these moments may, as shown, be identified separately.

NOTES

1. A function $f:[a,\infty)\to(0,\infty)$, $a>0$, is said to be slowly varying at infinity in the sense of Karamata if $\lim_{x\to\infty} f(\lambda x)/f(x)\to 1$ for $\lambda>0$. The constant function and the logarithmic function are trivially slowly varying.

2. It is easy to show that the limiting distribution of standardized Brownian local time is closely related to this result. Write

$$\begin{aligned}
\lim_{T\to\infty}\frac{1}{\sqrt{T}}L_B(T,x) &= \lim_{T\to\infty}\lim_{\varepsilon\to 0}\frac{1}{2\varepsilon\sqrt{T}}\int_0^T \mathbf{1}_{\{|B(s)-x|<\varepsilon\}}ds \\
&\overset{d}{=} \lim_{T\to\infty}\lim_{\varepsilon\to 0}\frac{\sqrt{T}}{2\varepsilon}\int_0^T \mathbf{1}_{\{|\sqrt{T}B(s/T)-x|<\varepsilon\}}\frac{ds}{T} \\
&\overset{d}{=} \lim_{T\to\infty}\lim_{\xi\to 0}\frac{1}{2\xi}\int_0^1 \mathbf{1}_{\{|B(u)-x/\sqrt{T}|<\xi\}}du \\
&\overset{d}{=} \lim_{T\to\infty}\lim_{\xi\to 0}\frac{1}{2\xi}\int_{-\infty}^{\infty} \mathbf{1}_{\{|a-x/\sqrt{T}|<\xi\}}L_B(1,a)da \\
&\overset{d}{=} \lim_{T\to\infty}\lim_{\xi\to 0}\frac{1}{2}\int_{-\infty}^{\infty} \mathbf{1}_{\{-1<u<1\}}L_B(1,\xi u+x/\sqrt{T})du \\
&\overset{d}{=} L_B(1,0).
\end{aligned}$$

However, $L_B(1,0)\overset{d}{=}|B(1)|$ (Lévy, 1939). See Pitman (1998), equation (12), for a more recent discussion.

3. Given the measure η, we recall that $\|g\|_r=(\int|f|^r d\eta)^{1/r}$ and $\|g\|_\infty=\inf\{x:\eta\{g>x\}=0\}$.

4. In the scalar case, equation (5) provides the invariant density.

5. Interestingly, one could work with any desired functional representation—*within* the range of the data—and assume that the driving functions are suitably modified *outside* of the range of the data so as to achieve boundedness. This is observationally equivalent to working with an unrestricted model over the entire domain. Bandi and Renò (2017) discuss this issue further and provide an illustrative example.

REFERENCES

Aït-Sahalia, Y. (1996) Testing continuous-time models of the spot interest rate. *Review of Financial Studies* 9, 385–426.

Aït-Sahalia, Y., L.P. Hansen, & J. Scheinkman (2010) Operator methods for continuous-time Markov processes. In Y. Aït-Sahalia & L.P. Hansen (eds.), *Handbook of Financial Econometrics* vol. 1, pp. 1–66. Elsevier.

https://doi.org/10.1017/S0266466617000305 Published online by Cambridge University Press

Aït-Sahalia, Y. & J. Park (2016) Bandwidth selection and asymptotic properties of local nonparametric estimators of possibly nonstationary continuous-time models. *Journal of Econometrics* 192, 119–138.

Athreya, K.B. & P. Ney (1978) A new approach to the limit theory of recurrent Markov chains. *Transactions of the American Mathematical Society* 245, 493–501.

Azéma, J., M. Kaplan-Duflo, & D. Revuz (1966) Mesure invariante sur les classes récurrentes des processus de Markov. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 8, 157–181.

Bandi, F.M., V. Corradi, & D. Wilhelm (2016) Possibly Nonstationary Cross-Validation. Working paper, Johns Hopkins University, University of Surrey and University College of London.

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

Bandi, F.M. & P.C.B. Phillips (2010) Nonstationary continuous-time processes. In Y. Aït-Sahalia & L.P. Hansen (eds.), *Handbook of Financial Econometrics* vol. 1, pp. 139–201. Elsevier.

Bandi, F.M. & R. Renò (2017) Nonparametric stochastic volatility. *Econometric Theory*, forthcoming.

Bickel, P.J. & M. Rosenblatt (1973) On some global measures of the deviations of density function estimates. *Annals of Statistics* 1, 1071–1095.

Bosq, D. (1995) Sur le comportement exotique de l'estimateur à noyau de la densité marginale d'un processus à temps continu. *Comptes Rendus de l'Académie des Sciences - Séries I* 320, 369–372.

Bosq, D. (1998) *Nonparametric Statistics for Stochastic Processes: Estimation and Prediction*, 2nd ed. Lecture Notes in Statistics, vol. 110. Springer-Verlag.

Boudoukh, J., C. Downing, M. Richardson, R. Stanton, & R.F. Whitelaw (2007) A multifactor, nonlinear, continuous-time model of interest rate volatility. In T. Bollerslev, J.R. Russell, & M.W. Watson, (eds.), *Volatility and Time Series Econometrics: Essays in Honor of Robert F. Engle* pp. 296–322. Oxford University Press.

Brugière, P. (1991) Estimation de la variance d'un processus de diffusion dans le cas multidimensionnel. *Comptes Rendus de l'Académie des Sciences - Séries I* 312, 999–1004.

Brugière, P. (1993) Théorème de limite centrale pour un estimateur nonparamétrique de la variance d'un processus de diffusion multidimensionnelle. *Annales de l'Institut Henri Poincaré* 29, 357–389.

Cai, Z. & Y. Hong (2003) Nonparametric methods in continuous-time finance: A selective review. In M.G. Akritas & D.N. Politis (eds.), *Recent Advances and Trends in Nonparametric Statistics*, pp. 283–302. Elsevier.

Castellana, J.V. & M.R. Leadbetter (1986) On smoothed probability density estimation for stationary processes. *Stochastic Processes and Their Applications* 21, 179–193.

Chen, X. (1999) How often does a Harris recurrent Markov chain recur? *The Annals of Probability* 27, 1324–1346.

Chen, X. (2000) On the limit laws of the second order for additive functionals of Harris recurrent Markov chains. *Probability Theory and Related Fields* 116, 89–123.

Chen, X. & L.P. Hansen (2002) Dependence Properties of Multivariate Reversible Diffusions. Working paper, London School of Economics and University of Chicago.

Conley, T., L.P. Hansen, E. Luttmer, & J. Scheinkman (1997) Short-term interest rates as subordinated diffusions. *Review of Financial Studies* 10, 525–577.

Darling, D.A. & M. Kac (1957) On occupation times for Markoff processes. *Transactions of the American Mathematical Society* 84, 444–458.

Fan, J. (1992) Design-adaptive nonparametric regression. *Journal of the American Statistical Association* 87, 998–1004.

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

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

Fan, Y. (1994) Testing the goodness of fit of a parametric density function by kernel method. *Econometric Theory* 10, 316–356.

https://doi.org/10.1017/S0266466617000305 Published online by Cambridge University Press

Feigin, P.D. (1985) Stable convergence of semimartingales. *Stochastic Processes and Their Applications* 19, 125–134.

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

Gallant, R. & G. Tauchen (2010) Simulated score inference and indirect inference for continuous-time models. In Y. Aït-Sahalia & L.P. Hansen (eds.), *Handbook of Financial Econometrics* vol. 1, pp. 427–477. Elsevier.

Geman, D. & J. Horowitz (1980) Occupation densities. *Annals of Probability* 8, 1–67.

Genon-Catalot, V. & J. Jacod (1993) On the estimation of the diffusion coefficient for multi-dimensional diffusion processes. *Annales de l'Institut Henri Poincaré* 29, 119–151.

Guerre, E. (2004) Design-Adaptive Point-Wise Nonparametric Regression Estimation for Recurrent Markov Time Series. Working paper, Centre de Recherche en Economie et Statistique.

Höpfner, R. & E. Löcherbach (2003) *Limit Theorems for Null Recurrent Markov Processes*. Memoirs of the American Mathematical Society, vol. 161, American Mathematical Society, No. 768.

Jiang, G.J. & J. Knight (1999) Finite sample comparison of alternative estimators of Itô diffusion processes: A Monte-Carlo study. *Journal of Computational Finance* 2, 5–38.

Johannes, M. & N. Polson (2010) MCMC methods for continuous-time financial econometrics. In Y. Aït-Sahalia & L.P. Hansen (eds.), *Handbook of Financial Econometrics* vol. 2, pp. 1–72. Elsevier.

Kallianpur, G. & H. Robbins (1953) Ergodic properties of the Brownian motion process. *Proceedings of the National Academy of Sciences of the United States of America* 39, 525–533.

Karatzas, I. & S.E. Shreve (1991) *Brownian Motion and Stochastic Calculus*. Springer-Verlag.

Karlsen, H.A. & D. Tjøstheim (2001) Nonparametric estimation in null recurrent time series. *Annals of Statistics* 29, 372–416.

Karlsen, H.A., T. Myklebust, & D. Tjøstheim (2007) Nonparametric estimation in a nonlinear cointegrating type model. *Annals of Statistics* 35, 252–299.

Kasahara, Y. (1984) Limit theorems for Levy processes and Poisson point processes and their applications to Brownian excursions. *Journal of Mathematics of Kyoto University* 24, 521–538.

Kusuoka, S. & D. Stroock (1985) Applications of the Malliavin calculus, Part II. *Journal of the Faculty of Science, University of Tokyo* 32, 1–76.

Kutoyants, A. (1995) On density estimation by the observations of ergodic diffusion process. Laboratoire de Statistique et Processus, Univ. du Maine et d'Angers, prepublication 95–8.

Lévy, P. (1939) Sur certains processus stochastiques homogènes. *Compositio Mathematica* 7, 283–339.

Löcherbach, E. & D. Loukianova (2008) On Nummelin splitting for continuous time Harris recurrent Markov processes and application to kernel estimation for multi-dimensional diffusions. *Stochastic Processes and Their Applications* 118, 1301–1321.

Löcherbach, E. & D. Loukianova (2009) The law of iterated logarithm for additive functionals and martingale additive functionals of Harris recurrent Markov processes. *Stochastic Processes and Their Applications* 119, 2312–2335.

Magnus, J.R. & H. Neudecker (1988) *Matrix Differential Calculus with Applications in Statistics and Econometrics.* Wiley.

Masry, E. (1996a) Multivariate local polynomial regression for time series: Uniform strong consistency and rates. *Journal of Time Series Analysis* 17, 571–599.

Masry, E. (1996b) Multivariate regression estimation: Local polynomial fitting for time series. *Journal of Stochastic Processes and their Applications* 65, 81–101.

Meyn, S.P. & R.L. Tweedie (1993) Stability of Markovian processes II. Continuous-time processes and sampled chains. *Advances in Applied Probability* 25, 487–517.

Moloche, G. (2004a) Local Nonparametric Estimation of Scalar Diffusions. Working paper, University of Chicago.

Moloche, G. (2004b) Kernel Regression for Nonstationary Harris Recurrent Processes. Working paper, University of Chicago.

Nummelin, E. (1978) A splitting technique for Harris recurrent Markov chains. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 43, 309–318.

https://doi.org/10.1017/S0266466617000305 Published online by Cambridge University Press

Pagan, A. & A. Ullah (1999) *Nonparametric Econometrics*. Cambridge University Press.
Park, J. (2006) Spatial Analysis of Time Series. Working paper, Texas A&M University.
Pitman, J. (1998) The Distribution of Local Times of a Brownian Bridge. Working paper, University of California, Berkeley.
Phillips, P.C.B. (2001) Descriptive econometrics for nonstationary time series with empirical illustrations. *Journal of Applied Econometrics* 16, 389–413.
Revuz, D. & M. Yor (1998) *Continuous Martingales and Brownian Motion*. Springer-Verlag.
Schienle, M. (2008) Nonparametric Nonstationary Regression. Doctoral thesis, Mannheim University.
Stanton, R. (1997) A nonparametric model of term structure dynamics and the market price of interest rate risk. *Journal of Finance* 52, 1973–2002.
Touati, A. (1992) On the functional convergence in distribution of sequences of semimartingales to a mixture of Brownian motions. *Theory of Probability and its Applications* 36, 752–771.
Veretennikov, A.Y. (1999) On Castellana-Leadbetter's condition for diffusion density estimation. *Statistical Inference for Stochastic Processes* 2, 1–9.
Wang, Q. (2015) *Limit Theorems for Nonlinear Cointegrating Regression*. World Scientific.
Wang, Q. & P.C.B. Phillips (2009a) Structural nonparametric cointegrating regression. *Econometrica* 77, 1901–1948.
Wang, Q. & P.C.B. Phillips (2009b) Asymptotic theory for local time density estimation and nonparametric cointegrating regression. *Econometric Theory* 25, 710–738.
Xu, K.-L. (2009) Empirical likelihood based inference for recurrent nonparametric diffusions. *Journal of Econometrics* 142, 265–280.

APPENDIX

Proof of Theorem 1. See Revuz and Yor (1998, Theorem 3.12, page 408) and Azéma, Kaplan-Duflo, and Revuz (1966, Remark 1, page 170). ■

Proof of Theorem 2. See Darling and Kac (1956). ■

We begin with two lemmas. In what follows, we use the notation $\mathbf{h} = h^d$ when convenient.

LEMMA A.1. *The function $\mathcal{S}(x)$ is a product function $\Pi_{j=1}^{d}\mathcal{S}_j(x^j)$. Each $\mathcal{S}_j$ is a nonnegative, bounded, and continuous function defined on a compact set* S. *The function $\mathcal{H}$ is also a nonnegative, bounded, and continuous function. Let Assumptions 1, 3.1, and 4 be satisfied. Let $v(T)\mathbf{h}^{2/r} \to \infty$. Then,*

$$\frac{1}{v(T)\mathbf{h}}\int_0^T \mathcal{S}\left(\frac{X_s - x}{\mathbf{h}}\right)\mathcal{H}(X_s)\,\mathrm{d}s \Rightarrow C_X\widetilde{\phi}(x)\,\mathcal{H}(x)\left[\Pi_{j=1}^{d}\left(\int_{\mathrm{S}}\mathcal{S}_j(u^j)\mathrm{d}u^j\right)\right]\mathrm{g}\alpha,$$

as $h \to 0$ and $T \to \infty$ jointly, where C_X is a process-specific constant.

Proof of Lemma A.1. Write $L_{T,\mathbf{h}}(x) := \frac{1}{\mathbf{h}}\int_0^T \mathcal{S}\left(\frac{X_s - x}{\mathbf{h}}\right)\mathcal{H}(X_s)\,\mathrm{d}s$. It would be helpful to break the integral into i.i.d. or stationary components $U_{n,\mathbf{h}}(x) := \frac{1}{\mathbf{h}}\int_{R_n}^{R_{n+1}} \mathcal{S}\left(\frac{X_s - x}{\mathbf{h}}\right)\mathcal{H}(X_s)\,\mathrm{d}s$. (The index n is used here and elsewhere without confusion with the number of discretely-sampled observations n.) By doing so, we would be able to appeal to limiting results for additive functionals of Markov processes. To this extent, if the process had a recurrent atom, x_0, then one could introduce a sequence of increasing stopping times R_n so that, among other properties, the following conditions would be satisfied:

(i) For all n, $X_{R_n} = x_0$;
(ii) For all n, the process $(X_{R_n+s})_{s>0}$ is independent of $\mathcal{F}_{R_n}$.

Being R_n a "regeneration time," not only would the components $U_{n,\mathbf{h}}(x)$ be independent (by property (ii)), they would also be identically distributed. Hence, our original problem

https://doi.org/10.1017/S0266466617000305 Published online by Cambridge University Press

of subdividing the integral into i.i.d or stationary components $U_{n,\mathbf{h}}(x)$ would be readily solved.

For general multidimensional diffusion processes of the kind we are interested in, however, a recurrent atom x_0 cannot be defined. In the absence of a recurrent atom for the process X, LL (2008) have shown how to "construct" a recurrent atom on a larger probability space by extending the (discrete-time) approach in Nummelin (1978) and Athreya and Ney (1978). They advocate "embedding" the original process X into a larger process Z with a definition of recurrent atom (LL, 2008, Proposition 2.8(a)). In their construction, the resulting process $Z = (Z^1, Z^2, Z^3)$ has three coordinates (with values in $I \times [0,1] \times I$, where I is the state space of the original process X) and a sequence of exponential-1 jump times T_n independent of X. The first coordinate is the same in law as the process X (LL, 2008, Proposition 2.8(b) and (c)). In other words, it is a "copy" of the original process. Over the interval $[T_n, T_{n+1})$, the second coordinate is an independent, uniformly (U) distributed, random variable on $[0,1]$. Over the same interval $[T_n, T_{n+1})$, the third coordinate is the original process X sampled at the (independent) exponential time T_{n+1}, i.e., $X_{T_{n+1}}$. Hence, both Z^2 and Z^3 are piece-wise constant. We refer the reader to Section 2.2 in LL (2008) for additional details.

Define by P_λ the probability measure of X for a generic, initial measure $\lambda(\mathrm{d}x)$. The corresponding expectation is denoted by E_λ. Define by $\mathbb{F}$ the filtration generated by the larger process Z. We call $\mathbb{P}_\lambda$ the probability measure of Z for an initial measure $\lambda(\mathrm{d}x) \otimes U(\mathrm{d}u) \otimes Q((x,u),\mathrm{d}y)$, where $Q((x,u),\mathrm{d}y)$ is a suitable transition kernel from $I \times [0,1]$ to I. Again, the corresponding expectation is denoted by $\mathbb{E}_\lambda$.

Because the multidimensional diffusion process X is a Harris recurrent (càdlàg) strong Markov process satisfying Assumption 4 in the main text, the construction in LL (2008) can be applied to our problem. By this, we mean that the process X, which does not have a recurrent atom x_0 in general, can be embedded into a larger Harris recurrent process Z with a definition of recurrent atom. By LL (Example 2.11, 2008), the larger process Z is also (càdlàg) strong Markov with respect to $\mathbb{F}$ when introduced for our problem. As said, a recurrent atom can be defined for Z. In fact, write

$$A = C \times [0,\beta] \times I,$$

where C is a compact set (with $\phi(C) > 0$) and $\beta \in (0,1]$.

1. We can define stopping times (S_n, R_n) satisfying $S_0 = R_0 := 0$, $S_{n+1} := \inf\{T_m > R_n : Z_{T_m} \in A\}$ and $R_{n+1} := \inf\{T_m : T_m > S_{n+1}\}$ for all n. Thus, $S_n < R_n < \infty$.
2. For all n, X_{R_n} is independent of $\sigma\{X_s,\ s \le S_n\}$.

The construction implies that we can now "split" the integral $L_{T,\mathbf{h}}(x)$ into components $U_{n,\mathbf{h}}(x) = \frac{1}{\mathbf{h}} \int_{R_n}^{R_{n+1}} \mathcal{S}\left(\frac{X_s - x}{\mathbf{h}}\right) \mathcal{H}(X_s)\,\mathrm{d}s$ for which the $\mathbb{F}$-stopping times $\{R_n : n \ge 1\}$ serve the same role as classical "regeneration times." In particular, the $U_{n,\mathbf{h}}(x)$s constitute a stationary and strong mixing sequence with $\sigma\left\{U_{j,\mathbf{h}}(x),\ j \le n\right\}$ independent of $\sigma\left\{U_{n+l,\mathbf{h}}(x),\ l \ge 2\right\}$ (by property 2). In other words, $U_{n,\mathbf{h}}(x)$ is not independent of $U_{n+1,\mathbf{h}}(x)$ but is independent of $U_{n+2,\mathbf{h}}(x)$. In sum, the i.i.d. property does not apply to all components (as in the case in which a recurrent atom exists for the original process X) but the strings $U_{1,\mathbf{h}}(x), U_{3,\mathbf{h}}(x), \ldots$ and $U_{0,\mathbf{h}}(x), U_{2,\mathbf{h}}(x), \ldots$ are, indeed, strings of i.i.d. random variables.

To conclude, we appeal to the construction in LL (2008) and apply it to X for which a "larger" process Z is defined. In what follows, by exploiting the construction, we will divide $\frac{L_{T,\mathbf{h}}(x)}{v(T)}$ into three terms, which will be studied in turn.

https://doi.org/10.1017/S0266466617000305 Published online by Cambridge University Press

Denote by $N_T = \sum_{n\geq 1} \mathbf{1}_{\{S_n \leq T\}}$ the number of regenerations of the process Z between 0 and T and write

$$\begin{aligned}\frac{1}{v(T)}L_{T,\mathbf{h}}(x) = &\underbrace{\frac{N_T}{v(T)}\mathbb{E}_\lambda\left(\frac{1}{\mathbf{h}}\int_{R_n}^{R_{n+1}}\mathcal{S}\left(\frac{X_s-x}{\mathbf{h}}\right)\mathcal{H}(X_s)\,\mathrm{d}s\right)}_{\alpha_{T,\mathbf{h}}(x)}\\ &+\underbrace{\frac{1}{v(T)\mathbf{h}}\int_0^T \mathcal{S}\left(\frac{X_s-x}{\mathbf{h}}\right)\mathcal{H}(X_s)\,\mathrm{d}s - \mathbb{E}_\lambda\left(\frac{1}{v(T)\mathbf{h}}\int_0^T \mathcal{S}\left(\frac{X_s-x}{\mathbf{h}}\right)\mathcal{H}(X_s)\,\mathrm{d}s\right)}_{\beta_{T,\mathbf{h}}(x)}\\ &+\underbrace{\mathbb{E}_\lambda\left(\frac{1}{v(T)\mathbf{h}}\int_0^T \mathcal{S}\left(\frac{X_s-x}{\mathbf{h}}\right)\mathcal{H}(X_s)\,\mathrm{d}s\right) - \frac{N_T}{v(T)}\mathbb{E}_\lambda\left(\frac{1}{\mathbf{h}}\int_{R_n}^{R_{n+1}}\mathcal{S}\left(\frac{X_s-x}{\mathbf{h}}\right)\mathcal{H}(X_s)\,\mathrm{d}s\right)}_{\gamma_{T,\mathbf{h}}(x)}.\end{aligned} \tag{A.1}$$

Under conditions on the bandwidth(s) listed in the statement of the lemma, we will show that $\alpha_{T,\mathbf{h}}(x)$ has a well-defined limiting distribution. The term $\beta_{T,\mathbf{h}}(x)$ and the term $\gamma_{T,\mathbf{h}}(x)$ will, instead, be shown to be asymptotically negligible.

We begin with $\alpha_{T,\mathbf{h}}(x)$. The quantity N_T is an integrable additive functional of the process Z. By Höpfner and Löcherbach (2003), Theorem 3.15 a) ii), we have

$$\frac{N_{zT}}{v(T)} = \frac{N_{zT}}{U(T)T^\alpha} \Rightarrow \mathrm{g}_\alpha(z), \tag{A.2}$$

where $\mathrm{g}_\alpha(z)$ is the Mittag-Leffler process with parameter α (Kasahara, 1984), i.e., the inverse process of a stable increasing process $\mathrm{S}_\alpha(z)$. (For analogous results in discrete time, see Chen (1999, Theorem 2.3(ii)) and Karlsen and Tjøstheim (2001, Theorem 3.2).)

We note, as an important subtlety of the construction, that $v(T) = U(T)T^\alpha$ is the probability order of additive functionals of the "larger" process Z, not only that of additive functionals of the original process X. Hence, the standardization in equation (A.2) is appropriate. This can be seen as follows. Any "special" function $f(.)$ of Z with positive measure (see LL, Definition 2.3, 2008, for a definition of special functions) can be used to define a "version" of the deterministic equivalent of Z, namely $\mathbb{E}_\lambda\left(\int_0^T f(Z_s)\mathrm{d}s\right)$. In fact, deterministic equivalents are unique (i.e., asymptotically of the same order) up to a positive constant. To this extent, use as the special function $f(.)$ of choice one which only depends on the first coordinate of the process, i.e., Z^1. Given the construction used for the splitting, write

$$\frac{\mathbb{E}_\lambda\left(\int_0^T f(Z_s)\mathrm{d}s\right)}{\mathrm{E}_\lambda\left(\int_0^T g(X_s)\mathrm{d}s\right)} = \frac{\mathbb{E}_\lambda\left(\int_0^T f(Z_s^1)\mathrm{d}s\right)}{\mathrm{E}_\lambda\left(\int_0^T g(X_s)\mathrm{d}s\right)} = \frac{\mathrm{E}_\lambda\left(\int_0^T f(X_s)\mathrm{d}s\right)}{\mathrm{E}_\lambda\left(\int_0^T g(X_s)\mathrm{d}s\right)} \underset{T\to\infty}{\to} \frac{\langle\phi,f\rangle}{\langle\phi,g\rangle},$$

where the first equality is due to the fact that $f(.)$ only depends on Z^1, the second equality depends on the fact that Z^1 is a "copy" of X, and convergence is due to the strong Chacon-Ornstein limit theorem (LL, Definition 2.3, 2008). The symbol $\langle\phi,f\rangle$, for any ϕ-integrable function $f(.)$, was defined in Theorem 1. Now, it is clear that, if $\mathrm{E}_\lambda\left(\int_0^T g(X_s)\mathrm{d}s\right) \sim v(T)$, then $\mathbb{E}_\lambda\left(\int_0^T f(Z_s)\mathrm{d}s\right) \sim v(T)$ as well.

We turn to $\mathbb{E}_\lambda\left(\frac{1}{\mathbf{h}}\int_{R_n}^{R_{n+1}}\mathcal{S}\left(\frac{X_s-x}{\mathbf{h}}\right)\mathcal{H}(X_s)\,\mathrm{d}s\right)$. Given the assumed construction, by Proposition 4.3 in LL (2009), we have

https://doi.org/10.1017/S0266466617000305 Published online by Cambridge University Press

$$\mathbb{E}_\lambda\left(\frac{1}{\mathbf{h}}\int_{R_n}^{R_{n+1}}\mathcal{S}\left(\frac{X_s-x}{\mathbf{h}}\right)\mathcal{H}(X_s)\,\mathrm{d}s\right)=\mathrm{E}_\lambda\left(\frac{1}{\mathbf{h}}\int_{R_n}^{R_{n+1}}\mathcal{S}\left(\frac{X_s-x}{\mathbf{h}}\right)\mathcal{H}(X_s)\,\mathrm{d}s\right)$$
$$=C_X\left\langle\phi,\frac{1}{\mathbf{h}}\mathcal{S}\left(\frac{.-x}{\mathbf{h}}\right)\mathcal{H}(.)\right\rangle, \tag{A.3}$$

where C_X is a process-specific constant. But, since

$$\left\langle\phi,\frac{1}{\mathbf{h}}\mathcal{S}\left(\frac{.-x}{\mathbf{h}}\right)\mathcal{H}(.)\right\rangle=\int\frac{1}{\mathbf{h}}\mathcal{S}\left(\frac{a-x}{\mathbf{h}}\right)\mathcal{H}(a)\,\widetilde{\phi}(a)\mathrm{d}a$$
$$=\int\mathcal{S}(u)\,\mathcal{H}(u\mathbf{h}+x)\widetilde{\phi}(u\mathbf{h}+x)\mathrm{d}u$$
$$\underset{\mathbf{h}\to 0}{\to}\widetilde{\phi}(x)\mathcal{H}(x)\Pi_{j=1}^{d}\left(\int_{\mathrm{S}}\mathcal{S}_j(u^j)\mathrm{d}u^j\right),$$

combining equations (A.2) and (A.3), we obtain

$$\alpha_{T,\mathbf{h}}(x)\Rightarrow C_X\widetilde{\phi}(x)\mathcal{H}(x)\Pi_{j=1}^{d}\left(\int_{\mathrm{S}}\mathcal{S}_j(u^j)\mathrm{d}u^j\right)\mathrm{g}\alpha,$$

as $h\to 0$ and $T\to\infty$ jointly.

As an aside, Proposition 4.3 in LL (2009) also implies that the constant C_X is

$$C_X=\frac{\mathrm{E}_\lambda\left(\int_{R_n}^{R_{n+1}}g(X_s)\,\mathrm{d}s\right)}{<\phi,g>}, \tag{A.4}$$

for any measurable ϕ-integrable function $g(.)$. This is easily obtained when replacing $\frac{1}{\mathbf{h}}\mathcal{S}\left(\frac{.-x}{\mathbf{h}}\right)\mathcal{H}(.)$ with $g(.)$ in equation (A.3). (An alternative way to see this is as follows. For a fixed bandwidth $\mathbf{h}$, the ratio-limit theorem (c.f., Theorem 1) gives

$$\lim_{T\to\infty}\frac{\sum_{n=1}^{N_T-1}\frac{1}{\mathbf{h}}\int_{R_n}^{R_{n+1}}\mathcal{S}\left(\frac{X_s-x}{\mathbf{h}}\right)\mathcal{H}(X_s)\mathrm{d}s}{\sum_{n=1}^{N_T-1}\int_{R_n}^{R_{n+1}}g(X_s)\mathrm{d}s}=\frac{\left\langle\phi,\frac{1}{\mathbf{h}}\mathcal{S}\left(\frac{.-x}{\mathbf{h}}\right)\mathcal{H}(.)\right\rangle}{<\phi,g>}\ \text{a.s.},$$

given a generic ϕ-integrable function $g(.)$. By the strong law of large numbers, because of the stationarity and strong mixing properties of the increments between R_n and R_{n+1}, again for a fixed bandwidth $\mathbf{h}$, we also have

$$\lim_{T\to\infty}\frac{\sum_{n=1}^{N_T-1}\frac{1}{\mathbf{h}}\int_{R_n}^{R_{n+1}}\mathcal{S}\left(\frac{X_s-x}{\mathbf{h}}\right)\mathcal{H}(X_s)\mathrm{d}s}{\sum_{n=1}^{N_T-1}\int_{R_n}^{R_{n+1}}g(X_s)\mathrm{d}s}=\frac{\mathrm{E}_\lambda\left(\frac{1}{\mathbf{h}}\int_{R_n}^{R_{n+1}}\mathcal{S}\left(\frac{X_s-x}{\mathbf{h}}\right)\mathcal{H}(X_s)\mathrm{d}s\right)}{\mathrm{E}_\lambda\left(\int_{R_n}^{R_{n+1}}g(X_s)\,\mathrm{d}s\right)}\ \text{a.s.}$$

Because the left-hand sides of the two equations are the same, by equating the right-hand sides, we obtain

$$\mathrm{E}_\lambda\left(\frac{1}{\mathbf{h}}\int_{R_n}^{R_{n+1}}\mathcal{S}\left(\frac{X_s-x}{\mathbf{h}}\right)\mathcal{H}(X_s)\mathrm{d}s\right)=\underbrace{\frac{C_X}{\frac{\mathrm{E}_\lambda\left(\int_{R_n}^{R_{n+1}}g(X_s)\mathrm{d}s\right)}{<\phi,g>}}}\left\langle\phi,\frac{1}{\mathbf{h}}\mathcal{S}\left(\frac{.-x}{\mathbf{h}}\right)\mathcal{H}(.)\right\rangle,$$

which leads, once more, to equation (A.3) and an explicit characterization of the process-specific constant C_X as in equation (A.4).)

https://doi.org/10.1017/S0266466617000305 Published online by Cambridge University Press

We now turn to $\beta_{T,\mathbf{h}}(x)$. Since

$$\mathbb{E}_\lambda\left(\frac{1}{v(T)\mathbf{h}}\int_0^T \mathcal{S}\left(\frac{X_s-x}{\mathbf{h}}\right)\mathcal{H}(X_s)\mathrm{d}s\right)=\mathrm{E}_\lambda\left(\frac{1}{v(T)\mathbf{h}}\int_0^T \mathcal{S}\left(\frac{X_s-x}{\mathbf{h}}\right)\mathcal{H}(X_s)\mathrm{d}s\right),$$

we have

$$\begin{aligned}
&P_\lambda\left(|\beta_{T,\mathbf{h}}(x)|\geq\varepsilon\right)\\
&\quad\leq\frac{1}{\varepsilon^2}\mathrm{E}_\lambda\left[\frac{1}{v(T)\mathbf{h}}\int_0^T \mathcal{S}\left(\frac{X_s-x}{\mathbf{h}}\right)\mathcal{H}(X_s)\mathrm{d}s-\mathrm{E}_\lambda\left(\frac{1}{v(T)\mathbf{h}}\int_0^T \mathcal{S}\left(\frac{X_s-x}{\mathbf{h}}\right)\mathcal{H}(X_s)\mathrm{d}s\right)\right]^2\\
&\quad=\frac{1}{\varepsilon^2}\frac{1}{v^2(T)}\int_0^T\int_0^T \mathrm{Cov}\left(\frac{1}{\mathbf{h}}\mathcal{S}\left(\frac{X_s-x}{\mathbf{h}}\right)\mathcal{H}(X_s),\frac{1}{\mathbf{h}}\mathcal{S}\left(\frac{X_t-x}{\mathbf{h}}\right)\mathcal{H}(X_t)\right)\mathrm{d}s\mathrm{d}t\\
&\quad=\frac{1}{\varepsilon^2}\frac{1}{v^2(T)}\int_0^T\int_0^T\left[\int\int\frac{1}{\mathbf{h}}\mathcal{S}\left(\frac{a-x}{\mathbf{h}}\right)\mathcal{H}(a)\frac{1}{\mathbf{h}}\mathcal{S}\left(\frac{b-x}{\mathbf{h}}\right)\mathcal{H}(b)\left(\widetilde{p}_{s,t}(a,b)-\widetilde{p}_s(a)\widetilde{p}_t(b)\right)\mathrm{d}a\mathrm{d}b\right]\mathrm{d}s\mathrm{d}t\\
&\quad\leq\frac{1}{\varepsilon^2}\frac{1}{v(T)h^{2d(1-1/q)}}\left(\int\int\frac{1}{h^d}\mathcal{S}^q\left(\frac{a-x}{\mathbf{h}}\right)\mathcal{H}^q(a)\frac{1}{h^d}\mathcal{S}^q\left(\frac{b-x}{\mathbf{h}}\right)\mathcal{H}^q(b)\mathrm{d}a\mathrm{d}b\right)^{1/q}\\
&\qquad\times\frac{1}{v(T)}\int_0^T\int_0^T\|\widetilde{g}_{s,t}(a,b)\|_r\mathrm{d}s\mathrm{d}t,
\end{aligned}$$

where the first inequality is Chebyshev's and the second is Hölder's for $r=\frac{1}{1-1/q}$. Now, since

$$\lim_{T\to\infty}\frac{1}{v(T)}\int_0^T\int_0^T\|\widetilde{g}_{s,t}(a,b)\|_r\mathrm{d}s\mathrm{d}t<\infty,$$

by Assumption 3.1, we have

$$P_\lambda\left(\left|\beta_{T,\mathbf{h}}(x)\right|\geq\varepsilon\right)\to 0,$$

as $v(T)h^{2d(1-1/q)}=v(T)h^{2d/r}\to\infty$.

Finally, we consider $\gamma_{T,\mathbf{h}}(x)$. After defining $R_T=\inf\{R_n:R_n>S_{N_T+1}\}$, given that $R_T>T$, we obtain

$$\begin{aligned}
&\mathbb{E}_\lambda\left(\frac{1}{v(T)\mathbf{h}}\int_0^T \mathcal{S}\left(\frac{X_s-x}{\mathbf{h}}\right)\mathcal{H}(X_s)\mathrm{d}s\right)-\frac{N_T}{v(T)}\mathbb{E}_\lambda\left(\frac{1}{\mathbf{h}}\int_{R_n}^{R_{n+1}}\mathcal{S}\left(\frac{X_s-x}{\mathbf{h}}\right)\mathcal{H}(X_s)\mathrm{d}s\right)\\
&\quad=\mathbb{E}_\lambda\left(\frac{1}{v(T)\mathbf{h}}\int_0^{R_1} \mathcal{S}\left(\frac{X_s-x}{\mathbf{h}}\right)\mathcal{H}(X_s)\mathrm{d}s\right)\\
&\qquad+\left[\mathbb{E}_\lambda\left(\frac{1}{v(T)\mathbf{h}}\int_{R_1}^{R_T} \mathcal{S}\left(\frac{X_s-x}{\mathbf{h}}\right)\mathcal{H}(X_s)\mathrm{d}s\right)-\frac{N_T}{v(T)}\mathbb{E}_\lambda\left(\frac{1}{\mathbf{h}}\int_{R_n}^{R_{n+1}}\mathcal{S}\left(\frac{X_s-x}{\mathbf{h}}\right)\mathcal{H}(X_s)\mathrm{d}s\right)\right]\\
&\qquad-\mathbb{E}_\lambda\left(\frac{1}{v(T)\mathbf{h}}\int_T^{R_T} \mathcal{S}\left(\frac{X_s-x}{\mathbf{h}}\right)\mathcal{H}(X_s)\mathrm{d}s\right)\\
&\quad=\mathbb{E}_\lambda\left(\frac{1}{v(T)\mathbf{h}}\int_0^{R_1} \mathcal{S}\left(\frac{X_s-x}{\mathbf{h}}\right)\mathcal{H}(X_s)\mathrm{d}s\right)\\
&\qquad+\left[\mathbb{E}_\lambda\left(\frac{1}{v(T)\mathbf{h}}\sum_{n\geq 1}\mathbf{1}_{\{S_n\leq T\}}\int_{R_n}^{R_{n+1}} \mathcal{S}\left(\frac{X_s-x}{\mathbf{h}}\right)\mathcal{H}(X_s)\mathrm{d}s\right)-\frac{N_T}{v(T)}\mathbb{E}_\lambda\left(\frac{1}{\mathbf{h}}\int_{R_n}^{R_{n+1}}\mathcal{S}\left(\frac{X_s-x}{\mathbf{h}}\right)\mathcal{H}(X_s)\mathrm{d}s\right)\right]\\
&\qquad-\mathbb{E}_\lambda\left(\frac{1}{v(T)\mathbf{h}}\int_T^{R_T} \mathcal{S}\left(\frac{X_s-x}{\mathbf{h}}\right)\mathcal{H}(X_s)\mathrm{d}s\right)\\
&\quad=\gamma^1_{T,\mathbf{h}}(x)+\gamma^2_{T,\mathbf{h}}(x)+\gamma^3_{T,\mathbf{h}}(x).
\end{aligned}\tag{A.5}$$

https://doi.org/10.1017/S0266466617000305 Published online by Cambridge University Press

The driving term is clearly $\gamma^2_{T,\mathbf{h}}(x)$ and we turn to it. Notice that

$$\begin{aligned}
&\mathbb{E}_\lambda\left(\frac{1}{v(T)\mathbf{h}}\sum_{n\geq 1}\mathbf{1}_{\{S_n\leq T\}}\int_{R_n}^{R_{n+1}}\mathcal{S}\left(\frac{X_s-x}{\mathbf{h}}\right)\mathcal{H}(X_s)\mathrm{d}s\right)-\frac{N_T}{v(T)}\mathbb{E}_\lambda\left(\frac{1}{\mathbf{h}}\int_{R_n}^{R_{n+1}}\mathcal{S}\left(\frac{X_s-x}{\mathbf{h}}\right)\mathcal{H}(X_s)\mathrm{d}s\right)\\
&=\frac{\mathbb{E}_\lambda\left(\sum_{n\geq 1}\mathbf{1}_{\{S_n\leq T\}}\right)}{v(T)}\mathbb{E}_\lambda\left(\frac{1}{\mathbf{h}}\int_{R_n}^{R_{n+1}}\mathcal{S}\left(\frac{X_s-x}{\mathbf{h}}\right)\mathcal{H}(X_s)\mathrm{d}s\right)-\frac{N_T}{v(T)}\mathbb{E}_\lambda\left(\frac{1}{\mathbf{h}}\int_{R_n}^{R_{n+1}}\mathcal{S}\left(\frac{X_s-x}{\mathbf{h}}\right)\mathcal{H}(X_s)\mathrm{d}s\right)\\
&=\frac{\mathbb{E}_\lambda\left(\frac{1}{\mathbf{h}}\int_{R_n}^{R_{n+1}}\mathcal{S}\left(\frac{X_s-x}{\mathbf{h}}\right)\mathcal{H}(X_s)\mathrm{d}s\right)}{v(T)}\left[\mathbb{E}_\lambda\left[N_T\right]-N_T\right],
\end{aligned}$$

where the first equality derives from the fact that (*i*) X_{R_n+s} is independent of $\{S_n\leq T\}$ for all n and $s>0$ by the construction leading to Nummelin splitting (Proposition 2.13 in LL, 2008) and (*ii*) $\mathbb{E}_\lambda\left(\frac{1}{\mathbf{h}}\int_{R_n}^{R_{n+1}}\mathcal{S}\left(\frac{X_s-x}{\mathbf{h}}\right)\mathcal{H}(X_s)\mathrm{d}s\right)$ is independent of n by the stationarity of the increments $U_{n,\mathbf{h}}(x)$, as discussed above.

Now, since $N_T-\mathbb{E}_\lambda\left[N_T\right]$ is a "compensated" additive functional of the process Z, we have $\left[N_T-\mathbb{E}_\lambda\left[N_T\right]\right]=O_p\left(\sqrt{v(T)}\right)$ by the second order law in Theorem 3.16 of Höpfner and Löcherbach (2003). (For an analogous result in discrete time, see Chen (2000, Theorem 1.3).) This, in turn, implies that

$$\gamma^2_{T,\mathbf{h}}(x)=O_p\left(\frac{1}{\sqrt{v(T)}}\right).$$

As for the end terms, we note that, given the nonnegativity of the $\mathcal{S}$ and $\mathcal{H}$ functions, by the same argument as for equation (A.3),

$$\gamma^3_{T,\mathbf{h}}(x)\leq\frac{1}{v(T)}\mathbb{E}_\lambda\left(\frac{1}{\mathbf{h}}\int_{R_{T-1}}^{R_T}\mathcal{S}\left(\frac{X_s-x}{\mathbf{h}}\right)\mathcal{H}(X_s)\mathrm{d}s\right)=O\left(\frac{1}{v(T)}\right),$$

where R_{T-1} is expressed as $R_{T-1}=\sup\{R_n:R_n<T\}$. The other end term $\gamma^1_{T,\mathbf{h}}(x)$ is similarly of order $O(1/v(T))$. One way to see this, using the properties of the construction explicitly, is the following. The measure λ is generic. Set $\lambda=\zeta$ with $\zeta\sim\phi(\cdot\cap C)$, where the set C was defined earlier. Then,

$$\begin{aligned}
&\mathbb{E}_\lambda\left(\frac{1}{\mathbf{h}}\int_0^{R_1}\mathcal{S}\left(\frac{X_s-x}{\mathbf{h}}\right)\mathcal{H}(X_s)\mathrm{d}s\right)\\
&=\mathrm{E}_\zeta\left(\frac{1}{\mathbf{h}}\int_0^{R_1}\mathcal{S}\left(\frac{X_s-x}{\mathbf{h}}\right)\mathcal{H}(X_s)\mathrm{d}s\right)\\
&=C_X\left\langle\phi,\frac{1}{\mathbf{h}}\mathcal{S}\left(\frac{.-x}{\mathbf{h}}\right)\mathcal{H}(.)\right\rangle\underset{\mathbf{h}\to 0}{\to}\widetilde{\phi}(x)\mathcal{H}(x)\Pi_{j=1}^d\left(\int_{\mathrm{S}}\mathcal{S}_j(u^j)\mathrm{d}u^j\right),
\end{aligned}$$

as in the case of equation (A.3), where the second equality derives from Proposition 4.3 in LL (2009). Thus, $\gamma^1_{T,\mathbf{h}}(x)=O\left(\frac{1}{v(T)}\right)$ and $\gamma^3_{T,\mathbf{h}}(x)=O\left(\frac{1}{v(T)}\right)$. Putting everything together, we have

$$\begin{aligned}
\frac{1}{v(T)}L_{T,\mathbf{h}}(x)=&\underbrace{\alpha_{T,\mathbf{h}}(x)}_{\Rightarrow C_X\widetilde{\phi}(x)\mathcal{H}(x)\Pi_{j=1}^d\left(\int_{\mathrm{S}}\mathcal{S}_j(u^j)\mathrm{d}u^j\right)\mathrm{g}_\alpha}+\underbrace{\beta_{T,\mathbf{h}}(x)}_{=O_p\left(\frac{1}{v^{1/2}(T)h^{d(1-1/q)}}\right)}+\underbrace{\gamma_{T,\mathbf{h}}(x)}_{=O_p\left(\frac{1}{v^{1/2}(T)}\right)}\\
&\Rightarrow C_X\widetilde{\phi}(x)\mathcal{H}(x)\Pi_{j=1}^d\left(\int_{\mathrm{S}}\mathcal{S}_j(u^j)\mathrm{d}u^j\right)\mathrm{g}_\alpha, \qquad \textbf{(A.6)}
\end{aligned}$$

as $h\to 0$ and $T\to\infty$ jointly with $v(T)h^{2d/r}\to\infty$. This concludes the proof. ■

https://doi.org/10.1017/S0266466617000305 Published online by Cambridge University Press

Remark A.1. Set $\mathcal{H}(x) = 1$. The theorem extends the Darling-Kac theorem (Theorem 2 in the article) to the case of a bounded function $f_{\mathbf{h}}(.) = \frac{1}{\mathbf{h}}\mathcal{S}\left(\frac{\cdot - x}{\mathbf{h}}\right)$ changing with the bandwidth. For this case, we provide conditions for joint limits over h and T. Just like for the Darling-Kac theorem, our results specialize to the ergodic case. For the ergodic case, the resulting bandwidth conditions are the same as those in the literature (see, e.g., Castellana and Leadbetter, 1986, or Bosq, 1995).

COROLLARY TO LEMMA A.1. *Let $\mathcal{H}(x) = 1$. Assume the function $\mathcal{S}(x)$ is the kernel function $\mathbf{K}(x)$ in Assumption 2. If $v(T)\mathbf{h}^{2/r} \to \infty$, then*

$$\frac{1}{\mathbf{h}v(T)}\int_0^T \mathbf{K}\left(\frac{X_s - x}{\mathbf{h}}\right)\mathrm{d}s \Rightarrow C_X\widetilde{\phi}(x)\,\mathrm{g}_\alpha,$$

as $h \to 0$ and $T \to \infty$ jointly.

This result, which is a weak convergence theorem for multivariate *occupation densities*, readily follows from observing that $\Pi_{j=1}^d \left(\int_{\mathrm{S}} K(u^j)\mathrm{d}u^j\right) = 1$.

Remark A.2 (The rate of divergence of the occupation densities)**.** As emphasized in, e.g., Remarks 3 and 4, the rate $v(T)$ is T for positive recurrent or strictly stationary processes irrespective of their dimension. In the null recurrent case, however, the rate is known only for a handful of processes. In addition, such a rate is expected to become slower, thereby generating the "curse of dimensionality" discussed in Remark 14, as the dimensionality of the system increases. In general, dimensionality increases slow down recurrence in null recurrent systems. To this extent, we recall that $v(T) = \sqrt{T}$ in the case of scalar Brownian motion but $v(T) = \log T$ *only* in the planar Brownian case. For dimensions equal to 3 or higher than 3, Brownian motion is known not to be recurrent. For a discussion regarding multivariate kernels in the presence of unit-root processes which reaches identical conclusions in discrete time, we refer the reader to Wang (2015), Example 5.1, page 208.

Remark A.3. The case with a generic product function $\mathcal{S}$ and a generic function $\mathcal{H}$ satisfying the properties of Lemma A.1 will be helpful in the remaining of this treatment.

LEMMA A.2. *The function $\mathcal{G}(x)$ is a product function $\Pi_{j=1}^d \mathcal{G}_j(x^j)$. Each $\mathcal{G}_j$ is a non-negative, bounded, and continuous function defined on a compact set* S. *Each $\mathcal{G}_j$ is twice-continuously differentiable with bounded derivatives over* S. *The function $\mathcal{N}$ is also a bounded, twice-continuously differentiable function with bounded derivatives over* S. *Let Assumptions 1, 3, and 4 be satisfied. Then,*

$$\frac{\Delta_{n,T}}{\mathbf{h}_{n,T}v(T)}\sum_{i=1}^n \mathcal{G}\left(\frac{X_{i\Delta_{n,T}} - x}{\mathbf{h}_{n,T}}\right)\mathcal{N}(X_{i\Delta_{n,T}}) - \frac{1}{\mathbf{h}_{n,T}v(T)}\int_0^T \mathcal{G}\left(\frac{X_s - x}{\mathbf{h}_{n,T}}\right)\mathcal{N}(X_s)\mathrm{d}s = O_p\left(\frac{\Delta_{n,T}}{h_{n,T}^2}\right).$$

Proof of Lemma A.2. Write

$$\begin{aligned}
&\frac{1}{\mathbf{h}_{n,T}}\sum_{i=1}^n \int_{(i-1)\Delta_{n,T}}^{i\Delta_{n,T}} \left[\mathcal{G}\left(\frac{X_{(i-1)\Delta_{n,T}} - x}{\mathbf{h}_{n,T}}\right)\mathcal{N}(X_{(i-1)\Delta_{n,T}}) - \mathcal{G}\left(\frac{X_s - x}{\mathbf{h}_{n,T}}\right)\mathcal{N}(X_s)\right]\mathrm{d}s \\
&= \frac{1}{\mathbf{h}_{n,T}}\sum_{i=1}^n \int_{(i-1)\Delta_{n,T}}^{i\Delta_{n,T}} \left[\prod_{j=1}^d \mathcal{G}_j\left(\frac{X^j_{(i-1)\Delta_{n,T}} - x}{h_{n,T}}\right)\mathcal{N}(X_{(i-1)\Delta_{n,T}}) - \prod_{j=1}^d \mathcal{G}_j\left(\frac{X^j_s - x}{h_{n,T}}\right)\mathcal{N}(X_s)\right]\mathrm{d}s.
\end{aligned}$$

https://doi.org/10.1017/S0266466617000305 Published online by Cambridge University Press

Given a multivariate Itô's expansion (see, e.g., Karatzas and Shreve, 1991, Theorem 3.6, page 153), the last expression becomes

$$\frac{1}{\mathbf{h}_{n,T}}\sum_{i=1}^{n}\int_{(i-1)\Delta_{n,T}}^{i\Delta_{n,T}}\left[\prod_{j=1}^{d}\mathcal{G}_j\left(\frac{X^j_{(i-1)\Delta_{n,T}}-x}{h_{n,T}}\right)\mathcal{N}(X_{(i-1)\Delta_{n,T}})-\prod_{j=1}^{d}\mathcal{G}_j\left(\frac{X^j_s-x}{h_{n,T}}\right)\mathcal{N}(X_s)\right]\mathrm{d}s$$
$$=(\Phi_1+\Phi_2+\Phi_3+\Phi_4)+\left(\overline{\Phi}_1+\overline{\Phi}_2+\overline{\Phi}_3+\overline{\Phi}_4\right)+\overline{\Phi}_5,$$

where

$$\begin{aligned}
&\Phi_1+\Phi_2+\Phi_3+\Phi_4\\
&=-\frac{1}{h^{d+1}_{n,T}}\sum_{i=1}^{n}\int_{(i-1)\Delta_{n,T}}^{i\Delta_{n,T}}\left[\int_{(i-1)\Delta_{n,T}}^{s}\mathcal{N}(X_v)\sum_{j=1}^{d}\mathcal{G}'_j\left(\frac{X^j_v-x}{h_{n,T}}\right)\prod_{u\neq j}^{d}\mathcal{G}_u\left(\frac{X^u_v-x}{h_{n,T}}\right)\mu_j(X_v)\,\mathrm{d}v\right]\mathrm{d}s\\
&\quad-\frac{1}{2h^{d+2}_{n,T}}\sum_{i=1}^{n}\int_{(i-1)\Delta_{n,T}}^{i\Delta_{n,T}}\left[\int_{(i-1)\Delta_{n,T}}^{s}\mathcal{N}(X_v)\sum_{j=1}^{d}\mathcal{G}''_{jj}\left(\frac{X^j_v-x}{h_{n,T}}\right)\prod_{u\neq j}^{d}\mathcal{G}_u\left(\frac{X^u_v-x}{h_{n,T}}\right)\left(\sum_{k=1}^{m}\sigma^2_{jk}(X_v)\,\mathrm{d}v\right)\right]\mathrm{d}s\\
&\quad-\frac{1}{2h^{d+2}_{n,T}}\sum_{i=1}^{n}\int_{(i-1)\Delta_{n,T}}^{i\Delta_{n,T}}\left[\int_{(i-1)\Delta_{n,T}}^{s}\mathcal{N}(X_v)\sum_{j=1}^{d}\sum_{q\neq j}^{d}\mathcal{G}'_j\left(\frac{X^j_v-x}{h_{n,T}}\right)\mathcal{G}'_q\left(\frac{X^q_v-x}{h_{n,T}}\right)\prod_{u\neq j,q}^{d}\mathcal{G}_u\left(\frac{X^u_v-x}{h_{n,T}}\right)\right.\\
&\qquad\left.\times\left(\sum_{k=1}^{m}\sigma_{jk}(X_v)\,\sigma_{qk}(X_v)\,\mathrm{d}v\right)\right]\mathrm{d}s\\
&\quad-\frac{1}{h^{d+1}_{n,T}}\sum_{i=1}^{n}\int_{(i-1)\Delta_{n,T}}^{i\Delta_{n,T}}\left[\int_{(i-1)\Delta_{n,T}}^{s}\mathcal{N}(X_v)\sum_{j=1}^{d}\mathcal{G}'_j\left(\frac{X^j_v-x}{h_{n,T}}\right)\prod_{u\neq j}^{d}\mathcal{G}_u\left(\frac{X^u_v-x}{h_{n,T}}\right)\sum_{k=1}^{m}\sigma_{jk}(X_v)\,\mathrm{d}B^k_v\right]\mathrm{d}s,
\end{aligned}$$

$$\begin{aligned}
&\overline{\Phi}_1+\overline{\Phi}_2+\overline{\Phi}_3+\overline{\Phi}_4\\
&=-\frac{1}{h^{d}_{n,T}}\sum_{i=1}^{n}\int_{(i-1)\Delta_{n,T}}^{i\Delta_{n,T}}\left[\int_{(i-1)\Delta_{n,T}}^{s}\prod_{j=1}^{d}\mathcal{G}_j\left(\frac{X^j_v-x}{h_{n,T}}\right)\sum_{j=1}^{d}\mathcal{N}'_j(X_v)\mu_j(X_v)\,\mathrm{d}v\right]\mathrm{d}s\\
&\quad-\frac{1}{2h^{d}_{n,T}}\sum_{i=1}^{n}\int_{(i-1)\Delta_{n,T}}^{i\Delta_{n,T}}\left[\int_{(i-1)\Delta_{n,T}}^{s}\prod_{j=1}^{d}\mathcal{G}_j\left(\frac{X^j_v-x}{h_{n,T}}\right)\sum_{j=1}^{d}\mathcal{N}''_{jj}(X_v)\left(\sum_{k=1}^{m}\sigma^2_{jk}(X_v)\,\mathrm{d}v\right)\right]\mathrm{d}s\\
&\quad-\frac{1}{2h^{d}_{n,T}}\sum_{i=1}^{n}\int_{(i-1)\Delta_{n,T}}^{i\Delta_{n,T}}\left[\int_{(i-1)\Delta_{n,T}}^{s}\prod_{j=1}^{d}\mathcal{G}_j\left(\frac{X^j_v-x}{h_{n,T}}\right)\sum_{j=1}^{d}\sum_{q\neq j}^{d}\mathcal{N}''_{jq}(X_v)\right.\\
&\qquad\left.\times\left(\sum_{k=1}^{m}\sigma_{jk}(X_v)\,\sigma_{qk}(X_v)\,\mathrm{d}v\right)\right]\mathrm{d}s\\
&\quad-\frac{1}{h^{d}_{n,T}}\sum_{i=1}^{n}\int_{(i-1)\Delta_{n,T}}^{i\Delta_{n,T}}\left[\int_{(i-1)\Delta_{n,T}}^{s}\prod_{j=1}^{d}\mathcal{G}_j\left(\frac{X^j_v-x}{h_{n,T}}\right)\sum_{j=1}^{d}\mathcal{N}'_j(X_v)\sum_{k=1}^{m}\sigma_{jk}(X_v)\,\mathrm{d}B^k_v\right]\mathrm{d}s,
\end{aligned}$$

and

$$\begin{aligned}
\overline{\Phi}_5&=-\frac{2}{2h^{d+1}_{n,T}}\sum_{i=1}^{n}\int_{(i-1)\Delta_{n,T}}^{i\Delta_{n,T}}\left[\int_{(i-1)\Delta_{n,T}}^{s}\sum_{j=1}^{d}\mathcal{G}'_j\left(\frac{X^j_v-x}{h_{n,T}}\right)\mathcal{N}'_j(X_v)\prod_{u\neq j}^{d}\mathcal{G}_u\left(\frac{X^u_v-x}{h_{n,T}}\right)\left(\sum_{k=1}^{m}\sigma^2_{jk}(X_v)\,\mathrm{d}v\right)\right]\mathrm{d}s\\
&\quad-\frac{2}{2h^{d+1}_{n,T}}\sum_{i=1}^{n}\int_{(i-1)\Delta_{n,T}}^{i\Delta_{n,T}}\left[\int_{(i-1)\Delta_{n,T}}^{s}\sum_{j=1}^{d}\sum_{q\neq j}^{d}\mathcal{G}'_j\left(\frac{X^j_v-x}{h_{n,T}}\right)\mathcal{N}'_q(X_v)\prod_{u\neq j}^{d}\mathcal{G}_u\left(\frac{X^u_v-x}{h_{n,T}}\right)\right.\\
&\qquad\left.\times\left(\sum_{k=1}^{m}\sigma_{jk}(X_v)\,\sigma_{qk}(X_v)\,\mathrm{d}v\right)\right]\mathrm{d}s.
\end{aligned}$$

https://doi.org/10.1017/S0266466617000305 Published online by Cambridge University Press

Now, write

$$|\Phi_1| \leq \frac{\Delta_{n,T}}{h_{n,T}^{d+1}} \sum_{i=1}^{n} \int_{(i-1)\Delta_{n,T}}^{i\Delta_{n,T}} \left[|\mathcal{N}(X_s)| \sum_{j=1}^{d} \left| \mathcal{G}_j' \left(\frac{X_s^j - x}{h_{n,T}} \right) \right| \prod_{u \neq j}^{d} \mathcal{G}_u \left(\frac{X_s^u - x}{h_{n,T}} \right) |\mu_j(X_s)| \right] \mathrm{d}s$$
$$= \frac{\Delta_{n,T}}{h_{n,T}^{d+1}} \int_0^T \left[|\mathcal{N}(X_s)| \sum_{j=1}^{d} \left| \mathcal{G}_j' \left(\frac{X_s^j - x}{h_{n,T}} \right) \right| \prod_{u \neq j}^{d} \mathcal{G}_u \left(\frac{X_s^u - x}{h_{n,T}} \right) |\mu_j(X_s)| \right] \mathrm{d}s.$$

Similarly,

$$|\Phi_2| \leq \frac{\Delta_{n,T}}{2h_{n,T}^{d+2}} \sum_{i=1}^{n} \int_{(i-1)\Delta_{n,T}}^{i\Delta_{n,T}} \left[|\mathcal{N}(X_s)| \sum_{j=1}^{d} \left| \mathcal{G}_{jj}'' \left(\frac{X_s^j - x}{h_{n,T}} \right) \right| \prod_{u \neq j}^{d} \mathcal{G}_u \left(\frac{X_s^u - x}{h_{n,T}} \right) \sum_{k=1}^{m} \sigma_{jk}^2(X_s) \right] \mathrm{d}s$$
$$= \frac{\Delta_{n,T}}{2h_{n,T}^{d+2}} \int_0^T \left[|\mathcal{N}(X_s)| \sum_{j=1}^{d} \left| \mathcal{G}_{jj}'' \left(\frac{X_s^j - x}{h_{n,T}} \right) \right| \prod_{u \neq j}^{d} \mathcal{G}_u \left(\frac{X_s^u - x}{h_{n,T}} \right) \sum_{k=1}^{m} \sigma_{jk}^2(X_s) \right] \mathrm{d}s$$

and

$$|\Phi_3| \leq \frac{\Delta_{n,T}}{2h_{n,T}^{d+2}} \sum_{i=1}^{n} \int_{(i-1)\Delta_{n,T}}^{i\Delta_{n,T}} \left[|\mathcal{N}(X_s)| \sum_{j=1}^{d} \sum_{q \neq j}^{d} \left| \mathcal{G}_j' \left(\frac{X_s^j - x}{h_{n,T}} \right) \right| \left| \mathcal{G}_q' \left(\frac{X_s^q - x}{h_{n,T}} \right) \right| \prod_{u \neq j,q}^{d} \mathcal{G}_u \left(\frac{X_s^u - x}{h_{n,T}} \right) \right.$$
$$\left. \times \sum_{k=1}^{m} \sigma_{jk}(X_s) \sigma_{qk}(X_s) \right] \mathrm{d}s$$
$$= \frac{\Delta_{n,T}}{2h_{n,T}^{d+2}} \int_0^T \left[|\mathcal{N}(X_s)| \sum_{j=1}^{d} \sum_{q \neq j}^{d} \left| \mathcal{G}_j' \left(\frac{X_s^j - x}{h_{n,T}} \right) \right| \left| \mathcal{G}_q' \left(\frac{X_s^q - x}{h_{n,T}} \right) \right| \prod_{u \neq j,q}^{d} \mathcal{G}_u \left(\frac{X_s^u - x}{h_{n,T}} \right) \right.$$
$$\left. \times \sum_{k=1}^{m} \sigma_{jk}(X_s) \sigma_{qk}(X_s) \right] \mathrm{d}s.$$

Before turning to Φ_4, we show that $\Phi_1 = O_p\left(\frac{\Delta_{n,T}}{h_{n,T}} v(T)\right)$ as $h_{n,T} \to 0$ with $\mathbf{h}_{n,T}^{2/r} v(T) \to \infty$. Given the boundedness of the first derivative of the functions $\mathcal{G}_j$, for all $1 \leq j \leq d$, writing $\mathcal{S}\left(\frac{X_s - x}{\mathbf{h}}\right) = \left| \mathcal{G}_j' \left(\frac{X_s^j - x}{h_{n,T}} \right) \right| \prod_{u \neq j}^{d} \left[\mathcal{G}_u \left(\frac{X_s^u - x}{h_{n,T}} \right) \right]$ and $\mathcal{H}(X_s) = |\mu_j(X_s)| |\mathcal{N}(X_s)|$, Lemma A.1 implies that, for all $1 \leq j \leq d$,

$$\frac{1}{\mathbf{h}_{n,T} v(T)} \int_0^T |\mathcal{N}(X_s)| \left| \mathcal{G}_j' \left(\frac{X_s^j - x}{h_{n,T}} \right) \right| \prod_{u \neq j}^{d} \mathcal{G}_u \left(\frac{X_s^u - x}{h_{n,T}} \right) |\mu_j(X_s)| \, \mathrm{d}s$$
$$\Rightarrow \mathrm{g}_\alpha(1) \widetilde{\phi}(x) |\mu_j(x)| |\mathcal{N}(x)| \left(\int_{\mathrm{S}} \left| \mathcal{G}_j' \left(v^j \right) \right| \mathrm{d}v^j \right) \left(\prod_{u \neq j}^{d} \int_{\mathrm{S}} \mathcal{G}_u(v^u) \, \mathrm{d}v^u \right),$$

provided $h_{n,T} \to 0$ and $T \to \infty$ jointly, with $\mathbf{h}_{n,T}^{2/r} v(T) \to \infty$. Hence,

$$\frac{1}{\mathbf{h}_{n,T} v(T)} \int_0^T |\mathcal{N}(X_s)| \left| \mathcal{G}_j' \left(\frac{X_s^j - x}{h_{n,T}} \right) \right| \prod_{u \neq j}^{d} \mathcal{G}_u \left(\frac{X_s^u - x}{h_{n,T}} \right) |\mu_j(X_s)| \, \mathrm{d}s = O_p(1),$$

which gives the probability order of Φ_1, namely $\frac{\Delta_{n,T}}{h_{n,T}} v(T)$.

https://doi.org/10.1017/S0266466617000305 Published online by Cambridge University Press

In the exact same way, we have $\Phi_2 = O_p\left(\frac{\Delta_{n,T}}{h_{n,T}^2} v(T)\right)$ and $\Phi_3 = O_p\left(\frac{\Delta_{n,T}}{h_{n,T}^2} v(T)\right)$. The first probability order is an implication of Lemma A.1 with

$$\mathcal{S}\left(\frac{X_s - x}{\mathbf{h}}\right) = \left|\mathcal{G}_{jj}''\left(\frac{X_s^j - x}{h_{n,T}}\right)\right| \prod_{u\neq j}^d \mathcal{G}_u\left(\frac{X_s^u - x}{h_{n,T}}\right),$$
$$\mathcal{H}(X_s) = |\mathcal{N}(X_s)| \sum_{k=1}^m \sigma_{jk}^2(X_s),$$

for all $1 \leq j \leq d$. The second is, again, an implication of Lemma A.1 with

$$\mathcal{S}\left(\frac{X_s - x}{\mathbf{h}}\right) = \left|\mathcal{G}_j'\left(\frac{X_s^j - x}{h_{n,T}}\right)\right| \left|\mathcal{G}_q'\left(\frac{X_s^q - x}{h_{n,T}}\right)\right| \prod_{u\neq j,q}^d \mathcal{G}_u\left(\frac{X_s^u - x}{h_{n,T}}\right),$$
$$\mathcal{H}(X_s) = |\mathcal{N}(X_s)| \sum_{k=1}^m \sigma_{jk}(X_s)\sigma_{qk}(X_s),$$

for all $1 \leq j \leq d$ and all $1 \leq q \leq d$ with $q \neq j$. Next, integrating by parts, and for all $1 \leq j \leq d$,

$$\Phi_4^j = -\frac{1}{h_{n,T}^{d+1}} \sum_{i=1}^n \int_{(i-1)\Delta_{n,T}}^{i\Delta_{n,T}} (i\Delta_{n,T} - s)\mathcal{N}(X_s)\mathcal{G}_j'\left(\frac{X_s^j - x}{h_{n,T}}\right) \prod_{u\neq j}^d \mathcal{G}_u\left(\frac{X_s^u - x}{h_{n,T}}\right) \sum_{k=1}^m \sigma_{jk}(X_s)\,\mathrm{d}B_s^k.$$

Now, define Φ_{4n}^j as a partial sum of the first $n-1$ terms of Φ_4^j plus a stochastic integral between $(n-1)\Delta_{n,T}$ and $n\Delta_{n,T}$:

$$\Phi_{4n}^j = \Phi_{4(n-1)}^j + \frac{1}{h_{n,T}^{d+1}} \int_{(n-1)\Delta_{n,T}}^{n\Delta_{n,T}} (n\Delta_{n,T} - s)\mathcal{N}(X_s)\mathcal{G}_j'\left(\frac{X_s^j - x}{h_{n,T}}\right) \prod_{u\neq j}^d \mathcal{G}_u\left(\frac{X_s^u - x}{h_{n,T}}\right) \sum_{k=1}^m \sigma_{jk}(X_s)\,\mathrm{d}B_s^k.$$

Equivalently, define Φ_{4t}^j as a partial sum of the first $b-1$ terms of Φ_4^j, with $b = 1, \ldots, n-1$, plus a stochastic integral between $(b-1)\Delta_{n,T}$ and t, where $(b-1)\Delta_{n,T} \leq t \leq b\Delta_{n,T}$:

$$\Phi_{4t}^j = \Phi_{4(b-1)}^j + \frac{1}{h_{n,T}^{d+1}} \int_{(b-1)\Delta_{n,T}}^{t} (b\Delta_{n,T} - s)\mathcal{N}(X_s)\mathcal{G}_j'\left(\frac{X_s^j - x}{h_{n,T}}\right) \prod_{u\neq j}^d \mathcal{G}_u\left(\frac{X_s^u - x}{h_{n,T}}\right) \sum_{k=1}^m \sigma_{jk}(X_s)\,\mathrm{d}B_s^k. \tag{A.7}$$

The quadratic variation of the continuous martingale Φ_{4t}^j can be expressed as follows:

$$[\Phi_4^j]_t = [\Phi_4^j]_{(b-1)} + \frac{1}{h_{n,T}^{2d+2}} \int_{(b-1)\Delta_{n,T}}^t (b\Delta_{n,T} - s)^2 \mathcal{N}^2(X_s) \left(\mathcal{G}_j'\left(\frac{X_s^j - x}{h_{n,T}}\right)\right)^2$$
$$\times \prod_{u\neq j}^d \mathcal{G}_u^2\left(\frac{X_s^u - x}{h_{n,T}}\right) \sum_{k=1}^m \sigma_{jk}^2(X_s)\,\mathrm{d}s$$
$$\leq \frac{\Delta_{n,T}^2}{h_{n,T}^{2d+2}} \int_0^t \mathcal{N}^2(X_s) \left(\mathcal{G}_j'\left(\frac{X_s^j - x}{h_{n,T}}\right)\right)^2 \prod_{u\neq j}^d \mathcal{G}_u^2\left(\frac{X_s^u - x}{h_{n,T}}\right) \sum_{k=1}^m \sigma_{jk}^2(X_s)\,\mathrm{d}s.$$

By Lemma A.1, with

$$\mathcal{S}\left(\frac{X_s - x}{\mathbf{h}}\right) = \left(\mathcal{G}_j'\left(\frac{X_s^j - x}{h_{n,T}}\right)\right)^2 \prod_{u\neq j}^d \mathcal{G}_u^2\left(\frac{X_s^u - x}{h_{n,T}}\right)$$

https://doi.org/10.1017/S0266466617000305 Published online by Cambridge University Press

and

$$\mathcal{H}(X_s) = \mathcal{N}^2(X_s) \sum_{k=1}^{m} \sigma_{jk}^2(X_s),$$

we have that $[\Phi_4^j]_T = O_p\left(\frac{h_{n,T}^d \Delta_{n,T}^2}{h_{n,T}^{2d+2}} v(T)\right)$ for all $1 \leq j \leq d$ and, equivalently, $\Phi_4 = O_p\left(\frac{\Delta_{n,T}}{h_{n,T}^{(d+2)/2}} v^{1/2}(T)\right)$ provided $\mathbf{h}_{n,T}^{2/r} v(T) \to \infty$.

Using the same method of proof as for the terms Φ_1, Φ_2, and Φ_3, it is now immediate to show that $\overline{\Phi}_1 = O_p(h_{n,T}\Phi_1)$, $\overline{\Phi}_2 = O_p(h_{n,T}^2\Phi_2)$, $\overline{\Phi}_3 = O_p(h_{n,T}^2\Phi_3)$, and $\overline{\Phi}_5 = O_p(\Phi_1)$. Using the same method of proof as for term Φ_4, it is also immediate to prove that $\overline{\Phi}_4 = O_p(h_{n,T}\Phi_4)$.

In sum,

$$\frac{\Delta_{n,T}}{\mathbf{h}_{n,T} v(T)} \sum_{i=1}^{n} \mathcal{G}\left(\frac{X_{(i-1)\Delta_{n,T}} - x}{\mathbf{h}_{n,T}}\right) \mathcal{N}(X_{(i-1)\Delta_{n,T}}) - \frac{1}{\mathbf{h}_{n,T} v(T)} \int_0^T \mathcal{G}\left(\frac{X_s - x}{\mathbf{h}_{n,T}}\right) \mathcal{N}(X_s) \mathrm{d}s$$
$$= O_p\left(\frac{\Delta_{n,T}}{h_{n,T}^2}\right) + O_p\left(\frac{\Delta_{n,T}}{h_{n,T}^{(d+2)/2} v^{1/2}(T)}\right). \tag{A.8}$$

The result also applies to $\frac{\Delta_{n,T}}{\mathbf{h}_{n,T} v(T)} \sum_{i=1}^{n} \mathcal{G}\left(\frac{X_{i\Delta_{n,T}} - x}{\mathbf{h}_{n,T}}\right) \mathcal{N}(X_{i\Delta_{n,T}})$ (which does not include X_0 but includes X_T) since the difference between

$$\frac{\Delta_{n,T}}{\mathbf{h}_{n,T} v(T)} \sum_{i=1}^{n} \mathcal{G}\left(\frac{X_{i\Delta_{n,T}} - x}{\mathbf{h}_{n,T}}\right) \mathcal{N}(X_{i\Delta_{n,T}})$$

and

$$\frac{\Delta_{n,T}}{\mathbf{h}_{n,T} v(T)} \sum_{i=1}^{n} \mathcal{G}\left(\frac{X_{(i-1)\Delta_{n,T}} - x}{\mathbf{h}_{n,T}}\right) \mathcal{N}\left(X_{(i-1)\Delta_{n,T}}\right)$$

is, given the boundedness of the functions $\mathcal{G}$ and $\mathcal{N}$, of order $O\left(\frac{\Delta_{n,T}}{\mathbf{h}_{n,T} v(T)}\right)$ and, hence, negligible asymptotically. It is negligible asymptotically since $\frac{\Delta_{n,T}}{\mathbf{h}_{n,T} v(T)} = \frac{\Delta_{n,T}}{h_{n,T}^2 h_{n,T}^{d-2} v(T)}$ goes to zero faster than $\frac{\Delta_{n,T}}{h_{n,T}^2}$ if $h_{n,T}^{d-2} v(T) \to \infty$, which we assume. Similarly, it should be noted that the assumption $h_{n,T}^{d-2} v(T) \to \infty$ implies that $\frac{\Delta_{n,T}}{h_{n,T}^2} \to 0$ is, also, more stringent than $\frac{\Delta_{n,T}}{h_{n,T}^{(d+2)/2} v^{1/2}(T)} = \frac{\Delta_{n,T}}{h_{n,T}^2 h_{n,T}^{d/2-1} v^{1/2}(T)} \to 0$. Hence, one only requires $\frac{\Delta_{n,T}}{h_{n,T}^2} \to 0$. Therefore,

$$\frac{\Delta_{n,T}}{\mathbf{h}_{n,T} v(T)} \sum_{i=1}^{n} \mathcal{G}\left(\frac{X_{i\Delta_{n,T}} - x}{\mathbf{h}_{n,T}}\right) \mathcal{N}(X_{i\Delta_{n,T}}) - \frac{1}{\mathbf{h}_{n,T} v(T)} \int_0^T \mathcal{G}\left(\frac{X_s - x}{\mathbf{h}_{n,T}}\right) \mathcal{N}(X_s) \mathrm{d}s = O_p\left(\frac{\Delta_{n,T}}{h_{n,T}^2}\right),$$

which proves the stated result. ■

https://doi.org/10.1017/S0266466617000305 Published online by Cambridge University Press

Proof of Theorem 3. By the Corollary to Lemma A.1, $\frac{1}{\mathbf{h}_{n,T}v(T)}\int_0^T \mathbf{K}\left(\frac{X_s-x}{\mathbf{h}_{n,T}}\right)\mathrm{d}s \Rightarrow C_X\widetilde{\phi}(x)\,\mathrm{g}_\alpha$ if $\mathbf{h}_{n,T}^{2/r}v(T)\to\infty$. By Lemma A.2, assuming $\mathcal{G}=\mathbf{K}$ and $\mathcal{N}=1$, we have

$$\frac{\Delta_{n,T}}{\mathbf{h}_{n,T}v(T)}\sum_{i=1}^{n}\mathbf{K}\left(\frac{X_{i\Delta_{n,T}}-x}{\mathbf{h}_{n,T}}\right)=\underbrace{\frac{1}{\mathbf{h}_{n,T}v(T)}\int_0^T\mathbf{K}\left(\frac{X_s-x}{\mathbf{h}_{n,T}}\right)\mathrm{d}s}_{\Rightarrow C_X\widetilde{\phi}(x)\mathrm{g}_\alpha}+\underbrace{O_p\left(\frac{\Delta_{n,T}}{h_{n,T}^2}\right)}_{\to 0},$$

provided $\mathbf{h}_{n,T}^{2/r}v(T)\to\infty$ and $\frac{\Delta_{n,T}}{h_{n,T}^2}\to 0$. ■

Proof of Theorem 4. Write

$$\widehat{\boldsymbol{\mu}}_{n,T}(x)=\frac{1}{\Delta_{n,T}}\frac{\sum_{i=1}^{n-1}\mathbf{K}\left(\frac{X_{i\Delta_{n,T}}-x}{\mathbf{h}_{n,T}}\right)\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}\boldsymbol{\mu}(X_s)\,\mathrm{d}s}{\sum_{i=1}^{n}\mathbf{K}\left(\frac{X_{i\Delta_{n,T}}-x}{\mathbf{h}_{n,T}}\right)} \tag{A.9}$$

$$+\frac{1}{\Delta_{n,T}}\frac{\sum_{i=1}^{n-1}\mathbf{K}\left(\frac{X_{i\Delta_{n,T}}-x}{\mathbf{h}_{n,T}}\right)\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}\boldsymbol{\sigma}(X_s)\,\mathrm{d}\mathbf{B}_s}{\sum_{i=1}^{n}\mathbf{K}\left(\frac{X_{i\Delta_{n,T}}-x}{\mathbf{h}_{n,T}}\right)}. \tag{A.10}$$

Express the estimation error decomposition as

$$\underbrace{(\mathrm{A.9})-\boldsymbol{\mu}(x)}_{\text{bias term}}+\underbrace{(\mathrm{A.10})}_{\text{variance term}}.$$

First, we concentrate on equation (A.10). The process

$$\Theta_{n,T}(u)=\frac{1}{\sqrt{\mathbf{h}_{n,T}}}\sum_{i=1}^{l-2}\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}\mathbf{K}\left(\frac{X_{i\Delta_{n,T}}-x}{\mathbf{h}_{n,T}}\right)\boldsymbol{\sigma}(X_s)\,\mathrm{d}\mathbf{B}_s$$
$$+\frac{1}{\sqrt{\mathbf{h}_{n,T}}}\int_{(l-1)\Delta_{n,T}}^{u}\mathbf{K}\left(\frac{X_{(l-1)\Delta_{n,T}}-x}{\mathbf{h}_{n,T}}\right)\sigma(X_s)\,\mathrm{d}\mathbf{B}_s,$$

with $u\in[(l-1)\Delta_{n,T},l\Delta_{n,T}]$ and $l=2,\ldots,n$ is a continuous martingale so that $\Theta_{n,T}(T)$ is equal to the numerator of equation (A.10) rescaled by $\sqrt{\mathbf{h}_{n,T}}$. Consider, now, the generic element g of the d-vector $\Theta_{n,T}(u)$, i.e., $\Theta_{n,T}^g(u)$, and its quadratic variation:

$$[\Theta_{n,T}^g]_u=\frac{1}{\mathbf{h}_{n,T}}\sum_{i=1}^{l-2}\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}\mathbf{K}^2\left(\frac{X_{i\Delta_{n,T}}-x}{\mathbf{h}_{n,T}}\right)\sum_{k=1}^{m}\sigma_{gk}^2(X_s)\,\mathrm{d}s$$
$$+\frac{1}{\mathbf{h}_{n,T}}\int_{(l-1)\Delta_{n,T}}^{u}\mathbf{K}^2\left(\frac{X_{(l-1)\Delta_{n,T}}-x}{\mathbf{h}_{n,T}}\right)\sum_{k=1}^{m}\sigma_{gk}^2(X_s)\,\mathrm{d}s.$$

Thus,

$$[\Theta_{n,T}^g]_T=\frac{1}{\mathbf{h}_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}^2\left(\frac{X_{i\Delta_{n,T}}-x}{\mathbf{h}_{n,T}}\right)\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}\sum_{k=1}^{m}\sigma_{gk}^2(X_s)\,\mathrm{d}s$$

or, equivalently,

https://doi.org/10.1017/S0266466617000305 Published online by Cambridge University Press

$$
\begin{aligned}
[\Theta^g_{n,T}]_T &= \frac{\Delta_{n,T}}{\mathbf{h}_{n,T}} \sum_{i=1}^{n-1} \mathbf{K}^2 \left(\frac{X_{i\Delta_{n,T}} - x}{\mathbf{h}_{n,T}} \right) \sum_{k=1}^{m} \sigma^2_{gk} \left(X_{i\Delta_{n,T}} \right) \\
&\quad + \frac{1}{\mathbf{h}_{n,T}} \sum_{i=1}^{n-1} \mathbf{K}^2 \left(\frac{X_{i\Delta_{n,T}} - x}{\mathbf{h}_{n,T}} \right) \int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}} \left[\left(\sum_{k=1}^{m} \sigma^2_{gk} \left(X_s \right) \right) - \left(\sum_{k=1}^{m} \sigma^2_{gk} \left(X_{i\Delta_{n,T}} \right) \right) \right] \mathrm{d}s \\
&= \left[\Theta^{g,I}_{n,T} \right]_T + \left[\Theta^{g,II}_{n,T} \right]_T . \qquad \textbf{(A.11)}
\end{aligned}
$$

Next, after writing $\sum_{k=1}^{m} \sigma^2_{gk} \left(X_s \right) = \mathcal{P}_g (X_s)$, by Itô's lemma, we have

$$
\begin{aligned}
[\Theta^{g,II}_{n,T}]_T &= \frac{1}{\mathbf{h}_{n,T}} \sum_{i=1}^{n-1} \mathbf{K}^2 \left(\frac{X_{i\Delta_{n,T}} - x}{\mathbf{h}_{n,T}} \right) \int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}} \left[\mathcal{P}_g (X_s) - \mathcal{P}_g (X_{i\Delta_{n,T}}) \right] \mathrm{d}s \\
&= \frac{1}{\mathbf{h}_{n,T}} \sum_{i=1}^{n-1} \mathbf{K}^2 \left(\frac{X_{i\Delta_{n,T}} - x}{\mathbf{h}_{n,T}} \right) \int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}} \left(\int_{i\Delta_{n,T}}^{s} \sum_{j=1}^{d} \mathcal{P}'_{g,j} (X_v) \mu_j (X_v) \, \mathrm{d}v \right) \mathrm{d}s \\
&\quad + \frac{1}{2\mathbf{h}_{n,T}} \sum_{i=1}^{n-1} \mathbf{K}^2 \left(\frac{X_{i\Delta_{n,T}} - x}{\mathbf{h}_{n,T}} \right) \int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}} \left(\int_{i\Delta_{n,T}}^{s} \sum_{j=1}^{d} \mathcal{P}''_{g,jj} (X_v) \sum_{k=1}^{m} \sigma^2_{jk} (X_v) \, \mathrm{d}v \right) \mathrm{d}s \\
&\quad + \frac{1}{2\mathbf{h}_{n,T}} \sum_{i=1}^{n-1} \mathbf{K}^2 \left(\frac{X_{i\Delta_{n,T}} - x}{\mathbf{h}_{n,T}} \right) \int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}} \left(\int_{i\Delta_{n,T}}^{s} \sum_{j=1}^{d} \sum_{q \neq j}^{d} \mathcal{P}''_{g,jq} (X_v) \sum_{k=1}^{m} \sigma_{jk} (X_v) \sigma_{qk} (X_v) \, \mathrm{d}v \right) \mathrm{d}s \\
&\quad + \frac{1}{\mathbf{h}_{n,T}} \sum_{i=1}^{n-1} \mathbf{K}^2 \left(\frac{X_{i\Delta_{n,T}} - x}{\mathbf{h}_{n,T}} \right) \int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}} \left(\int_{i\Delta_{n,T}}^{s} \sum_{j=1}^{d} \mathcal{P}'_{g,j} (X_v) \sum_{k=1}^{m} \sigma_{jk} (X_v) \, \mathrm{d}B^k_v \right) \mathrm{d}s .
\end{aligned}
$$

And, integrating by parts,

$$
\begin{aligned}
[\Theta^{g,II}_{n,T}]_T &= \frac{1}{\mathbf{h}_{n,T}} \sum_{i=1}^{n-1} \mathbf{K}^2 \left(\frac{X_{i\Delta_{n,T}} - x}{\mathbf{h}_{n,T}} \right) \int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}} ((i+1)\Delta_{n,T} - s) \sum_{j=1}^{d} \mathcal{P}'_{g,j} (X_s) \mu_j (X_s) \, \mathrm{d}s \\
&\quad + \frac{1}{2\mathbf{h}_{n,T}} \sum_{i=1}^{n-1} \mathbf{K}^2 \left(\frac{X_{i\Delta_{n,T}} - x}{\mathbf{h}_{n,T}} \right) \int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}} ((i+1)\Delta_{n,T} - s) \sum_{j=1}^{d} \mathcal{P}''_{g,jj} (X_s) \sum_{k=1}^{m} \sigma^2_{jk} (X_s) \, \mathrm{d}s \\
&\quad + \frac{1}{2\mathbf{h}_{n,T}} \sum_{i=1}^{n-1} \mathbf{K}^2 \left(\frac{X_{i\Delta_{n,T}} - x}{\mathbf{h}_{n,T}} \right) \int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}} ((i+1)\Delta_{n,T} - s) \sum_{j=1}^{d} \sum_{q \neq j}^{d} \mathcal{P}''_{g,jq} (X_s) \\
&\quad \times \sum_{k=1}^{m} \sigma_{jk} (X_s) \sigma_{qk} (X_s) \, \mathrm{d}s \\
&\quad + \frac{1}{\mathbf{h}_{n,T}} \sum_{i=1}^{n-1} \mathbf{K}^2 \left(\frac{X_{i\Delta_{n,T}} - x}{\mathbf{h}_{n,T}} \right) \int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}} ((i+1)\Delta_{n,T} - s) \sum_{j=1}^{d} \mathcal{P}'_{g,j} (X_s) \sum_{k=1}^{m} \sigma_{jk} (X_s) \, \mathrm{d}B^k_s .
\end{aligned}
$$

Now, notice that, for all $1 \leq j \leq d$, since

$$
\mathcal{P}'_{g,j} (X_s) \mu_j (X_s) \leq |\mathcal{P}'_{g,j} (X_s)| |\mu_j (X_s)| \leq C_{|\mathcal{P}'_{g,j}| |\mu_j|},
$$

where $C_{|\mathcal{P}'_{g,j}| |\mu_j|}$ is a suitable constant by Assumption 4.1, the first term in $[\Theta^{g,II}_{n,T}]_T$ can be bounded by

$$
\begin{aligned}
&\frac{1}{\mathbf{h}_{n,T}} \sum_{i=1}^{n-1} \mathbf{K}^2 \left(\frac{X_{i\Delta_{n,T}} - x}{\mathbf{h}_{n,T}} \right) \int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}} \left((i+1)\Delta_{n,T} - s \right) \mathcal{P}'_j (X_s) \mu_j (X_s) \, \mathrm{d}s \\
&\quad \leq C_{|\mathcal{P}'_{g,j}| |\mu_j|} \frac{\Delta^2_{n,T}}{\mathbf{h}_{n,T}} \sum_{i=1}^{n-1} \mathbf{K}^2 \left(\frac{X_{i\Delta_{n,T}} - x}{\mathbf{h}_{n,T}} \right) .
\end{aligned}
$$

https://doi.org/10.1017/S0266466617000305 Published online by Cambridge University Press

However, by Lemmas A.1 and A.2, $C_{|\mathcal{P}'_{g,j}\|\mu_j|}\frac{\Delta^2_{n,T}}{\mathbf{h}_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}^2\left(\frac{X_{i\Delta_{n,T}}-x}{\mathbf{h}_{n,T}}\right)$ is of order $O_p(\Delta_{n,T}v(T))$. The second and the third term in $[\Theta^{g,II}_{n,T}]_T$ can be treated in just the same way.

As for the fourth term in $[\Theta^{g,II}_{n,T}]_T$, after expressing each component (indexed by j with $1\leq j\leq d$) as a continuous martingale like in the case of equation (A.7) above, the j-th component's quadratic variation can be represented as follows:

$$\frac{1}{\mathbf{h}^2_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}^4\left(\frac{X_{i\Delta_{n,T}}-x}{\mathbf{h}_{n,T}}\right)\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}\left((i+1)\Delta_{n,T}-s\right)^2\left(\mathcal{P}'_{g,j}(X_s)\right)^2\sum_{k=1}^{m}\sigma^2_{jk}(X_s)\,\mathrm{d}s$$
$$\leq C_{\left(\mathcal{P}'_{g,j}\right)^2\sum_{k=1}^m\sigma^2_{jk}}\frac{\Delta^3_{n,T}}{\mathbf{h}^2_{n,T}}\sum_{i=1}^{n}\mathbf{K}^4\left(\frac{X_{i\Delta_{n,T}}-x}{\mathbf{h}_{n,T}}\right).$$

Given Lemmas A.1 and A.2, the probability order of this term is $O_p\left(\frac{\Delta^2_{n,T}v(T)}{\mathbf{h}_{n,T}}\right)$. Hence, the probability order of the fourth term in $[\Theta^{g,II}_{n,T}]_T$ is $O_p\left(\frac{\Delta_{n,T}v^{1/2}(T)}{h^{d/2}_{n,T}}\right)$. Therefore,

$$[\Theta^{g,II}_{n,T}]_T=O_p(\Delta_{n,T}v(T))+O_p\left(\frac{\Delta_{n,T}v^{1/2}(T)}{h^{d/2}_{n,T}}\right). \tag{A.12}$$

In sum,

$$\begin{aligned}
&\frac{[\Theta^{g}_{n,T}]_T}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}\left(\frac{X_{i\Delta_{n,T}}-x}{\mathbf{h}_{n,T}}\right)}\\
&=\frac{[\Theta^{g,I}_{n,T}]_T+[\Theta^{g,II}_{n,T}]_T}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}\left(\frac{X_{i\Delta_{n,T}}-x}{\mathbf{h}_{n,T}}\right)}\\
&=\frac{\frac{\Delta_{n,T}}{\mathbf{h}_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}^2\left(\frac{X_{i\Delta_{n,T}}-x}{\mathbf{h}_{n,T}}\right)\sum_{k=1}^{m}\sigma^2_{gk}\left(X_{i\Delta_{n,T}}\right)+O_p(\Delta_{n,T}v(T))+O_p\left(\frac{\Delta_{n,T}v^{1/2}(T)}{h^{d/2}_{n,T}}\right)}{\frac{\Delta_{n,T}}{\mathbf{h}_{n,T}}\sum_{i=1}^{n}\mathbf{K}\left(\frac{X_{i\Delta_{n,T}}-x}{\mathbf{h}_{n,T}}\right)}\\
&=\frac{\frac{1}{\mathbf{h}_{n,T}}\int_0^T\mathbf{K}^2\left(\frac{X_s-x}{\mathbf{h}_{n,T}}\right)\sum_{k=1}^{m}\sigma^2_{gk}(X_s)\,\mathrm{d}s+O_p\left(\frac{\Delta_{n,T}v(T)}{h^2_{n,T}}\right)}{\frac{1}{\mathbf{h}_{n,T}}\int_0^T\mathbf{K}\left(\frac{X_s-x}{\mathbf{h}_{n,T}}\right)\mathrm{d}s+O_p\left(\frac{\Delta_{n,T}v(T)}{h^2_{n,T}}\right)},
\end{aligned} \tag{A.13}$$

where equation (A.13) follows from Lemma A.2. One simply has to apply Lemma A.2 to the numerator (with $\mathcal{G}(.)=\mathbf{K}^2(.)$ and $\mathcal{N}(.)=\sum_{k=1}^m\sigma^2_{gk}(.)$) and to the denominator (with $\mathcal{G}(.)=\mathbf{K}(.)$ and $\mathcal{N}(.)=1$).

We can now invoke the same decomposition as in equation (A.1) with $\mathcal{S}(.)=\mathbf{K}^2(.)$ and $\mathcal{H}(.)=\sum_{k=1}^m\sigma^2_{gk}(.)$ in the numerator and with $\mathcal{S}(.)=\mathbf{K}(.)$ and $\mathcal{H}(.)=1$ in the denominator, to obtain

https://doi.org/10.1017/S0266466617000305 Published online by Cambridge University Press

$$\frac{\frac{1}{\mathbf{h}_{n,T}}\int_0^T \mathbf{K}^2\left(\frac{X_s-x}{\mathbf{h}_{n,T}}\right)\sum_{k=1}^m \sigma_{gk}^2(X_s)\,\mathrm{d}s}{\frac{1}{\mathbf{h}_{n,T}}\int_0^T \mathbf{K}\left(\frac{X_s-x}{\mathbf{h}_{n,T}}\right)\mathrm{d}s}$$

$$= \frac{N_T \mathrm{E}_\lambda\left(\frac{1}{\mathbf{h}}\int_{R_n}^{R_{n+1}} \mathbf{K}^2\left(\frac{X_s-x}{\mathbf{h}}\right)\sum_{k=1}^m \sigma_{gk}^2(X_s)\,\mathrm{d}s\right) + O_p\left(\frac{v^{1/2}(T)}{h^{d(1-1/q)}}\right) + O_p\left(v^{1/2}(T)\right)}{N_T \mathrm{E}_\lambda\left(\frac{1}{\mathbf{h}}\int_{R_n}^{R_{n+1}} \mathbf{K}\left(\frac{X_s-x}{\mathbf{h}}\right)\mathrm{d}s\right) + O_p\left(\frac{v^{1/2}(T)}{h^{d(1-1/q)}}\right) + O_p\left(v^{1/2}(T)\right)}$$

$$= \frac{\mathrm{E}_\lambda\left(\frac{1}{\mathbf{h}}\int_{R_n}^{R_{n+1}} \mathbf{K}^2\left(\frac{X_s-x}{\mathbf{h}}\right)\sum_{k=1}^m \sigma_{gk}^2(X_s)\,\mathrm{d}s\right) + \frac{v(T)}{N_T}O_p\left(\frac{1}{v^{1/2}(T)h^{d(1-1/q)}}\right) + \frac{v(T)}{N_T}O_p\left(\frac{1}{v^{1/2}(T)}\right)}{\mathrm{E}_\lambda\left(\frac{1}{\mathbf{h}}\int_{R_n}^{R_{n+1}} \mathbf{K}\left(\frac{X_s-x}{\mathbf{h}}\right)\mathrm{d}s\right) + \frac{v(T)}{N_T}O_p\left(\frac{1}{v^{1/2}(T)h^{d(1-1/q)}}\right) + \frac{v(T)}{N_T}O_p\left(\frac{1}{v^{1/2}(T)}\right)} \quad \textbf{(A.14)}$$

$$\xrightarrow{p} \left(\int_{\mathrm{S}} \mathbf{K}^2(u)\,\mathrm{d}u\right)\left(\sum_{k=1}^m \sigma_{gk}^2(x)\right)$$

$$= V, \quad \textbf{(A.15)}$$

if $\mathbf{h}_{n,T}^{2/r}v(T) \to \infty$. To further justify equation (A.14), we recall that $\frac{N_{sT}}{v(T)}$ converges weakly to a strictly positive Mittag-Leffler process (c.f., equation (A.2)) and the order terms derive from the analogues of $\gamma_{T,\mathbf{h}}(x)$ and $\beta_{T,\mathbf{h}}(x)$ in equation (A.6).

Under Assumption 4.1, using Theorem 3.16 of Höpfner and Löcherbach (2003) or Proposition 1 of Feigin (1985), we now have stable convergence of pairs, (See, also, Touati (1992), Theorem 1.) so that

$$\left(\frac{[\Theta_{n,T}^g]_T}{v(T)}, \frac{\Theta_{n,T}^g(T)}{\sqrt{v(T)}}\right) \Rightarrow \left(C_X \mathbf{b}^g(x)\widetilde{\phi}(x)\,\mathrm{g}_\alpha, \sqrt{C_X \mathbf{b}^g(x)\widetilde{\phi}(x)}\,(\mathrm{N}\circ\mathrm{g}_\alpha)\right),$$

where $\mathbf{b}^g(x) = \left(\int_{\mathrm{S}} \mathbf{K}^2(u)\,\mathrm{d}u\right)\left(\sum_{k=1}^m \sigma_{gk}^2(x)\right)$ and $\mathrm{N}\circ\mathrm{g}_\alpha$ is a Gaussian mixture with variance g_α. Clearly, $\frac{[\Theta_{n,T}^g]_T}{v(T)} \Rightarrow C_X \mathbf{b}^g(x)\widetilde{\phi}(x)\,\mathrm{g}_\alpha$ is, also, an implication of Lemma A.1 with $\mathcal{S}(.) = \mathbf{K}^2(.)$ and $\mathcal{H}(.) = \sum_{k=1}^m \sigma_{gk}^2(.)$. Similarly, using equation (A.15),

$$\left(\frac{[\Theta_{n,T}^g]_T}{\frac{\Delta_{n,T}}{\mathbf{h}_{n,T}}\sum_{i=1}^n \mathbf{K}\left(\frac{X_{i\Delta_{n,T}}-x}{\mathbf{h}_{n,T}}\right)}, \frac{\Theta_{n,T}^g(T)}{\sqrt{\frac{\Delta_{n,T}}{\mathbf{h}_{n,T}}\sum_{i=1}^n \mathbf{K}\left(\frac{X_{i\Delta_{n,T}}-x}{\mathbf{h}_{n,T}}\right)}}\right) \Rightarrow \left(V, V^{1/2}\mathrm{N}\right), \quad \textbf{(A.16)}$$

where N is a standard Gaussian distribution. This proves the result for a generic element g.

We now turn to the multivariate case. Write

$$\frac{\Theta_{n,T}(u)}{\sqrt{\frac{\Delta_{n,T}}{\mathbf{h}_{n,T}}\sum_{i=1}^n \mathbf{K}\left(\frac{X_{i\Delta_{n,T}}-x}{\mathbf{h}_{n,T}}\right)}} = \frac{1}{\sqrt{\frac{\Delta_{n,T}}{\mathbf{h}_{n,T}}\sum_{i=1}^n \mathbf{K}\left(\frac{X_{i\Delta_{n,T}}-x}{\mathbf{h}_{n,T}}\right)}}\left(\Theta_{n,T}^{g=1}(u), \Theta_{n,T}^{g=2}(u), \ldots, \Theta_{n,T}^{g=d}(u)\right)^\intercal,$$

where the generic $\Theta_{n,T}^g(u)$ was defined earlier. Again, by Lemmas A.1 and A.2, we have

$$\frac{\left[\Theta_{n,T}^f, \Theta_{n,T}^s\right]_T}{\frac{\Delta_{n,T}}{\mathbf{h}_{n,T}}\sum_{i=1}^n \mathbf{K}\left(\frac{X_{i\Delta_{n,T}}-x}{\mathbf{h}_{n,T}}\right)} \xrightarrow{p} \left(\int_{\mathrm{S}} \mathbf{K}^2(u)\,\mathrm{d}u\right)\left(\sum_{k=1}^m \sigma_{fk}(x)\,\sigma_{sk}(x)\right) = V_{f,s}, \qquad \forall f,s = 1,\ldots,d. \quad \textbf{(A.17)}$$

Next, we orthogonalize the continuous martingales in $\Theta_{n,T}(u)$ by writing

$$\mathbf{C}^\intercal(x)\left(\Theta_{n,T}^{g=1}(u), \Theta_{n,T}^{g=2}(u), \ldots, \Theta_{n,T}^{g=d}(u)\right)^\intercal = \left(\overline{\Theta}_{n,T}^{g=1}(u), \overline{\Theta}_{n,T}^{g=2}(u), \ldots, \overline{\Theta}_{n,T}^{g=d}(u)\right)^\intercal,$$

where $\mathbf{C}(x)$ is a $d \times d$ matrix so that

https://doi.org/10.1017/S0266466617000305 Published online by Cambridge University Press

$$\mathbf{C}^{\intercal}(x)\underbrace{\left(\left(\int_{\mathrm{S}}\mathbf{K}^2(u)\,\mathrm{d}u\right)\mathbf{a}(x)\right)}_{\mathbf{b}(x)}\mathbf{C}(x)=\Lambda(x)$$

with $\Lambda(x)$ diagonal. $\mathbf{C}(x)$ and $\Lambda(x)$ are the matrices containing the eigenvectors and the eigenvalues of $\mathbf{b}(x)$, respectively. Since

$$\frac{\left[\overline{\Theta}^f_{n,T},\overline{\Theta}^s_{n,T}\right]_T}{\frac{\Delta_{n,T}}{\mathbf{h}_{n,T}}\sum_{i=1}^n\mathbf{K}\left(\frac{X_{i\Delta_{n,T}}-x}{\mathbf{h}_{n,T}}\right)}\xrightarrow{p}0,\qquad\forall f,s=1,\ldots,d,$$

we use equation (A.17) and apply, again, Theorem 3.16 of Höpfner and Löcherbach (2003) or Proposition 1 of Feigin (1985) to obtain

$$\left(\frac{1}{\frac{\Delta_{n,T}}{\mathbf{h}_{n,T}}\sum_{i=1}^n\mathbf{K}\left(\frac{X_{i\Delta_{n,T}}-x}{\mathbf{h}_{n,T}}\right)}[\overline{\Theta}_{n,T}]_T,\frac{1}{\sqrt{\frac{\Delta_{n,T}}{\mathbf{h}_{n,T}}\sum_{i=1}^n\mathbf{K}\left(\frac{X_{i\Delta_{n,T}}-x}{\mathbf{h}_{n,T}}\right)}}\left(\overline{\Theta}^{g=1}_{n,T}(T),\overline{\Theta}^{g=2}_{n,T}(T),\ldots,\overline{\Theta}^{g=d}_{n,T}(T)\right)^{\intercal}\right)$$
$$\Rightarrow\left(\Lambda,\Lambda^{1/2}\mathrm{N}\right),$$

where $[\overline{\Theta}_{n,T}]_T$ is naturally used here to denote the quadratic covariation matrix with generic element $\left[\overline{\Theta}^f_{n,T},\overline{\Theta}^s_{n,T}\right]_T$ and N is a vector standard normal. This implies

$$\frac{1}{\sqrt{\frac{\Delta_{n,T}}{\mathbf{h}_{n,T}}\sum_{i=1}^n\mathbf{K}\left(\frac{X_{i\Delta_{n,T}}-x}{\mathbf{h}_{n,T}}\right)}}\mathbf{C}(x)\left(\overline{\Theta}^{g=1}_{n,T}(T),\overline{\Theta}^{g=2}_{n,T}(T),\ldots,\overline{\Theta}^{g=d}_{n,T}(T)\right)^{\intercal}$$
$$\Rightarrow\mathrm{N}\left(0,\mathbf{C}(x)\Lambda(x)\mathbf{C}^{\intercal}(x)\right)\overset{d}{=}\mathrm{N}\left(0,\mathbf{b}(x)\right).$$

Finally,

$$\left(\sqrt{h^d_{n,T}\widehat{L}_{n,T}(x)}\right)\frac{\sum_{i=1}^{n-1}\mathbf{K}\left(\frac{X_{i\Delta_{n,T}}-x}{\mathbf{h}_{n,T}}\right)\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}\boldsymbol{\sigma}(X_s)\,\mathrm{d}\mathbf{B}_s}{\Delta_{n,T}\sum_{i=1}^n\mathbf{K}\left(\frac{X_{i\Delta_{n,T}}-x}{\mathbf{h}_{n,T}}\right)}\Rightarrow\mathrm{N}\left(0,\left(\int_{\mathrm{S}}K^2(u)\,\mathrm{d}s\right)^d\mathbf{a}(x)\right),$$

as $h_{n,T}\to0$ so that $\frac{\Delta_{n,T}}{h^2_{n,T}}\to0$ and $h^{d-2}_{n,T}v(T)\to\infty$ with $\mathbf{h}^{2/r}_{n,T}v(T)\to\infty$.

Next, we discuss the bias term. Note that

$$\underbrace{(\mathrm{A}.9)-\boldsymbol{\mu}(x)}_{\text{bias term}}$$
$$=\frac{1}{\Delta_{n,T}}\frac{\sum_{i=1}^{n-1}\mathbf{K}\left(\frac{X_{i\Delta_{n,T}}-x}{\mathbf{h}_{n,T}}\right)\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}\boldsymbol{\mu}(X_s)\,\mathrm{d}s}{\sum_{i=1}^n\mathbf{K}\left(\frac{X_{i\Delta_{n,T}}-x}{\mathbf{h}_{n,T}}\right)}-\boldsymbol{\mu}(x)$$
$$=\frac{\frac{\Delta_{n,T}}{\mathbf{h}_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}\left(\frac{X_{i\Delta_{n,T}}-x}{\mathbf{h}_{n,T}}\right)\boldsymbol{\mu}\left(X_{i\Delta_{n,T}}\right)+\frac{1}{\mathbf{h}_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}\left(\frac{X_{i\Delta_{n,T}}-x}{\mathbf{h}_{n,T}}\right)\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}\left(\boldsymbol{\mu}(X_s)-\boldsymbol{\mu}\left(X_{i\Delta_{n,T}}\right)\right)\mathrm{d}s}{\frac{\Delta_{n,T}}{\mathbf{h}_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}\left(\frac{X_{i\Delta_{n,T}}-x}{\mathbf{h}_{n,T}}\right)}$$

https://doi.org/10.1017/S0266466617000305 Published online by Cambridge University Press

$$
\begin{aligned}
&-\boldsymbol{\mu}(x)\\
&=\frac{\frac{\Delta_{n,T}}{\mathbf{h}_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}\left(\frac{X_{i\Delta_{n,T}}-x}{\mathbf{h}_{n,T}}\right)\boldsymbol{\mu}\left(X_{i\Delta_{n,T}}\right)+O_p(\Delta_{n,T}v(T))+O_p\left(\frac{\Delta_{n,T}v^{1/2}(T)}{h_{n,T}^{d/2}}\right)}{\frac{\Delta_{n,T}}{\mathbf{h}_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}\left(\frac{X_{i\Delta_{n,T}}-x}{\mathbf{h}_{n,T}}\right)} \qquad \text{(A.18)}\\
&-\boldsymbol{\mu}(x)\\
&=\frac{\frac{1}{\mathbf{h}_{n,T}}\int_0^T\mathbf{K}\left(\frac{X_s-x}{\mathbf{h}_{n,T}}\right)\boldsymbol{\mu}(X_s)\,\mathrm{d}s+O_p\left(\frac{\Delta_{n,T}v(T)}{h_{n,T}^2}\right)}{\frac{1}{\mathbf{h}_{n,T}}\int_0^T\mathbf{K}\left(\frac{X_s-x}{\mathbf{h}_{n,T}}\right)\mathrm{d}s+O_p\left(\frac{\Delta_{n,T}v(T)}{h_{n,T}^2}\right)} \qquad \text{(A.19)}\\
&-\boldsymbol{\mu}(x),
\end{aligned}
$$

where equation (A.18) derives from the same method of proof leading to equation (A.12) (and leading to the probability order of $[\Theta_{n,T}^{g,II}]_T$) and equation (A.19) derives from Lemma A.2 applied to both the numerator $\left(\text{with } \mathcal{G}(.)=\mathbf{K}(.) \text{ and } \mathcal{N}(.)=\boldsymbol{\mu}(.)\right)$ and the denominator (with $\mathcal{G}(.)=\mathbf{K}(.)$ and $\mathcal{N}(.)=1$). Thus,

$$
\text{bias term}=\frac{\frac{1}{\mathbf{h}_{n,T}}\int_0^T\mathbf{K}\left(\frac{X_s-x}{\mathbf{h}_{n,T}}\right)\left(\boldsymbol{\mu}(X_s)-\boldsymbol{\mu}(x)\right)\mathrm{d}s+(1-\boldsymbol{\mu}(x))O_p\left(\frac{\Delta_{n,T}v(T)}{h_{n,T}^2}\right)}{\frac{1}{\mathbf{h}_{n,T}}\int_0^T\mathbf{K}\left(\frac{X_s-x}{\mathbf{h}_{n,T}}\right)\mathrm{d}s+O_p\left(\frac{\Delta_{n,T}v(T)}{h_{n,T}^2}\right)}.
$$

Focusing now, as earlier, on the generic component g of the bias term, we obtain

$$
\begin{aligned}
&\frac{\frac{1}{\mathbf{h}_{n,T}}\int_0^T\mathbf{K}\left(\frac{X_s-x}{\mathbf{h}_{n,T}}\right)\left(\mu_g(X_s)-\mu_g(x)\right)\mathrm{d}s}{\frac{1}{\mathbf{h}_{n,T}}\int_0^T\mathbf{K}\left(\frac{X_s-x}{\mathbf{h}_{n,T}}\right)\mathrm{d}s}\\
&=\frac{\frac{1}{\mathbf{h}_{n,T}}\int_0^T\mathbf{K}\left(\frac{X_s-x}{\mathbf{h}_{n,T}}\right)\left(\mu_g(X_s)-\mu_g(x)\right)\mathrm{d}s-N_T\mathrm{E}_\lambda\left(\int_{R_n}^{R_{n+1}}\frac{1}{\mathbf{h}_{n,T}}\mathbf{K}\left(\frac{X_s-x}{\mathbf{h}_{n,T}}\right)\left(\mu_g(X_s)-\mu_g(x)\right)\mathrm{d}s\right)}{\frac{1}{\mathbf{h}_{n,T}}\int_0^T\mathbf{K}\left(\frac{X_s-x}{\mathbf{h}_{n,T}}\right)\mathrm{d}s}\\
&\quad+\frac{N_T\mathrm{E}_\lambda\left(\int_{R_n}^{R_{n+1}}\frac{1}{\mathbf{h}_{n,T}}\mathbf{K}\left(\frac{X_s-x}{\mathbf{h}_{n,T}}\right)\left(\mu_g(X_s)-\mu_g(x)\right)\mathrm{d}s\right)}{\frac{1}{\mathbf{h}_{n,T}}\int_0^T\mathbf{K}\left(\frac{X_s-x}{\mathbf{h}_{n,T}}\right)\mathrm{d}s}\\
&=\underbrace{\frac{\frac{1}{\mathbf{h}_{n,T}}\int_0^T\mathbf{K}\left(\frac{X_s-x}{\mathbf{h}_{n,T}}\right)\left(\mu_g(X_s)-\mu_g(x)\right)\mathrm{d}s-N_T\mathrm{E}_\lambda\left(\int_{R_n}^{R_{n+1}}\frac{1}{\mathbf{h}_{n,T}}\mathbf{K}\left(\frac{X_s-x}{\mathbf{h}_{n,T}}\right)\left(\mu_g(X_s)-\mu_g(x)\right)\mathrm{d}s\right)}{\frac{1}{\mathbf{h}_{n,T}}\int_0^T\mathbf{K}\left(\frac{X_s-x}{\mathbf{h}_{n,T}}\right)\mathrm{d}s}}_{\mathcal{C}_2(x)}\\
&\quad+\underbrace{\frac{N_TC_X\left(\int_{S^d}\frac{1}{\mathbf{h}_{n,T}}\mathbf{K}\left(\frac{a-x}{\mathbf{h}_{n,T}}\right)\left(\mu_g(a)-\mu_g(x)\right)\widetilde{\phi}(a)\mathrm{d}a\right)}{\frac{1}{\mathbf{h}_{n,T}}\int_0^T\mathbf{K}\left(\frac{X_s-x}{\mathbf{h}_{n,T}}\right)\mathrm{d}s}}_{\mathcal{C}_1(x)},
\end{aligned}
$$

where N_T was defined in the proof of Lemma A.1 and the second equality follows from equation (A.3). Using the symmetry of the kernel function, write

$$
\begin{aligned}
&\mathcal{C}_1(x)\\
&=\frac{\frac{1}{\mathbf{h}_{n,T}}C_X\int_{S^d}\mathbf{K}\left(\frac{a-x}{\mathbf{h}_{n,T}}\right)\left(\mu_g(a)-\mu_g(x)\right)\widetilde{\phi}(a)\mathrm{d}a}{\left(\frac{1}{\mathbf{h}_{n,T}}\int_0^T\mathbf{K}\left(\frac{X_s-x}{\mathbf{h}_{n,T}}\right)\mathrm{d}s\right)/(N_T)}
\end{aligned}
$$

https://doi.org/10.1017/S0266466617000305 Published online by Cambridge University Press

$$
\begin{aligned}
&= \frac{C_X \int_{S^d} \mathbf{K}(u)\left(\mu_g(x+u\mathbf{h}_{n,T})-\mu_g(x)\right)\widetilde{\phi}(x+u\mathbf{h}_{n,T})\mathrm{d}u}{\left(\frac{1}{\mathbf{h}_{n,T}}\int_0^T \mathbf{K}\left(\frac{X_s-x}{\mathbf{h}_{n,T}}\right)\mathrm{d}s\right)/(N_T)} \\
&= \frac{C_X \int_{S^d} \prod_{j=1}^d K(u^j)\left[h_{n,T}\sum_{j=1}^d \frac{\partial\mu_g(x)}{\partial x^j}u^j + \frac{1}{2}h_{n,T}^2\sum_{p,j=1}^d \frac{\partial^2\mu_g(x)}{\partial x^p\partial x^j}u^p u^j + o\right]\left[\widetilde{\phi}(x)+h_{n,T}\sum_{j=1}^d\frac{\partial\widetilde{\phi}(x)}{\partial x^j}u^j+o\right]\mathrm{d}u}{\left(\frac{1}{\mathbf{h}_{n,T}}\int_0^T \mathbf{K}\left(\frac{X_s-x}{\mathbf{h}_{n,T}}\right)\mathrm{d}s\right)/(N_T)} \\
&= \frac{C_X \int_{S^d} \prod_{j=1}^d K(u^j)\left[h_{n,T}^2\sum_{j=1}^d \frac{\partial\mu_g(x)}{\partial x^j}\frac{\partial\widetilde{\phi}(x)}{\partial x^j}\left(u^j\right)^2 + h_{n,T}^2\frac{1}{2}\sum_{j=1}^d\frac{\partial^2\mu_g(x)}{\partial x^j\partial x^j}\widetilde{\phi}(x)\left(u^j\right)^2\right]\mathrm{d}u}{\left(\frac{1}{\mathbf{h}_{n,T}}\int_0^T \mathbf{K}\left(\frac{X_s-x}{\mathbf{h}_{n,T}}\right)\mathrm{d}s\right)/(N_T)} + o_p(h_{n,T}^2) \\
&= h_{n,T}^2\left(\int_S s^2K(s)\mathrm{d}s\right)\left(\sum_{j=1}^d\frac{\partial\mu_g(x)}{\partial x^j}\frac{\partial\widetilde{\phi}(x)}{\partial x^j}\frac{1}{\widetilde{\phi}(x)}+\frac{1}{2}\sum_{j=1}^d\frac{\partial^2\mu_g(x)}{\partial x^j\partial x^j}\right)+o_p(h_{n,T}^2), \qquad \textbf{(A.20)}
\end{aligned}
$$

since, from equation (A.1), we have

$$
\begin{aligned}
&\frac{1}{N_T}\left(\frac{1}{\mathbf{h}_{n,T}}\int_0^T \mathbf{K}\left(\frac{X_s-x}{\mathbf{h}_{n,T}}\right)\mathrm{d}s\right) \\
&= \frac{N_T}{N_T}\mathrm{E}_\lambda\left(\frac{1}{\mathbf{h}}\int_{R_n}^{R_{n+1}}\mathbf{K}\left(\frac{X_s-x}{\mathbf{h}}\right)\mathrm{d}s\right)+\frac{v(T)}{N_T}\beta_{T,\mathbf{h}}^*(x)+\frac{v(T)}{N_T}\gamma_{T,\mathbf{h}}^*(x) \\
&= \mathrm{E}_\lambda\left(\frac{1}{\mathbf{h}}\int_{R_n}^{R_{n+1}}\mathbf{K}\left(\frac{X_s-x}{\mathbf{h}}\right)\mathrm{d}s\right)+\frac{v(T)}{N_T}O_p\left(\frac{1}{v^{1/2}(T)h^{d(1-1/q)}}\right)+\frac{v(T)}{N_T}O_p\left(\frac{1}{v^{1/2}(T)}\right) \\
&\xrightarrow{p} C_X\widetilde{\phi}(x),
\end{aligned}
$$

given $\mathbf{h}_{n,T}^{2/r}v(T)\to\infty$, where $\beta_{T,\mathbf{h}}^*(x)$ and $\gamma_{T,\mathbf{h}}^*(x)$ are the analogues of $\beta_{T,\mathbf{h}}(x)$ and $\gamma_{T,\mathbf{h}}(x)$ in equation (A.6) of Lemma A.1 with $\mathcal{S}(.)=\mathbf{K}(.)$ and $\mathcal{H}(.)=1$. Finally, we note that

$$
\begin{aligned}
&\mathcal{C}_2(x) \\
&= \frac{\frac{1}{v(T)\mathbf{h}_{n,T}}\int_0^T\mathbf{K}\left(\frac{X_s-x}{\mathbf{h}_{n,T}}\right)\left(\mu_g(X_s)-\mu_g(x)\right)\mathrm{d}s-\frac{N_T}{v(T)}\mathrm{E}_\lambda\left(\int_{R_n}^{R_{n+1}}\frac{1}{\mathbf{h}_{n,T}}\mathbf{K}\left(\frac{X_s-x}{\mathbf{h}_{n,T}}\right)\left(\mu_g(X_s)-\mu_g(x)\right)\mathrm{d}s\right)}{\frac{1}{v(T)\mathbf{h}_{n,T}}\int_0^T\mathbf{K}\left(\frac{X_s-x}{\mathbf{h}_{n,T}}\right)\mathrm{d}s} \\
&= \frac{\mathrm{E}_\lambda\left(\frac{1}{v(T)\mathbf{h}_{n,T}}\int_0^T\mathbf{K}\left(\frac{X_s-x}{\mathbf{h}_{n,T}}\right)\left(\mu_g(X_s)-\mu_g(x)\right)\mathrm{d}s\right)-\frac{N_T}{v(T)}\mathrm{E}_\lambda\left(\int_{R_n}^{R_{n+1}}\frac{1}{\mathbf{h}_{n,T}}\mathbf{K}\left(\frac{X_s-x}{\mathbf{h}_{n,T}}\right)\left(\mu_g(X_s)-\mu_g(x)\right)\mathrm{d}s\right)}{\frac{1}{v(T)\mathbf{h}_{n,T}}\int_0^T\mathbf{K}\left(\frac{X_s-x}{\mathbf{h}_{n,T}}\right)\mathrm{d}s} \\
&\quad+\frac{\frac{1}{v(T)\mathbf{h}_{n,T}}\int_0^T\mathbf{K}\left(\frac{X_s-x}{\mathbf{h}_{n,T}}\right)\left(\mu_g(X_s)-\mu_g(x)\right)\mathrm{d}s-\mathrm{E}_\lambda\left(\frac{1}{v(T)\mathbf{h}_{n,T}}\int_0^T\mathbf{K}\left(\frac{X_s-x}{\mathbf{h}_{n,T}}\right)\left(\mu_g(X_s)-\mu_g(x)\right)\mathrm{d}s\right)}{\frac{1}{v(T)\mathbf{h}_{n,T}}\int_0^T\mathbf{K}\left(\frac{X_s-x}{\mathbf{h}_{n,T}}\right)\mathrm{d}s} \\
&= \frac{num\mathcal{C}_2^1(x)+num\mathcal{C}_2^2(x)}{\frac{1}{v(T)\mathbf{h}_{n,T}}\int_0^T\mathbf{K}\left(\frac{X_s-x}{\mathbf{h}_{n,T}}\right)\mathrm{d}s}.
\end{aligned}
$$

Now, $\frac{1}{\mathbf{h}_{n,T}v(T)}\int_0^T\mathbf{K}\left(\frac{X_s-x}{\mathbf{h}_{n,T}}\right)\mathrm{d}s=O_p(1)$ given the Corollary to Lemma A.1. Also, $num\mathcal{C}_2^1(x)$ and $num\mathcal{C}_2^2(x)$ are, again, the analogues of $\gamma_{T,\mathbf{h}}(x)$ and $\beta_{T,\mathbf{h}}(x)$ in equation (A.6), respectively, with the term $\mu_g(X_s)-\mu_g(x)$ adding, just like in the case of equation (A.20) above, $h_{n,T}^2$ to their convergence rate in light of the symmetry of the kernel function. We have, in fact, $num\mathcal{C}_2^1(x)=O_p\left(\frac{h_{n,T}^2}{v^{1/2}(T)}\right)=o_p(h_{n,T}^2)$ and $num\mathcal{C}_2^2(x)=O_p\left(\frac{h_{n,T}^2}{v^{1/2}(T)h^{d(1-1/q)}}\right)=o_p(h_{n,T}^2)$. In sum,

https://doi.org/10.1017/S0266466617000305 Published online by Cambridge University Press

$$\frac{\frac{1}{\mathbf{h}_{n,T}}\int_0^T \mathbf{K}\left(\frac{X_s-x}{\mathbf{h}_{n,T}}\right)\left(\mu_g(X_s)-\mu_g(x)\right)\mathrm{d}s}{\frac{1}{\mathbf{h}_{n,T}}\int_0^T \mathbf{K}\left(\frac{X_s-x}{\mathbf{h}_{n,T}}\right)\mathrm{d}s}$$
$$= h_{n,T}^2\left(\int_S s^2K(s)\mathrm{d}s\right)\left(\sum_{j=1}^d \frac{\partial\mu_g(x)}{\partial x^j}\frac{\partial\widetilde{\phi}(x)}{\partial x^j}\frac{1}{\widetilde{\phi}(x)}+\frac{1}{2}\sum_{j=1}^d\frac{\partial^2\mu_g(x)}{\partial x^j\partial x^j}\right)+o_p(h_{n,T}^2),$$

which yields the stated result. ■

Proof of Theorem 5. Using Itô's lemma, write

$$\left(X_{(i+1)\Delta_{n,T}}-X_{i\Delta_{n,T}}\right)\left(X_{(i+1)\Delta_{n,T}}-X_{i\Delta_{n,T}}\right)^{\intercal}$$
$$=\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}\mathcal{D}_X(X_s-X_{i\Delta_{n,T}})\otimes\boldsymbol{\mu}(X_s)\mathrm{d}s+\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}\mathcal{D}_X(X_s-X_{i\Delta_{n,T}})\otimes\boldsymbol{\sigma}(X_s)\mathrm{d}\mathbf{B}_s$$
$$+\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}\mathbf{a}(X_s)\mathrm{d}s,$$

where $\mathcal{D}_X(X_s-X_{i\Delta_{n,T}})$ is a $d\times d$ matrix of $1\times d$ vectors. The i-th element of the ab-th vector is given by the derivative $\frac{\partial\left((X_s-X_{i\Delta_{n,T}})(X_s-X_{i\Delta_{n,T}})^{\intercal}\right)_{ab}}{\partial X_s^i}$. Now, write

$$vech\widehat{\mathbf{a}}_{n,T}(x)=vech\left(\frac{1}{\Delta_{n,T}}\frac{\sum_{i=1}^{n-1}\mathbf{K}\left(\frac{X_{i\Delta_{n,T}}-x}{\mathbf{h}_{n,T}}\right)\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}\mathcal{D}_X(X_s-X_{i\Delta_{n,T}})\otimes\boldsymbol{\mu}(X_s)\mathrm{d}s}{\sum_{i=1}^{n}\mathbf{K}\left(\frac{X_{i\Delta_{n,T}}-x}{\mathbf{h}_{n,T}}\right)}\right)$$
$$+vech\left(\frac{1}{\Delta_{n,T}}\frac{\sum_{i=1}^{n-1}\mathbf{K}\left(\frac{X_{i\Delta_{n,T}}-x}{\mathbf{h}_{n,T}}\right)\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}\mathcal{D}_X(X_s-X_{i\Delta_{n,T}})\otimes\boldsymbol{\sigma}(X_s)\mathrm{d}\mathbf{B}_s}{\sum_{i=1}^{n}\mathbf{K}\left(\frac{X_{i\Delta_{n,T}}-x}{\mathbf{h}_{n,T}}\right)}\right)$$
$$+vech\left(\frac{1}{\Delta_{n,T}}\frac{\sum_{i=1}^{n-1}\mathbf{K}\left(\frac{X_{i\Delta_{n,T}}-x}{\mathbf{h}_{n,T}}\right)\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}\mathbf{a}(X_s)\mathrm{d}s}{\sum_{i=1}^{n}\mathbf{K}\left(\frac{X_{i\Delta_{n,T}}-x}{\mathbf{h}_{n,T}}\right)}\right)$$
$$=vech\boldsymbol{\Psi}_1+vech\boldsymbol{\Psi}_2+vech\boldsymbol{\Psi}_3.$$

We focus on the generic element a,b of the matrices $\boldsymbol{\Psi}_1+\boldsymbol{\Psi}_2$, namely

$$(\boldsymbol{\Psi}_1+\boldsymbol{\Psi}_2)_{ab}$$
$$=\frac{\frac{1}{\mathbf{h}_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}\left(\frac{X_{i\Delta_{n,T}}-x}{\mathbf{h}_{n,T}}\right)\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}(X_s^a-X_{i\Delta_{n,T}}^a)\mu_b(X_s)\mathrm{d}s+\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}(X_s^b-X_{i\Delta_{n,T}}^b)\mu_a(X_s)\mathrm{d}s}{\frac{\Delta_{n,T}}{\mathbf{h}_{n,T}}\sum_{i=1}^{n}\mathbf{K}\left(\frac{X_{i\Delta_{n,T}}-x}{\mathbf{h}_{n,T}}\right)}$$
$$+\frac{\frac{1}{\mathbf{h}_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}\left(\frac{X_{i\Delta_{n,T}}-x}{\mathbf{h}_{n,T}}\right)\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}(X_s^a-X_{i\Delta_{n,T}}^a)\boldsymbol{\sigma}_b(X_s)\mathrm{d}\mathbf{B}_s+\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}(X_s^b-X_{i\Delta_{n,T}}^b)\boldsymbol{\sigma}_a(X_s)\mathrm{d}\mathbf{B}_s}{\frac{\Delta_{n,T}}{\mathbf{h}_{n,T}}\sum_{i=1}^{n}\mathbf{K}\left(\frac{X_{i\Delta_{n,T}}-x}{\mathbf{h}_{n,T}}\right)}$$
$$=\boldsymbol{\Psi}_1^{\mu_b}+\boldsymbol{\Psi}_1^{\mu_a}+\boldsymbol{\Psi}_2^{\sigma_b}+\boldsymbol{\Psi}_2^{\sigma_a},$$

where the notation should not create confusion. We will find the stochastic orders of $\boldsymbol{\Psi}_1^{\mu_b}$ and $\boldsymbol{\Psi}_2^{\sigma_b}$. These will give us the stochastic order of $vech\boldsymbol{\Psi}_1+vech\boldsymbol{\Psi}_2$. We begin with the numerator of $\boldsymbol{\Psi}_1^{\mu_b}$, i.e., $num\boldsymbol{\Psi}_1^{\mu_b}$,

https://doi.org/10.1017/S0266466617000305 Published online by Cambridge University Press

$$
\begin{aligned}
num\Psi_1^{\mu_b} &= \frac{1}{\mathbf{h}_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}\left(\frac{X_{i\Delta_{n,T}}-x}{\mathbf{h}_{n,T}}\right)\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}\left(\int_{i\Delta_{n,T}}^{s}\mu_a(X_v)\mathrm{d}v\right)\mu_b(X_s)\mathrm{d}s \\
&\quad+\frac{1}{\mathbf{h}_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}\left(\frac{X_{i\Delta_{n,T}}-x}{\mathbf{h}_{n,T}}\right)\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}\left(\int_{i\Delta_{n,T}}^{s}\boldsymbol{\sigma}_a(X_v)\mathrm{d}\mathbf{B}_v\right)\mu_b(X_s)\mathrm{d}s \\
&= num\Psi_{1,1}^{\mu_b}+num\Psi_{1,2}^{\mu_b}.
\end{aligned}
$$

We have

$$
\begin{aligned}
num\Psi_{1,1}^{\mu_b} &\le \frac{1}{\mathbf{h}_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}\left(\frac{X_{i\Delta_{n,T}}-x}{\mathbf{h}_{n,T}}\right)\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}\left(\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}|\mu_a(X_s)|\,\mathrm{d}s\right)|\mu_b(X_s)|\,\mathrm{d}s \\
&\le C_{|\mu_a|}C_{|\mu_b|}\frac{\Delta_{n,T}^2}{\mathbf{h}_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}\left(\frac{X_{i\Delta_{n,T}}-x}{\mathbf{h}_{n,T}}\right) \\
&= C_{|\mu_a|}C_{|\mu_b|}\frac{\Delta_{n,T}}{\mathbf{h}_{n,T}}\int_0^T\mathbf{K}\left(\frac{X_{i\Delta_{n,T}}-x}{\mathbf{h}_{n,T}}\right)\mathrm{d}s+\Delta_{n,T}O_p\left(\frac{\Delta_{n,T}v(T)}{h_{n,T}^2}\right) \\
&= O_p(\Delta_{n,T}v(T))+O_p\left(\frac{\Delta_{n,T}^2v(T)}{h_{n,T}^2}\right),
\end{aligned}
$$

where the first inequality derives from Assumption 4, the discretization order in the first equality derives from Lemma A.2 and the first order term in the second equality derives from Lemma A.1.

We now turn to $num\Psi_{1,2}^{\mu_b}$. Integrating by parts,

$$
\begin{aligned}
num\Psi_{1,2}^{\mu_b} &= -\frac{1}{\mathbf{h}_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}\left(\frac{X_{i\Delta_{n,T}}-x}{\mathbf{h}_{n,T}}\right)\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}\left(\int_{i\Delta_{n,T}}^{s}\mu_b(X_v)\mathrm{d}v\right)\boldsymbol{\sigma}_a(X_s)\mathrm{d}\mathbf{B}_s \\
&\quad+\frac{1}{\mathbf{h}_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}\left(\frac{X_{i\Delta_{n,T}}-x}{\mathbf{h}_{n,T}}\right)\left(\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}\mu_b(X_s)\mathrm{d}s\right)\left(\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}\boldsymbol{\sigma}_a(X_s)\mathrm{d}\mathbf{B}_s\right) \\
&= \Psi_{1,2}^{\mu_b,A}+\Psi_{1,2}^{\mu_b,B}.
\end{aligned}
$$

Next, the quadratic variation of $\Psi_{1,2}^{\mu_b,A}$ can be written as

$$
\begin{aligned}
&[\Psi_{1,2}^{\mu_b,A}] \\
&= \frac{1}{\mathbf{h}_{n,T}^2}\sum_{i=1}^{n-1}\mathbf{K}^2\left(\frac{X_{i\Delta_{n,T}}-x}{\mathbf{h}_{n,T}}\right)\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}\left(\int_{i\Delta_{n,T}}^{s}\mu_b(X_v)\mathrm{d}v\right)^2(\boldsymbol{\sigma}_a(X_s)^\intercal\boldsymbol{\sigma}_a(X_s))\,\mathrm{d}s \\
&\le \frac{1}{\mathbf{h}_{n,T}^2}\sum_{i=1}^{n-1}\mathbf{K}^2\left(\frac{X_{i\Delta_{n,T}}-x}{\mathbf{h}_{n,T}}\right)\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}\left(\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}|\mu_b(X_s)|\,\mathrm{d}s\right)^2(\boldsymbol{\sigma}_a(X_s)^\intercal\boldsymbol{\sigma}_a(X_s))\,\mathrm{d}s \\
&\le \frac{1}{\mathbf{h}_{n,T}^2}\sum_{i=1}^{n-1}\mathbf{K}^2\left(\frac{X_{i\Delta_{n,T}}-x}{\mathbf{h}_{n,T}}\right)\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}\Delta_{n,T}\left(\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}\mu_b^2(X_s)\mathrm{d}s\right)(\boldsymbol{\sigma}_a(X_s)^\intercal\boldsymbol{\sigma}_a(X_s))\,\mathrm{d}s \\
&\le \frac{C_{\mu_b^2}\Delta_{n,T}^2}{\mathbf{h}_{n,T}}\frac{1}{\mathbf{h}_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}^2\left(\frac{X_{i\Delta_{n,T}}-x}{\mathbf{h}_{n,T}}\right)\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}(\boldsymbol{\sigma}_a(X_s)^\intercal\boldsymbol{\sigma}_a(X_s))\,\mathrm{d}s \\
&= \frac{C_{\mu_b^2}\Delta_{n,T}^2}{\mathbf{h}_{n,T}}\left(\frac{\Delta_{n,T}}{\mathbf{h}_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}^2\left(\frac{X_{i\Delta_{n,T}}-x}{\mathbf{h}_{n,T}}\right)\boldsymbol{\sigma}_a(X_{i\Delta_{n,T}})^\intercal\boldsymbol{\sigma}_a(X_{i\Delta_{n,T}})\right)+\frac{C_{\mu_b^2}\Delta_{n,T}^2}{\mathbf{h}_{n,T}}
\end{aligned}
$$

https://doi.org/10.1017/S0266466617000305 Published online by Cambridge University Press

$$\times \left(\frac{1}{\mathbf{h}_{n,T}} \sum_{i=1}^{n-1} \mathbf{K}^2 \left(\frac{X_{i\Delta_{n,T}} - x}{\mathbf{h}_{n,T}} \right) \int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}} \left[\boldsymbol{\sigma}_a(X_s)^\intercal \boldsymbol{\sigma}_a(X_s) - \boldsymbol{\sigma}_a(X_{i\Delta_{n,T}})^\intercal \boldsymbol{\sigma}_a(X_{i\Delta_{n,T}}) \right] \mathrm{d}s \right)$$

$$= O_p \left(\frac{\Delta_{n,T}^2 v(T)}{\mathbf{h}_{n,T}} \right) + O_p \left(\frac{\Delta_{n,T}^2}{\mathbf{h}_{n,T}} \frac{\Delta_{n,T}}{h_{n,T}^2} v(T) \right), \tag{A.21}$$

where the second inequality is Jensen's and the order terms derive from the same method of proof used for equation (A.11). Hence, $\boldsymbol{\Psi}_{1,2}^{\mu_b,A} = O_p \left(\frac{\Delta_{n,T} v^{1/2}(T)}{\mathbf{h}_{n,T}^{1/2}} \right)$ since $\frac{\Delta_{n,T}}{h_{n,T}^2} \to 0$. Next,

$$\begin{aligned}
\boldsymbol{\Psi}_{1,2}^{\mu_b,B} &= \frac{\Delta_{n,T}}{\mathbf{h}_{n,T}} \sum_{i=1}^{n-1} \mathbf{K} \left(\frac{X_{i\Delta_{n,T}} - x}{\mathbf{h}_{n,T}} \right) \mu_b(X_{i\Delta_{n,T}}) \left(\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}} \boldsymbol{\sigma}_a(X_s) \mathrm{d}\mathbf{B}_s \right) \\
&\quad + \frac{1}{\mathbf{h}_{n,T}} \sum_{i=1}^{n-1} \mathbf{K} \left(\frac{X_{i\Delta_{n,T}} - x}{\mathbf{h}_{n,T}} \right) \left(\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}} \left(\mu_b(X_s) - \mu_b(X_{i\Delta_{n,T}}) \right) \mathrm{d}s \right) \\
&\quad \times \left(\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}} \boldsymbol{\sigma}_a(X_s) \mathrm{d}\mathbf{B}_s \right) \\
&= \boldsymbol{\Psi}_{1,2}^{\mu_b,B,1} + \boldsymbol{\Psi}_{1,2}^{\mu_b,B,2}.
\end{aligned}$$

Again, like with equation (A.21), the quadratic variation of $\boldsymbol{\Psi}_{1,2}^{\mu_b,B,1}$ can be written as

$$\begin{aligned}
[\boldsymbol{\Psi}_{1,2}^{\mu_b,B,1}] &= \frac{\Delta_{n,T}^2}{\mathbf{h}_{n,T}^2} \sum_{i=1}^{n-1} \mathbf{K}^2 \left(\frac{X_{i\Delta_{n,T}} - x}{\mathbf{h}_{n,T}} \right) \mu_b^2(X_{i\Delta_{n,T}}) \int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}} \boldsymbol{\sigma}_a(X_s)^\intercal \boldsymbol{\sigma}_a(X_s) \mathrm{d}s \\
&= \frac{\Delta_{n,T}^2}{\mathbf{h}_{n,T}} \left(\frac{\Delta_{n,T}}{\mathbf{h}_{n,T}} \sum_{i=1}^{n-1} \mathbf{K}^2 \left(\frac{X_{i\Delta_{n,T}} - x}{\mathbf{h}_{n,T}} \right) \mu_b^2(X_{i\Delta_{n,T}}) \boldsymbol{\sigma}_a(X_{i\Delta_{n,T}})^\intercal \boldsymbol{\sigma}_a(X_{i\Delta_{n,T}}) \mathrm{d}s \right) \\
&\quad + \frac{\Delta_{n,T}^2}{\mathbf{h}_{n,T}} \left(\frac{1}{\mathbf{h}_{n,T}} \sum_{i=1}^{n-1} \mathbf{K}^2 \left(\frac{X_{i\Delta_{n,T}} - x}{\mathbf{h}_{n,T}} \right) \mu_b^2(X_{i\Delta_{n,T}}) \right. \\
&\qquad \left. \times \int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}} \left[\boldsymbol{\sigma}_a(X_s)^\intercal \boldsymbol{\sigma}_a(X_s) - \boldsymbol{\sigma}_a(X_{i\Delta_{n,T}})^\intercal \boldsymbol{\sigma}_a(X_{i\Delta_{n,T}}) \right] \mathrm{d}s \right) \\
&= O_p \left(\frac{\Delta_{n,T}^2 v(T)}{\mathbf{h}_{n,T}} \right),
\end{aligned}$$

if $\frac{\Delta_{n,T}}{h_{n,T}^2} \to 0$. Thus, $\boldsymbol{\Psi}_{1,2}^{\mu_b,B,1} = O_p \left(\frac{\Delta_{n,T} v^{1/2}(T)}{\mathbf{h}_{n,T}^{1/2}} \right)$. Now,

$$\begin{aligned}
&\mathrm{E}_\lambda \left(\left| \boldsymbol{\Psi}_{1,2}^{\mu_b,B,2} \right| \right) \\
&\quad \leq \mathrm{E}_\lambda \left(\frac{1}{\mathbf{h}_{n,T}} \sum_{i=1}^{n-1} \mathbf{K} \left(\frac{X_{i\Delta_{n,T}} - x}{\mathbf{h}_{n,T}} \right) \right. \\
&\qquad \left. \times \mathrm{E}_\lambda \left(\left| \int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}} \left(\mu_b(X_s) - \mu_b(X_{i\Delta_{n,T}}) \right) \mathrm{d}s \right| \left| \int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}} \boldsymbol{\sigma}_a(X_v) \mathrm{d}\mathbf{B}_v \right| \mathcal{F}_i \right) \right)
\end{aligned}$$

https://doi.org/10.1017/S0266466617000305 Published online by Cambridge University Press

$$\leq \mathrm{E}_\lambda\left(\frac{1}{\mathbf{h}_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}\left(\frac{X_{i\Delta_{n,T}}-x}{\mathbf{h}_{n,T}}\right)\left(\mathrm{E}_\lambda\left(\left|\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}\left(\mu_b(X_s)-\mu_b(X_{i\Delta_{n,T}})\right)\mathrm{d}s\right|^2|\mathcal{F}_i\right)\right)^{1/2}\right.$$
$$\left.\times\left(\mathrm{E}_\lambda\left(\left(\sup_{i\Delta_{n,T}\leq s\leq(i+1)\Delta_{n,T}}\left|\int_{i\Delta_{n,T}}^{s}\boldsymbol{\sigma}_a(X_v)\mathrm{d}\mathbf{B}_v\right|\right)^2|\mathcal{F}_i\right)\right)^{1/2}\right),$$

where the second inequality is Hölder's. Burkholder, Davis, Gundy (BDG) inequality (see, e.g., Karatzas and Shreve, 1991, Theorem 3.28, page 166) now gives

$$\mathrm{E}_\lambda\left(\left(\sup_{i\Delta_{n,T}\leq s\leq(i+1)\Delta_{n,T}}\left|\int_{i\Delta_{n,T}}^{s}\boldsymbol{\sigma}_a(X_v)\mathrm{d}\mathbf{B}_v\right|\right)^2|\mathcal{F}_i\right)$$
$$\leq C_1\mathrm{E}_\lambda\left(\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}\boldsymbol{\sigma}_a(X_s)^\intercal\boldsymbol{\sigma}_a(X_s)\mathrm{d}s|\mathcal{F}_i\right)$$
$$\leq C_2\Delta_{n,T},$$

for some constant C_2. Similarly, expressing the infinitesimal generator $\mathcal{L}$ of X applied to the function $\mu_b(.)$ as

$$\mathcal{L}\mu_b(X_s)=\mathcal{Q}(X_s)$$
$$=\sum_{j=1}^{d}\mu'_{b,j}(X_s)\mu_j(X_s)+\frac{1}{2}\sum_{j=1}^{d}\mu''_{b,jj}(X_s)\sum_{k=1}^{m}\sigma^2_{jk}(X_s)+\frac{1}{2}\sum_{j=1}^{d}\sum_{q\neq j}^{d}\mu''_{b,jq}(X_s)\sum_{k=1}^{m}\sigma_{jk}(X_s)\sigma_{qk}(X_s),$$

we have

$$\mathrm{E}_\lambda\left(\left|\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}\left(\mu_b(X_s)-\mu_b(X_{i\Delta_{n,T}})\right)\mathrm{d}s\right|^2|\mathcal{F}_i\right)$$
$$\leq \mathrm{E}_\lambda\left(\Delta_{n,T}\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}\left[\mu_b(X_s)-\mu_b(X_{i\Delta_{n,T}})\right]^2\mathrm{d}s|\mathcal{F}_i\right)$$
$$\leq \mathrm{E}_\lambda\left(\Delta_{n,T}\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}\left[\int_{i\Delta_{n,T}}^{s}\mathcal{Q}(X_v)\mathrm{d}v\right]^2\mathrm{d}s|\mathcal{F}_i\right)$$
$$+\mathrm{E}_\lambda\left(\Delta_{n,T}\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}\left[\int_{i\Delta_{n,T}}^{s}\sum_{j=1}^{d}\mu'_{b,j}(X_v)\sum_{k=1}^{m}\sigma_{jk}(X_v)\mathrm{d}B^k_v\right]^2\mathrm{d}s|\mathcal{F}_i\right)$$
$$+2\mathrm{E}_\lambda\left(\Delta_{n,T}\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}\left[\int_{i\Delta_{n,T}}^{s}\mathcal{Q}(X_v)\mathrm{d}v\right]\left[\int_{i\Delta_{n,T}}^{s}\sum_{j=1}^{d}\mu'_{b,j}(X_v)\sum_{k=1}^{m}\sigma_{jk}(X_v)\mathrm{d}B^k_v\right]\mathrm{d}s|\mathcal{F}_i\right).$$

Now, by Jensen's inequality again, the first term can be expressed as follows:

$$\mathrm{E}_\lambda\left(\Delta_{n,T}\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}\left[\int_{i\Delta_{n,T}}^{s}\mathcal{Q}(X_v)\mathrm{d}v\right]^2\mathrm{d}s|\mathcal{F}_i\right)$$
$$\leq \mathrm{E}_\lambda\left(\Delta_{n,T}\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}\left[\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}|\mathcal{Q}(X_s)|\mathrm{d}s\right]^2\mathrm{d}s|\mathcal{F}_i\right)$$
$$\leq \mathrm{E}_\lambda\left(\Delta^2_{n,T}\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}\left[\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}|\mathcal{Q}(X_s)|^2\mathrm{d}s\right]\mathrm{d}s|\mathcal{F}_i\right)$$
$$\leq C_{\mathcal{Q}^2}\Delta^4_{n,T}.$$

https://doi.org/10.1017/S0266466617000305 Published online by Cambridge University Press

By BDG inequality, applied after the second inequality, the second term becomes

$$
\begin{aligned}
&\mathrm{E}_\lambda\left(\Delta_{n,T}\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}\left[\int_{i\Delta_{n,T}}^{s}\sum_{j=1}^{d}\mu'_{b,j}(X_v)\sum_{k=1}^{m}\sigma_{jk}(X_v)\,\mathrm{d}B_v^k\right]^2\mathrm{d}s|\mathcal{F}_i\right)\\
&\quad\le \mathrm{E}_\lambda\left(\Delta_{n,T}\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}\left[\sup_{i\Delta_{n,T}\le s\le(i+1)\Delta_{n,T}}\left|\int_{i\Delta_{n,T}}^{s}\sum_{j=1}^{d}\mu'_{b,j}(X_v)\sum_{k=1}^{m}\sigma_{jk}(X_v)\,\mathrm{d}B_v^k\right|\right]^2\mathrm{d}s|\mathcal{F}_i\right)\\
&\quad\le C_3\mathrm{E}_\lambda\left(\Delta_{n,T}\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}\left|\left(\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}\sum_{j=1}^{d}\left(\mu'_{b,j}(X_s)\right)^2\sum_{k=1}^{m}\sigma_{jk}^2(X_s)\,\mathrm{d}s\right)\right|\mathrm{d}s|\mathcal{F}_i\right)\\
&\qquad+C_3\mathrm{E}_\lambda\left(\Delta_{n,T}\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}\left|\left(\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}\sum_{j=1}^{d}\sum_{q\ne1}^{d}\mu'_{b,j}(X_s)\mu'_{b,q}(X_s)\sum_{k=1}^{m}\sigma_{jk}(X_s)\sigma_{qk}(X_s)\,\mathrm{d}s\right)\right|\mathrm{d}s|\mathcal{F}_i\right)\\
&\quad\le C_4\Delta_{n,T}^3,
\end{aligned}
$$

for some constant C_4. Finally,

$$
\begin{aligned}
&\mathrm{E}_\lambda\left(\left(\Delta_{n,T}\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}\left[\int_{i\Delta_{n,T}}^{s}\mathcal{Q}(X_v)\mathrm{d}v\right]\left[\int_{i\Delta_{n,T}}^{s}\sum_{j=1}^{d}\mu'_{b,j}(X_v)\sum_{k=1}^{m}\sigma_{jk}(X_v)\,\mathrm{d}B_v^k\right]\mathrm{d}s\right)|\mathcal{F}_i\right)\\
&\quad\le\Delta_{n,T}\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}\left(\mathrm{E}_\lambda\left[\int_{i\Delta_{n,T}}^{s}\mathcal{Q}(X_v)\mathrm{d}v\right]^2|\mathcal{F}_i\right)^{1/2}\left(\mathrm{E}_\lambda\left[\int_{i\Delta_{n,T}}^{s}\sum_{j=1}^{d}\mu'_{b,j}(X_v)\sum_{k=1}^{m}\sigma_{jk}(X_v)\,\mathrm{d}B_v^k\right]^2|\mathcal{F}_i\right)^{1/2}\mathrm{d}s\\
&\quad\le\Delta_{n,T}\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}\left(\Delta_{n,T}\mathrm{E}_\lambda\left[\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}|\mathcal{Q}(X_s)|^2\mathrm{d}s\right]|\mathcal{F}_i\right)^{1/2}\\
&\qquad\times\left(\mathrm{E}_\lambda\left[\sup_{i\Delta_{n,T}\le s\le(i+1)\Delta_{n,T}}\left|\int_{i\Delta_{n,T}}^{s}\sum_{j=1}^{d}\mu'_{b,j}(X_v)\sum_{k=1}^{m}\sigma_{jk}(X_v)\,\mathrm{d}B_v^k\right|\right]^2|\mathcal{F}_i\right)^{1/2}\mathrm{d}s\\
&\quad\le C_5\Delta_{n,T}\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}\left(\Delta_{n,T}\mathrm{E}_\lambda\left[\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}|\mathcal{Q}(X_s)|^2\mathrm{d}s\right]|\mathcal{F}_i\right)^{1/2}\\
&\qquad\times\left(\mathrm{E}_\lambda\left[\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}\left(\sum_{j=1}^{d}\left(\mu'_{b,j}(X_s)\right)^2\sum_{k=1}^{m}\sigma_{jk}^2(X_s)+\sum_{j=1}^{d}\sum_{q\ne1}^{d}\mu'_{b,j}(X_s)\mu'_{b,q}(X_s)\sum_{k=1}^{m}\sigma_{jk}\sigma_{qk}\right)\mathrm{d}s\right]|\mathcal{F}_i\right)^{1/2}\mathrm{d}s\\
&\quad\le C_6\Delta_{n,T}^{7/2},
\end{aligned}
$$

for some constant C_6. Thus,

$$
\begin{aligned}
&\mathrm{E}_\lambda\left(\boldsymbol{\Psi}_{1,2}^{\mu_b,B,2}\right)\\
&\quad\le\mathrm{E}_\lambda\left(\frac{1}{\mathbf{h}_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}\left(\frac{X_{i\Delta_{n,T}}-x}{\mathbf{h}_{n,T}}\right)\left(\mathrm{E}_\lambda\left(\left|\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}\left(\mu_b(X_s)-\mu_b(X_{i\Delta_{n,T}})\right)\mathrm{d}s\right|^2|\mathcal{F}_i\right)\right)^{1/2}\right.\\
&\qquad\left.\times\left(\mathrm{E}_\lambda\left(\left(\sup_{i\Delta_{n,T}\le s\le(i+1)\Delta_{n,T}}\left|\int_{i\Delta_{n,T}}^{s}\boldsymbol{\sigma}_a(X_v)\mathrm{d}\mathbf{B}_v\right|\right)^2|\mathcal{F}_i\right)\right)^{1/2}\right)\\
&\quad\le C_7\mathrm{E}_\lambda\left(\frac{\Delta_{n,T}^{3/2+1/2}}{\mathbf{h}_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}\left(\frac{X_{i\Delta_{n,T}}-x}{\mathbf{h}_{n,T}}\right)\right)\\
&\quad=O(\Delta_{n,T}v(T)).
\end{aligned}
$$

https://doi.org/10.1017/S0266466617000305 Published online by Cambridge University Press

Therefore, $\Psi_{1,2}^{\mu_b,B,2} = O_p(\Delta_{n,T} v(T))$. In the end,

$$\Psi_1^{\mu_b} = \frac{num\Psi_{1,1}^{\mu_b} + num\Psi_{1,2}^{\mu_b}}{\frac{\Delta_{n,T}}{\mathbf{h}_{n,T}} \sum_{i=1}^n \mathbf{K}\left(\frac{X_{i\Delta_{n,T}} - x}{\mathbf{h}_{n,T}}\right)} = O_p(\Delta_{n,T}) + O_p\left(\frac{\Delta_{n,T}^2}{h_{n,T}^2}\right) + O_p\left(\frac{\Delta_{n,T}}{\mathbf{h}_{n,T}^{1/2} v^{1/2}(T)}\right)$$
$$= O_p(\Delta_{n,T}) + O_p\left(\frac{\Delta_{n,T}}{\mathbf{h}_{n,T}^{1/2} v^{1/2}(T)}\right),$$

since $\frac{\Delta_{n,T}}{h_{n,T}^2} \to 0$, which we assume. We now turn to $\Psi_2^{\sigma_b}$. Write

$$\frac{\sqrt{\mathbf{h}_{n,T}}}{\sqrt{\Delta_{n,T}}} num\Psi_2^{\sigma_b}$$
$$= \frac{1}{\sqrt{\mathbf{h}_{n,T}\Delta_{n,T}}} \sum_{i=1}^{n-1} \mathbf{K}\left(\frac{X_{i\Delta_{n,T}} - x}{\mathbf{h}_{n,T}}\right) \int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}} (X_s^a - X_{i\Delta_{n,T}}^a) \boldsymbol{\sigma}_b(X_s) d\mathbf{B}_s$$
$$= \frac{1}{\sqrt{\mathbf{h}_{n,T}\Delta_{n,T}}} \sum_{i=1}^{n-1} \mathbf{K}\left(\frac{X_{i\Delta_{n,T}} - x}{\mathbf{h}_{n,T}}\right) \int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}} \left(\int_{i\Delta_{n,T}}^s \mu_a(X_v) dv\right) \boldsymbol{\sigma}_b(X_s) d\mathbf{B}_s$$
$$+ \frac{1}{\sqrt{\mathbf{h}_{n,T}\Delta_{n,T}}} \sum_{i=1}^{n-1} \mathbf{K}\left(\frac{X_{i\Delta_{n,T}} - x}{\mathbf{h}_{n,T}}\right) \int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}} \left(\int_{i\Delta_{n,T}}^s \boldsymbol{\sigma}_a(X_v) d\mathbf{B}_v\right) \boldsymbol{\sigma}_b(X_s) d\mathbf{B}_s$$
$$= \frac{\sqrt{\mathbf{h}_{n,T}}}{\sqrt{\Delta_{n,T}}} (num\Psi_{2,1}^{\sigma_b} + num\Psi_{2,2}^{\sigma_b}).$$

Since

$$\left[\frac{\sqrt{\mathbf{h}_{n,T}}}{\sqrt{\Delta_{n,T}}} num\Psi_{2,1}^{\sigma_b}\right] = \frac{1}{\mathbf{h}_{n,T}\Delta_{n,T}} \sum_{i=1}^{n-1} \mathbf{K}^2\left(\frac{X_{i\Delta_{n,T}} - x}{\mathbf{h}_{n,T}}\right)$$
$$\times \int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}} \left(\int_{i\Delta_{n,T}}^s \mu_a(X_s) ds\right)^2 (\boldsymbol{\sigma}_b(X_s)^\intercal \boldsymbol{\sigma}_b(X_s))\, ds$$
$$= O_p\left(\Delta_{n,T} v(T)\right),$$

then

$$\frac{\sqrt{\mathbf{h}_{n,T}}}{\sqrt{\Delta_{n,T}}} (num\Psi_{2,1}^{\sigma_b}) = O_p\left(\sqrt{\Delta_{n,T} v(T)}\right),$$

which implies

$$\frac{\sqrt{\mathbf{h}_{n,T}\widehat{L}_{n,T}(x)}}{\sqrt{\Delta_{n,T}}} \left(\frac{num\Psi_{2,1}^{\sigma_b}}{\frac{\Delta_{n,T}}{\mathbf{h}_{n,T}} \sum_{i=1}^n \mathbf{K}\left(\frac{X_{i\Delta_{n,T}} - x}{\mathbf{h}_{n,T}}\right)}\right) = O_p\left(\sqrt{\Delta_{n,T}}\right).$$

We now turn to

$$\frac{\sqrt{\mathbf{h}_{n,T}}}{\sqrt{\Delta_{n,T}}} (num\Psi_{2,2}^{\sigma_b}) = \frac{1}{\sqrt{\mathbf{h}_{n,T}\Delta_{n,T}}} \sum_{i=1}^{n-1} \mathbf{K}\left(\frac{X_{i\Delta_{n,T}} - x}{\mathbf{h}_{n,T}}\right) \int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}} \left(\int_{i\Delta_{n,T}}^s \boldsymbol{\sigma}_a(X_v) d\mathbf{B}_v\right) \boldsymbol{\sigma}_b(X_s) d\mathbf{B}_s.$$

https://doi.org/10.1017/S0266466617000305 Published online by Cambridge University Press

Writing $\boldsymbol{\sigma}_a(X_v) = \boldsymbol{\sigma}_{a,v}$ for conciseness, the quadratic variation $\left[\frac{\sqrt{\mathbf{h}_{n,T}}}{\sqrt{\Delta_{n,T}}}(num\boldsymbol{\Psi}^{\boldsymbol{\sigma}_b}_{2,2})\right]$ is

$$\begin{aligned}
&\frac{1}{\mathbf{h}_{n,T}\Delta_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}^2\left(\frac{X_{i\Delta_{n,T}}-x}{\mathbf{h}_{n,T}}\right)\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}\left(\int_{i\Delta_{n,T}}^{s}\boldsymbol{\sigma}_a(X_v)\mathrm{d}\mathbf{B}_v\right)^2(\boldsymbol{\sigma}_b(X_s)^\intercal\boldsymbol{\sigma}_b(X_s))\,\mathrm{d}s\\
&\quad=\frac{1}{\mathbf{h}_{n,T}\Delta_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}^2\left(\frac{X_{i\Delta_{n,T}}-x}{\mathbf{h}_{n,T}}\right)\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}\left(\int_{i\Delta_{n,T}}^{s}\boldsymbol{\sigma}^\intercal_{a,v}\boldsymbol{\sigma}_{a,v}\mathrm{d}v\right)\boldsymbol{\sigma}^\intercal_{b,s}\boldsymbol{\sigma}_{b,s}\mathrm{d}s\\
&\qquad+\frac{1}{\mathbf{h}_{n,T}\Delta_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}^2\left(\frac{X_{i\Delta_{n,T}}-x}{\mathbf{h}_{n,T}}\right)\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}\left[2\left(\int_{i\Delta_{n,T}}^{s}\left(\int_{i\Delta_{n,T}}^{u}\boldsymbol{\sigma}_{a,v}\mathrm{d}\mathbf{B}_v\right)\boldsymbol{\sigma}_{a,u}\mathrm{d}\mathbf{B}_u\right)\right]\boldsymbol{\sigma}^\intercal_{b,s}\boldsymbol{\sigma}_{b,s}\mathrm{d}s,
\end{aligned}$$

where the equality derives from Doob-Meyer's decomposition (Karatzas and Shreve, 1991, Theorem 4.10, page 24).

We begin with the second term. Denoting the continuous local martingale $\int_{i\Delta_{n,T}}^{t}\boldsymbol{\sigma}_{a,v}\mathrm{d}\mathbf{B}_v$ by $\widetilde{M}_{a,t}$ and integrating by parts, we obtain

$$\begin{aligned}
&\frac{1}{\mathbf{h}_{n,T}\Delta_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}^2\left(\frac{X_{i\Delta_{n,T}}-x}{\mathbf{h}_{n,T}}\right)\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}2\left[\left(\int_{i\Delta_{n,T}}^{s}\widetilde{M}_{a,u}\boldsymbol{\sigma}_{a,u}\mathrm{d}\mathbf{B}_u\right)\right]\boldsymbol{\sigma}^\intercal_{b,s}\boldsymbol{\sigma}_{b,s}\mathrm{d}s\\
&\quad=-\frac{1}{\mathbf{h}_{n,T}\Delta_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}^2\left(\frac{X_{i\Delta_{n,T}}-x}{\mathbf{h}_{n,T}}\right)\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}2\left(\int_{i\Delta_{n,T}}^{s}\boldsymbol{\sigma}^\intercal_{b,v}\boldsymbol{\sigma}_{b,v}\mathrm{d}v\right)\widetilde{M}_{a,s}\boldsymbol{\sigma}_{a,s}\mathrm{d}\mathbf{B}_s\\
&\qquad+\frac{1}{\mathbf{h}_{n,T}\Delta_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}^2\left(\frac{X_{i\Delta_{n,T}}-x}{\mathbf{h}_{n,T}}\right)2\left(\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}\boldsymbol{\sigma}^\intercal_{b,v}\boldsymbol{\sigma}_{b,v}\mathrm{d}v\right)\left(\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}\widetilde{M}_{a,v}\boldsymbol{\sigma}_{a,v}\mathrm{d}\mathbf{B}_v\right).
\end{aligned}$$

The quadratic variation of $\frac{1}{\mathbf{h}_{n,T}\Delta_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}^2\left(\frac{X_{i\Delta_{n,T}}-x}{\mathbf{h}_{n,T}}\right)\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}2\left(\int_{i\Delta_{n,T}}^{s}\boldsymbol{\sigma}^\intercal_{b,v}\boldsymbol{\sigma}_{b,v}\mathrm{d}v\right)\widetilde{M}_{a,s}\boldsymbol{\sigma}_{a,s}\mathrm{d}\mathbf{B}_s$ is

$$\frac{1}{\mathbf{h}^2_{n,T}\Delta^2_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}^4\left(\frac{X_{i\Delta_{n,T}}-x}{\mathbf{h}_{n,T}}\right)\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}4\left(\int_{i\Delta_{n,T}}^{s}\boldsymbol{\sigma}^\intercal_{b,v}\boldsymbol{\sigma}_{b,v}\mathrm{d}v\right)^2\widetilde{M}^2_{a,s}\boldsymbol{\sigma}^\intercal_{a,s}\boldsymbol{\sigma}_{a,s}\mathrm{d}s.$$

Now,

$$\begin{aligned}
&\mathrm{E}_\lambda\left(\frac{1}{\mathbf{h}^2_{n,T}\Delta^2_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}^4\left(\frac{X_{i\Delta_{n,T}}-x}{\mathbf{h}_{n,T}}\right)\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}4\left(\int_{i\Delta_{n,T}}^{s}\boldsymbol{\sigma}^\intercal_{b,v}\boldsymbol{\sigma}_{b,v}\mathrm{d}v\right)^2\widetilde{M}^2_{a,s}\boldsymbol{\sigma}^\intercal_{a,s}\boldsymbol{\sigma}_{a,s}\mathrm{d}s\right)\\
&\quad\le C_{\boldsymbol{\sigma}^\intercal_b\boldsymbol{\sigma}_b}\left(\frac{\Delta^2_{n,T}}{\Delta^2_{n,T}}\right)\mathrm{E}_\lambda\left(\frac{1}{\mathbf{h}^2_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}^4\left(\frac{X_{i\Delta_{n,T}}-x}{\mathbf{h}_{n,T}}\right)\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}4\widetilde{M}^2_{a,s}\boldsymbol{\sigma}^\intercal_{a,s}\boldsymbol{\sigma}_{a,s}\mathrm{d}s\right)\\
&\quad\le C_{\boldsymbol{\sigma}^\intercal_b\boldsymbol{\sigma}_b}\mathrm{E}_\lambda\left(\frac{1}{\mathbf{h}^2_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}^4\left(\frac{X_{i\Delta_{n,T}}-x}{\mathbf{h}_{n,T}}\right)\mathrm{E}_\lambda\left(\left(\sup_{i\Delta_{n,T}\le s\le(i+1)\Delta_{n,T}}\left|\widetilde{M}_{a,s}\right|\right)^2\left(\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}4\boldsymbol{\sigma}^\intercal_{a,s}\boldsymbol{\sigma}_{a,s}\mathrm{d}s\right)|\mathcal{F}_i\right)\right)\\
&\quad\le C_8\frac{\Delta_{n,T}}{\mathbf{h}_{n,T}}\mathrm{E}_\lambda\left(\frac{1}{\mathbf{h}_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}^4\left(\frac{X_{i\Delta_{n,T}}-x}{\mathbf{h}_{n,T}}\right)\mathrm{E}_\lambda\left(\left(\sup_{i\Delta_{n,T}\le s\le(i+1)\Delta_{n,T}}\left|\widetilde{M}_{a,s}\right|\right)^2|\mathcal{F}_i\right)\right)\\
&\quad\le C_9\frac{\Delta_{n,T}}{\mathbf{h}_{n,T}}\mathrm{E}_\lambda\left(\frac{1}{\mathbf{h}_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}^4\left(\frac{X_{i\Delta_{n,T}}-x}{\mathbf{h}_{n,T}}\right)\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}\boldsymbol{\sigma}^\intercal_{a,s}\boldsymbol{\sigma}_{a,s}\mathrm{d}s\right)\\
&\quad=O\left(\frac{\Delta_{n,T}}{\mathbf{h}_{n,T}}v(T)\right).
\end{aligned}\tag{A.22}$$

Therefore,

$$\frac{1}{\mathbf{h}_{n,T}\Delta_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}^2\left(\frac{X_{i\Delta_{n,T}}-x}{\mathbf{h}_{n,T}}\right)\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}2\left(\int_{i\Delta_{n,T}}^{s}\boldsymbol{\sigma}^\intercal_{b,v}\boldsymbol{\sigma}_{b,v}\mathrm{d}v\right)\widetilde{M}_{a,s}\boldsymbol{\sigma}_{a,s}\mathrm{d}\mathbf{B}_s=O_p\left(\frac{\Delta^{1/2}_{n,T}}{\mathbf{h}^{1/2}_{n,T}}v^{1/2}(T)\right).$$

https://doi.org/10.1017/S0266466617000305 Published online by Cambridge University Press

The additional term is

$$\begin{aligned}
&\frac{1}{\mathbf{h}_{n,T}\Delta_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}^2\left(\frac{X_{i\Delta_{n,T}}-x}{\mathbf{h}_{n,T}}\right)2\left(\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}\boldsymbol{\sigma}_{b,s}^{\intercal}\boldsymbol{\sigma}_{b,s}\mathrm{d}s\right)\left(\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}\widetilde{M}_{a,s}\boldsymbol{\sigma}_{a,s}\mathrm{d}\mathbf{B}_s\right)\\
&\quad\leq\frac{\Delta_{n,T}}{\mathbf{h}_{n,T}\Delta_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}^2\left(\frac{X_{i\Delta_{n,T}}-x}{\mathbf{h}_{n,T}}\right)2\boldsymbol{\sigma}_{b,i\Delta_{n,T}}^{\intercal}\boldsymbol{\sigma}_{b,i\Delta_{n,T}}\left(\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}\widetilde{M}_{a,s}\boldsymbol{\sigma}_{a,s}\mathrm{d}\mathbf{B}_s\right)\\
&\qquad+\frac{1}{\mathbf{h}_{n,T}\Delta_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}^2\left(\frac{X_{i\Delta_{n,T}}-x}{\mathbf{h}_{n,T}}\right)2\left(\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}\left(\boldsymbol{\sigma}_{b,s}^{\intercal}\boldsymbol{\sigma}_{b,s}-\boldsymbol{\sigma}_{b,i\Delta_{n,T}}^{\intercal}\boldsymbol{\sigma}_{b,i\Delta_{n,T}}\right)\mathrm{d}s\right)\\
&\qquad\times\left(\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}\widetilde{M}_{a,s}\boldsymbol{\sigma}_{a,s}\mathrm{d}\mathbf{B}_s\right).
\end{aligned}$$

Using the same method of proof leading to equation (A.22), the first term has a quadratic variation given by

$$\frac{\Delta_{n,T}^2}{\mathbf{h}_{n,T}^2\Delta_{n,T}^2}\sum_{i=1}^{n-1}\mathbf{K}^4\left(\frac{X_{i\Delta_{n,T}}-x}{\mathbf{h}_{n,T}}\right)4\left(\boldsymbol{\sigma}_{b,i\Delta_{n,T}}^{\intercal}\sigma_{b,i\Delta_{n,T}}\right)^2\left(\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}\widetilde{M}_s^2\left(\boldsymbol{\sigma}_{b,s}^{\intercal}\boldsymbol{\sigma}_{b,s}\right)\mathrm{d}s\right)=O_p\left(\frac{\Delta_{n,T}}{\mathbf{h}_{n,T}}v(T)\right).$$

Hence, the order of the first term is $O_p\left(\frac{\Delta_{n,T}^{1/2}}{\mathbf{h}_{n,T}^{1/2}}v^{1/2}(T)\right)$. As for the second term, we can treat it in just the same way as $\boldsymbol{\Psi}_{1,2}^{\mu_b,B,2}$ above and obtain an asymptotic order equal to $O_p(\Delta_{n,T}^{1/2}v(T))$.

Thus,

$$\begin{aligned}
&\frac{1}{\mathbf{h}_{n,T}\Delta_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}^2\left(\frac{X_{i\Delta_{n,T}}-x}{\mathbf{h}_{n,T}}\right)\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}\left(\int_{i\Delta_{n,T}}^{s}\boldsymbol{\sigma}_a(X_v)\mathrm{d}\mathbf{B}_v\right)^2(\boldsymbol{\sigma}_b(X_s)^{\intercal}\boldsymbol{\sigma}_b(X_s))\,\mathrm{d}s\\
&\quad=\frac{1}{\mathbf{h}_{n,T}\Delta_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}^2\left(\frac{X_{i\Delta_{n,T}}-x}{\mathbf{h}_{n,T}}\right)\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}\left(\int_{i\Delta_{n,T}}^{s}\boldsymbol{\sigma}_{a,v}^{\intercal}\boldsymbol{\sigma}_{a,v}\mathrm{d}v\right)\boldsymbol{\sigma}_{b,s}^{\intercal}\boldsymbol{\sigma}_{b,s}\mathrm{d}s\\
&\qquad+O_p\left(\frac{\Delta_{n,T}^{1/2}}{\mathbf{h}_{n,T}^{1/2}}v^{1/2}(T)\right)+O_p(\Delta_{n,T}^{1/2}v(T))\\
&\quad=\frac{\Delta_{n,T}}{2\mathbf{h}_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}^2\left(\frac{X_{i\Delta_{n,T}}-x}{\mathbf{h}_{n,T}}\right)\sigma_a(X_{i\Delta_{n,T}})^{\intercal}\boldsymbol{\sigma}_a(X_{i\Delta_{n,T}})\boldsymbol{\sigma}_b(X_{i\Delta_{n,T}})^{\intercal}\boldsymbol{\sigma}_b(X_{i\Delta_{n,T}})\\
&\qquad+O_p\left(\frac{\Delta_{n,T}^{1/2}}{\mathbf{h}_{n,T}^{1/2}}v^{1/2}(T)\right)+O_p(\Delta_{n,T}^{1/2}v(T)),
\end{aligned}\tag{A.23}$$

where the second identity derives from the fact that

$$\begin{aligned}
&\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}\left(\int_{i\Delta_{n,T}}^{s}\boldsymbol{\sigma}_{a,v}^{\intercal}\boldsymbol{\sigma}_{a,v}\mathrm{d}v\right)\boldsymbol{\sigma}_{b,s}^{\intercal}\boldsymbol{\sigma}_{b,s}\mathrm{d}s\\
&\quad=\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}\left(\int_{i\Delta_{n,T}}^{s}\left(\boldsymbol{\sigma}_{a,v}^{\intercal}\boldsymbol{\sigma}_{a,v}-\boldsymbol{\sigma}_a(X_{i\Delta_{n,T}})^{\intercal}\boldsymbol{\sigma}_a(X_{i\Delta_{n,T}})\right)\mathrm{d}v\right)\boldsymbol{\sigma}_{b,s}^{\intercal}\boldsymbol{\sigma}_{b,s}\mathrm{d}s\\
&\qquad+\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}\left(\int_{i\Delta_{n,T}}^{s}\boldsymbol{\sigma}_a(X_{i\Delta_{n,T}})^{\intercal}\boldsymbol{\sigma}_a(X_{i\Delta_{n,T}})\mathrm{d}v\right)\left(\boldsymbol{\sigma}_{b,s}^{\intercal}\boldsymbol{\sigma}_{b,s}-\boldsymbol{\sigma}_b(X_{i\Delta_{n,T}})^{\intercal}\boldsymbol{\sigma}_b(X_{i\Delta_{n,T}})\right)\mathrm{d}s\\
&\qquad+\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}\left(\int_{i\Delta_{n,T}}^{s}\boldsymbol{\sigma}_a(X_{i\Delta_{n,T}})^{\intercal}\boldsymbol{\sigma}_a(X_{i\Delta_{n,T}})\mathrm{d}v\right)\boldsymbol{\sigma}_b(X_{i\Delta_{n,T}})^{\intercal}\boldsymbol{\sigma}_b(X_{i\Delta_{n,T}})\mathrm{d}s.
\end{aligned}\tag{A.24}$$

https://doi.org/10.1017/S0266466617000305 Published online by Cambridge University Press

Notice that

$$\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}} \left(\int_{i\Delta_{n,T}}^{s} \boldsymbol{\sigma}_a(X_{i\Delta_{n,T}})^{\intercal} \boldsymbol{\sigma}_a(X_{i\Delta_{n,T}}) \mathrm{d}v \right) \boldsymbol{\sigma}_b(X_{i\Delta_{n,T}})^{\intercal} \boldsymbol{\sigma}_b(X_{i\Delta_{n,T}}) \mathrm{d}s$$
$$= \boldsymbol{\sigma}_a(X_{i\Delta_{n,T}})^{\intercal} \boldsymbol{\sigma}_a(X_{i\Delta_{n,T}}) \boldsymbol{\sigma}_b(X_{i\Delta_{n,T}})^{\intercal} \boldsymbol{\sigma}_b(X_{i\Delta_{n,T}}) \left(\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}} \left(\int_{i\Delta_{n,T}}^{s} \mathrm{d}v \right) \mathrm{d}s \right)$$
$$= \frac{\Delta_{n,T}^2}{2} \boldsymbol{\sigma}_a(X_{i\Delta_{n,T}})^{\intercal} \boldsymbol{\sigma}_a(X_{i\Delta_{n,T}}) \boldsymbol{\sigma}_b(X_{i\Delta_{n,T}})^{\intercal} \boldsymbol{\sigma}_b(X_{i\Delta_{n,T}}),$$

which is the dominant term in equation (A.24). Using the usual arguments based on Itô's lemma (see, e.g., equation (A.11)), the remaining two terms in equation (A.24) lead to probability orders dominated by $O_p\left(\frac{\Delta_{n,T}^{1/2}}{\mathbf{h}_{n,T}^{1/2}} v^{1/2}(T)\right) + O_p(\Delta_{n,T}^{1/2} v(T))$. Equation (A.23), therefore, gives

$$\frac{\sqrt{\mathbf{h}_{n,T} \widehat{L}_{n,T}(x)}}{\sqrt{\Delta_{n,T}}} \left(\frac{num \boldsymbol{\Psi}_{2,2}^{\boldsymbol{\sigma}_b}}{\frac{\Delta_{n,T}}{\mathbf{h}_{n,T}} \sum_{i=1}^{n} \mathbf{K}\left(\frac{X_{i\Delta_{n,T}} - x}{\mathbf{h}_{n,T}}\right)} \right) = O_p(1)$$

since

$$O_p\left(\frac{\Delta_{n,T}^{1/2}}{v^{1/2}(T) \mathbf{h}_{n,T}^{1/2}} \right) + O_p(\Delta_{n,T}^{1/2}) = o_p(1).$$

We now write the estimation error decomposition as

$$vech \widehat{\mathbf{a}}_{n,T}(x) - vech \mathbf{a}(x) = vech \boldsymbol{\Psi}_3 - vech \mathbf{a}(x) + vech \boldsymbol{\Psi}_1 + vech \boldsymbol{\Psi}_2. \tag{A.25}$$

We use the same procedure as in Theorem 4, i.e., we orthogonalize the vector of martingales $vech \boldsymbol{\Psi}_2$, and apply Theorem 3.16 in Höpfner and Löcherbach (2003) or Proposition 1 of Feigin (1985) to find that

$$\sqrt{\frac{\mathbf{h}_{n,T} \widehat{L}_{n,T}(x)}{\Delta_{n,T}}} vech \boldsymbol{\Psi}_2 \Rightarrow \mathbf{N}\left(0, \mathbf{K}_2 P_D \left(2\mathbf{a}(x) \otimes \mathbf{a}(x)\right) P_D^{\intercal}\right) \tag{A.26}$$

since

$$\frac{\sqrt{\mathbf{h}_{n,T} \widehat{L}_{n,T}(x)}}{\sqrt{\Delta_{n,T}}} \left(\frac{num \boldsymbol{\Psi}_{2,1}^{\boldsymbol{\sigma}_b}}{\frac{\Delta_{n,T}}{\mathbf{h}_{n,T}} \sum_{i=1}^{n} \mathbf{K}\left(\frac{X_{i\Delta_{n,T}} - x}{\mathbf{h}_{n,T}}\right)} \right) = O_p\left(\sqrt{\Delta_{n,T}}\right) = o_p(1)$$

and

$$\frac{\sqrt{\mathbf{h}_{n,T} \widehat{L}_{n,T}(x)}}{\sqrt{\Delta_{n,T}}} \left(\frac{num \boldsymbol{\Psi}_{2,2}^{\boldsymbol{\sigma}_b}}{\frac{\Delta_{n,T}}{\mathbf{h}_{n,T}} \sum_{i=1}^{n} \mathbf{K}\left(\frac{X_{i\Delta_{n,T}} - x}{\mathbf{h}_{n,T}}\right)} \right) = O_p(1),$$

https://doi.org/10.1017/S0266466617000305 Published online by Cambridge University Press

if $\frac{\Delta_{n,T}}{\mathbf{h}_{n,T}\widehat{L}_{n,T}(x)} \xrightarrow{p} 0$ and $\frac{\Delta_{n,T}}{h_{n,T}^2} \to 0$, where $P_D = (D^\intercal D)^{-1} D^\intercal$ is the Moore-Penrose generalized inverse of D (see, e.g., Magnus and Neudecker, 1988). Also, we note that

$$\frac{\sqrt{\mathbf{h}_{n,T}\widehat{L}_{n,T}(x)}}{\sqrt{\Delta_{n,T}}}(vech\Psi_1) = \frac{\sqrt{\mathbf{h}_{n,T}\widehat{L}_{n,T}(x)}}{\sqrt{\Delta_{n,T}}}O_p(\Delta_{n,T}) \tag{A.27}$$

$$+\frac{\sqrt{\mathbf{h}_{n,T}\widehat{L}_{n,T}(x)}}{\sqrt{\Delta_{n,T}}}O_p\left(\frac{\Delta_{n,T}}{\mathbf{h}_{n,T}^{1/2}v^{1/2}(T)}\right) \tag{A.28}$$

$$= o_p(1). \tag{A.29}$$

Using the derivations leading to equation (A.20) above, and the twice differentiability properties of the diffusion elements and the invariant density, we have

$$vech\Psi_3 - vech\mathbf{a}(x) = vech\mathbf{s}(x) + O_p\left(\frac{\Delta_{n,T}}{h_{n,T}^2}\right) + o_p(h_{n,T}^2),$$

where $\mathbf{s}(x)$ is the $d \times d$ matrix with elements

$$s_{gq}(x) = h_{n,T}^2\left(\int_S s^2K(s)ds\right)\left(\sum_{j=1}^{d}\frac{\partial a_{gq}(x)}{\partial x^j}\frac{\frac{\partial\tilde{\phi}(x)}{\partial x^j}}{\tilde{\phi}(x)} + \frac{1}{2}\sum_{j=1}^{d}\frac{\partial^2 a_{gq}(x)}{\partial x^j\partial x^j}\right), \qquad g,q = (1,1),\ldots,(d,d). \tag{A.30}$$

We obtain the stated result by combining the estimation error decomposition in equation (A.25) with equations (A.26), (A.27), and (A.30), provided $\frac{h_{n,T}^d\widehat{L}_{n,T}(x)}{\Delta_{n,T}} \xrightarrow{p} \infty$, $\frac{h_{n,T}^{d+4}\widehat{L}_{n,T}(x)}{\Delta_{n,T}} \xrightarrow{p} C$, and $h_{n,T}^{d-4}\widehat{L}_{n,T}(x)\Delta_{n,T} \xrightarrow{p} 0$. Notice, in fact, that

$$\begin{aligned}\sqrt{\frac{h_{n,T}^d\widehat{L}_{n,T}(x)}{\Delta_{n,T}}}\{vech\widehat{\mathbf{a}}_{n,T}(x) - vech\mathbf{a}(x)\} &= \sqrt{\frac{h_{n,T}^d\widehat{L}_{n,T}(x)}{\Delta_{n,T}}}vech\Psi_2 \\ &+\sqrt{\frac{h_{n,T}^d\widehat{L}_{n,T}(x)}{\Delta_{n,T}}}\left\{O_p\left(\Delta_{n,T} + \frac{\Delta_{n,T}}{\mathbf{h}_{n,T}^{1/2}v^{1/2}(T)}\right)\right\} \\ &+\sqrt{\frac{h_{n,T}^d\widehat{L}_{n,T}(x)}{\Delta_{n,T}}}\left\{O(h^2) + O_p\left(\frac{\Delta_{n,T}}{h_{n,T}^2}\right)\right\} \\ &\Rightarrow \mathbf{N}\left(0, \mathbf{K}_2P_D\left(2\mathbf{a}(x)\otimes\mathbf{a}(x)\right)P_D^\intercal\right),\end{aligned}$$

if $\frac{h_{n,T}^d\widehat{L}_{n,T}(x)}{\Delta_{n,T}} \xrightarrow{p} \infty$, $\frac{h_{n,T}^{d+4}\widehat{L}_{n,T}(x)}{\Delta_{n,T}} \xrightarrow{p} 0$ and $\sqrt{\frac{h_{n,T}^d\widehat{L}_{n,T}(x)}{\Delta_{n,T}}}\frac{\Delta_{n,T}}{h_{n,T}^2} \xrightarrow{p} 0$. ■

https://doi.org/10.1017/S0266466617000305 Published online by Cambridge University Press