On the functional estimation of multivariate diffusion processes

Federico Bandi and Guillermo Moloche

University of Chicago Booth School of Business, Johns Hopkins University Carey Business School, University of Chicago, Pontificia Universidad Católica del Perú

1. July 2008

ON THE FUNCTIONAL ESTIMATION OF MULTIVARIATE DIFFUSION PROCESSES*

Federico M. Bandi†
Graduate School of Business
University of Chicago

Guillermo Moloche‡
Economics Department
University of Chicago

This draft: July 2008

Abstract

We propose a fully nonparametric estimation theory for 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. Minimal assumptions are imposed on the statistical properties of the multivariate system to obtain limiting results. Harris recurrence is all that we require to show strong consistency and asymptotic (mixed) normality of the functional estimates. Hence, the estimation method and asymptotic theory apply to both stationary and nonstationary multivariate diffusion processes of the recurrent type.

Keywords: Harris recurrence, Multivariate diffusion processes, Nonparametric estimation, Double curse of dimensionality.

*We are grateful to Xiaohong Chen for her useful comments. We thank the seminar participants at ESSEC, UPenn, the University of Indiana, the Workshop on Financial Mathematics and Econometrics in Montréal, the Econometric Society Winter Meetings in Atlanta, and the Conference "Current Advances and Trends in Nonparametric Statistics" in Crete for discussions. Bandi acknowledges financial support from the IBM Corporation Faculty Research Fund at the Graduate School of Business, University of Chicago.

† *5807 South Woodlawn Avenue, Chicago, IL 60637.*

‡ *1126 East 59th Street, Chicago IL, 60637.*

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 et al. (2008), Bandi and Phillips (2008), Cai and Hong (2003), Fan (2005), Gallant and Tauchen (2008), and Johannes and Polson (2008) 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., Aït-Sahalia, 2008, 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 development appears to be important in virtue 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 and diffusion 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 estimates. Importantly, when dealing with multidimensional diffusion processes, a complete theory of inference for these traditional estimates has yet to be established, to the best of our knowledge, even in the stationary case.

To this extent, this paper derives the limiting properties of Nadaraya-Watson kernel estimates 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 non-zero Lebesgue measure in its admissible range an infinite number of times over time. Thus, it represents a sufficient 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).

We prove strong 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 Moloche (2004)). 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 non-zero Lebesgue measure. Unlike the concept of local time, the notion of occupation density extends to the multivariate framework (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 non-existence 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 paper provides a solution to this problem while offering additional insights about the simpler scalar diffusion case which, as discussed, may be viewed as a sub-case 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 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 occupation measure. The occupation measure is known to diverge linearly with the time span when examining positive Harris recurrent (ergodic) 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 course 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 has been made by, e.g., Guerre (2007), Karlsen and Tjøstheim (2001), Karlsen et al. (2007), and Moloche (2004b). We refer the reader to Park and Phillips (1998), Wang and Phillips (2008, 2009), and the reference therein for a promising alternative approach based on Skorohod embedding and nonlinear transformations of the embedded process.

Previous, stimulating 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, 2003). 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 et al. (2003) extend the univariate procedure in Stanton (1997) to propose nonparametric kernel analogues to drift and diffusion matrices for multidimensional diffusions. The asymptotic properties of their proposed Nadaraya-Watson-style estimators are not discussed, thereby rendering statistical inference difficult to implement and interpret in their framework. Downing (2003) evaluates the finite-sample properties of Boudouck et al.'s approach through simulations.

The paper is organized as follows. Section 2 introduces the model and the nonparametric estimates. Section 3 discusses Harris recurrence for multidimensional diffusion processes. Section 4 presents important preliminaries about Harris recurrent processes. These results are used in the development 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 $\Rightarrow$, $\stackrel{a.s.}{\to}$ and $:\stackrel{d}{=}$ stand for weak convergence, convergence with probability one, and distributional equivalence, respectively. When applied to a generic matrix A the operator *vec* stacks the column of A. 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 $(\Omega, \Im, P)$, the filtration of sub-σ-fields $\Im_t$, and the continuous adapted process $\{X_t, \Im_t; 0 \leq t < \infty\}$ with

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

where X_0 is a given initial condition, $\mathbf{B} = \{B_t, \Im_t; 0 \leq t < \infty\}$ is an m-dimensional standard Brownian motion, $\boldsymbol{\mu}(.) = \{\mu_i(.)\}_{1\leq i\leq d}$ is a $d \times 1$ Borel measurable drift vector and $\boldsymbol{\sigma}(.) = \{\sigma_{ij}(.)\}_{\substack{1\leq i\leq d \\ 1\leq j\leq m}}$ is a $d \times m$ Borel measurable matrix. Assume X_0 is taken to be independent of $\mathbf{B}$ and X_t takes values in $I \subseteq \Re^d$. Each coordinate X_t^i of the process can be written as

$$X_t^i = X_0^i + \int_0^t \mu_i(X_s)\, ds + \sum_{j=1}^m \int_0^t \sigma_{ij}(X_s)\, dB_s^j, \qquad 0 \leq t < \infty, 1 \leq i \leq d. \tag{2}$$

Define the $d \times d$ symmetric and non-negative (diffusion) matrix $\mathbf{a}(x) = \boldsymbol{\sigma}(x)\boldsymbol{\sigma}(x)'$ with generic element $a_{ij}(x) = \sum_{s=1}^m \sigma_{is}(x)\sigma_{sj}(x)$ $1 \leq i \leq d, 1 \leq j \leq d, \forall x \in I \subseteq \Re^d$. Write the conditional expectation on x, where $x = (x^1, x^2, ..., x^d)$ is a d-dimensional initial condition, as $\mathbf{E}^x[.]$. Hence, the drift vector $\boldsymbol{\mu}(.)$ and the diffusion matrix $\mathbf{a}(.)$ have classical representations in terms of infinitesimal conditional moments, i.e.,

$$\begin{aligned} \mathbf{E}^x\left[X_t^i - x^i\right] &= t\mu_i(x) + o(t) \\ \mathbf{E}^x\left[\left(X_t^i - x^i\right)\left(X_t^j - x^j\right)\right] &= ta_{ij}(x) + o(t), \end{aligned}$$

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

Now assume the process $\{X_t : t \geq 0\}$ is sampled at equispaced times $\{t = t_1, t_2, .., t_n\}$ in the interval $[0, T]$, where T is a strictly positive number. It readily follows that $\{X_t = X_{\Delta_{n,T}}, X_{2\Delta_{n,T}}, X_{3\Delta_{n,T}}, ..., X_{n\Delta_{n,T}}\}$ are n observations on the process X_t at $\{t_1 = \Delta_{n,T}, t_2 = 2\Delta_{n,T}, t_3 = 3\Delta_{n,T}, ..., t_n = n\Delta_{n,T}\}$, where $\Delta_{n,T} = T/n$.

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

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

and

$$\widehat{\mathbf{a}}_{n,T}(x) = \frac{1}{\Delta_{n,T}} \frac{\sum_{k=1}^{n-1} \mathbf{K}_{h_{n,T}}\left(X_{k\Delta_{n,T}} - x\right)\left(X_{(k+1)\Delta_{n,T}} - X_{k\Delta_{n,T}}\right)\left(X_{(k+1)\Delta_{n,T}} - X_{k\Delta_{n,T}}\right)'}{\sum_{k=1}^{n} \mathbf{K}_{h_{n,T}}\left(X_{k\Delta_{n,T}} - x\right)}, \tag{4}$$

where $\mathbf{K}_h(X_{k\Delta_{n,T}} - x) = \frac{1}{h^d}\prod_{i=1}^d k\left(\frac{X_{k\Delta_{n,T}}^i - x^i}{h}\right)$ is a product kernel function whose properties are laid out in Assumption 2 below and $h_{n,T}$ is a bandwidth sequence. The estimators in Eq. (3) and Eq. (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 strong 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 important to approximate the continuous trajectory of the process with a sample of discretely-sampled observations while replicating the infinitesimal properties of the relevant moments. More generally, the assumption is crucial for nonparametric identification in the absence of a time-invariant stationary density.

We view our asymptotic design as a realistic approximation in fields, such as finance, where data sets comprise observations sampled at relatively high frequencies over sufficiently long spans of time. To this extent, several simulation studies (see, e.g., Jiang and Knight (1999)) have shown that daily data, for instance, are valid 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 scopes of the present paper.

3 Harris Recurrence

This section discusses our assumed conditions on the underlying continuous-time process. Under Assumption 1 (a), the d-dimensional process $\{X_t : t \geq 0\}$ in (1) exists and is unique up to null sets. Under Assumption 1 (b), the process is Harris recurrent. Assumption 1 (a) and (b) are sufficient for the derivation of our limiting results.

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,* $\mathfrak{B}$*-measurable functions on* $I \subseteq \Re^d$ *where* $\mathfrak{B}$ *is the* σ*-field generated by Borel sets on* I*. Both functions satisfy local Lipschitz and linear growth conditions. Thus, for* $H > 0$ *there exists constants* $C_1(H)$ *and* $C_2(H)$ *such that*

$$||\boldsymbol{\mu}(x) - \boldsymbol{\mu}(y)|| + ||\boldsymbol{\sigma}(x) - \boldsymbol{\sigma}(y)|| \leq C_1(H)||x - y||,$$

and

$$||\boldsymbol{\mu}(x)|| + ||\boldsymbol{\sigma}(x)|| \leq C_2(H)\{1 + ||x||\},$$

where $||\boldsymbol{\sigma}||^2 = \sum_{i=1}^{d}\sum_{j=1}^{m}\sigma_{ij}^2$ *and* $||\boldsymbol{\mu}||^2 = \sum_{i=1}^{d}\mu_i^2$.

(b) *(Recurrence) Denote the closure of a generic set* A *by* $\overline{A}$. *Assume that, for every open and bounded set* $A \subset I$,

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

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

$$\mathfrak{L}\varphi(.) = \sum_{i=1}^{d}\mu_i(.)\frac{\partial\varphi(.)}{\partial x^i} + \frac{1}{2}\sum_{i=1}^{d}\sum_{k=1}^{d}a_{ik}(.)\frac{\partial\varphi(.)}{\partial x^i\partial x^k}.$$

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

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

and is such that $\Psi(r) := \min_{||x||=r}\varphi(.)$ *is strictly increasing with* $\lim_{r\to\infty}\Psi(r) = \infty$ *(c.f. Karatzas and Shreve, 1991, Exercise 7.13, part (i), page 370).*

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

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

and is such that $\Psi(r) := \min_{||x||=r}\varphi(.)$ *is strictly increasing with* $\lim_{r\to\infty}\Psi(r) = \infty$ *(c.f. Karatzas and Shreve, 1991, Exercise 7.13, part (ii), page 370).*

Let A be a measurable set of $I \subseteq \mathfrak{R}^d$ 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 $\overline{A}$. The process X_t is null Harris recurrent if $P^x\left[\tau_{\overline{A}} < \infty\right] = 1$ for every $x \in I\backslash\overline{A}$. The process X_t is positive Harris recurrent if $\mathbf{E}^x\left[\tau_{\overline{A}}\right] < \infty$ for every $x \in I\backslash\overline{A}$.

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

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

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, 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 ($d = 1$). The "speed measure," i.e.,

$$m(dx) = \frac{2dx}{\nu'(x)\sigma^2(x)} \qquad \forall x \in I \subseteq \Re$$

where $\nu(.)$ is the "scale function," i.e.,

$$\int_c^x \exp\left\{-2\int_c^\xi \frac{\mu(\varepsilon)}{\sigma^2(\varepsilon)}d\varepsilon\right\}d\xi \qquad \forall x \in I \subseteq \Re$$

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., $m(dx)/m(I) = p(dx)$, is the time-invariant probability measure of X, namely

$$\lim_{t\to\infty} P^x(X_t < u) = \frac{m((l,u))}{m(I)} \qquad \forall x, u \in I \subseteq \Re, \tag{5}$$

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,[1] it does of course apply to all strictly stationary or ergodic systems regardless of their linearity properties and dimensionality.[2] As discussed, Harris recurrence is a weaker assumption than stationarity and mixing (see, e.g., Meyne and Tweedie (1993)). In general, it is in fact the weakest assumption that one could impose to show point-wise identification of nonparametric estimates.[3] Intuitively, point-wise identification requires returns of the sample path of the process to local neighborhoods. This is precisely what Harris recurrence yields.

[1] Chen and Hansen (2002) provide an example of multidimensional diffusion displaying various recurrence properties depending on the relation between dimensionality of the system and parameters of the invariant measure and diffusion function. Assume $\phi(dx) = c_1(1 + ||x||^2)^{-\vartheta}dx$ 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 $\vartheta \geq \theta - 1 + \frac{d}{2}$.

[2] Thus, the well-known class of multivariate affine diffusions (linear drift vector $\boldsymbol{\mu}$ and linear diffusion matrix $\mathbf{a}$) is, trivially, positive recurrent under standard assumptions.

[3] Diffusion estimation is an important exception. Since the diffusion matrix can be estimated over a fixed time span (see, e.g., Brugiére (1991) and Remark 15 below), recurrence is not required and the process can be transient for identification. Should the focus be on both the diffusion matrix and the drift vector (i.e., on the full system's dynamics), as in this paper, then recurrence represents a necessary and sufficient condition for the identification of the full system. Similarly, the point-wise estimation of general conditional moments in discrete time requires recurrence at the minimum.

4 Preliminaries about Harris recurrent processes

We now present two theorems which will be useful in our subsequent analysis. Both theorems apply to Harris recurrent continuous-time Markov processes potentially more general than multivariate diffusion processes.

Theorem 1 (The Quotient Limit Theorem) *Consider the continuous-time Markov process X_t defined on the filtered probability space $(\Omega, \Im, (\Im_t)_{t\geq 0}, P)$. Assume X_t is Harris recurrent with invariant measure ϕ. Then, for any Borel measurable pair $f(.)$ and $g(.)$ that is integrable with respect to ϕ, the ratio of the functionals $\int_0^T f(X_s)ds$ and $\int_0^T g(X_s)ds$ is so that*

$$P^x\left(\lim_{T\to\infty}\frac{\int_0^T f(X_s)ds}{\int_0^T g(X_s)ds} = \frac{<\phi, f>}{<\phi, g>}\right) = 1, \tag{6}$$

provided $<\phi, g> = \int g(x)\phi(dx) > 0$.

Theorem 2 (The Darling-Kac Theorem) *Consider the continuous-time Markov process X_t defined on the filtered probability space $(\Omega, \Im, (\Im_t)_{t\geq 0}, P)$. Assume X_t is Harris recurrent with stationary transition densities and invariant measure ϕ. If, for a given non-negative function $f(.)$, there exists a function $v(s)$ such that*

$$\lim_{s\to 0}\frac{1}{v(s)}\mathbf{E}^x\left[\int_0^\infty e^{-st}f(X_t)dt\right] = C_X$$

for a positive (process-specific) constant C_X, and

$$v(s) = \frac{U(1/s)}{s^\alpha} \qquad 0 \leq \alpha \leq 1$$

so that $U(1/s)$ is slowly-varying as $s \to 0$,[4] *then it follows that*

$$\lim_{T\to\infty} P^x\left(\frac{1}{C_X v(1/T)}\int_0^T f(X_s)ds < u\right) = G_\alpha(u), \tag{7}$$

where

$$G_\alpha(u) = \frac{1}{\pi\alpha}\int_0^u \sum_{j=1}^\infty \frac{(-1)^{j-1}}{j!}\sin(\pi\alpha j)\,\Gamma(\alpha j+1)y^{j-1}dy$$

and $\Gamma(.)$ is the Gamma function.

[4] 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.

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 et al. (1966) for additional details.)

Remark 2 Theorem 2 assumes the existence of a regularly-varying function $v(.)$ satisfying certain properties. For diffusion processes the existence of this function is guaranteed (see, e.g., Kasahara (1975)).

Remark 3 Theorem 2 provides a weak convergence result for additive functionals of potentially-nonstationary continuous-time Markov processes. The function $G_\alpha(u)$ is the cumulative distribution of the Mittag-Leffler density, $g_\alpha(u)$. For $\alpha = 0$ the Mittag-Leffler density becomes the exponential density with parameter 1 and $G_0(u) = 1 - e^{-u}$ with $u \geq 0$. For $\alpha = \frac{1}{2}$, the Mittag-Leffler density corresponds to the truncated standard normal density and $G_{\frac{1}{2}}(u) = \frac{2}{\sqrt{2\pi}} \int_0^u e^{-y^2/2} dy$ with $u \geq 0$.

Remark 4 If $f(.)$ is the characteristic function of the generic set A, then the additive functional $\int_0^T f(X_s)ds$ defines the occupation time of the set A, i.e.,

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

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, $\mathbb{L}_X(T, A) \to \infty$ with probability one $\forall A \subset I$. The 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., $v(1/T) = \sqrt{T}$) yielding

$$\lim_{T \to \infty} P^x \left(\frac{1}{C_X \sqrt{T}} \int_0^T \mathbf{1}_{\{X_s \in A\}} ds < u \right) = G_{\frac{1}{2}}(u) = \frac{2}{\sqrt{2\pi}} \int_0^u e^{-y^2/2} dy \qquad u \geq 0.^5 \tag{8}$$

In the planar Brownian motion case, $\alpha = 0$ and $U(.) = \log(.)$ (i.e., $v(1/T) = \log T$) yielding,

$$\lim_{T \to \infty} P^x \left(\frac{1}{C_X \log T} \int_0^T \mathbf{1}_{\{X_s \in A\}} ds < u \right) = G_0(u) = 1 - e^{-u} \qquad u \geq 0. \tag{9}$$

[5] The well-known limiting distribution of the local time of a scalar Brownian motion is readily implied by this result, i.e., $\frac{1}{\sqrt{T}} L_X(T, x) = \frac{1}{\sqrt{T}} \lim_{\varepsilon \to 0} \frac{1}{2\varepsilon} \int_0^T \mathbf{1}_{\{|X_s - x| < \varepsilon\}} ds \underset{T \to \infty}{\Rightarrow} |B(1)| = |N(1, 0)|$ (see, e.g., Revuz and Yor (1998)).

The weak convergence results in Eq. (8) and Eq. (9) are versions of the Kallianpur-Robbins Theorem (Kallianpur and Robbins (1953)).

Remark 5 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., $v(1/T) = T$) yielding

$$\frac{1}{T}\int_0^T f(X_s)ds \xrightarrow{p} C_X = \int_{-\infty}^{\infty} f(x)p(dx), \tag{10}$$

which is a weak ergodic theorem. This result can be strengthened to almost sure convergence using the Quotient limit theorem. Eq. (10) 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 null recurrent process as shown in the previous remark. This observation will be important to understand the convergence properties of our functional estimates.

5 Asymptotics

Before discussing our limiting results, we present the assumption on the kernel function $\mathbf{K}(.)$ appearing in the definitions of the estimators in Eq. (3) and Eq. (4).

Assumption 2 *The function $\mathbf{K}(x)$ is a product kernel function $\Pi_{i=1}^d k(x^i)$. $k(.)$ is a nonnegative, bounded, continuous, and symmetric function on $\Re$ with $\int k(s)ds = 1$, $\int k^2(s)ds < \infty$, and $\int s^2 k(s)ds < \infty$. Additionally, there exists a nonnegative function $D(v, \varepsilon)$ such that*

$$|\mathbf{K}(x) - \mathbf{K}(v)| \leq D(v, \varepsilon)||x - v|| \tag{11}$$

$\forall\ x, v \in \Re^d$ so that $||x - v|| < \varepsilon$. Furthermore,

$$\lim_{\varepsilon \to 0} \int D(v, \varepsilon)dv < \infty, \tag{12}$$

and

$$\int D(v, \varepsilon)\phi(dv) < \infty \qquad \forall \varepsilon < \infty. \tag{13}$$

We start with the convergence properties of the averaged kernel function. In what follows, the symbol $\widetilde{\phi}(x)$ signifies $\frac{\phi(dx)}{dx}$, where ϕ is, as earlier, the invariant measure of the process. We also write $\widetilde{\phi}\,(x + hu)$ to signify $\widetilde{\phi}\left(x^1 + hu^1, ..., x^d + hu^d\right)$, where $u, x \in I \subseteq \Re^d$.

Theorem 3 *Define*

$$\widehat{L}_{n,T}(T,x) = \Delta_{n,T} \sum_{k=1}^{n} \mathbf{K}_{h_{n,T}}(X_{k\Delta_{n,T}} - x) \tag{14}$$

and suppose that $h_{n,T}$ *is such that*

$$\widehat{L}_{n,T}(T,x) \stackrel{a.s.}{\to} \infty$$

and

$$(\Delta_{n,T} \log(1/\Delta_{n,T}))^{1/2}\, \widehat{L}_{n,T}(T,x)/h_{n,T}^{d} \stackrel{a.s.}{\to} 0 \tag{15}$$

when $n, T \to \infty$ *so that* $\Delta_{n,T} \to 0$.

(i) Assume T *is fixed. Let* $n \to \infty$ *and fix* $h_{n,T}$*. Then,* $\widehat{L}_{n,T}(T,x)$ *converges with probability one to the random process* $\Phi(T,x,h_{n,T})$ *defined as*

$$\Phi(T,x,h_{n,T}) := \int_0^T \mathbf{K}_{h_{n,T}}(X_s - x)ds \quad \forall x \in I \subseteq \mathfrak{R}^d.$$

(ii) $\Phi(T,x,h_{n,T})$ *is such that*

$$\frac{\Phi(T,x,h_{n,T})}{v(1/T)} \Rightarrow C_X \left(\int \mathbf{K}\,(u)\, \widetilde{\phi}\,(x+hu)\, du \right) g_\alpha \quad \forall x \in I \subseteq \mathfrak{R}^d$$

as $h_{n,T} \to h > 0$ *when* $n, T \to \infty$*, for some function* $v(1/T)$ *which is regularly-varying at infinity with parameter* α *so that* $0 \le \alpha \le 1$*, where* g_α *is the Mittag-Leffler density with the same parameter* α*.* C_X *is a process-specific constant.*

(iii)

$$\frac{\Phi(T,x,h_{n,T})}{v(1/T)} \Rightarrow C_X \widetilde{\phi}\,(x)\, g_\alpha \quad \forall x \in I \subseteq \mathfrak{R}^d$$

if $n, T \to \infty$ *and* $h_{n,T} \to 0$ *with* $\Delta_{n,T} \to 0$.

(iv)

$$\frac{\widehat{L}_{n,T}(T,x)}{v(1/T)} \Rightarrow C_X \widetilde{\phi}\,(x)\, g_\alpha \quad \forall x \in I \subseteq \mathfrak{R}^d \tag{16}$$

as $h_{n,T} \to 0$ *with* $n, T \to \infty$ *so that* $\Delta_{n,T} \to 0$.

Remark 6 ($d = 1$) In the univariate case, Theorem 3 gives the (almost sure) convergence of the averaged kernel estimator to the local time of the process (from result (i) with $h_{n,T} \to 0$ together with a straightforward application of the occupation time formula, see, e.g., Revuz and Yor (1998, page 222)) as well as a weak convergence result for the local time estimator. The later result is particularly important since the growth rate of $\widehat{L}_{n,T}(T,x)$ has been shown to affect the rate of convergence of the drift and diffusion function estimators in the case of scalar recurrent diffusions (BP (2003)). For further discussions using regular variation the reader is referred to Moloche (2004). 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 4 above) and

$$\widehat{L}_{n,T}(T,x) = \sqrt{T} O_p(1). \tag{17}$$

If X is positive recurrent (or stationary) as in Remark 5, then $\alpha = 1$ and

$$\widehat{L}_{n,T}(T,x) \stackrel{a.s.}{\to} T p(x). \tag{18}$$

Similar findings were previously discussed by BP (2003) and Moloche (2004). In both papers they were obtained following different routes. Moloche (2004) and Park (2006) provide an interesting discussion of the asymptotic properties of the expected local time and its estimates.

Remark 7 ($d > 1$) In the more general multivariate case, local time is not defined but $\widehat{L}_{n,T}(T,x)$ is shown to converge to a random process (c.f. result (i)) whose rate of divergence to infinity is driven by a deterministic function of time which is regularly varying at infinity (c.f. result (iv)). In particular, the averaged kernel function $\widehat{L}_{n,T}(T,x)$ has divergence properties which mimic those of the additive functionals of the underlying process (from Theorem 2 above). This result will prove particularly 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 $\widehat{L}_{n,T}(T,x)$.

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

$$\frac{1}{T}\widehat{L}_{n,T}(T,x) = \frac{1}{n}\sum_{k=1}^{n} \mathbf{K}_{h_{n,T}}\left(X_{k\Delta_{n,T}} - x\right) \stackrel{p}{\to} p(x) \tag{19}$$

provided $\left(\Delta_{n,T}\log(1/\Delta_{n,T})\right)^{1/2} T/h^d_{n,T} \to 0$ 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 averaged kernel function represents a well-defined density estimator for every dimension. Formula (19) readily derives from (16). Consistently with Remark 5 above, this result can be sharpened in the sense that strong consistency can be proved under the same assumptions. Thus, we obtain a classical result in the nonparametric estimation of multivariate density functions (see, e.g., the review in Pagan and Ullah (1999)) as a sub-case of the more general theory discussed in this paper.

5.1 Estimating the drift vector

Theorem 4 and 5 below discuss the consistency and limiting distribution of the drift vector estimator in Eq. (3).

Theorem 4 (Consistency of the drift vector estimator) *If*

$$(\Delta_{n,T}\log(1/\Delta_{n,T}))^{1/2}\,\widehat{L}_{n,T}(T,x)/h_{n,T}^{d}\overset{a.s.}{\to}0$$

and $\widehat{L}_{n,T}(T,x)h_{n,T}^{d}\overset{a.s.}{\to}\infty$ *as* $h_{n,T}\to 0$ *with* $n,T\to\infty$ *and* $\Delta_{n,T}\to 0$, *then*

$$\widehat{\boldsymbol{\mu}}_{n,T}(x)\overset{a.s.}{\to}\boldsymbol{\mu}(x)\qquad\forall x\in I\subseteq\Re^{d}.$$

Theorem 5 (The asymptotic distribution of the drift vector estimator) *If*

$$(\Delta_{n,T}\log(1/\Delta_{n,T}))^{1/2}\,\widehat{L}_{n,T}(T,x)/h_{n,T}^{d}\overset{a.s.}{\to}0,$$

$$\widehat{L}_{n,T}(T,x)h_{n,T}^{d}\overset{a.s.}{\to}\infty,$$

and $h_{n,T}=O_{a.s.}\left(\widehat{L}_{n,T}(T,x)^{-\frac{1}{d+4}}\right)$ *as* $h_{n,T}\to 0$ *with* $n,T\to\infty$ *and* $\Delta_{n,T}\to 0$, *then*

$$\begin{aligned}&\sqrt{\widehat{L}_{n,T}(T,x)h_{n,T}^{d}}\,(\widehat{\boldsymbol{\mu}}(x)-\boldsymbol{\mu}(x)-\boldsymbol{\Gamma}^{\mu}(x))\\ \Rightarrow\quad&(\mathbf{a}(x))^{1/2}\,\mathbf{N}\left(0,\left(\int k^{2}(u)\,du\right)^{d}\mathbf{I}\right),\quad\forall x\in I\subseteq\Re^{d}\end{aligned}\tag{20}$$

where

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

$$bias_{g}(x)=h_{n,T}^{2}\left(\int s^{2}k(s)ds\right)\left(\sum_{i=1}^{d}\frac{\partial\mu_{g}(x)}{\partial x_{i}}\frac{\frac{\partial\widetilde{\phi}(x)}{\partial x_{i}}}{\widetilde{\phi}(x)}+\frac{1}{2}\sum_{i=1}^{d}\frac{\partial^{2}\mu_{g}(x)}{\partial x_{i}\partial x_{i}}\right)\qquad\forall g=1,...,d,$$

and $\phi(dx) = \widetilde{\phi}(x)dx$ *is the* σ*-finite invariant measure of the process. If* $h_{n,T}^{d+4}\widehat{L}_{n,T}(T,x) \overset{a.s.}{\to} 0$ *as* $h_{n,T} \to 0$ *with* $n, T \to \infty$ *and* $\Delta_{n,T} \to 0$,

$$(\Delta_{n,T}\log(1/\Delta_{n,T}))^{1/2}\,\widehat{L}_{n,T}(T,x)/h_{n,T}^{d} \overset{a.s.}{\to} 0,$$

and

$$\widehat{L}_{n,T}(T,x)h_{n,T}^{d} \overset{a.s.}{\to} \infty,$$

then

$$\begin{aligned} &\sqrt{\widehat{L}_{n,T}(T,x)h_{n,T}^{d}}\,(\widehat{\boldsymbol{\mu}}(x) - \boldsymbol{\mu}(x)) \\ \Rightarrow\quad &(\mathbf{a}(x))^{1/2}\,\mathbf{N}\left(0, \left(\int k^2(u)\,du\right)^{d}\mathbf{I}\right) \quad \forall x \in I \subseteq \mathfrak{R}^d. \end{aligned} \tag{21}$$

Remark 9 The quantity $\widehat{L}_{n,T}(T,x)$ plays the same role here played by the number of data points in the more standard estimation of conditional expectations in the stationary, discrete-time, context. What matters in our framework 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. For consistency, we require $\widehat{L}_{n,T}(T,x) \overset{a.s.}{\to} \infty$ but this result is guaranteed by the Harris recurrence of the process, as shown in Theorem 3 above.

Remark 10 The smoothing sequence $h_{n,T}$ has to accommodate the divergence properties of the quantity $\widehat{L}_{n,T}(T,x)$ so that $h_{n,T}^{d}\widehat{L}_{n,T}(T,x) \overset{a.s.}{\to} \infty$. By virtue of Theorem 3, in the case $d = 1$, such a condition collapses to the standard assumption $h_{n,T}\overline{L}_X(T,x) \overset{a.s.}{\to} \infty\ \forall x \in I \subseteq \mathfrak{R}$, where $\overline{L}_X(T,x)$ is the chronological local time of the process X.

Remark 11 (Bandwidth choice) The asymptotic mean-squared error (AMSE) of the drift vector estimator is of order $O_p(h_{n,T}^4) + O_p\left(\frac{1}{\widehat{L}_{n,T}(T,x)h_{n,T}^{d}}\right)$. In consequence, the optimal drift bandwidth in the mean-squared error sense should be proportional to $\widehat{L}_{n,T}(T,x)^{-\frac{1}{d+4}}$ where d is the number of equations in the system and $\widehat{L}_{n,T}(T,x)$ is defined as in (14). Typically, the smoothing sequence may be set equal to

$$h_{n,T}^{drift}(x) = c^{drift}\left(\frac{1}{\log \widehat{L}_{n,T}(T,x)}\right)\widehat{L}_{n,T}(T,x)^{-\frac{1}{d+4}}. \tag{22}$$

This choice allows us 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.

It is noted that computation of $\widehat{L}_{n,T}(T,x)$ requires choice of an additional smoothing parameter. Furthermore, the constant c^{drift} should be evaluated using automated methods. The design of data-driven procedures for selecting the proportionality factor playing a role in (22) and the optimal bandwidth for estimating $\widehat{L}_{n,T}(T,x)$ is of apparent importance but goes beyond the scope of the present paper and is left for future research.

The form of (22) clarifies the potential relevance of local adaptation when estimating the drift vector. In particular, we expect the optimal drift bandwidth to depend inversely on the number of observations in the local neighborhood of a point. In stationary kernel regression, bias reduction is a conventional justification for employing smoothing sequences that are inversely related to the availability of observations as summarized by the estimated density function of the underlying data (c.f. Pagan and Ullah (1999)). Differently from more standard problems in the nonparametric estimation of conditional expectations in discrete time, in our framework the potentially important role played by local adaptation emerges directly from the asymptotic conditions which the drift bandwidth ought to satisfy.

Remark 12 (The ergodic and strictly stationary case) In the positive Harris recurrent case $\widehat{L}_{n,T}(T,x) \overset{a.s.}{\to} Tp(x)$ and

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

since the invariant measure is integrable, i.e. $\phi(I) < \infty$ (see comments in Section 3). In consequence, if

$$\left(\Delta_{n,T}\log(1/\Delta_{n,T})\right)^{1/2} T/h_{n,T}^d \to 0,$$

$Th_{n,T}^d \to \infty$ and $h_{n,T} = O_{a.s.}\left(T^{-\frac{1}{d+4}}\right)$ as $h_{n,T} \to 0$ with $n,T \to \infty$ and $\Delta_{n,T} \to 0$, then

$$\begin{aligned} &\sqrt{Th_{n,T}^d}\left(\widehat{\mu}_g(x) - \mu_g(x) - bias_g(x)\right) \\ \Rightarrow\quad &\mathbf{N}\left(0, \left(\int k^2(u)\,du\right)^d \frac{\sum_{j=1}^m \sigma_{gj}^2(x)}{p(x)}\right), \quad \forall x \in I \subseteq \mathfrak{R}^d, \forall g = 1,...,d, \end{aligned}$$

and

$$bias_g(x) = h_{n,T}^2\left(\int s^2 k(s)ds\right)\left(\sum_{i=1}^d \frac{\partial \mu_g(x)}{\partial x_i}\frac{\frac{\partial p(x)}{\partial x_i}}{p(x)} + \frac{1}{2}\sum_{i=1}^d \frac{\partial^2 \mu_g(x)}{\partial x_i \partial x_i}\right).$$

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

Remark 13 (The *two* curses of dimensionality) Consistently 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 exponentially on d, and a *stochastic* effect which depends on the speed of divergence to infinity of the quantity $\widehat{L}_{n,T}(T, x)$. 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 exponentially on d. We recall that the AMSE-optimal bandwidth is $h_{n,T} = O_{a.s.}\left(T^{-\frac{1}{d+4}}\right)$. 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 $\widehat{L}_{n,T}(T, x)$ which are inversely related to the dimensionality of the problem. Importantly, while these rates cannot be quantified in general, practical implementation of our procedures does not require their evaluation a-priori, as we discuss in Remark 14. The scalar and planar Brownian motion cases are notable exceptions for which the rates can be computed in closed form (c.f. Remark 4).

To summarize, under stationarity (or positive Harris recurrence), the curse of dimensionality operates through only one channel, i.e. the dimension d, since $\widehat{L}_{n,T}(T, x)$ diverges at the constant rate T independently of the number of equations in the system (c.f. Remark 5). 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 $\widehat{L}_{n,T}(T, x)$. The optimal (in an AMSE sense) bandwidth $h_{n,T} = O_{a.s.}\left(\widehat{L}_{n,T}(T, x)^{-\frac{1}{d+4}}\right)$ accounts for both effects.

We expect the "double course 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 is of course not specific to the study of continuous time processes.

Remark 14 The asymptotic distribution in (21) depends on quantities which can be estimated from the data, i.e. $\widehat{L}_{n,T}(T, x)$ and $\mathbf{a}(.)$ (c.f. Theorem 3 and Theorem 6 below). 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 easily estimable random variances.

Of course, from a theoretical standpoint, the rates of convergence are 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 $\widehat{L}_{n,T}(T,x)$ (c.f. Theorem 3). Importantly, however, such a constant does not have to be identified empirically for statistical inference to be conducted.

Similar arguments to those in Remark 9 through 14 apply to the estimation of the diffusion matrix to which we now turn. Below we will place emphasis on those aspects that are specific to diffusion matrix evaluation.

5.2 Estimating the diffusion matrix

Theorem 6 and 7 below discuss the consistency and limiting distribution of the diffusion matrix estimator in Eq. (4).

Theorem 6 (Consistency of the diffusion matrix estimator) *If*

$$\widehat{L}_{(n,T)}(T,x) \overset{a.s.}{\to} \infty$$

and

$$\frac{(\Delta_{n,T}\log(1/\Delta_{n,T}))^{1/2}\,\widehat{L}_{(n,T)}(T,x)}{h^d_{n,T}} \overset{a.s.}{\to} 0$$

with $h_{n,T} \to 0$ *when* $n, T \to \infty$ *so that* $\Delta_{n,T} \to 0$, *then*

$$\widehat{\mathbf{a}}_{(n,T)}(x) \overset{a.s.}{\to} \mathbf{a}(x) \qquad \forall x \in I \subseteq \Re^d.$$

Theorem 7 (The limiting distribution of the diffusion matrix estimator) *Assume*

$$\widehat{L}_{(n,T)}(T,x) \overset{a.s.}{\to} \infty$$

and

$$\frac{(\Delta_{n,T}\log(1/\Delta_{n,T}))^{1/2}\,\widehat{L}_{(n,T)}(T,x)}{h^d_{n,T}} \overset{a.s.}{\to} 0$$

with $n, T \to \infty$ *so that* $\Delta_{n,T} \to 0$. *If*

$$\sqrt{\frac{h_{n,T}^{d+4}\widehat{L}_{(n,T)}(T,x)}{\Delta_{n,T}}} \overset{a.s.}{\to} 0$$

as $h_{n,T} \to 0$ *with* $n, T \to \infty$ *so that* $\Delta_{n,T} \to 0$, *then*

$$\sqrt{\frac{h_{n,T}^{d}\widehat{L}_{(n,T)}(T,x)}{\Delta_{n,T}}} \left(vech\widehat{\mathbf{a}}_{(n,T)}(x) - vech\mathbf{a}(x)\right) \Rightarrow \quad (\mathbf{\Xi}(x))^{1/2}\,\mathbf{N}\left(0, \left(\int k^2(s)ds\right)^d \mathbf{I}\right), \quad \forall x \in I \subseteq \Re^d, \tag{23}$$

where

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

$$P_D = (D^{'}D)^{-1}D^{'},$$

and D *is the standard duplication matrix, i.e., the unique* $d^2 \times (d(d+1))/2$ *matrix such that*

$$vech\mathbf{a}(x) = P_D vec\mathbf{a}(x) = \begin{bmatrix} a_{1,1} \\ a_{2,1} \\ a_{2,2} \\ a_{3,1} \\ \ldots \\ a_{d,d} \end{bmatrix}.$$

If

$$\sqrt{\frac{h_{n,T}^{d+4}\widehat{L}_{(n,T)}(T,x)}{\Delta_{n,T}}} = O_{a.s.}(1),$$

then

$$\sqrt{\frac{h_{n,T}^{d}\widehat{L}_{(n,T)}(T,a)}{\Delta_{n,T}}} \left(vech\widehat{\mathbf{a}}_{(n,T)}(x) - vech\mathbf{a}(x) - \mathbf{\Gamma}^{\sigma^2}(x)\right) \Rightarrow \quad (\mathbf{\Xi}(x))^{1/2}\,\mathbf{N}\left(0, \left(\int k^2(s)ds\right)^d \mathbf{I}\right), \quad \forall x \in I \subseteq \Re^d \tag{24}$$

where

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

with

$$bias_{i,j}(x) = h_{n,T}^2\left(\int s^2 k(s)ds\right)\left(\sum_{k=1}^{d}\frac{\partial a_{i,j}(x)}{\partial x_k}\frac{\frac{\partial\widetilde{\phi}(x)}{\partial x_k}}{\widetilde{\phi}(x)} + \frac{1}{2}\sum_{k=1}^{d}\frac{\partial^2 a_{i,j}(x)}{\partial x_k \partial x_k}\right) \qquad for\ i,j = (1,1), ..., (d,d).$$

Remark 15 (The fixed T case) Contrary to drift estimation, the features of Theorem 6 clarify that the diffusion matrix can be consistently estimated over a fixed span of data $T = \overline{T}$ (c.f. Florens-Zmirou (1993) and Brugiére (1991, 1993)). In fact, provided

$$\frac{\left(\Delta_{n,\overline{T}} \log(1/\Delta_{n,\overline{T}})\right)^{1/2} \widehat{L}_{(n,\overline{T})}(\overline{T}, x)}{h^d_{n,T}} \overset{a.s.}{\to} 0$$

with $h_{n,\overline{T}} \to 0$ and $n \to \infty$, then

$$\widehat{\mathbf{a}}_{\left(n,\overline{T}\right)}(x) \overset{a.s.}{\to} \mathbf{a}(x) \qquad \forall x \in I \subseteq \Re^d.$$

This finding complements the consistency results in the multivariate context studied by Brugiére (1991) where convergence is in probability.

Remark 16 (The bivariate case) It is worth being explicit about the form of the limiting variance in (23). We consider the simple bivariate case, i.e. $d = 2$. Write

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

where $\sum = \sum_{i=1}^m$. 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 $vec\mathbf{a}(x)$ can be written as

$$\begin{aligned}
& D\boldsymbol{\Xi}(x)D^{'} \\
= \;& DP_D\left(2\mathbf{a}(x)\otimes\mathbf{a}(x)\right)P_D^{'}D \\
= \;& 2D(D^{'}D)^{-1}D^{'}\left(\mathbf{a}(x)\otimes\mathbf{a}(x)\right)D^{'}(D^{'}D)^{-1}D \\
= \;& 2\overline{P}_D\left(\mathbf{a}(x)\otimes\mathbf{a}(x)\right)\overline{P}_D^{'} \\
= \;& 2\overline{P}_D\left(\mathbf{a}(x)\otimes\mathbf{a}(x)\right)\overline{P}_D \\
= \;& 2\overline{P}_D\left(\mathbf{a}(x)\otimes\mathbf{a}(x)\right)
\end{aligned}$$

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 $\Re^{d^2}$ 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 (i,j) 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

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

which implies

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

Remark 17 (Bandwidth choice) A "nearly" optimal selection rule for the bandwidth in the diffusion case is given by:

$$h_{n,T}^{diff}(x) \approx c^{diff}\left(\frac{1}{\log\left(\widehat{L}_{n,T}(T,x)/\Delta_{n,T}\right)}\right)\left(\widehat{L}_{n,T}(T,x)/\Delta_{n,T}\right)^{-\frac{1}{d+4}}. \tag{25}$$

Coherently with our previous discussion in the drift case (c.f. Remark 11), this choice allows us to eliminate the influence of the bias term from the limiting distribution of the diffusion estimates and obtain centering about zero, while achieving a close-to-optimal speed of convergence.

Being the diffusion function estimable over a fixed span of observations (c.f. Remark 14), there is 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

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

which is standard in nonparametric statistics and does not depend on the spatial level x.

Remark 18 (Non-vanishing bandwidths)

(1) For scalar and planar Brownian motion (which have constant drift, diffusion and invariant measure) all the results in this work go through unmodified with constant or explosive bandwidths (i.e. $h_{n,T} \to \infty$) provided the bandwidths satisfy the admissibility conditions in the statements of the corresponding theorems. Phillips and Park (1998) find a similar result when estimating nonparametrically the (constant) conditional first moment of a standard random walk embeddable in Brownian motion.

(2) If either one of the two functions of interest is constant irrespective of the shape of the other function, then the consistency result for that function is valid in the presence of a constant bandwidth. The weak convergence result is also valid with the caveat that the asymptotic variance has a form which depends on the constant smoothing sequence being used. Such a variance can be easily deduced from the proofs of Theorems 5 and 7.

(3) If either one of the two functions of interest is constant irrespective of the shape of the other function, then the consistency result for that function is valid in the presence of explosive bandwidths provided the relevant bandwidth conditions are satisfied.

6 Conclusion

This paper 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 strong consistency and asymptotic (mixed) normality of the functional estimates of drift vector and diffusion matrix. On the one hand, this assumption is known to be milder than stationarity and mixing and might prove useful to study multivariate (discrete- or continuous-) time processes whose stationarity can neither be guaranteed nor ruled out a priori. On the other hand, 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,b) and, for interesting work in the scalar diffusion case, Fan and Zhang (2003) and Moloche (2004)).

Even though the focus of this paper 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) and Fan (1994), among others, is a possible criterion. Interestingly, such a comparison might be conducted separately for drift and diffusion since these moments may, as shown, be identified separately.

We leave the study of alternative kernel methods for multivariate diffusions and the design of testing procedures for multivariate parametric specifications for future work.

7 Appendix

Proof of Theorem 1 See Revuz and Yor (1998, Theorem 3.12, page 408) and Azéma et al. (Remark 1, page 170, 1966).

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

Proof of Theorem 3 In what follows, for convenience, we use the notation $\frac{1}{\mathbf{h}}\mathbf{K}\left(\frac{a}{\mathbf{h}}\right) = \frac{1}{h^d}\mathbf{K}\left(\frac{a_1}{h},...,\frac{a_d}{h}\right)$, $\widetilde{\phi}(x+\mathbf{h}u) = \widetilde{\phi}(x_1+hu_1,...,x_d+hu_d)$, where $a,u,x \in I \subseteq \Re^d$ and $\mathbf{h} = h^d$. We begin with part (i). We wish to show that

$$\frac{\Delta_{n,T}}{\mathbf{h}_{n,T}}\sum_{i=1}^{n}\mathbf{K}\left(\frac{X_{i\Delta_{n,T}}-x}{\mathbf{h}_{n,T}}\right) \underset{n\to\infty}{\overset{a.s.}{\to}} \frac{1}{\mathbf{h}_{n,T}}\int_0^T \mathbf{K}\left(\frac{X_s-x}{\mathbf{h}_{n,T}}\right)ds$$

for a fixed $\mathbf{h}_{n,T}$. Let

$$\delta_{i\Delta_{n,T}} = \sup_{|t-s|\le\Delta_{n,T}\ t,s\in[i\Delta_{n,T},(i+1)\Delta_{n,T}]} ||X_t - X_s||.$$

Using the modulus of continuity of multivariate Brownian semimartingales (see, e.g., McKean (1969)), write

$$P\left[\limsup_{\Delta_{n,T}\downarrow 0}\frac{\delta_{i\Delta_{n,T}}}{\sqrt{\Delta_{n,T}\ln(1/\Delta_{n,T})}} = \max_{t\in[i\Delta_{n,T},(i+1)\Delta_{n,T}]}\sqrt{2\gamma(X_t)}\right] = 1,$$

where $\gamma(x)$ is the largest eigenvalue of the diffusion matrix $\mathbf{a}(x) = \boldsymbol{\sigma}(x)\boldsymbol{\sigma}(x)'$. Since this matrix is positive definite, all its eigenvalues are positive. Moreover, since $trace[\mathbf{a}(X_t)] < \infty\ \forall t \in [i\Delta_{n,T},(i+1)\Delta_{n,T}]$, then all its eigenvalues are bounded in compact subsets. Hence,

$$\limsup_{\Delta_{n,T}\downarrow 0}\frac{\kappa_{n,T}}{\sqrt{\Delta_{n,T}\ln(1/\Delta_{n,T})}} < \infty \qquad a.s., \tag{26}$$

where

$$\kappa_{n,T} = \max_{1\le i\le n}\delta_{i\Delta_{n,T}}.$$

Now, write

$$\begin{aligned}
&\left|\frac{\Delta_{n,T}}{\mathbf{h}_{n,T}}\sum_{i=1}^{n}\mathbf{K}\left(\frac{X_{i\Delta_{n,T}}-x}{\mathbf{h}_{n,T}}\right) - \frac{1}{\mathbf{h}_{n,T}}\int_0^T\mathbf{K}\left(\frac{X_s-x}{\mathbf{h}_{n,T}}\right)ds\right| \\
= &\left|\sum_{i=0}^{n-1}\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}\frac{1}{\mathbf{h}_{n,T}}\left[\mathbf{K}\left(\frac{X_{i\Delta_{n,T}}-x}{\mathbf{h}_{n,T}}\right) - \mathbf{K}\left(\frac{X_s-x}{\mathbf{h}_{n,T}}\right)\right]ds\right. \\
&\left. - \frac{\Delta_{n,T}}{\mathbf{h}_{n,T}}\mathbf{K}\left(\frac{X_{0\Delta_{n,T}}-x}{\mathbf{h}_{n,T}}\right) + \frac{\Delta_{n,T}}{\mathbf{h}_{n,T}}\mathbf{K}\left(\frac{X_{n\Delta_{n,T}}-x}{\mathbf{h}_{n,T}}\right)\right| \\
\le &\left|\sum_{i=0}^{n-1}\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}\frac{1}{\mathbf{h}_{n,T}}\left[\mathbf{K}\left(\frac{X_{i\Delta_{n,T}}-x}{\mathbf{h}_{n,T}}\right) - \mathbf{K}\left(\frac{X_s-x}{\mathbf{h}_{n,T}}\right)\right]ds\right| \\
&+ \left|\frac{\Delta_{n,T}}{\mathbf{h}_{n,T}}\mathbf{K}\left(\frac{X_{0\Delta_{n,T}}-x}{\mathbf{h}_{n,T}}\right)\right| + \left|\frac{\Delta_{n,T}}{\mathbf{h}_{n,T}}\mathbf{K}\left(\frac{X_{n\Delta_{n,T}}-x}{\mathbf{h}_{n,T}}\right)\right| \\
\le &\frac{\kappa_{n,T}}{\mathbf{h}_{n,T}}\left|\int_0^T\frac{1}{\mathbf{h}_{n,T}}D\left(\frac{X_s-x}{\mathbf{h}_{n,T}},\frac{\kappa_{n,T}}{\mathbf{h}_{n,T}}\right)ds\right| + O_{a.s.}\left(\frac{\Delta_{n,T}}{\mathbf{h}_{n,T}}\right)
\end{aligned}$$

by the triangle inequality and the regularity conditions of the kernel function from Assumption 2 above. But,

$$\int_0^T \frac{1}{\mathbf{h}_{n,T}} D\left(\frac{X_s - x}{\mathbf{h}_{n,T}}, \frac{\kappa_{n,T}}{\mathbf{h}_{n,T}}\right) ds = O_{a.s.}\left(\frac{1}{\mathbf{h}_{n,T}} \int_0^T \mathbf{K}\left(\frac{X_s - x}{\mathbf{h}_{n,T}}\right) ds\right),$$

by the Quotient limit theorem. Then,

$$\left| \frac{\Delta_{n,T}}{\mathbf{h}_{n,T}} \sum_{i=1}^{n} \mathbf{K}\left(\frac{X_{i\Delta_{n,T}} - x}{\mathbf{h}_{n,T}}\right) - \frac{1}{\mathbf{h}_{n,T}} \int_0^T \mathbf{K}\left(\frac{X_s - x}{\mathbf{h}_{n,T}}\right) ds \right| \overset{a.s.}{\to} 0$$

using the assumptions of the theorem. This proves the result in (i). Results (ii) and (iii) are easily proved using arguments contained in the proof of (iv). For brevity, we only focus on (iv). We wish to show that

$$\lim_{n,T\to\infty} \left| \mathbf{E}\left[\xi\left(\frac{\Delta_{n,T}}{v(1/T)\mathbf{h}_{n,T}} \sum_{i=1}^{n} \mathbf{K}\left(\frac{X_{i\Delta_{n,T}} - x}{\mathbf{h}_{n,T}}\right)\right)\right] - \mathbf{E}\left[\xi\left(C_X \widetilde{\phi}(x)\, g_\alpha\right)\right] \right| = 0,$$

for any bounded and continuous function ξ. We will require the window width to be chosen from a totally bounded, complete, and non-empty set and vanish slowly enough as to guarantee uniform convergence over bandwidth sequences. This requirement is formalized as follows. The bandwidth sequence $h_{n,T}$ on $\mathfrak{R}$ is $\mathcal{F}_T$-adapted, bounded, and such that $h_{n,T} \in \mathcal{H}_{T,n}(\varepsilon)$ where

$$\mathcal{H}_{T,n}(\varepsilon) = \left\{ h : \max\left(\widehat{L}_{(n,T)}(T,x)\sqrt{\Delta_{n,T}\log(1/\Delta_{n,T})}, 1/\widehat{L}_{(n,T)}(T,x)\right)/\varepsilon < h^d < \varepsilon \right\}.$$

Then, given (i), we simply need to prove that

$$\lim_{n,T\to\infty} \left| \underbrace{\mathbf{E}\left[\xi\left(\frac{1}{v(1/T)\mathbf{h}_{n,T}} \int_0^T \mathbf{K}\left(\frac{X_s - x}{\mathbf{h}_{n,T}}\right) ds\right)\right] - \mathbf{E}\left[\xi\left(C_X \widetilde{\phi}(x)\, g_\alpha\right)\right]}_{\alpha_T} \right| = 0. \tag{27}$$

This is true if, for any $\delta > 0$, there exists $\varepsilon > 0$ and $\widetilde{T}, \widetilde{n} > 0$ so that, for $T > \widetilde{T}$, $n > \widetilde{n}$ and $h \in \mathcal{H}_{\widetilde{T},\widetilde{n}}(\varepsilon)$, $\alpha_T < \delta$. Equivalently, one could verify whether for any $\delta > 0$, there exists $\varepsilon > 0$ and $\widetilde{T}, \widetilde{n}$ such that, for $T > \widetilde{T}$, $n > \widetilde{n}$ and $h \in \mathcal{H}_{\widetilde{T},\widetilde{n}}(\varepsilon)$, we obtain

$$\sup_{h\in\mathcal{H}_{\widetilde{T},\widetilde{n}}(\varepsilon)} \left| \mathbf{E}\left[\xi\left(\frac{1}{v(1/T)\mathbf{h}} \int_0^T \mathbf{K}\left(\frac{X_s - x}{\mathbf{h}}\right) ds\right)\right] - \mathbf{E}\left[\xi\left(C_X \int \mathbf{K}(u)\widetilde{\phi}(u\mathbf{h} + x)du g_\alpha\right)\right] \right| < \frac{\delta}{2} \tag{28}$$

and

$$\sup_{h\in\mathcal{H}_{\widetilde{T},\widetilde{n}}(\varepsilon)} \left| \mathbf{E}\left[\xi\left(C_X \int \mathbf{K}(u)\phi(u\mathbf{h} + x)du g_\alpha\right)\right] - \mathbf{E}\left[\xi\left(C_X \widetilde{\phi}(x) g_\alpha\right)\right] \right| < \frac{\delta}{2}. \tag{29}$$

Expression (29) is immediate based on the continuity of the invariant measure. Expression (28) requires additional care. We will show that pointwise convergence over a dense set, along with an asymptotic equicontinuity condition, leads to the desired conclusion.

Since $\mathcal{H}_{\tilde{T},\widetilde{n}}(\varepsilon)$ is totally bounded and complete, it has a cover $\{S_{\tilde{T},\widetilde{n}}(\varepsilon, h_i, \gamma/2), i = 1, ..., q\}$. Let $\mathcal{H}^0_{\tilde{T},\widetilde{n}}(\varepsilon)$ be a dense subset of $\mathcal{H}_{\tilde{T},\widetilde{n}}(\varepsilon)$. Choose $\{\tilde{h}_1, ..., \tilde{h}_q\}$ such that $|h_i - \tilde{h}_i| < \gamma/2$, $i = 1, ..., q$, $\tilde{h}_i \in \mathcal{H}^0_{\tilde{T}}(\varepsilon)$. $\{S_{\tilde{T},\widetilde{n}}(\varepsilon, \tilde{h}_i, \gamma), i = 1, ..., q\}$ is also a cover for $\mathcal{H}_{\tilde{T},\widetilde{n}}(\varepsilon)$. For all $h \in \mathcal{H}_{\tilde{T},\widetilde{n}}(\varepsilon)$, there is some i such that $h \in S_{\tilde{T},\widetilde{n}}(\varepsilon, \tilde{h}_i, \gamma)$. Hence, there exists some i so that

$$
\begin{aligned}
&\left|\mathbf{E}\left[\xi\left(\frac{1}{v(1/T)\mathbf{h}}\int_0^T \mathbf{K}\left(\frac{X_s-x}{\mathbf{h}}\right)ds\right)\right]-\mathbf{E}\left[\xi\left(C_X\int \mathbf{K}(u)\widetilde{\phi}(u\mathbf{h}+x)dug_\alpha\right)\right]\right| \\
\leq &\sup_{h'\in S_{\tilde{T},\tilde{n}}(\varepsilon,\tilde{h}_i,\gamma)}\left|\mathbf{E}\left[\xi\left(\frac{1}{v(1/T)\mathbf{h}'}\int_0^T \mathbf{K}\left(\frac{X_s-x}{\mathbf{h}'}\right)ds\right)\right]-\mathbf{E}\left[\xi\left(C_X\int \mathbf{K}(u)\widetilde{\phi}(u\mathbf{h}'+x)dug_\alpha\right)\right]\right| \\
\leq &\sup_{h'\in S_{\tilde{T},\tilde{n}}(\varepsilon,\tilde{h}_i,\gamma)}\left|\left(\mathbf{E}\left[\xi\left(\frac{1}{v(1/T)\mathbf{h}'}\int_0^T \mathbf{K}\left(\frac{X_s-x}{\mathbf{h}'}\right)ds\right)\right]-\mathbf{E}\left[\xi\left(C_X\int \mathbf{K}(u)\widetilde{\phi}(u\mathbf{h}'+x)dug_\alpha\right)\right]\right)\right. \\
&\left.-\left(\mathbf{E}\left[\xi\left(\frac{1}{v(1/T)\widetilde{\mathbf{h}}_i}\int_0^T \mathbf{K}\left(\frac{X_s-x}{\widetilde{\mathbf{h}}_i}\right)ds\right)\right]+\mathbf{E}\left[\xi\left(C_X\int \mathbf{K}(u)\widetilde{\phi}(u\widetilde{\mathbf{h}}_i+x)dug_\alpha\right)\right]\right)\right| \\
&+\left|\mathbf{E}\left[\xi\left(\frac{1}{v(1/T)\widetilde{\mathbf{h}}_i}\int_0^T \mathbf{K}\left(\frac{X_s-x}{\widetilde{\mathbf{h}}_i}\right)ds\right)\right]-\mathbf{E}\left[\xi\left(C_X\int \mathbf{K}(u)\widetilde{\phi}(u\widetilde{\mathbf{h}}_i+x)dug_\alpha\right)\right]\right|.
\end{aligned}
$$

Thus,

$$
\begin{aligned}
&\sup_{h\in\mathcal{H}_{\tilde{T},\tilde{n}}(\varepsilon)}\left|\mathbf{E}\left[\xi\left(\frac{1}{v(1/T)\mathbf{h}}\int_0^T \mathbf{K}\left(\frac{X_s-x}{\mathbf{h}}\right)ds\right)\right]-\mathbf{E}\left[\xi\left(C_X\int \mathbf{K}(u)\widetilde{\phi}(u\mathbf{h}+x)dug_\alpha\right)\right]\right| \\
\leq &\max_{1\leq i\leq q}\sup_{h'\in S_{\tilde{T},\tilde{n}}(\varepsilon,\tilde{h}_i,\gamma)}\left|\left(\mathbf{E}\left[\xi\left(\frac{1}{v(1/T)\mathbf{h}'}\int_0^T \mathbf{K}\left(\frac{X_s-x}{\mathbf{h}'}\right)ds\right)\right]-\mathbf{E}\left[\xi\left(\frac{1}{v(1/T)\widetilde{\mathbf{h}}_i}\int_0^T \mathbf{K}\left(\frac{X_s-x}{\widetilde{\mathbf{h}}_i}\right)ds\right)\right]\right)\right. \\
&\left.\left(-\mathbf{E}\left[\xi\left(C_X\int \mathbf{K}(u)\phi(u\mathbf{h}'+x)dug_\alpha\right)\right]+\mathbf{E}\left[\xi\left(C_X\int \mathbf{K}(u)\widetilde{\phi}(u\widetilde{\mathbf{h}}_i+x)dug_\alpha\right)\right]\right)\right| \\
&+\max_{1\leq i\leq q}\left|\mathbf{E}\left[\xi\left(\frac{1}{v(1/T)\widetilde{\mathbf{h}}_i}\int_0^T \mathbf{K}\left(\frac{X_s-x}{\widetilde{\mathbf{h}}_i}\right)ds\right)\right]-\mathbf{E}\left[\xi\left(C_X\int \mathbf{K}(u)\widetilde{\phi}(u\widetilde{\mathbf{h}}_i+x)dug_\alpha\right)\right]\right|.
\end{aligned}
$$

The first maximum can be bounded by an arbitrarily small η (such that $\eta < \delta/3$) provided γ is chosen accordingly. This results from a stochastic equicontinuity property. Stochastic equicontinuity derives here from the modulus of continuity of ξ and the properties of the moments of the Mittag-Leffler distribution. See the proof of Theorems 4 and 5 for similar arguments worked out at length. The second maximum can be bounded by an arbitrarily small η ($< \delta/6$) for $T > \widetilde{T}$ and $n > \widetilde{n}$ by virtue of the Darling-Kac theorem and the pointwise weak convergence that this theorem implies. Finally, we show that for some $n > \widetilde{n}$ and $T > \widetilde{T}$ there exists an arbitrarily small ε such that

$$
\mathbf{E}\left[\xi\left(\frac{1}{v(1/T)}\frac{\Delta_{n,T}}{\mathbf{h}_{n,T}}\sum_{i=1}^n \mathbf{K}\left(\frac{X_{i\Delta_{n,T}}-x}{\mathbf{h}_{n,T}}\right)\right)\right]-\mathbf{E}\left[\xi\left(C_X\widetilde{\phi}\left(x\right)g_\alpha\right)\right]<\varepsilon.
$$

Write,

$$
\begin{aligned}
&\mathbf{E}\left[\xi\left(\frac{1}{v(1/T)}\frac{\Delta_{n,T}}{\mathbf{h}_{n,T}}\sum_{i=1}^n \mathbf{K}\left(\frac{X_{i\Delta_{n,T}}-x}{\mathbf{h}_{n,T}}\right)\right)\right]-\mathbf{E}\left[\xi\left(C_X\widetilde{\phi}\left(x\right)g_\alpha\right)\right] \\
\leq &\sup_{T,n}\left|\mathbf{E}\left[\xi\left(\frac{1}{v(1/T)}\frac{\Delta_{n,T}}{\mathbf{h}_{n,T}}\sum_{i=1}^n \mathbf{K}\left(\frac{X_{i\Delta_{n,T}}-x}{\mathbf{h}_{n,T}}\right)\right)\right]-\mathbf{E}\left[\xi\left(\frac{1}{v(1/T)}\frac{1}{\mathbf{h}_{n,T}}\int_0^T \mathbf{K}\left(\frac{X_s-x}{\mathbf{h}_{n,T}}\right)ds\right)\right]\right| \quad (30) \\
&+\sup_{h\in\mathcal{H}_{\widetilde{T},\widetilde{n}}(\varepsilon)}\left|\mathbf{E}\left[\xi\left(\frac{1}{v(1/T)}\frac{1}{\mathbf{h}}\int_0^T \mathbf{K}\left(\frac{X_s-x}{\mathbf{h}}\right)ds\right)\right]-\mathbf{E}\left[\xi\left(C_X\widetilde{\phi}\left(x\right)g_\alpha\right)\right]\right|. \quad (31)
\end{aligned}
$$

But (30) and (31) can be bounded by an arbitrarily small number, say $\frac{\varepsilon}{2}$ ($= \delta$), provided $\widetilde{n}$ and $\widetilde{T}$ are chosen accurately, for $n > \widetilde{n}$ and $T > \widetilde{T}$, using result (i) in the theorem and (27), respectively. This concludes the proof.

Proof of Theorem 4 Write

$$\begin{aligned}
&\widehat{\boldsymbol{\mu}}_{n,T}(x) \\
= \; & \frac{1}{\Delta_{n,T}} \frac{\sum_{i=1}^{n-1} \mathbf{K}\left(\frac{X_{i\Delta_{n,T}}-x}{\mathbf{h}_{n,T}}\right) \int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}} \boldsymbol{\mu}\left(X_s\right) ds}{\sum_{i=1}^{n} \mathbf{K}\left(\frac{X_{i\Delta_{n,T}}-x}{\mathbf{h}_{n,T}}\right)} \qquad (32) \\
& + \frac{1}{\Delta_{n,T}} \frac{\sum_{i=1}^{n-1} \mathbf{K}\left(\frac{X_{i\Delta_{n,T}}-x}{\mathbf{h}_{n,T}}\right) \int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}} \boldsymbol{\sigma}\left(X_s\right) d\mathbf{B}_s}{\sum_{i=1}^{n} \mathbf{K}\left(\frac{X_{i\Delta_{n,T}}-x}{\mathbf{h}_{n,T}}\right)}. \qquad (33)
\end{aligned}$$

We start with (32). Using the Quotient limit theorem and the method of proof of Theorem 3, we obtain

$$(32) = \frac{\frac{1}{\mathbf{h}_{n,T}} \int_0^T \mathbf{K}\left(\frac{X_s-x}{\mathbf{h}_{n,T}}\right) \boldsymbol{\mu}\left(X_s\right) ds + O_{a.s.}\left(\sqrt{\Delta_{n,T}\log(1/\Delta_{n,T})}\widehat{L}_{(n,T)}(T,x)/\mathbf{h}_{n,T}\right)}{\frac{1}{\mathbf{h}_{n,T}} \int_0^T \mathbf{K}\left(\frac{X_s-x}{\mathbf{h}_{n,T}}\right) ds + O_{a.s.}\left(\sqrt{\Delta_{n,T}\log(1/\Delta_{n,T})}\widehat{L}_{(n,T)}(T,x)/\mathbf{h}_{n,T}\right)} \qquad (34)$$

for a fixed $\mathbf{h}_{n,T}$. We now wish to prove that

$$\frac{\frac{1}{\mathbf{h}_{n,T}} \int_0^T \mathbf{K}\left(\frac{X_s-x}{\mathbf{h}_{n,T}}\right) \boldsymbol{\mu}\left(X_s\right) ds}{\frac{1}{\mathbf{h}_{n,T}} \int_0^T \mathbf{K}\left(\frac{X_s-x}{\mathbf{h}_{n,T}}\right) ds} \underset{\mathbf{h}_{n,T}\to 0, T, n\to\infty}{\overset{a.s.}{\longrightarrow}} \boldsymbol{\mu}(x),$$

jointly over $\mathbf{h}_{n,T}$ and T. Define

$$\begin{aligned}
\Pi_{T,\mathbf{h}}(x) &= \boldsymbol{\mu}_{T,\mathbf{h}}(x) - \boldsymbol{\mu}_{\mathbf{h}}(x) \\
&= \frac{\int_0^T \mathbf{K}\left(\frac{X_s-x}{\mathbf{h}}\right) \boldsymbol{\mu}\left(X_s\right) ds}{\int_0^T \mathbf{K}\left(\frac{X_s-x}{\mathbf{h}}\right) ds} - \frac{\int \mathbf{K}(u)\boldsymbol{\mu}\left(\mathbf{h}u+x\right)\widetilde{\phi}(\mathbf{h}u+x)du}{\int \mathbf{K}(u)\widetilde{\phi}(\mathbf{h}u+x)du} \\
&= \frac{\int_0^T \mathbf{K}\left(\frac{X_s-x}{\mathbf{h}}\right) \mathbf{z}_{s,\mathbf{h}} ds}{\int_0^T \mathbf{K}\left(\frac{X_s-x}{\mathbf{h}}\right) ds},
\end{aligned}$$

where

$$\mathbf{z}_{s,\mathbf{h}}(x) = \boldsymbol{\mu}\left(X_s\right) - \frac{\int \mathbf{K}(u)\boldsymbol{\mu}\left(\mathbf{h}u+x\right)\widetilde{\phi}(\mathbf{h}u+x)du}{\int \mathbf{K}(u)\widetilde{\phi}(\mathbf{h}u+x)du}.$$

The value of x is constant throughout and will be omitted in what follows. In other words, we will write $\Pi_{T,\mathbf{h}_{n,T}}$, $\boldsymbol{\mu}_{\mathbf{h}}$ and so on, for brevity. We wish to show that

$$\boldsymbol{\mu}_{T,\mathbf{h}_{n,T}} - \boldsymbol{\mu} \underset{\mathbf{h}_{n,T}\to 0, T, n\to\infty}{\overset{a.s.}{\longrightarrow}} 0. \qquad (35)$$

We need to guarantee that the bandwidth sequence is chosen from a totally bounded and non-empty set. As in the proof of Theorem 3, we consider the set

$$\mathcal{H}_{T,n}(\varepsilon) = \left\{h : \max\left(\widehat{L}_{(n,T)}(T,x)\sqrt{\Delta_{n,T}\log\left(1/\Delta_{n,T}\right)}, 1/\widehat{L}_{(n,T)}(T,x)\right)/\varepsilon < h^d < \varepsilon\right\}.$$

The expression (35) holds if, for any $\delta > 0$, there exists $\varepsilon, \tilde{T}$ and $\widetilde{n}$ so that $\mathcal{H}_{\tilde{T},\widetilde{n}}(\varepsilon)$ is non-empty and, for $T > \tilde{T}$, $n > \widetilde{n}$ and $h \in \mathcal{H}_{\tilde{T},\widetilde{n}}(\varepsilon)$, we obtain

$$\sup_{h\in\mathcal{H}_{\tilde{T},\widetilde{n}}(\varepsilon)} \left|\boldsymbol{\mu}_{\mathbf{h}} - \boldsymbol{\mu}\right| < \frac{\delta}{2} \qquad (36)$$

and

$$\sup_{h\in\mathcal{H}_{\tilde{T},\tilde{n}}(\varepsilon)} |\Pi_{T,\mathbf{h}}| < \frac{\delta}{2}. \tag{37}$$

The conditions (36) and (37) imply that

$$\left|\boldsymbol{\mu}_{T,\mathbf{h}_{n,T}} - \boldsymbol{\mu}\right| \leq \sup_{h\in\mathcal{H}_{\tilde{T},\tilde{n}}(\varepsilon)} \left(\left|\boldsymbol{\mu}_{T,\mathbf{h}} - \boldsymbol{\mu}_{\mathbf{h}}\right| + \left|\boldsymbol{\mu}_{\mathbf{h}} - \boldsymbol{\mu}\right|\right) < \delta.$$

Expression (36) is immediate given the continuity of the drift function and invariant measure. We now show (37). We proceed as in the proof of Theorem 3. Since $\mathcal{H}_{\tilde{T},\tilde{n}}(\varepsilon)$ is totally bounded and dense, it has a cover $\{S_{\tilde{T},\tilde{n}}(\varepsilon, h_i, \gamma/2), i = 1, ..., q\}$. Let $\mathcal{H}^0_{\tilde{T},\tilde{n}}(\varepsilon)$ be a dense subset of $\mathcal{H}_{\tilde{T},\tilde{n}}(\varepsilon)$. Choose $\{\tilde{h}_1, ..., \tilde{h}_q\}$ such that $|h_i - \tilde{h}_i| < \gamma/2$, $i = 1, ..., q$ and $\tilde{h}_i \in \mathcal{H}^0_{\tilde{T},\tilde{n}}(\varepsilon)$. $\{S_{\tilde{T},\tilde{n}}(\varepsilon, \tilde{h}_i, \gamma), i = 1, ..., q\}$ is also a cover for $\mathcal{H}_{\tilde{T},\tilde{n}}(\varepsilon)$. For all $h \in \mathcal{H}_{\tilde{T},\tilde{n}}(\varepsilon)$, there is some i such that $h \in S_{\tilde{T},\tilde{n}}(\varepsilon, \tilde{h}_i, \gamma)$. Thus, there exists some i so that

$$\begin{aligned}
|\Pi_{T,\mathbf{h}}| &\leq \sup_{h'\in S_{\tilde{T},\tilde{n}}(\varepsilon,\tilde{h}_i,\gamma)} |\Pi_{T,\mathbf{h}'}| \\
&\leq \sup_{h'\in S_{\tilde{T},\tilde{n}}(\varepsilon,\tilde{h}_i,\gamma)} |\Pi_{T,\mathbf{h}'} - \Pi_{T,\tilde{h}_i}| + |\Pi_{T,\tilde{h}_i}| \text{ a.s.}
\end{aligned}$$

Hence,

$$\sup_{h\in\mathcal{H}_{\tilde{T},\tilde{n}}(\varepsilon)} |\Pi_{T,\mathbf{h}}| \leq \max_{1\leq i\leq q} \sup_{h'\in S_{\tilde{T},\tilde{n}}(\varepsilon,\tilde{h}_i,\gamma)} |\Pi_{T,\mathbf{h}'} - \Pi_{T,\tilde{h}_i}| + \max_{1\leq i\leq q} |\Pi_{T,\tilde{h}_i}| \text{ a.s.} \tag{38}$$

for all $T > \tilde{T}$ and $n > \tilde{n}$. The first term on the right hand side converges to zero if an asymptotic equicontinuity condition is satisfied while the second term vanishes if pointwise convergence holds. The latter follows from the Quotient limit theorem for any fixed bandwidth. In fact, for a fixed x and a fixed $\mathbf{h}$ (and given the integrability properties that are sufficient for the Quotient limit theorem to hold), we obtain

$$\begin{aligned}
&\frac{\frac{1}{\mathbf{h}}\int_0^T \mathbf{K}\left(\frac{X_s - x}{\mathbf{h}}\right)\boldsymbol{\mu}(X_s)\,ds}{\frac{1}{\mathbf{h}}\int_0^T \mathbf{K}\left(\frac{X_s - x}{\mathbf{h}}\right)ds} \overset{a.s.}{\to} \frac{\frac{1}{\mathbf{h}}\int \mathbf{K}\left(\frac{u-x}{\mathbf{h}}\right)\boldsymbol{\mu}(u)\,\widetilde{\phi}(u)du}{\frac{1}{\mathbf{h}}\int \mathbf{K}\left(\frac{u-x}{\mathbf{h}}\right)\widetilde{\phi}(u)du} \\
= \;& \frac{\int \mathbf{K}(s)\,\boldsymbol{\mu}(x + \mathbf{h}s)\,\widetilde{\phi}(x + \mathbf{h}s)ds}{\int \mathbf{K}(s)\,\widetilde{\phi}(x + \mathbf{h}s)ds}.
\end{aligned}$$

As for the former, first we have to verify that for all T, n and some $\varepsilon > 0$ so that $\mathcal{H}_{\tilde{T},\tilde{n}}(\varepsilon)$ is non-empty and compact, the quantity

$$\sup_{h\in\mathcal{H}_{\tilde{T},\tilde{n}}(\varepsilon)} |\Pi_{T,\mathbf{h}}|$$

is indeed measurable with respect to $\mathcal{F}_T$. For a fixed h, the numerator and the denominator of $\boldsymbol{\mu}_{T,\mathbf{h}}$ are additive functionals and it is a standard result that such functionals are $\mathcal{F}_T$-adapted. To extend this property to the supremum over $h \in \mathcal{H}_{\tilde{T},\tilde{n}}(\varepsilon)$, we need to verify that $\mathcal{H}_{T,n}(\varepsilon)$ is compact and non-empty for all T and n, almost surely. This is true by construction. Then, since $\Pi_{T,\mathbf{h}}$ is continuous in $\mathbf{h}$, the main result of Stinchcombe and White (1992) assures that $\Pi_{T,\mathbf{h}}$ is at least "nearly" $\mathcal{F}_T$-measurable, and we can proceed as if the supremum of $|\Pi_{T,\mathbf{h}}|$ is a well-defined stochastic process and compute the limit supremum. We now turn to asymptotic equicontinuity. Fix a positive and finite $\mathbf{h}$ and take any $\mathbf{h}'$ such that $|\mathbf{h} - \mathbf{h}'| < v$. Write

$$
\begin{aligned}
&\left|\Pi_{T,\mathbf{h}'}-\Pi_{T,\mathbf{h}}\right| \\
= &\left|\frac{\int_0^T \mathbf{K}\left(\frac{X_s-x}{\mathbf{h}'}\right)\mathbf{z}_{s,\mathbf{h}'}ds}{\int_0^T \mathbf{K}\left(\frac{X_s-x}{\mathbf{h}'}\right)ds}-\frac{\int_0^T \mathbf{K}\left(\frac{X_s-x}{\mathbf{h}}\right)\mathbf{z}_{s,\mathbf{h}}ds}{\int_0^T \mathbf{K}\left(\frac{X_s-x}{\mathbf{h}}\right)ds}\right| \\
= &\left|\frac{\int_0^T\left(\mathbf{K}\left(\frac{X_s-x}{\mathbf{h}}\right)-\mathbf{K}\left(\frac{X_s-x}{\mathbf{h}'}\right)\frac{\int_0^T \mathbf{K}\left(\frac{X_s-x}{\mathbf{h}}\right)ds}{\int_0^T \mathbf{K}\left(\frac{X_s-x}{\mathbf{h}'}\right)ds}\right)\boldsymbol{\mu}(X_s)ds}{\int_0^T \mathbf{K}\left(\frac{X_s-x}{\mathbf{h}}\right)ds}-\boldsymbol{\mu}_{\mathbf{h}}+\boldsymbol{\mu}_{\mathbf{h}'}\right| \\
= &|\frac{\int_0^T\left(\mathbf{K}\left(\frac{X_s-x}{\mathbf{h}}\right)-\mathbf{K}\left(\frac{X_s-x}{\mathbf{h}'}\right)\right)\boldsymbol{\mu}(X_s)ds}{\int_0^T \mathbf{K}\left(\frac{X_s-x}{\mathbf{h}}\right)ds} \\
&+\left(1-\frac{\int_0^T \mathbf{K}\left(\frac{X_s-x}{\mathbf{h}}\right)ds}{\int_0^T \mathbf{K}\left(\frac{X_s-x}{\mathbf{h}'}\right)ds}\right)\frac{\int_0^T \mathbf{K}\left(\frac{X_s-x}{\mathbf{h}'}\right)\boldsymbol{\mu}(X_s)ds}{\int_0^T \mathbf{K}\left(\frac{X_s-x}{\mathbf{h}}\right)ds}-\boldsymbol{\mu}_{\mathbf{h}}+\boldsymbol{\mu}_{\mathbf{h}'}| \\
\leq &\left|\frac{\int_0^T\left(\mathbf{K}\left(\frac{X_s-x}{\mathbf{h}}\right)-\mathbf{K}\left(\frac{X_s-x}{\mathbf{h}'}\right)\right)\boldsymbol{\mu}(X_s)ds}{\int_0^T \mathbf{K}\left(\frac{X_s-x}{\mathbf{h}}\right)ds}\right| \\
&+\left|1-\frac{\int_0^T \mathbf{K}\left(\frac{X_s-x}{\mathbf{h}}\right)ds}{\int_0^T \mathbf{K}\left(\frac{X_s-x}{\mathbf{h}'}\right)ds}\right|\left|\frac{\int_0^T \mathbf{K}\left(\frac{X_s-x}{\mathbf{h}'}\right)\boldsymbol{\mu}(X_s)ds}{\int_0^T \mathbf{K}\left(\frac{X_s-x}{\mathbf{h}}\right)ds}\right|+|\boldsymbol{\mu}_{\mathbf{h}}-\boldsymbol{\mu}_{\mathbf{h}'}|.
\end{aligned}
$$

Applying the Quotient limit theorem repeatedly we find that

$$
\begin{aligned}
&\lim_{T\to\infty}\left|\frac{\int_0^T\left(\mathbf{K}\left(\frac{X_s-x}{\mathbf{h}}\right)-\mathbf{K}\left(\frac{X_s-x}{\mathbf{h}'}\right)\right)\boldsymbol{\mu}(X_s)ds}{\int_0^T \mathbf{K}\left(\frac{X_s-x}{\mathbf{h}}\right)ds}\right| \\
= &\left|\frac{\int \mathbf{K}(u)\left(\boldsymbol{\mu}(\mathbf{h}u+x)\widetilde{\phi}(\mathbf{h}u+x)-\frac{\mathbf{h}'}{\mathbf{h}}\boldsymbol{\mu}(\mathbf{h}'u+x)\widetilde{\phi}(\mathbf{h}'u+x)\right)du}{\int \mathbf{K}(u)\widetilde{\phi}(\mathbf{h}u+x)du}\right| \\
= &O_{a.s}(\upsilon),
\end{aligned}
$$

$$
\begin{aligned}
&\lim_{T\to\infty}\left|1-\frac{\int_0^T \mathbf{K}\left(\frac{X_s-x}{\mathbf{h}}\right)ds}{\int_0^T \mathbf{K}\left(\frac{X_s-x}{\mathbf{h}'}\right)ds}\right| \\
= &\left|\frac{\int \mathbf{K}(u)\left(\widetilde{\phi}(\mathbf{h}'u+x)-\frac{\mathbf{h}}{\mathbf{h}'}\widetilde{\phi}(\mathbf{h}u+x)\right)du}{\int \mathbf{K}(u)\widetilde{\phi}(\mathbf{h}'u+x)du}\right| \\
= &O_{a.s.}(\upsilon),
\end{aligned}
$$

$$
\begin{aligned}
&\lim_{T\to\infty}\left|\frac{\int_0^T \mathbf{K}\left(\frac{X_s-x}{\mathbf{h}'}\right)\boldsymbol{\mu}(X_s)ds}{\int_0^T \mathbf{K}\left(\frac{X_s-x}{\mathbf{h}}\right)ds}\right| \\
= &\left|\frac{\frac{\mathbf{h}'}{\mathbf{h}}\int \mathbf{K}(u)\mu(\mathbf{h}'u+x)\widetilde{\phi}(\mathbf{h}'u+x)du}{\int \mathbf{K}(u)\widetilde{\phi}(\mathbf{h}u+x)du}\right|<\infty \qquad a.s.,
\end{aligned}
$$

and

$$|\boldsymbol{\mu}_{\mathbf{h}} - \boldsymbol{\mu}_{\mathbf{h}'}| = O_{a.s.}(v).$$

Then, for all $\frac{\delta}{4} > \eta > 0$, there exists $\gamma > 0$, $\widetilde{T}$, $\widetilde{n}$ and $\varepsilon > 0$ such that for $T > \widetilde{T}$ and $\widetilde{n} > n$ we obtain

$$\sup_{h \in \mathcal{H}_{\widetilde{T},\widetilde{n}}(\varepsilon)} \sup_{h' \in \{b: |\mathbf{h} - b| < \gamma\}} \left| \Pi_{T,\mathbf{h}'} - \Pi_{T,\mathbf{h}} \right| < \eta \qquad a.s. \tag{39}$$

In consequence, we only need to show that for some $n \geq \widetilde{n}$ and $T > \widetilde{T}$ there exists an arbitrarily small ε such that

$$\begin{aligned} & |(32) - \boldsymbol{\mu}| \\ \leq\ & \sup_{T,n} \left| (32) - \frac{\frac{1}{\mathbf{h}_{n,T}} \int_0^T \mathbf{K}\left(\frac{X_s - x}{\mathbf{h}_{n,T}}\right) \boldsymbol{\mu}(X_s)\, ds}{\frac{1}{\mathbf{h}_{n,T}} \int_0^T \mathbf{K}\left(\frac{X_s - x}{\mathbf{h}_{n,T}}\right) ds} \right| & (40) \\ & + \sup_{h \in \mathcal{H}_{\widetilde{T},\widetilde{n}}(\varepsilon)} \left| \boldsymbol{\mu}_{T,\mathbf{h}_{n,T}} - \boldsymbol{\mu} \right| & (41) \\ <\ & \varepsilon. & (42) \end{aligned}$$

But (40) and (41) can be bounded by $\frac{\varepsilon}{2} = \delta$ using (34) and (38) above, thereby giving (42). We now turn to (33). Each component of the vector (33) converges to zero almost surely (as $n, T \to \infty$) by the law of large numbers for martingale difference arrays (c.f. the proof of Theorem 5) along with the requirement that $\widehat{L}_{n,T}(T,x)\mathbf{h}_{n,T} \overset{a.s.}{\to} \infty$. This proves the stated result.

Proof of Theorem 5 Write the estimation error decomposition as

$$\underbrace{(32) - \boldsymbol{\mu}}_{bias} + \underbrace{(33)}_{variance}.$$

First, we concentrate on (33). Write

$$\begin{aligned} & \sup_n \left| \frac{\frac{1}{\sqrt{\mathbf{h}_{n,T}}} \int_{T/n}^T \mathbf{K}\left(\frac{X_{i\Delta_{n,T}} - x}{\mathbf{h}_{n,T}}\right) \boldsymbol{\sigma}(X_s) d\mathbf{B}_s}{\sqrt{\frac{\Delta_{n,T}}{\mathbf{h}_{n,T}} \sum_{i=1}^n \mathbf{K}\left(\frac{X_{i\Delta_{n,T}} - x}{\mathbf{h}_{n,T}}\right)}} - \frac{\frac{1}{\sqrt{\mathbf{h}_{n,T}}} \int_0^T \mathbf{K}\left(\frac{X_s - x}{\mathbf{h}_{n,T}}\right) \boldsymbol{\sigma}(X_s) d\mathbf{B}_s}{\sqrt{\frac{1}{\mathbf{h}_{n,T}} \int_0^T \mathbf{K}\left(\frac{X_s - x}{\mathbf{h}_{n,T}}\right) ds}} \right| \\ \leq\ & \sup_n \left| \frac{\frac{1}{\sqrt{\mathbf{h}_{n,T}}} \int_0^T \mathbf{K}\left(\frac{X_{i\Delta_{n,T}} - x}{\mathbf{h}_{n,T}}\right) \boldsymbol{\sigma}(X_s) d\mathbf{B}_s}{\sqrt{\frac{\Delta_{n,T}}{\mathbf{h}_{n,T}} \sum_{i=1}^n \mathbf{K}\left(\frac{X_{i\Delta_{n,T}} - x}{\mathbf{h}_{n,T}}\right)}} - \frac{\frac{1}{\sqrt{\mathbf{h}_{n,T}}} \int_0^T \mathbf{K}\left(\frac{X_s - x}{\mathbf{h}_{n,T}}\right) \boldsymbol{\sigma}(X_s) d\mathbf{B}_s}{\sqrt{\frac{1}{\mathbf{h}_{n,T}} \int_0^T \mathbf{K}\left(\frac{X_s - x}{\mathbf{h}_{n,T}}\right) ds}} \right| \\ & + \sup_n \left| \frac{\frac{1}{\sqrt{\mathbf{h}_{n,T}}} \int_0^{T/n} \mathbf{K}\left(\frac{X_{i\Delta_{n,T}} - x}{\mathbf{h}_{n,T}}\right) \boldsymbol{\sigma}(X_s) d\mathbf{B}_s}{\sqrt{\frac{\Delta_{n,T}}{\mathbf{h}_{n,T}} \sum_{i=1}^n \mathbf{K}\left(\frac{X_{i\Delta_{n,T}} - x}{\mathbf{h}_{n,T}}\right)}} \right| = O_{a.s.}\left(\sqrt{\Delta_{n,T} \log(1/\Delta_{n,T})} \widehat{L}_{(n,T)}(T,x) \right), \end{aligned} \tag{43}$$

which follows from the continuity of the kernel function (from Assumption 2) and Theorem 3 (i) for a fixed $\mathbf{h}_{n,T}$. We now derive the limiting distribution of

$$\frac{\frac{1}{\sqrt{\mathbf{h}_{n,T}}} \int_0^T \mathbf{K}\left(\frac{X_s - x}{\mathbf{h}_{n,T}}\right) \boldsymbol{\sigma}(X_s) d\mathbf{B}_s}{\sqrt{\frac{1}{\mathbf{h}_{n,T}} \int_0^T \mathbf{K}\left(\frac{X_s - x}{\mathbf{h}_{n,T}}\right) ds}}.$$

We wish to show that

$$\lim_{n,T\to\infty}\left|\mathbf{E}\left[\lambda\left(\frac{\frac{1}{\sqrt{\mathbf{h}_{n,T}}}\int_0^T \mathbf{K}\left(\frac{X_s-x}{\mathbf{h}_{n,T}}\right)\boldsymbol{\sigma}(X_s)d\mathbf{B}_s}{\sqrt{\frac{1}{\mathbf{h}_{n,T}}\int_0^T \mathbf{K}\left(\frac{X_s-x}{\mathbf{h}_{n,T}}\right)ds}}\right)\right]-\mathbf{E}\left[\lambda\left(\mathbf{N}\left(0,\mathbf{a}(x)\left(\int k^2\left(u\right)du\right)^d\right)\right)\right]\right|$$
$$= \quad 0$$

for any bounded and continuous function λ. As earlier, we need to prove that for any $\delta>0$, there exists $\varepsilon>0$, $\widetilde{n}>0$ and $\widetilde{T}>0$ such that, for $T>\widetilde{T}$, $n>\widetilde{n}$ and $h\in\mathcal{H}_{\widetilde{T},\widetilde{n}}(\varepsilon)$, we obtain

$$\begin{aligned}\sup_{h\in\mathcal{H}_{\widetilde{T},\widetilde{n}}(\varepsilon)}&|\mathbf{E}\left[\lambda\left(\frac{\frac{1}{\sqrt{\mathbf{h}}}\int_0^T \mathbf{K}\left(\frac{X_s-x}{\mathbf{h}}\right)\boldsymbol{\sigma}(X_s)d\mathbf{B}_s}{\sqrt{\frac{1}{\mathbf{h}}\int_0^T \mathbf{K}\left(\frac{X_s-x}{\mathbf{h}}\right)ds}}\right)\right]\\&-\mathbf{E}\left[\lambda\left(\mathbf{N}\left(0,\frac{\int_{\Re^d}\mathbf{K}^2\left(u\right)\mathbf{a}(u\mathbf{h}+x)\widetilde{\phi}(u\mathbf{h}+x)du}{\int\mathbf{K}\left(u\right)\widetilde{\phi}(u\mathbf{h}+x)du}\right)\right)\right]|\\\leq\quad&\frac{\delta}{2},\end{aligned}\tag{44}$$

$$\begin{aligned}\sup_{h\in\mathcal{H}_{\widetilde{T},\widetilde{n}}(\varepsilon)}&\left|\mathbf{E}\left[\lambda\left(\mathbf{N}\left(0,\frac{\int_{\Re^d}\mathbf{K}^2\left(u\right)\mathbf{a}(u\mathbf{h}+x)\widetilde{\phi}(u\mathbf{h}+x)du}{\int\mathbf{K}\left(u\right)\widetilde{\phi}(u\mathbf{h}+x)du}\right)\right)\right]-\mathbf{E}\left[\lambda\left(\mathbf{N}(0,\mathbf{a}(x)\left(\int_{\Re}k^2\left(u\right)du\right)^d\right)\right]\right|\\\leq\quad&\frac{\delta}{2},\end{aligned}\tag{45}$$

where

$$\begin{aligned}\mathcal{H}_{T,n}(\varepsilon)\quad&=\quad\{h:\max\left(\widehat{L}_{(n,T)}(T,x)\sqrt{\Delta_{n,T}\log\left(1/\Delta_{n,T}\right)},1/\widehat{L}_{(n,T)}(T,x)\right)/\varepsilon<h^d<\varepsilon,\\h^{d+4}\quad&<\quad\varepsilon/\widehat{L}_{(n,T)}(T,x)\},\end{aligned}$$

or

$$\begin{aligned}\mathcal{H}_{T,n}(\varepsilon)\quad&=\quad\{h:\max\left(\widehat{L}_{(n,T)}(T,x)\sqrt{\Delta_{n,T}\log\left(1/\Delta_{n,T}\right)},1/\widehat{L}_{(n,T)}(T,x)\right)/\varepsilon<h^d<\varepsilon,\\h^{d+4}\quad&<\quad J/\widehat{L}_{(n,T)}(T,x)\qquad \textit{for some } J\}.\end{aligned}$$

Expression (45) is immediate given the continuity of the diffusion matrix and invariant measure. As for (44), as earlier we need to show that for all $\frac{\delta}{6}>\eta>0$, there exists $\gamma>0$, $\widetilde{T}>0$, $\widetilde{n}>0$ and $\varepsilon>0$ so that, for $T>\widetilde{T}$ and $n>\widetilde{n}$, we obtain

$$\begin{aligned}\sup_{h\in\mathcal{H}_{\widetilde{T}}(\varepsilon)}\sup_{h'\in\{b:|\mathbf{h}-b|<\gamma\}}&\left|\mathbf{E}\left[\lambda\left(\frac{\frac{1}{\sqrt{\mathbf{h}'}}\int_0^T \mathbf{K}\left(\frac{X_s-x}{\mathbf{h}'}\right)\boldsymbol{\sigma}(X_s)d\mathbf{B}_s}{\sqrt{\frac{1}{\mathbf{h}'}\int_0^T \mathbf{K}\left(\frac{X_s-x}{\mathbf{h}'}\right)ds}}\right)\right]\right.\\&\left.-\mathbf{E}\left[\lambda\left(\frac{\frac{1}{\sqrt{\mathbf{h}}}\int_0^T \mathbf{K}\left(\frac{X_s-x}{\mathbf{h}}\right)\boldsymbol{\sigma}(X_s)d\mathbf{B}_s}{\sqrt{\frac{1}{\mathbf{h}}\int_0^T \mathbf{K}\left(\frac{X_s-x}{\mathbf{h}}\right)ds}}\right)\right]\right|\\<\quad&\eta,\end{aligned}\tag{46}$$

$$\begin{aligned}&\sup_{h\in\mathcal{H}_{\widetilde{T}}(\varepsilon)}\sup_{h'\in\{b:|\mathbf{h}-b|<\gamma\}}\left|\mathbf{E}\left[\lambda\left(\mathbf{N}\left(0,\frac{\int\mathbf{K}^2(u)\,\mathbf{a}(u\mathbf{h}+x)\widetilde{\phi}(u\mathbf{h}+x)du}{\int\mathbf{K}(u)\,\widetilde{\phi}(u\mathbf{h}+x)du}\right)\right)\right]\right.\\&\qquad\left.-\mathbf{E}\left[\lambda\left(\mathbf{N}\left(0,\frac{\int\mathbf{K}^2(u)\,\mathbf{a}(u\mathbf{h}'+x)\widetilde{\phi}(u\mathbf{h}'+x)du}{\int\mathbf{K}(u)\,\widetilde{\phi}(u\mathbf{h}'+x)du}\right)\right)\right]\right|\\&<\quad\eta,\end{aligned}\tag{47}$$

and

$$\begin{aligned}&\max_{1\le j\le q}\mathbf{E}\left[\lambda\left(\frac{\frac{1}{\sqrt{\mathbf{h}_j}}\int_0^T\mathbf{K}\left(\frac{X_s-x}{\mathbf{h}_j}\right)\boldsymbol{\sigma}(X_s)d\mathbf{B}_s}{\sqrt{\frac{1}{\mathbf{h}_j}\int_0^T\mathbf{K}\left(\frac{X_s-x}{\mathbf{h}_j}\right)ds}}\right)\right]\\&\qquad-\mathbf{E}\left[\lambda\left(\mathbf{N}\left(\frac{\int\mathbf{K}^2(u)\,\mathbf{a}(u\mathbf{h}_j+x)\widetilde{\phi}(u\mathbf{h}_j+x)du}{\int\mathbf{K}(u)\,\widetilde{\phi}(u\mathbf{h}_j+x)du}\right)\right)\right]\\&<\quad\eta.\end{aligned}\tag{48}$$

We start with (48). We simply need to prove that

$$\frac{\frac{1}{\sqrt{\mathbf{h}}}\int_0^T\mathbf{K}\left(\frac{X_s-x}{\mathbf{h}}\right)\boldsymbol{\sigma}(X_s)d\mathbf{B}_s}{\sqrt{\frac{1}{\mathbf{h}}\int_0^T\mathbf{K}\left(\frac{X_s-x}{\mathbf{h}}\right)ds}}\Rightarrow\mathbf{N}\left(0,\frac{\int\mathbf{K}^2(u)\,\mathbf{a}(u\mathbf{h}+x)\widetilde{\phi}(u\mathbf{h}+x)du}{\int\mathbf{K}(u)\,\widetilde{\phi}(u\mathbf{h}+x)du}\right)\tag{49}$$

pointwise, i.e. for every $\mathbf{h}\in\mathcal{H}_{\widetilde{T},\widetilde{n}}(\varepsilon)$, as we proceed to illustrate. Consider the generic element g of the $d-$vector

$$\frac{1}{\sqrt{\mathbf{h}}}\int_0^T\mathbf{K}\left(\frac{X_s-x}{\mathbf{h}}\right)\boldsymbol{\sigma}(X_s)d\mathbf{B}_s.$$

i.e.

$$\frac{1}{\sqrt{\mathbf{h}}}\int_0^T\mathbf{K}\left(\frac{X_s-x}{\mathbf{h}}\right)\sum_{j=1}^m\sigma_{gj}(X_s)\,dB_s^j\qquad 1\le g\le d.$$

Define the sequence of martingales

$$M_r^{(T)g}=\frac{1}{\sqrt{\mathbf{h}}}\int_0^{[rT]}\mathbf{K}\left(\frac{X_s-x}{\mathbf{h}}\right)\sum_{j=1}^m\sigma_{gj}(X_s)\,dB_s^j.$$

Its quadratic variation is

$$[M^{(T)g}]_r=\frac{1}{\mathbf{h}}\int_0^{[rT]}\mathbf{K}^2\left(\frac{X_s-x}{\mathbf{h}}\right)\sum_{j=1}^m\sigma_{gj}^2(X_s)\,ds.$$

Clearly, $[M^{(T)g}]_\infty=\infty\ \forall T$. Call $\tau^{(T)}$ the time-change associated with $[M^{(T)g}]$ and $\beta^{(T)g}$ the *Dambis, Dubins-Schwarz* Brownian motion of $M^{(T)g}$ (see, e.g., Revuz and Yor (1998), Theorem 1.6, page 173), then

$$M_r^{(T)g}=\beta_{[M^{(T)g}]_r}\overset{d}{=}\mathbf{N}(0,[M^{(T)g}]_r).$$

In consequence,

$$\frac{M_1^{(T)g}}{\sqrt{\frac{1}{\mathbf{h}}\int_0^T\mathbf{K}\left(\frac{X_s-x}{\mathbf{h}}\right)ds}}\overset{d}{=}\mathbf{N}\left(0,\frac{[M^{(T)g}]_1}{\frac{1}{\mathbf{h}}\int_0^T\mathbf{K}\left(\frac{X_s-x}{\mathbf{h}}\right)ds}\right).$$

But, by the Quotient limit theorem, we obtain

$$\begin{aligned}
& \frac{[M^{(T)g}]_1}{\frac{1}{\mathbf{h}}\int_0^T \mathbf{K}\left(\frac{X_s-x}{\mathbf{h}}\right)ds} \\
= \quad & \frac{\frac{1}{\mathbf{h}}\int_0^T \mathbf{K}^2\left(\frac{X_s-x}{\mathbf{h}}\right)\sum_{j=1}^m \sigma_{gj}^2\left(X_s\right)ds}{\frac{1}{\mathbf{h}}\int_0^T \mathbf{K}\left(\frac{X_s-x}{\mathbf{h}}\right)ds} \\
& \overset{a.s.}{\to} \frac{\int \mathbf{K}^2\left(u\right)\left(\sum_{j=1}^m \sigma_{gj}^2\left(x+u\mathbf{h}\right)\right)\widetilde{\phi}(x+u\mathbf{h})du}{\int \mathbf{K}\left(u\right)\widetilde{\phi}(x+u\mathbf{h})du}.
\end{aligned}$$

This proves the result for a generic element g. We now turn to the multivariate case. Write

$$\frac{\mathbf{M}_r^{(T)}}{\sqrt{\frac{1}{\mathbf{h}}\int_0^T \mathbf{K}\left(\frac{X_s-x}{\mathbf{h}}\right)ds}} = \frac{1}{\sqrt{\frac{1}{\mathbf{h}}\int_0^T \mathbf{K}\left(\frac{X_s-x}{\mathbf{h}}\right)ds}}\left(M_r^{(T)1}, M_r^{(T)2}, ..., M_r^{(T)d}\right),$$

where the generic $M_r^{(T)g}$ was defined earlier. Using the same procedure as above, we can prove that

$$\frac{\left[M^{(T)f}, M^{(T)s}\right]_r}{\sqrt{\frac{1}{\mathbf{h}}\int_0^T \mathbf{K}\left(\frac{X_s-x}{\mathbf{h}}\right)ds}} \overset{a.s.}{\to} \frac{\int \mathbf{K}^2\left(u\right)\sum_{j=1}^m \sigma_{fj}\left(x+u\mathbf{h}\right)\sigma_{js}\left(x+u\mathbf{h}\right)\widetilde{\phi}(x+u\mathbf{h})du}{\int \mathbf{K}\left(u\right)\widetilde{\phi}(x+u\mathbf{h})du} \qquad \forall f, s = 1, ..., d.$$

We now orthogonalize the martingales in $\mathbf{M}_r^{(T)}$ by writing

$$\mathbf{C}^{'}(x)\left(M_r^{(T)1}, M_r^{(T)2}, ..., M_r^{(T)d}\right) = \left(\overline{M}_r^{(T)1}, \overline{M}_r^{(T)2}, ..., \overline{M}_r^{(T)d}\right)$$

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

$$\mathbf{C}^{'}(x)\underbrace{\left(\frac{\int \mathbf{K}^2\left(u\right)\mathbf{a}(u\mathbf{h}+x)\widetilde{\phi}(u\mathbf{h}+x)du}{\int \mathbf{K}\left(u\right)\widetilde{\phi}(u\mathbf{h}+x)du}\right)}_{\mathbf{b}(x)}\mathbf{C}(x) = \Lambda(x)$$

with $\Lambda(x)$ diagonal. $\mathbf{C}(x)$ and $\Lambda(x)$ are the matrices containing the eigenvectors and eigenvalues of $\mathbf{b}(x)$, respectively. We can now apply a variation of the multivariate limiting Knight Theorem (see, e.g., Revuz and Yor (1998), Corollary 2.4, page 497) since

$$\frac{\left[\overline{M}^{(T)f}, \overline{M}^{(T)s}\right]_{\tau_f^n}}{\frac{1}{\mathbf{h}}\int_0^T \mathbf{K}\left(\frac{X_s-x}{\mathbf{h}}\right)ds} = \frac{\left[\overline{M}^{(T)f}, \overline{M}^{(T)s}\right]_{\tau_s^n}}{\frac{1}{\mathbf{h}}\int_0^T \mathbf{K}\left(\frac{X_s-x}{\mathbf{h}}\right)ds} \overset{a.s.}{\to} 0.$$

Then,

$$\frac{1}{\sqrt{\frac{1}{\mathbf{h}}\int_0^T \mathbf{K}\left(\frac{X_s-x}{\mathbf{h}}\right)ds}}\left(\overline{M}_1^{(T)1}, \overline{M}_1^{(T)2}, ..., \overline{M}_1^{(T)d}\right) \Rightarrow \mathbf{B}\left(\Lambda(x)\right),$$

where $\mathbf{B}$ is a $d-$dimensional Brownian motion, which implies

$$\begin{aligned}
\frac{1}{\sqrt{\frac{1}{\mathbf{h}}\int_0^T \mathbf{K}\left(\frac{X_s-x}{\mathbf{h}}\right)ds}}\mathbf{C}(x)\left(\overline{M}_1^{(T)1}, \overline{M}_1^{(T)2}, ..., \overline{M}_1^{(T)d}\right) \quad \Rightarrow \quad & \left(\mathbf{B}\left(\mathbf{C}(x)\Lambda(x)\mathbf{C}^{'}(x)\right)\right) \\
& \overset{d}{=} \mathbf{N}\left(\frac{\int \mathbf{K}^2\left(u\right)\mathbf{a}(u\mathbf{h}+x)\widetilde{\phi}(u\mathbf{h}+x)du}{\int \mathbf{K}\left(u\right)\widetilde{\phi}(u\mathbf{h}+x)du}\right),
\end{aligned}$$

but this proves (49). The proof of (47) is straightforward by the continuity of the diffusion function and invariant measure. To prove (46) we notice that λ is continuous and bounded with modulus of continuity ω^{λ}

$$
\begin{aligned}
&\left|\mathbf{E}\left[\lambda\left(\frac{\frac{1}{\sqrt{\mathbf{h}}}\int_0^T \mathbf{K}\left(\frac{X_s-x}{\mathbf{h}}\right)\boldsymbol{\sigma}(X_s)d\mathbf{B}_s}{\sqrt{\frac{1}{\mathbf{h}}\int_0^T \mathbf{K}\left(\frac{X_s-x}{\mathbf{h}}\right)ds}}\right)\right]-\mathbf{E}\left[\lambda\left(\frac{\frac{1}{\sqrt{\mathbf{h}'}}\int_0^T \mathbf{K}\left(\frac{X_s-x}{\mathbf{h}'}\right)\boldsymbol{\sigma}(X_s)d\mathbf{B}_s}{\sqrt{\frac{1}{\mathbf{h}'}\int_0^T \mathbf{K}\left(\frac{X_s-x}{\mathbf{h}'}\right)ds}}\right)\right]\right| \\
\leq\;& \mathbf{E}\left[w^{\lambda}\left(\frac{\frac{1}{\sqrt{\mathbf{h}}}\int_0^T \mathbf{K}\left(\frac{X_s-x}{\mathbf{h}}\right)\boldsymbol{\sigma}(X_s)d\mathbf{B}_s}{\sqrt{\frac{1}{\mathbf{h}}\int_0^T \mathbf{K}\left(\frac{X_s-x}{\mathbf{h}}\right)ds}},\gamma\right)\right. \\
&\left.\times\left|\frac{\frac{1}{\sqrt{\mathbf{h}}}\int_0^T \mathbf{K}\left(\frac{X_s-x}{\mathbf{h}}\right)\boldsymbol{\sigma}(X_s)d\mathbf{B}_s}{\sqrt{\frac{1}{\mathbf{h}}\int_0^T \mathbf{K}\left(\frac{X_s-x}{\mathbf{h}}\right)ds}}-\frac{\frac{1}{\sqrt{\mathbf{h}'}}\int_0^T \mathbf{K}\left(\frac{X_s-x}{\mathbf{h}'}\right)\boldsymbol{\sigma}(X_s)d\mathbf{B}_s}{\sqrt{\frac{1}{\mathbf{h}'}\int_0^T \mathbf{K}\left(\frac{X_s-x}{\mathbf{h}'}\right)ds}}\right|\right] \\
\leq\;& O(\gamma)\,\mathbf{E}\left[\left|\frac{\frac{1}{\sqrt{\mathbf{h}}}\int_0^T \mathbf{K}\left(\frac{X_s-x}{\mathbf{h}}\right)\boldsymbol{\sigma}(X_s)d\mathbf{B}_s}{\sqrt{\frac{1}{\mathbf{h}}\int_0^T \mathbf{K}\left(\frac{X_s-x}{\mathbf{h}}\right)ds}}-\frac{\frac{1}{\sqrt{\mathbf{h}'}}\int_0^T \mathbf{K}\left(\frac{X_s-x}{\mathbf{h}'}\right)\boldsymbol{\sigma}(X_s)d\mathbf{B}_s}{\sqrt{\frac{1}{\mathbf{h}'}\int_0^T \mathbf{K}\left(\frac{X_s-x}{\mathbf{h}'}\right)ds}}\right|\right] \\
\leq\;& O(\gamma)\left(\mathbf{E}\left[\left|\frac{\frac{1}{\sqrt{\mathbf{h}'}}\int_0^T \mathbf{K}\left(\frac{X_s-x}{\mathbf{h}'}\right)\boldsymbol{\sigma}(X_s)d\mathbf{B}_s}{\sqrt{\frac{1}{\mathbf{h}'}\int_0^T \mathbf{K}\left(\frac{X_s-x}{\mathbf{h}'}\right)ds}}\right|\right]+\mathbf{E}\left[\left|\frac{\frac{1}{\sqrt{\mathbf{h}}}\int_0^T \mathbf{K}\left(\frac{X_s-x}{\mathbf{h}}\right)\boldsymbol{\sigma}(X_s)d\mathbf{B}_s}{\sqrt{\frac{1}{\mathbf{h}}\int_0^T \mathbf{K}\left(\frac{X_s-x}{\mathbf{h}}\right)ds}}\right|\right]\right)
\end{aligned}
$$

But we know that

$$
\mathbf{E}\left[\left|\frac{\frac{1}{\sqrt{\mathbf{h}'}}\int_0^T \mathbf{K}\left(\frac{X_s-x}{\mathbf{h}'}\right)\boldsymbol{\sigma}(X_s)d\mathbf{B}_s}{\sqrt{\frac{1}{\mathbf{h}'}\int_0^T \mathbf{K}\left(\frac{X_s-x}{\mathbf{h}'}\right)ds}}\right|\right]<\infty
$$

for all $0<h<\infty$, giving the desired result. Now write

$$
\mathbf{E}\left[\lambda\left(\frac{\frac{1}{\sqrt{\mathbf{h}_{n,T}}}\int_{T/n}^T \mathbf{K}\left(\frac{X_{i\Delta_{n,T}}-x}{\mathbf{h}_{n,T}}\right)\boldsymbol{\sigma}(X_s)d\mathbf{B}_s}{\sqrt{\frac{\Delta_{n,T}}{\mathbf{h}_{n,T}}\sum_{i=1}^n \mathbf{K}\left(\frac{X_{i\Delta_{n,T}}-x}{\mathbf{h}_{n,T}}\right)}}\right)\right]-\mathbf{E}\left[\lambda\left(\mathbf{N}\left(0,\left(\int k^2(u)\,du\right)^d\mathbf{a}(x)\right)\right)\right] \tag{50}
$$

$$
\leq \sup_{T,n}\left|\mathbf{E}\left[\lambda\left(\frac{\frac{1}{\sqrt{\mathbf{h}_{n,T}}}\int_{T/n}^T \mathbf{K}\left(\frac{X_{i\Delta_{n,T}}-x}{\mathbf{h}_{n,T}}\right)\boldsymbol{\sigma}(X_s)d\mathbf{B}_s}{\sqrt{\frac{\Delta_{n,T}}{\mathbf{h}_{n,T}}\sum_{i=1}^n \mathbf{K}\left(\frac{X_{i\Delta_{n,T}}-x}{\mathbf{h}_{n,T}}\right)}}\right)\right]-\mathbf{E}\left[\lambda\left(\frac{\frac{1}{\sqrt{\mathbf{h}_{T,n}}}\int_0^T \mathbf{K}\left(\frac{X_s-x}{\mathbf{h}_{T,n}}\right)\boldsymbol{\sigma}(X_s)d\mathbf{B}_s}{\sqrt{\frac{1}{\mathbf{h}_{T,n}}\int_0^T \mathbf{K}\left(\frac{X_s-x}{\mathbf{h}_{T,n}}\right)ds}}\right)\right]\right| \tag{51}
$$

$$
+\sup_{h\in\mathcal{H}_{\widetilde{T},\widetilde{n}}(\varepsilon)}\left|\mathbf{E}\left[\lambda\left(\frac{\frac{1}{\sqrt{\mathbf{h}}}\int_0^T \mathbf{K}\left(\frac{X_s-x}{\mathbf{h}}\right)\boldsymbol{\sigma}(X_s)d\mathbf{B}_s}{\sqrt{\frac{1}{\mathbf{h}}\int_0^T \mathbf{K}\left(\frac{X_s-x}{\mathbf{h}}\right)ds}}\right)\right]-\mathbf{E}\left[\lambda\left(\mathbf{N}\left(0,\left(\int \mathbf{K}^2(u)\,du\right)^d\mathbf{a}(x)\right)\right)\right]\right|. \tag{52}
$$

Formulae (51) and (52) are bounded by an arbitrarily small number, $\frac{\varepsilon}{2}$ $(<\delta)$ say, for $n\geq\widetilde{n}$ and $T>\widetilde{T}$ provided $\widetilde{n}$ and $\widetilde{T}$ are chosen accurately, using (44), (45) and (43). This implies that (50) can be bounded by ε giving the weak convergence result in the statement of the theorem.

We now turn to the bias term. Write

$$
\begin{aligned}
&\frac{\frac{1}{\mathbf{h}}\int_0^T \mathbf{K}\left(\frac{X_s-x}{\mathbf{h}}\right)\left(\mu_g(X_s)-\mu_g(x)\right)ds}{\frac{1}{\mathbf{h}}\int_0^T \mathbf{K}\left(\frac{X_s-x}{\mathbf{h}}\right)ds} \\
&\overset{a.s.}{\to}\frac{\frac{1}{\mathbf{h}}\int_{-\infty}^{\infty}\mathbf{K}\left(\frac{a-x}{\mathbf{h}}\right)\left(\mu_g(a)-\mu_g(x)\right)\widetilde{\phi}(a)da}{\frac{1}{\mathbf{h}}\int_{-\infty}^{\infty}\mathbf{K}\left(\frac{a-x}{\mathbf{h}}\right)\widetilde{\phi}(a)da},
\end{aligned}
$$

by the Quotient limit Theorem. But,

$$
\begin{aligned}
& \frac{\frac{1}{\mathbf{h}}\int_{\Re^d}\mathbf{K}\left(\frac{a-x}{\mathbf{h}}\right)\left(\mu_g(a)-\mu_g(x)\right)\widetilde{\phi}(a)da}{\frac{1}{\mathbf{h}}\int_{\Re^d}\mathbf{K}\left(\frac{a-x}{\mathbf{h}}\right)\widetilde{\phi}(a)da} \\
= \quad & \frac{\int_{\Re^d}\mathbf{K}(u)\left(\mu_g(x+u\mathbf{h})-\mu_g(x)\right)\widetilde{\phi}(x+u\mathbf{h})du}{\int_{\Re^d}\mathbf{K}(u)\widetilde{\phi}(x+u\mathbf{h})du} \\
= \quad & \frac{\int_{\Re^d}\prod_{i=1}^d k(u_i)\left[h\sum_{i=1}^d \frac{\partial\mu_g(x)}{\partial x_i}u_i+\frac{1}{2}h^2\sum_{i,j=1}^d \frac{\partial^2\mu_g(x)}{\partial x_i\partial x_j}u_iu_j+o\right]\left[\widetilde{\phi}(x)+h\sum_{i=1}^d\frac{\partial\widetilde{\phi}(x)}{\partial x_i}u_i+o\right]du}{\int_{\Re^d}\prod_{i=1}^d k(u_i)\left[\widetilde{\phi}(x)+h\sum_{i=1}^d\frac{\partial\widetilde{\phi}(x)}{\partial x_i}u_i+o\right]du} \\
= \quad & \frac{\int_{\Re^d}\prod_{i=1}^d k(u_i)\left[h^2\sum_{i=1}^d \frac{\partial\mu_g(x)}{\partial x_i}\frac{\partial\widetilde{\phi}(x)}{\partial x_i}u_i^2+h^2\frac{1}{2}\sum_{i=1}^d \frac{\partial^2\mu_g(x)}{\partial x_i\partial x_i}\widetilde{\phi}(x)u_i^2\right]du}{\int_{\Re^d}\prod_{i=1}^d k(u_i)\left[\widetilde{\phi}(x)+h\sum_{i=1}^d\frac{\partial\widetilde{\phi}(x)}{\partial x_i}u_i+o\right]du} \\
= \quad & h^2\left(\int_{\Re}s^2k(s)ds\right)\left(\sum_{i=1}^d\frac{\partial\mu_g(x)}{\partial x_i}\frac{\frac{\partial\widetilde{\phi}(x)}{\partial x_i}}{\widetilde{\phi}(x)}+\frac{1}{2}\sum_{i=1}^d\frac{\partial^2\mu_g(x)}{\partial x_i\partial x_i}\right) \qquad (53)
\end{aligned}
$$

by the symmetry of the kernel function. Thus,

$$
bias_g(x)=h^2\left(\int_{\Re}s^2k(s)ds\right)\left(\sum_{i=1}^d\frac{\partial\mu_g(x)}{\partial x_i}\frac{\frac{\partial\widetilde{\phi}(x)}{\partial x_i}}{\widetilde{\phi}(x)}+\frac{1}{2}\sum_{i=1}^d\frac{\partial^2\mu_g(x)}{\partial x_i\partial x_i}\right). \qquad (54)
$$

This proves the stated result.

Proof of Theorem 6 Using Itô's lemma, write

$$
\begin{aligned}
& \left(X_{(i+1)\Delta_{n,T}}-X_{i\Delta_{n,T}}\right)\left(X_{(i+1)\Delta_{n,T}}-X_{i\Delta_{n,T}}\right)' \\
= \quad & \int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}\mathcal{D}_X(X_s-X_{i\Delta_{n,T}})\otimes\boldsymbol{\mu}(X_s)ds \\
& +\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}\mathcal{D}_X(X_s-X_{i\Delta_{n,T}})\otimes\boldsymbol{\sigma}(X_s)d\mathbf{B}_s \\
& +\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}\mathbf{a}(X_s)ds,
\end{aligned}
$$

where $\mathcal{D}_X(u)$ is a $d\times d$ array of $1\times d$ vectors. The i-th element of the mn-th vector is given by $\frac{\partial\left((x-y)(x-y)'\right)_{mn}}{\partial x_i}$. We use the duplication matrix, i.e., the unique $d^2\times(d(d+1))/2$ matrix D so that

$$
Dvech\mathbf{A}=vec\mathbf{A}
$$

for a generic symmetric matrix $\mathbf{A}$, to knock $\mathbf{A}$ down to non-redundant elements by stacking its upper triangular components. Now write

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

The term $vech\boldsymbol{\Psi}_2$ averages martingale difference sequences and converges to zero at speed $\sqrt{\frac{\mathbf{h}_{n,T}\widehat{L}_{n,T}(T,x)}{\Delta_{n,T}}}$ (c.f. the proof of Theorem 7). The term $vech\boldsymbol{\Psi}_1$ is clearly so that

$$
vech\boldsymbol{\Psi}_1 = o_{a.s.}(vech\boldsymbol{\Psi}_2) = o_{a.s.}\left(\sqrt{\frac{\Delta_{n,T}}{\mathbf{h}_{n,T}\widehat{L}_{n,T}(T,x)}}\right).
$$

As for $vech\boldsymbol{\Psi}_3$, the same steps as in the proof of Theorem 3 allow us to show that

$$
vech\boldsymbol{\Psi}_3 \overset{a.s.}{\to} vech\mathbf{a}(x)
$$

provided $(\Delta_{n,T}\log(1/\Delta_{n,T}))^{1/2}\,\widehat{L}_{n,T}(T,x)/\mathbf{h}_{n,T} \overset{a.s.}{\to} 0$, but this proves the stated result.

Proof of Theorem 7 We can write the estimation error decomposition as

$$
\begin{aligned}
& vech\widehat{\mathbf{a}}_{n,T}(x) - vech\mathbf{a}(x) \\
= \; & vech\boldsymbol{\Psi}_3 - vech\mathbf{a}(x) + vech\boldsymbol{\Psi}_1 + vech\boldsymbol{\Psi}_2. \qquad (55)
\end{aligned}
$$

We use the same procedure as in Theorem 5, i.e., we orthogonalize the vector of martingales $vech\boldsymbol{\Psi}_2$ and apply the asymptotic multivariate Knight theorem to show that

$$
\sqrt{\frac{\mathbf{h}_{n,T}\widehat{L}_{n,T}(T,x)}{\Delta_{n,T}}}vech\boldsymbol{\Psi}_2 \Rightarrow \mathbf{N}\left(0,\mathbf{K}_2P_D\left(2\mathbf{a}(x)\otimes\mathbf{a}(x)\right)P_D\right), \qquad (56)
$$

where

$$
\mathbf{K}_2 = \left(\int_{\mathfrak{R}}k^2(s)ds\right)^d
$$

and

$$
P_D = \left(D'D\right)^{-1}D
$$

is the Moore-Penrose generalized inverse of D (see, e.g., Magnus and Neudecker (1988)). Also, we note that

$$
vech\boldsymbol{\Psi}_1 = o_{a.s.}\left(\sqrt{\frac{\Delta_{n,T}}{\mathbf{h}_{n,T}\widehat{L}_{n,T}(T,x)}}\right) \qquad (57)
$$

and, using the derivations leading to (54) above, that

$$vech\mathbf{\Psi}_3 - vech\mathbf{a}(x) = vech\mathbf{s}(x)$$

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

$$s_{ij}(x) = h_{n,T}^2 \left(\int s^2 k(s) ds \right) \left(\sum_{k=1}^{d} \frac{\partial a_{ij}(x)}{\partial x_k} \frac{\frac{\partial \widetilde{\phi}(x)}{\partial x_k}}{\widetilde{\phi}(x)} + \frac{1}{2} \sum_{k=1}^{d} \frac{\partial^2 a_{ij}(x)}{\partial x_k \partial x_k} \right) \qquad i,j = (1,1), ..., (d,d). \tag{58}$$

We obtain the stated result by combining the estimation error decomposition in (55) with (56), (57), and (58).

References

[1] Aït-Sahalia, Y., 2008. Closed-form likelihood expansions for multivariate diffusions. *Annals of Statistics* 36, 906-937.

[2] Aït-Sahalia, Y., L.P. Hansen, and J. Scheinkman, 2008. Operator methods for continuous-time Markov processes. In *Handbook of Financial Econometrics* (Y. Aït-Sahalia and L. P. Hansen, eds.) Elsevier, *forthcoming.*

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

[4] Bandi, F.M. and P.C.B. Phillips, 2003. Fully nonparametric estimation of scalar diffusion models. *Econometrica* 71, 241-283.

[5] Bandi, F.M. and P.C.B. Phillips, 2008. Nonstationary continuous-time models. In *Handbook of Financial Econometrics* (Y. Aït-Sahalia and L. P. Hansen, eds.) Elsevier, *forthcoming.*

[6] Bickel, P.J. and M. Rosenblatt, 1973. On some global measures of the deviations of density function estimates. *Annals of Statistics* 1, 1071-1095.

[7] Bosq, D., 1998. *Nonparametric Statistics for Stochastic Processes.* Springer-Verlag.

[8] Boudoukh, J., M. Richardson, R. Stanton, and R.F. Whitelaw, 2003. The stochastic behavior of interest rates: implications from a nonlinear, continuous-time, multifactor model. *Working paper.*

[9] Brugière, P., 1991. Estimation de la variance d'un processus de diffusion dans le cas multidimensionnel. *C. R. Acad. Sci. Paris Série I* 312, 999–1004.

[10] Brugière, P., 1993. Théorème de limite centrale pour un estimateur nonparamétrique de la variance d'un processus de diffusion multidimensionnelle. *Ann. Inst. Henri Poincaré* 29, 357–389.

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

[12] Chen, X. and L.P. Hansen, 2002. Dependence properties of multivariate reversible diffusions. *Working paper.*

[13] Darling, D.A. and M. Kac, 1957. On occupation times for Markoff processes. *Transactions of the American Mathematical Society* 84, 444–458.

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

[15] Fan, J., 2005. A selective overview of nonparametric methods in financial econometrics. *Statistical Science* 20, 317-337.

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

[17] Fan, Y., 1994. Testing the goodness of fit of a parametric density function by kernel method. *Econometric Theory* 10, 316-356.

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

[19] Gallant, R. and G. Tauchen, 2008. Simulated methods and indirect inference for continuous-time models. In *Handbook of Financial Econometrics* (Y. Aït-Sahalia and L.P. Hansen, eds.) Elsevier, *forthcoming.*

[20] Geman, D. and J. Horowitz, 1980. Occupation densities. *Annals of Probability* 8, 1–67.

[21] Genon-Catalot, V. and J. Jacod, 1993. On the estimation of the diffusion coefficient for multi-dimensional diffusion processes. *Ann. Inst. H. Poincaré Probab. Statist.* 29, 119–151.

[22] Guerre, E., 2007. Design-adaptive point-wise nonparametric regression estimation for recurrent Markov time series. *Working paper.*

[23] Jiang, G.J. and J. Knight, 1999. Finite sample comparison of alternative estimators of Itô diffusion processes: a Monte-Carlo study. *Journal of Computational Finance* 2, 5-38.

[24] Johannes, M. and N. Polson, 2008. MCMC methods for financial econometrics. In *Handbook of Financial Econometrics* (Y. Aït-Sahalia and L.P. Hansen, eds.) Elsevier, *forthcoming.*

[25] Kallianpur, G. and H. Robbins, 1953. Ergodic properties of the Brownian motion Process. *Proc. Nat. Acad.* 39, 525-533.

[26] Karatzas, I. and S. E. Shreve, 1988. *Brownian Motion and Stochastic Calculus.* Springer-Verlag.

[27] Karlsen, H.A. and D. Tjøstheim, 2001. Nonparametric estimation in null recurrent time series. *Annals of Statistics* 29, 372-416.

[28] Karlsen, H.A., T. Myklebust, and D. Tjøstheim, 2007. Nonparametric estimation in a nonlinear cointegrating type model. *Annals of Statistics* 35, 252-299.

[29] Kasahara, Y., 1975. Spectral theory of generalized second order differential operators and its applications to Markov processes. *Japan J. Math.* 1, 67–84.

[30] Kaspi, H. and A. Mandelbaum, 1994. On Harris recurrence in continuous time. *Mathematics of Operation Research* 19, 211–222.

[31] Kliemann, W., 1983. Transience, recurrence and invariant measures for diffusions. In *Nonlinear Stochastic Problems* (R.S. Bucy and M.F. Moura, eds.) Reidel Publishing Company, 437–454.

[32] Magnus, J. R. and H. Neudecker, 1988. *Matrix Differential Calculus with Applications in Statistics and Econometrics.* Wiley.

[33] Masry, E., 1996a. Multivariate local polynomial regression for time series: uniform strong consistency and rates. *Journal of Time Series Analysis* 17, 571-599.

[34] Masry, E., 1996b. Multivariate regression estimation: local polynomial fitting for time series. *Journal of Stochastic Processes and their Applications* 65, 81-101.

[35] McKean, H. P., 1969. *Stochastic Integrals.* Springer-Verlag.

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

[37] Moloche, G., 2004. Local nonparametric estimation of scalar diffusions. *Working paper.*

[38] Moloche, G., 2004b. Kernel regression for non-stationary Harris recurrent processes. *Working paper.*

[39] Pagan, A. and A. Ullah, 1999. *Nonparametric Econometrics.* Cambridge University Press.

[40] Park, J., 2006. Spatial analysis of time series. *Working paper.*

[41] Phillips, P.C.B. and J. Park, 1998. Nonstationary density estimation and kernel autoregression. *Working paper.*

[42] Revuz, D. and M. Yor, 1998. *Continuous Martingales and Brownian Motion*, Springer-Verlag.

[43] Stanton, R., 1997. A nonparametric model of term structure dynamics and the market price of interest rate risk. *Journal of Finance* 52, 1973–2002.

[44] Wang, Q., and P.C.B. Phillips, 2008. Structural nonparametric cointegrating regression. *Working paper.*

[45] Wang, Q., and P.C.B. Phillips, 2009. Asymptotic theory for local time density estimation and nonparametric cointegrating regression. *Econometric Theory,* forthcoming.