Local Nonparametric Estimation of Scalar Diffusions

Moloche, Guillermo

University of Chicago

23 September 2001

Local Nonparametric Estimation of Scalar Diffusions*

Guillermo Moloche
M.I.T.

September 23rd, 2001
First Draft: March 31st, 2000

Abstract

This paper studies the functional estimation of the drift and diffusion functions for recurrent scalar diffusion processes from equally spaced observations using the local polynomial kernel approach. Almost sure convergence and a CLT for the estimators are established as the sampling frequency and the time span go to infinity. The asymptotic distributions follow a mixture of normal laws. This theory covers both positive and null recurrent diffusions.

Almost sure convergence rates are sometimes path dependent but expected rates can always be characterized in terms of regularly varying functions.

The general theory is specialized for positive recurrent diffusion processes, and it is shown in this case that the asymptotic distributions are normal.

We also obtain the limit theory for kernel density estimators when the process is positive recurrent, namely, requiring only that the invariant probability measure exists. Nonetheless, it is also shown that such an estimator paradoxically vanishes almost surely when the invariant measure is fat tailed and nonintegrable, that is, in the null recurrent case.

*Email: gmoloche@mit.edu. This paper benefited from comments by Federico Bandi.

1 Introduction.

This paper studies the functional estimation of the drift and diffusion functions for recurrent scalar diffusion processes from equally spaced observations using the local polynomial kernel approach. Almost sure convergence and a central limit theorem for the drift and diffusion estimators are established as the sampling frequency and the time span go to infinity. The asymptotic distributions follow a mixture of normal laws. This theory is valid for both null and positive recurrent diffusions. In the latter case, it is not necessary to initialize the process with the invariant probability measure to obtain the results in this paper.

The theory presented here builds on a methodology recently introduced by Bandi and Phillips [3] to obtain the asymptotics of kernel estimators for recurrent diffusion processes. Their methodology showed how to obtain asymptotic theory without assuming the existence of an invariant probability measure. Recurrence is all that is needed in their framework. Further, it is one of a handful of papers that are able to derive the statistical properties of econometric estimators without imposing awkward mixing conditions. Geman [8] first suggested that this is possible in the continuous time case. More recently, other authors conjecture that this is possible in the discrete time case, as well, cf. Yakowitz [18].

This paper extends the Bandi-Phillips theory in three directions: first, showing how the local polynomial approach is able to reduce the small sample bias, second, further analyzing the convergence rates and showing how they depend on the infinitesimal coefficients, and third, studying completely the positive recurrent case.

The improvement of local polynomial kernel estimation with respect to simple kernel estimators arises from the fact that the bias terms are of smaller order of magnitude even when the slower martingale terms in both methods are of similar order of magnitude. Both estimators have similar asymptotic rates of convergence, but the local polynomial kernel estimator has a sample bias of smaller order of magnitude, and consequently it is expected to perform better in small, discrete samples. In addition, local polynomial estimators do not need additional bias corrections at the boundaries. While convergence rates are path dependent in the null recurrent case, expected rates can be still characterized by means of regularly varying functions. We show how to obtain these rates from the scale function and the speed measure.

The general theory is specialized for positive recurrent diffusion processes, and it is shown in this case that the asymptotic distributions are normal and no longer mixed normal. It is not necessary to initialize the process with the invariant probability measure. Bounds for the convergence rate to the true infinitesimal parameters are also obtained; these bounds are no longer path dependent. This particular theory applies to the estimation of any positive recurrent diffusion with continuous infinitesimal parameters. Conventional estimators require not only positive recurrence: they also require to impose strict stationarity, that is, an appropriate initialization, and sometimes additional conditions to ensure that functionals of the process to be estimated satisfy the same regularity as the original process.

As we are not imposing constraints on mixing conditions, this theory applies when they are not satisfied, as long as recurrence still holds. In such a case, the process would be strong-dependent in the sense of not having decaying (at a fast enough rate) mixing coefficients.

The asymptotic distributions of the estimators in the stationary case depend on the marginal density and this has to be estimated. We derive the limit theory for kernel density estimators when the process is positive recurrent, namely, requiring only that the invariant probability measure exists. The process might still be nonstationary due to a initialization with a probability distribution that is not invariant. We only study the case of stationary density estimation with both infill and long span asymptotics. It is known that the density estimators are consistent even when the sampling frequency is finite, but faster rates of convergence are achieved by sampling at increasing frequencies.

Nonetheless it is also shown that such an estimator paradoxically vanishes almost surely when the invariant measure is fat tailed and nonintegrable, that is, in the null recurrent case. This is a consequence of Theorem 1, below. The kernel density estimator converges to the local time estimator under the stated conditions, but the local time estimator vanishes a.s. when the invariant measure is not finite.

We begin by showing the definitions, notation, assumptions that will be used in the paper, and some preliminar results, basically concerned with the asymptotics of the local time of scalar diffusion processes. The estimators will be described before proceeding to section 3, where the results are presented. The positive recurrent case is analyzed in section 4. Proofs are in section 5.

2 Preliminars.

We study a scalar diffusion X defined on an open interval in the real line $I = (l, r) \subseteq \mathbb{R}$, where $-\infty \leq l < r \leq +\infty$. Given a standard Brownian motion B defined on the filtered probability space $(\Omega, \mathcal{F}^B, (\mathcal{F}_t^B)_{t\geq 0}, P)$, the process X solves the time homogeneous stochastic differential equation

$$X_t = X_0 + \int_0^t \mu(X_s)ds + \int_0^t \sigma(X_s)dB_s \tag{1}$$

almost surely, where $\sigma : I \to I$, $\mu : I \to I$. The initial condition $X_0 = \overline{X} \in \mathcal{L}^2$ is independent of B. The process X is adapted to the augmented filtration $\widetilde{\mathcal{F}}_t^X$, which is constructed by defining the left-continuous filtration and the collection of null sets

$$\mathcal{F}_t^X = \sigma(\overline{X}, B_s; 0 \leq s \leq t),\ 0 \leq t < \infty,$$

$$\mathcal{X} = \{F \subseteq \Omega; \exists G \in \mathcal{F}_\infty^X \text{ with } F \subseteq G \text{ and } P(G) = 0\}\ ,$$

and setting $\widetilde{\mathcal{F}}_t^X := \sigma(\mathcal{F}_t^X \cup \mathcal{X})$, $0 \leq t < \infty$.

Define the exit time of the process X: $S = \inf\{t \geq 0 : X_t \notin I\}$. We say that a process is explosive if $P[S < +\infty] > 0$.

Fixing some $c \in I$, the scale function is

$$p(x) = \int_c^x \exp\left\{-2\int_c^\xi \frac{\mu(\zeta)}{\sigma^2(\zeta)}d\zeta\right\} d\xi, \quad x \in I$$

and the speed measure

$$m(dx) = \frac{dx}{p'(x)\sigma^2(x)}, \quad x \in I,$$

which is finite when the diffusion is positive recurrent. On the other hand, when the diffusion is null recurrent, m is not integrable but it is still σ-finite. In any case, m is always an invariant measure. Let also

$$v(x) = \int_c^x (p(x) - p(y))\, m(dy), \quad x \in I.$$

We first ensure the existence and pathwise uniqueness of a non-explosive solution to (1) that is adapted to the augmented filtration $\left\{\widetilde{\mathcal{F}}_t^X\right\}$:

Assumption 1.

i. σ^2, $\mu \in \mathcal{C}^{p+1}(I)$, *some* $p \geq 0$.

ii. *(Non degeneracy):* $\sigma^2 > 0$, $\forall x \in I$.

iii. *(Local Integrability):* $\forall x \in I$, *there exists some* $\varepsilon > 0$ *such that*

$$\int_{x-\varepsilon}^{x+\varepsilon} \frac{1+|\mu(y)|}{\sigma^2(y)} dy < \infty; \ [x-\varepsilon, x+\varepsilon] \subset I.$$

iv. *There exist locally integrable functions* g, *a constant* c *and a number* $\delta > 0$ *such that for* $\forall x \in I$, $\forall y \in [x-\delta, x+\delta] \subset I$

$$(\sigma(x) - \sigma(y))^2 \leq \left(c + g(x)\sigma^2(x)\right) \rho(|x-y|)$$

where $\rho : (0, \infty) \to (0, \infty)$ *is such that*

$$\int_0^\infty \frac{dy}{\rho(y)} = +\infty.$$

μ *satisfies*

$$|\mu(x) - \mu(y)| \leq \kappa\left(|x-y|\right)$$

where $\kappa = [0, \infty) \to [0, \infty)$ *is a strictly increasing concave function with* $\kappa(0) = 0$ *and*

$$\int_0^\varepsilon \frac{dy}{\kappa(y)} < \infty; \qquad \forall \varepsilon > 0.$$

v. $v(l+) = v(r-) = \infty$.

Condition *(1.i)* implies that the infinitesimal parameters and their derivatives are bounded on compact subsets of I. Assumption *(1.v)* ensures that the process is nonexplosive. It turns out that nonexplosive one dimensional diffusion processes are also recurrent. Pointwise recurrence is a strong assumption for a Markov process, but it is satisfied nonetheless by all nonexplosive diffusion processes due to the continuity of paths. We will not need to discuss here weaker notions of recurrence such as Harris recurrence, this will be necessary only for diffusions of several dimensions and more general Markov processes. Once we have imposed enough conditions to ensure that the process is nonexplosive, we can limit the discussion of the properties of the drift and diffusion functions to compact subsets as in condition *(1.i)*.

Conditions *(1.ii)-(1.iii)* are sufficient for existence of a weak solution up to an explosion time, given any initial distribution. This solution is unique in the sense of probability law. Condition *(1.iv)* guarantees the pathwise uniqueness of the solution. Then, applying a theorem by Yamada and Watanabe, weak existence and pathwise uniqueness imply strong existence. These conditions are only sufficient but still weaker than, for example, the global boundedness of σ and μ and their derivatives, or global growth and Lipschitz condition.

Condition *(1.v)* is Feller's test for explosions which ensures that the solution is nonexplosive, that the boundaries l and r are nonattainable, $P[S = \infty] = 1$, and that X is pointwise recurrent:

$$\forall x \in I,\ P[X_t = x,\ \text{some } 0 \leq t < \infty] = 1.$$

Verifying that the scale function diverges at the boundaries, $p(l+) = -\infty$ and $p(r-) = \infty$ is a sufficient but not necessary test. There are diffusions with a bounded scale function that nevertheless are nonexplosive. Feller's test is sufficient and necessary.

When the process is known to never cross the origin, i.e. when the left boundary is zero ($l = 0$) and unattainable, there exists a simple set of sufficient conditions: it is enough to ensure the continuity and differentiability of σ and μ, nondegeneracy, and the nonattainability of the boundaries. Then, standard local Lipschitz conditions are obtained from an elementary application of the mean value theorem.

After ensuring that a nonexplosive solution for the SDE exists on an interval I, for notational simplicity and without loss of generality we can assume henceforth that $I = \mathbb{R}$. We impose some regularity on the kernel functions.

Assumption 2.

i. *The kernel function K is bounded and continuous on $\mathbb{R}$ with $\int K(u)du = 1$, $\int K^j(u)du < \infty$, $\int u^j K(u)du < \infty$, for $j = 1, 2, \ldots$.*

ii. *There exist functions $D^j(u, \varepsilon)$ such that:*

1.

$$\left|v^j K(v) - u^j K(u)\right| \leq D^j(u, \varepsilon)\,|v - u|$$

all $u \in \mathbb{R}$ *for* $j = 0, 1, 2, ...,$ $|v - u| < \varepsilon$, *some* $\varepsilon > 0$, *clearly,* $D^j(u, \varepsilon) \leq D^j(u, \varepsilon')$, *if* $\varepsilon < \varepsilon'$.

2.

$$\lim_{\varepsilon \downarrow 0} \int D^j(u, \varepsilon) du < \infty, \forall j \ .$$

These conditions are satisfied by the most common kernel functions including the Gaussian, the Epanechnikov and the indicator function. The existence of functions $D^j(u, \varepsilon)$ satisfying the required conditions can be easily verified by applying dominated convergence arguments.

We now define local time. It is customary to define the local time $L_X(t, x)$ of a continuous semimartingale through the Tanaka-Meyer formula:

$$|X_t - x| = |X_0 - x| + \int_0^t \mathrm{sgn}\,(X_s - x)\, dX_s + L_X(t, x)$$

$$(X_t - x)^+ = (X_0 - x)^+ + \int_0^t 1_{\{X_s > x\}} dX_s + \frac{1}{2} L_X(t, x)$$

$$(X_t - x)^- = (X_0 - x)^- - \int_0^t 1_{\{X_s \leq x\}} dX_s + \frac{1}{2} L_X(t, x),$$

almost surely. The asymptotic distributions for the estimators developed below involve the local time of the diffusion process as in Florens-Zmirou [7], and Bandi and Phillips [3]. In the latter, the local time of X in x during $[0, T]$ is defined as

$$\overline{L}_X(T, a) = \frac{1}{\sigma^2(a)} L_X(T, a),\ a \in \mathbb{R}$$

We can also define local times as an occupation times density. We prefer the last interpretation because it can be extended to multivariate diffusions, and more general processes. The occupation times measure $\pi_T : \mathcal{B}(\mathbb{R}) \to \mathbb{R}$ is defined as

$$\pi_T(B) = \int_0^T 1_B(X_s) ds; \ \ \forall B \in \mathcal{B}(\mathbb{R})$$

Then, if π_T is absolutely continuous with respect to Lebesgue measure, its density, the occupation density is $\overline{L}_X(T, x)$ satisfying

$$\pi_T(B) = \int_B \overline{L}_X(T, x) dx; \ \ \forall B \in \mathcal{B}(\mathbb{R}).$$

Note that $\overline{L}_X(T,x)$ is a version of the Radon-Nikodym derivative of the occupation measure with respect to Lebesgue measure.

Next, we recall some asymptotics for occupation densities (cf. Itô and McKean [9]) which will be useful to study the positive recurrent case and rates of convergence in the general case.

Theorem 1. *If X is a diffusion satisfying Assumption 1, then*

$$\lim_{T\to\infty}\frac{\overline{L}_X(T,a)}{T}=\lim_{T\to\infty}\frac{E.\left[\overline{L}_X(T,a)\right]}{T}=\frac{m(a)}{m\left(I\right)}\ a.s.,\ \forall a\in I.$$

This theorem is valid even if $m\left(I\right)=\infty$, that is, when the diffusion is null recurrent. Note that if $m\left(I\right)<\infty$, the limit is the density of the invariant probability measure $f\left(x\right)=m\left(x\right)/m\left(I\right)$. Since the infinitesimal parameters are assumed to be continuous, the absolute continuity of the invariant probability measure and hence the existence of the stationary density are guaranteed when $m(I)<\infty$. It is clear from this theorem that when the speed measure is not integrable, local time is $o_{a.s.}\left(T\right)$; but it is $O_{a.s.}\left(T\right)$ when the process is positive recurrent. Paradoxically, when the tails of the invariant measure are extremely fat and thus, the invariant measure is nonintegrable, the local time estimator of the invariant density $\overline{L}_X(T,x)/T$ vanishes almost surely.

Note that $E_a\left[\overline{L}_X\left(T,b\right)\right]=\int_0^T p\left(s,a,b\right)ds$, $a\in I$, $T\in(0,\infty)$. Then, the ergodic theory in Theorem 1 says also that

$$\lim_{T\to\infty}\frac{\int_0^T p\left(s,a_1,b\right)ds}{\int_0^T p\left(s,a_2,b\right)ds}=1,\ a_1,a_2\in I,$$

$$\lim_{T\to\infty}\frac{\int_0^T p\left(s,a,b_1\right)ds}{\int_0^T p\left(s,a,b_2\right)ds}=\frac{m\left(b_1\right)}{m\left(b_2\right)},\ b_1,b_2\in I;$$

and the expected rate of divergence of local time is independent of both the initial condition and the space variable.

We proceed to characterize the growth rate of $E.\left[\overline{L}_X\left(T,a\right)\right]$ for an arbitrary $a\in I$ in the null recurrent case. It turns out that this rate can be characterized by analyzing the rate of divergence of the Laplace transform of the time integral of transition densities by using the classic Karamata's Tauberian theorem. Before the statement of the next theorem let us recall that a positive monotone function defined on $(0,\infty)$ is said to be regularly varying at infinity with exponent δ if

$$\lim_{x\to\infty}\frac{h\left(\lambda x\right)}{h\left(x\right)}=\lambda^{\delta},\ \forall\lambda>0$$

where $|\delta|<\infty$, see Feller [6], Chapter VIII. Such functions are of the form $h\left(x\right)=x^{\delta}L\left(x\right)$, where $L\left(x\right)$ is a slowly varying function, that is, a positive, not necessarily monotone function defined on $(0,\infty)$ such that

$$\lim_{x\to\infty}\frac{L\left(\lambda x\right)}{L\left(x\right)}=1,\ \forall\lambda>0.$$

An example of regularly varying function is $h\left(x\right)=x^{\delta}$, for $\delta\in[0,1]$; a slowly varying function is $L\left(x\right)=\ln\left(x\right)$.

Theorem 2. *If the diffusion X satisfies Assumption 1, then there exists a regularly varying function h with exponent $0<\delta\leq 1$ such that*

$$\lim_{T\to\infty}\frac{E.\left[\overline{L}_X(T,a)\right]}{h\left(T\right)}=m\left(a\right)\ \ a.s.,\ \forall a\in I, \tag{2}$$

The function h is the slower of the regularly varying functions h_1, h_2 satisfying

$$\lim_{a\uparrow r}h_1\left(p\left(a\right)m\left(\left[c,a\right)\right)\right)/p\left(a\right)=C_1,$$

$$\lim_{a\downarrow l}h_2\left(-p\left(a\right)m\left(\left(a,c\right]\right)\right)/p\left(a\right)=C_2,$$

where $c\in I$ is fixed, and C_1, C_2 are two positive constants. The function h that makes (2) true is unique. If the process is positive recurrent then $h\left(T\right)=T$, m is the invariant probability measure and $\delta=1$. If it is null recurrent, $\delta\in(0,1)$, and h is of the form $h\left(T\right)=T^{\delta}L\left(T\right)$, for some slowly varying function L.

Theorem 1 implied that in the null recurrent case $h\left(T\right)=o\left(T\right)$. Theorem 2 says that expected local times have a positive, nonstochastic limit when normalized by $h(T)$, even in the null recurrent case, when there is no initialization of the process that makes it stationary. Such normalization depends only on the scale and the speed measure, and in consequence, only on the drift and diffusion function. The limit is always a rescaled speed measure, which in the recurrent case is always equal (up to scale) to the invariant measure.

However, $\lim_{T\to\infty} \overline{L}_X(T,x)/h(T)$ is stochastic when $0<\delta<1$, and a.s. rates of convergence cannot be characterized independently of the sample path in the null recurrent case.

Let us illustrate how to apply the last two theorems to Brownian motion. It is not always possible to find the function h explicitly, but it is well known that for Brownian motion $h(T)=\sqrt{T}$, cf. Revuz and Yor [16]. Let us derive this function using the Tauberian theorem. Since Brownian motion is null recurrent, its speed measure is not integrable; then $L_B(T,a)/T\to 0$ a.s. $\forall a\in I$, as stated in Theorem 1. We want to find a deterministic function $h(T)\to\infty$ such that $E_.[L_B(T,a)]/h(T)\to c$, for some constant $c(a)>0$. Note first that without loss of generality it is enough to analyze the case $E_a[L_X(T,a)]=\int_0^T p(s,a,a)\,ds=\int_0^T \frac{1}{\sqrt{2\pi s}}e^{-(a-a)^2/2s}ds=\sqrt{2T/\pi}$. Then, it follows that $\lim_{T\to\infty} E_.[L_B(T,a)]/\sqrt{T}=\sqrt{2/\pi}$, $a\in I$ and this is indeed the density of a Lebesgue measure. To apply the Tauberian theorem to Brownian motion, first compute the Laplace transform $\int_0^\infty \frac{1}{\sqrt{2\pi s}}e^{-\alpha s}ds=1/\sqrt{2\alpha}$ implying that $\lim_{\alpha\downarrow 0}\int_0^\infty e^{-\alpha s}p_s(a,b)\,ds/\sqrt{(1/\alpha)}=\sqrt{1/2}$ and $h(T)=\sqrt{T}$. Theorem 2 is applied by recalling that for Brownian motion $p(x)=x$ and $m(x)=1$, $m([a,b])=b-a$. Then $\lim_{a\uparrow\infty} h(a^2-ca)/a=1$ only if $h(a)=a^{1/2}$ and h is clearly regularly varying with $\delta=1/2$. The h-function for the left boundary is identical. Then $h(T)=\sqrt{T}$.

We now introduce the estimators. Assume that we observe the process X at $\{t=t_0,t_1,t_2,...,t_n\}$ in $[0,T]$, with $T>0$. The observations are equispaced.

Then, $\{X_t=X_0,X_{\Delta_{n,T}},X_{2\Delta_{n,T}},X_{3\Delta_{n,T}},...,X_{n\Delta_{n,T}}\}$ are $n+1$ observations of the process X_t at $\{t_0=0,t_1=\Delta_{n,T},t_2=2\Delta_{n,T},,t_3=3\Delta_{n,T},...,t_n=n\Delta_{n,T},\}$ where $\Delta_{n,T}=T/n$.

We explain how to obtain the local fitting estimator for the diffusion function. The method attempts to estimate the diffusion function and its derivatives $\{\sigma^{2(\nu)}=\beta_\nu \nu!\}_{\nu=0}^p$ at an arbitrary point x by solving the following problem:

$$\begin{aligned}\beta_{n,T} &= \arg\min_\beta \sum_{i=0}^{n-1}\frac{\Delta_{n,T}}{h_{n,T}}K\left(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}}\right)\\ &\quad\times\left\{\frac{1}{\Delta_{n,T}}\left(X_{(i+1)\Delta_{n,T}}-X_{i\Delta_{n,T}}\right)^2-\sum_{\nu=0}^p\beta_\nu(X_{i\Delta_{n,T}}-x)^\nu\right\}^2 . \quad (3)\end{aligned}$$

The solution to the minimization problem (3) gives estimates for the diffusion function and its derivatives:

$$\sigma^{2(\nu)}_{n,T}(x) = \beta^{\nu}_{n,T}\nu!,\ \nu = 0, ..., p\ .$$

Similarly, the drift function and its derivatives $\{\mu^{(\nu)} = \alpha_\nu \nu!\}_{\nu=0}^{p}$ are estimated at an arbitrary point x by solving:

$$\begin{aligned}\alpha_{n,T} &= \arg\min_{\alpha} \sum_{i=0}^{n-1} \frac{\Delta_{n,T}}{g_{n,T}} K\left(\frac{X_{i\Delta_{n,T}} - x}{g_{n,T}}\right) \\ &\times \left\{ \frac{1}{\Delta_{n,T}} \left(X_{(i+1)\Delta_{n,T}} - X_{i\Delta_{n,T}}\right) - \sum_{\nu=0}^{p} \alpha_\nu (X_{i\Delta_{n,T}} - x)^\nu \right\}^2 \ . \quad (4)\end{aligned}$$

The solution to the minimization problem (4) gives estimates for the drift function and its derivatives:

$$\mu^{(\nu)}_{n,T}(x) = \alpha^{j}_{n,T}\nu!\ ,\ \nu = 0, ..., p\ .$$

The local polynomial kernel estimators can be written as a weighted least squares problem. See Fan and Gijbels [5]:

$$\mathbf{X}_{n,T} = \begin{bmatrix} 1 & (X_0 - x) & ... & (X_0 - x)^p \\ ... & & & ... \\ 1 & (X_{(n-1)\Delta_{n,T}} - x) & ... & (X_{(n-1)\Delta_{n,T}} - x)^p \end{bmatrix}\ ,$$

$$\mathbf{y}_{n,T} = \begin{bmatrix} \frac{1}{\Delta_{n,T}}(X_{\Delta_{n,T}} - X_0)^2 \\ ... \\ \frac{1}{\Delta_{n,T}}(X_{n\Delta_{n,T}} - X_{(n-1)\Delta_{n,T}})^2 \end{bmatrix}\ ,$$

$$\mathbf{W}_{n,T} = \operatorname{diag}\left(\frac{\Delta_{n,T}}{h_{n,T}} K\left(\frac{X_0 - x}{h_{n,T}}\right), ..., \frac{\Delta_{n,T}}{h_{n,T}} K\left(\frac{X_{n\Delta_{n,T}} - x}{h_{n,T}}\right)\right)\ ,$$

then, we can write

$$\beta_{n,T} = (\mathbf{X}'_{n,T}\mathbf{W}_{n,T}\mathbf{X}_{n,T})^{-1}\mathbf{X}'_{n,T}\mathbf{W}_{n,T}\mathbf{y}_{n,T}\ .$$

If in addition

$$\mathbf{z}_{n,T} = \begin{bmatrix} \frac{1}{\Delta_{n,T}}(X_{\Delta_{n,T}} - X_0) \\ ... \\ \frac{1}{\Delta_{n,T}}(X_{n\Delta_{n,T}} - X_{(n-1)\Delta_{n,T}}) \end{bmatrix}\ ,$$

and assuming that the matrix $\mathbf{W}_{n,T}$ is constructed using the drift bandwidths $g_{n,T}$, instead of $h_{n,T}$ we can write

$$\alpha_{n,T} = (\mathbf{X}'_{n,T}\mathbf{W}_{n,T}\mathbf{X}_{n,T})^{-1}\mathbf{X}'_{n,T}\mathbf{W}_{n,T}\mathbf{z}_{n,T}\ .$$

3 Results: the recurrent case.

Define

$$s_{n,T}^{k} = \sum_{i=0}^{n-1} \frac{\Delta_{n,T}}{h_{n,T}} \left(\frac{X_{i\Delta_{n,T}} - x}{h_{n,T}} \right)^{k} K\left(\frac{X_{i\Delta_{n,T}} - x}{h_{n,T}} \right) ,$$

$$t_{n,T}^{k} = \sum_{i=0}^{n-1} \frac{\Delta_{n,T}}{h_{n,T}} \left(\frac{X_{i\Delta_{n,T}} - x}{h_{n,T}} \right)^{k} \times K\left(\frac{X_{i\Delta_{n,T}} - x}{h_{n,T}} \right) \frac{1}{\Delta_{n,T}} \left(X_{(i+1)\Delta_{n,T}} - X_{i\Delta_{n,T}} \right)^{2} ,$$

and

$$\mathbf{S}_{n,T} = \begin{bmatrix} s_{n,T}^{0} & \dots & s_{n,T}^{p} \\ \dots & & \dots \\ s_{n,T}^{p} & \dots & s_{n,T}^{2p} \end{bmatrix}, \ \mathbf{t}_{n,T} = \begin{bmatrix} t_{n,T}^{0} \\ \dots \\ t_{n,T}^{p} \end{bmatrix}, \ \mathbf{H} = \text{diag}\left(1, \dots, h^{p}\right) .$$

Then, $\beta_{n,T}$ can be written as

$$\beta_{n,T} = \mathbf{H}^{-1} \mathbf{S}_{n,T}^{-1} \mathbf{t}_{n,T} .$$

Theorem 3. *Under Assumptions 1-2, if $n \to \infty$, $T = \overline{T} < \infty$, $\Delta_{n,\overline{T}} = T/n \to 0$, and $h_{n,\overline{T}} \to 0$, as $n \to \infty$, such that $\left(\Delta_{n,\overline{T}} \log\left(1/\Delta_{n,\overline{T}}\right)\right)^{1/2} / h_{n,\overline{T}} = o_{a.s.}(1)$, then $\mathbf{S}_{n,\overline{T}} \overset{a.s.}{\to} \mathbf{S}\ \overline{L}_X(\overline{T}, x)$, where*

$$\mathbf{S} = \begin{bmatrix} s^{0} & \dots & s^{p} \\ \dots & & \dots \\ s^{p} & \dots & s^{2p} \end{bmatrix}, \ s^{k} = \int u^{k} K(u) du.$$

The results in this paper apply for vanishing sequences of sampling intervals bounded above by $1/e$: $\Delta_{n,T} < 1/e$. Under this convention, tight bounds for the local increments of the diffusion can be obtained using the modulus of continuity of Brownian motion. Sharp asymptotic convergence rates for the bias terms are also obtained. This is done without loss of generality because other values of $\Delta_{n,T}$ can be accomodated by properly rescaling the infinitesimal parameters of the diffusion.

We now present a long-span analog to the previous theorem.

Corollary 1. *Under Assumptions 1-2, if* $T \to \infty$, $n \to \infty$, $\Delta_{n,T} \to 0$ *and* $\left(\Delta_{n,T} \log\left(1/\Delta_{n,T}\right)\right)^{1/2} \overline{L}_X(T,x)/\ h_{n,T} = o_{a.s.}(1)$ *then* $\mathbf{S}_{n,T} \overset{a.s.}{\to} \mathbf{S}\overline{L}_X(\sup\{t : X_t = x\}, x)$. *Since the diffusion process is recurrent, local time at* x *diverges* $\overline{L}_X(\sup\{t : X_t = x\}, x) = \infty$ *almost surely.*

The last corollary is an extension of Theorem 1 in Bandi and Phillips [3]. Now, let

$$\begin{aligned} t_{n,T}^{*k} &= \sum_{i=0}^{n-1} \frac{\Delta_{n,T}}{h_{n,T}} \left(\frac{X_{i\Delta_{n,T}} - x}{h_{n,T}}\right)^k K\left(\frac{X_{i\Delta_{n,T}} - x}{h_{n,T}}\right) \\ &\times \left[\frac{1}{\Delta_{n,T}} \left(X_{(i+1)\Delta_{n,T}} - X_{i\Delta_{n,T}}\right)^2 - \sigma^2\left(X_{i\Delta_{n,T}}\right)\right] \end{aligned}$$

$$\mathbf{t}_{n,T}^{*} = \begin{bmatrix} t_{n,T}^{*0} \\ \dots \\ t_{n,T}^{*p} \end{bmatrix}$$

note that

$$\begin{aligned} \sigma^2\left(X_{i\Delta_{n,T}}\right) &= \sum_{\nu=0}^{p} \frac{h_{n,T}^{\nu}}{\nu!} \sigma^{2(\nu)}(x) \left(\frac{X_{i\Delta_{n,T}} - x}{h_{n,T}}\right)^{\nu} \\ &+ \frac{h_{n,T}^{p+1}}{(p+1)!} \sigma^{2(p+1)}(\widetilde{x}_i) \left(\frac{X_{i\Delta_{n,T}} - x}{h_{n,T}}\right)^{p+1} \end{aligned}$$

where $\widetilde{x}_i \in \left(X_{i\Delta_{n,T}}, x\right)$. Then we can write

$$\mathbf{H}\left(\beta_{n,T} - \beta\right) = \mathbf{S}_{n,T}^{-1} \mathbf{t}_{n,T}^{*} + O_{a.s.}\left(h_{n,T}^{p+1}\right) \mathbf{1}^{p+1}\ , \tag{5}$$

where $\mathbf{1}^{p+1}$ is a $(p+1)-$vector of ones. The derivation of equation (5) is in section 5. Approximate closed form expressions for equation (5) can be obtained after lengthy calculations as in the proof of Theorem 3 in Moloche [14], and Bandi and Moloche [2]. These calculations involve mathematical arguments not used elsewhere in this paper, so we will not include the proofs here. For $h_{n,T}$ small, the approximation can be written as:

$$\begin{aligned} &\mathbf{H}\left(\beta_{n,T} - \beta\right) = \\ &\quad \left(O_{a.s.}\left(\left(\Delta_{n,T} \log\left(1/\Delta_{n,T}\right)\right)^{1/2}\right) + O_{a.s.}\left(\Delta_{n,T}/h_{n,T}\right)\right. \\ &\quad \left.+ O_{a.s.}\left(\Delta_{n,T}/\left(h_{n,T}\overline{L}_X(T,x)\right)\right)\right) \mathbf{1}^{p+1} \\ &\quad + \mathbf{S}^{-1} \frac{h_{n,T}^{p+1}}{(p+1)!} \mathbf{m}\ , \end{aligned} \tag{6}$$

where $m^k = \int u^{k+p+1}K(u)du\sigma^{2(p+1)}(x) + o_{a.s.}(1)$, if $p-k$ is odd, and $m^k = h_{n,T}\int u^{k+p+2}K(u)du\left[\sigma^{2(p+1)}(x)m'(x)/m(x) + \sigma^{2(p+2)}(x)/(p+2)\right] + o_{a.s.}(h_{n,T})$, if $p-k$ is even. The expression above makes clear the smaller bias of local linear Kernel estimators in comparison with simple Kernel estimators. Both bias terms are of order $h_{n,T}^2$ but the simple Kernel estimator has an additional term that depends on the invariant measure, cf. Bandi and Phillips [3]. The difference in the terms for $p-k$ odd and even arises from the vanishing odd moments of symmetric Kernels. The bias has clearly two parts, one arising from time discretization and the other one arising from the nonlinearity of the diffusion function. For example, assuming that the discretization bias is negligible, the bias term of the local linear Kernel estimator can be approximated by

$$\frac{h_{n,T}^2}{2}\int u^2K(u)du\sigma^{2''}(x) \ .$$

Note also than when higher order derivatives are zero, that is, when the diffusion function is of polynomial order, the bias term vanishes, and bandwidths can explode without affecting consistency. The following theorem presents a set of sufficient conditions for consistency.

Theorem 4. *Under Assumptions 1-2, if* $T \to \infty$*,* $n \to \infty$*,* $h_{n,T} \to 0$*,* $\Delta_{n,T} \to 0$ *and* $\left(\Delta_{n,T}\log\left(1/\Delta_{n,T}\right)\right)^{1/2}\,\overline{L}_X(T,x)/h_{n,T} = o_{a.s.}(1)$*. then* $\mathbf{H}\left(\beta_{n,T} - \beta\right) \overset{a.s.}{\to} 0$*. If the diffusion function has derivatives of order higher than p equal to zero:* $\sigma^{2(\nu)}(x) = 0$*;* $\nu \geq (p+1)$*, then the last statement is also valid when* $h_{n,T} \to \infty$*.*

Since bandwidths eliminate the bias asymptotically by converging to zero, when the bias is unexistent, bandwidths do not have to vanish and in fact, they can explode. Unlike the regression case, cf. Park and Phillips [15], Moloche [14], exploding bandwidths do not require to change additional conditions because the estimators proposed in this paper are obtained by differencing.

Once consistency has been verified, we derive the asymptotic distribution. Define the matrix

$$\mathbf{R} = \begin{bmatrix} R^0 & \dots & R^p \\ \dots & & \dots \\ R^p & \dots & R^{2p} \end{bmatrix},\ R^k = \int u^kK^2(u)du \ .$$

Theorem 5. *(CLT for the diffusion function estimator) Under Assumptions 1-2, if* $T \to \infty$, $n \to \infty$, $\Delta_{n,T} \to 0$, $h_{n,T} \to 0$, $\left(\Delta_{n,T} \log\left(1/\Delta_{n,T}\right)\right)^{1/2} \overline{L}_X(T,x)/h_{n,T} = o_{a.s.}(1)$ *and* $h_{n,T}^{2p+3}\overline{L}_X(T,x)/\ \Delta_{n,T} = o_{a.s.}(1)$ *then*

$$\sqrt{\frac{h_{n,T}^{2\nu+1}\overline{L}_X(T,x)}{\Delta_{n,T}\nu!}}\left[\sigma_{n,T}^{2(\nu)}(x) - \sigma^{2(\nu)}(x)\right] \xrightarrow{d} N\left(0, 4s^{\nu,\nu}\sigma^4(x)\right),\ x \in \mathbb{R}\ ,$$

$\nu = 0, ..., p$, *where* N *denotes a Normal distribution and* $\{s^{ij}\}_{ij} = \mathbf{S}^{-1}\mathbf{R}\mathbf{S}^{-1}$. *If the diffusion function has derivatives of order higher than* p *equal to zero:* $\sigma^{2(\nu)}(x) = 0$; $\nu \geq (p+1)$, *then the last statement is also valid when* $h_{n,T} \to \infty$.

Theorem 5 is also valid when $T < \infty$. See Bandi and Phillips [3]. The advantage of the approach described here with respect to conventional kernel estimators is as mentioned above, that bias terms have a faster rate of convergence to zero.

When $T < \infty$ the drift cannot be identified, in general. In fact, kernel estimators for the drift diverge when $T < \infty$, see Bandi [1].

Next, we describe the asymptotic properties of the functional estimator for the drift function when $T \to \infty$. Assume in what follows that the matrix $\mathbf{S}_{n,T}$ was constructed using the drift bandwidths $g_{n,t}$, instead of $h_{n,t}$. Define

$$\begin{aligned} u_{n,T}^k &= \sum_{i=0}^{n-1} \frac{\Delta_{n,T}}{g_{n,T}} \left(\frac{X_{i\Delta_{n,T}} - x}{g_{n,T}}\right)^k \\ &\quad \times K\left(\frac{X_{i\Delta_{n,T}} - x}{g_{n,T}}\right) \frac{1}{\Delta_{n,T}} \left(X_{(i+1)\Delta_{n,T}} - X_{i\Delta_{n,T}}\right)\ , \end{aligned}$$

and

$$\mathbf{u}_{n,T} = \begin{bmatrix} u_{n,T}^0 \\ \cdots \\ u_{n,T}^p \end{bmatrix} .$$

Then, $\alpha_{n,T}$ can be written as

$$\alpha_{n,T} = \mathbf{H}^{-1}\mathbf{S}_{n,T}^{-1}\mathbf{u}_{n,T}\ .$$

Now, let

$$\begin{aligned} u_{n,T}^{*k} &= \sum_{i=0}^{n-1} \frac{\Delta_{n,T}}{g_{n,T}} \left(\frac{X_{i\Delta_{n,T}} - x}{g_{n,T}}\right)^k K\left(\frac{X_{i\Delta_{n,T}} - x}{g_{n,T}}\right) \\ &\quad \times \left[\frac{1}{\Delta_{n,T}}(X_{(i+1)\Delta_{n,T}} - X_{i\Delta_{n,T}}) - \mu(X_{i\Delta_{n,T}})\right], \end{aligned}$$

$$\mathbf{u}_{n,T}^{*}=\begin{bmatrix} u_{n,T}^{*0} \\ \dots \\ u_{n,T}^{*p} \end{bmatrix}$$

note that

$$\mu(X_{i\Delta_{n,T}})=\sum_{\nu=0}^{p}\frac{g_{n,T}}{\nu!}\mu^{(\nu)}(x)\left(\frac{X_{i\Delta_{n,T}}-x}{g_{n,T}}\right)^{\nu}$$
$$+\frac{g_{n,T}^{p+1}}{(p+1)!}\mu^{(p+1)}(\widetilde{x}_i)\left(\frac{X_{i\Delta_{n,T}}-x}{g_{n,T}}\right)^{p+1}$$

where $\widetilde{x}_i\in\left(X_{i\Delta_{n,T}},x\right)$. Then we can write

$$\mathbf{H}\left(\alpha_{n,T}-\alpha\right)=\mathbf{S}_{n,T}^{-1}\mathbf{u}_{n,T}^{*}+O_{a.s.}\left(g_{n,T}^{p+1}\right)\mathbf{1}^{p+1}\ .$$

Approximate closed form expressions for the last equation are available:

$$\begin{aligned}\mathbf{H}\left(\alpha_{n,T}-\alpha\right)&=\\ &\left(O_{a.s.}\left(\left(\Delta_{n,T}\log\left(1/\Delta_{n,T}\right)\right)^{1/2}\right)+O_{a.s.}\left(\Delta_{n,T}/g_{n,T}\right)\right.\\ &\left.+O_{a.s.}\left(1/\left(g_{n,T}\overline{L}_X(T,x)\right)\right)+O_{a.s.}\left(g_{n,T}^{p+2}\right)\right)\mathbf{1}^{p+1}\\ &+\mathbf{S}^{-1}\frac{g_{n,T}^{p+1}}{(p+1)!}\mathbf{m}\ ,\end{aligned}\tag{7}$$

where $m^k=\int u^{k+p+1}K(u)du\mu^{(p+1)}(x)+o_{a.s.}(1)$, if $p-k$ is odd, and $m^k=g_{n,T}\int u^{k+p+2}K(u)du\left[\mu^{(p+1)}(x)m'(x)/m(x)+\mu^{(p+2)}(x)/(p+2)!\right]+o_{a.s.}(g_{n,T})$, if $p-k$ is even, here. The bias term of the drift estimator can be approximated by

$$\frac{g_{n,T}^{2}}{2}\int u^2K(u)du\mu''(x)\ ,$$

for the local linear Kernel estimator. Note also than when higher order derivatives are zero, that is, when the drift function is of polynomial order, the bias term vanishes, and bandwidths can explode without affecting consistency.

Theorem 6. *Under Assumptions 1-2, if* $T\to\infty$, $n\to\infty$, $g_{n,T}\to 0$, $\Delta_{n,T}\to 0$, $\left(\Delta_{n,T}\log\left(1/\Delta_{n,T}\right)\right)^{1/2}\overline{L}_X(T,\,x)/g_{n,T}=o_{a.s.}(1)$, *and* $g_{n,T}\overline{L}_X(T,x)\to\infty$ *then* $\mathbf{H}\left(\alpha_{n,T}-\alpha\right)\to 0$. *If the drift function has derivatives of order higher than p equal to zero:* $\mu^{2(\nu)}(x)=0$; $\nu\geq(p+1)$, *then the last statement is also valid when* $g_{n,T}\to\infty$.

Theorem 7. *(CLT for the drift function estimator) Under Assumptions 1-2, if* $T \to \infty$, $n \to \infty$, $\Delta_{n,T} \to 0$, $g_{n,T} \to 0$, $(\Delta_{n,T} \log(1/\Delta_{n,T}))^{1/2} \overline{L}_X(T,x)/g_{n,T} = o_{a.s.}(1)$ *and* $g_{n,T}^{2p+3}\overline{L}_X(T,x) \to \infty$ *then*

$$\sqrt{\frac{g_{n,T}^{2\nu+1}\overline{L}_X(T,x)}{\nu!}}\left[\mu_{n,T}^{(\nu)}(x) - \mu^{(\nu)}(x)\right] \xrightarrow{d} N\left(0, s^{\nu,\nu}\sigma^2(x)\right), \, x \in \mathbb{R},$$

$\nu = 0, ..., p$, *where* N *denotes a Normal distribution and* $\{s^{ij}\}_{ij} = \mathbf{S}^{-1}\mathbf{R}\mathbf{S}^{-1}$. *If the drift function has derivatives of order higher than* p *equal to zero:* $\mu^{2(\nu)}(x) = 0$; $\nu \geq (p+1)$, *then the last statement is also valid when* $g_{n,T} \to \infty$.

The a.s rate of convergence of the estimators are characterized in terms of local times. We have mentioned in the discussion following Theorem 2 that the growth rates of local times cannot be characterized deterministically when the process is null recurrent. Nonetheless, expected growth rates always exist and they are deterministic, and in this sense, rates of convergence are always slower when the process is null recurrent. These rates can be characterized with slowly varying functions. Actual rates for null recurrent diffusions will always be slower than for positive recurrent diffusions, but they depend on the state of nature. Almost sure rates of convergence can be precisely characterized only when the invariant probability measure exists.

4 The positive recurrent case.

Theorems 4-7 and Corollary 1 can be strengthened when the process X is positive recurrent.

Assumption 3. *The speed measure* $m(dx)$ *is integrable:*

$$\int_I m(dx) < \infty$$

Under this assumption, the process is positive recurrent and, when initialized with the invariant probability measure, stationary. The density of the invariant probability measure $f(x)$ is proportional to the speed measure

$$f(x) = m(x)/m(I), \, x \in I.$$

The estimation of stationary scalar diffusions is usually done under mixing assumptions or weakly dependence. We do not need to impose further mixing constraints other than those already implied by the recurrence of the process.

Let us present the results. By Theorem 1, positive recurrence is sufficient to set $\overline{L}_X(T,x)/T = O_{a.s.}(1)$. If the process X is initialized with the invariant probability measure, the quantity $\overline{L}_X(T,x)/T$ is also an unbiased estimator of $f(x)$. This case is analized by Kutoyants [13].

The Kernel density estimator from an equally spaced discrete sample is

$$f_{n,T}(x) = \frac{\Delta_{n,T}}{Tb_{n,T}} \sum_{i=0}^{n} K\left(\frac{X_{i\Delta_{n,T}} - x}{b_{n,T}}\right) = \frac{1}{nb_{n,T}} \sum_{i=0}^{n} K\left(\frac{X_{i\Delta_{n,T}} - x}{b_{n,T}}\right)$$

This estimator is consistent under the sufficient conditions of the following theorem.

Theorem 8. *Under Assumptions 1-3, if* $T \to \infty$, $n \to \infty$, $\Delta_{n,T} \to 0$, *and* $b_{n,T} \to 0$ *with* $\left(\Delta_{n,T} \log\left(1/\Delta_{n,T}\right)\right)^{1/2}/b_{n,T} \to 0$ *then*

$$f_{n,T}(x) \overset{a.s.}{\to} f(x).$$

Counterintuitively, the density estimator vanishes in the case of fat-tailed nonintegrable invariant measures as the time span and the sample frequency goes to infinity:

Corollary 2. *Under Assumptions 1-2, if Assumption 3 is not satisfied, and if* $T \to \infty$, $n \to \infty$, $\Delta_{n,T} \to 0$, *and* $b_{n,T} \to 0$ *with* $\left(\Delta_{n,T} \log\left(1/\Delta_{n,T}\right)\right)^{1/2}/b_{n,T} \to 0$ *then*

$$f_{n,T}(x) \overset{a.s.}{\to} 0.$$

The results in the previous section and Theorem 1 imply directly the following:

Corollary 3. *Under Assumptions 1-3, if* $T \to \infty$, $n \to \infty$, $\Delta_{n,T} \to 0$, $\left(\Delta_{n,T} \log\left(1/\Delta_{n,T}\right)\right)^{1/2}/h_{n,T} = o_{a.s.}(1)$ *then* $\mathbf{S}_{n,T}/T \overset{a.s.}{\to} \mathbf{S}f(x)$.

Corollary 4. *Under Assumptions 1-3, if* $T \to \infty$, $n \to \infty$, $\Delta_{n,T} \to 0$, $h_{n,T} \to 0$, $\left(\Delta_{n,T} \log\left(1/\Delta_{n,T}\right)\right)^{1/2} T/\ h_{n,T} = o_{a.s.}(1)$ *then* $\mathbf{H}\left(\beta_{n,T} - \beta\right) \overset{a.s.}{\to} 0$. *If the diffusion function has derivatives of order higher than p equal to zero:* $\sigma^{2(\nu)}(x) = 0$; $\nu \geq (p{+}1)$, *then the last statement is also valid when* $h_{n,T} \to \infty$.

Corollary 5. *(CLT for the diffusion function estimator) Under Assumptions 1-3, if* $T \to \infty$, $n \to \infty$, $\Delta_{n,T} \to 0$, $h_{n,T} \to 0$, $\left(\Delta_{n,T} \log\left(1/\Delta_{n,T}\right)\right)^{1/2} T/ h_{n,T} = o_{a.s.}(1)$ *and* $h_{n,T}^{2p+3} n \to \infty$ *then*

$$\sqrt{h_{n,T}^{2\nu+1} n}\left[\sigma_{n,T}^{2(\nu)}(x) - \sigma^{2(\nu)}(x)\right] \xrightarrow{d} N\left(0, 4\nu! s^{\nu,\nu} \sigma^4(x)/f(x)\right),$$

$\nu = 0, ..., p$, *where* $\{s^{ij}\}_{ij} = \mathbf{S}^{-1}\mathbf{R}\mathbf{S}^{-1}$. *If the diffusion function has derivatives of order higher than p equal to zero:* $\sigma^{2(\nu)}(x) = 0$; $\nu \geq (p+1)$, *then the last statement is also valid when* $h_{n,T} \to \infty$.

Note that the rate of convergence is deterministic and no longer path dependent, and that the asymptotic distribution for the estimation error $\sigma_{n,T}^{2(\nu)}(x) - \sigma^{2(\nu)}(x)$ is Normal and no longer mixed Normal. This is hardly surprising, since according to Theorem 1, when the diffusion is positive recurrent, mean local times have a positive "degenerate" limit. Corollary 4 is valid also if $T < +\infty$, as long as the joint conditions on $\Delta_{n,T}$ and $h_{n,T}$ are satisfied. Of course, practical implementation of the corollaries above requires that $T \to +\infty$, since otherwise the density cannot be estimated consistently. This is not a problem for the general asymptotic theory in the previous section because it does not require to estimate the density of the invariant probability measure. For the drift we have

Corollary 6. *Under Assumptions 1-3, if* $T \to \infty$, $n \to \infty$, $\Delta_{n,T} \to 0$, $g_{n,T} \to 0$, $\left(\Delta_{n,T} \log\left(1/\Delta_{n,T}\right)\right)^{1/2} T/\ g_{n,T} = o_{a.s.}(1)$ *then* $\mathbf{H}\left(\alpha_{n,T} - \alpha\right) \xrightarrow{a.s.} 0$. *If the drift function has derivatives of order higher than p equal to zero:* $\mu^{2(\nu)}(x) = 0$; $\nu \geq (p+1)$, *then the last statement is also valid when* $g_{n,T} \to \infty$.

Corollary 7. *(CLT for the drift function estimator) Under Assumptions 1-3, if* $T \to \infty$, $n \to \infty$, $\Delta_{n,T} \to 0$, $g_{n,T} \to 0$, $\left(\Delta_{n,T} \log\left(1/\Delta_{n,T}\right)\right)^{1/2} T/g_{n,T} = o_{a.s.}(1)$ *and* $g_{n,T}^{2p+3} T \to \infty$ *then*

$$\sqrt{g_{n,T}^{2\nu+1} T}\left[\mu_{n,T}^{(\nu)}(x) - \mu^{(\nu)}(x)\right] \xrightarrow{d} N\left(0, \nu! s^{\nu,\nu} \sigma^2(x)/f(x)\right),$$

$\nu = 0, ..., p$, *where N denotes a Normal distribution and* $\{s^{ij}\}_{ij} = \mathbf{S}^{-1}\mathbf{R}\mathbf{S}^{-1}$. *If the drift function has derivatives of order higher than p equal to zero:* $\mu^{2(\nu)}(x) = 0$; $\nu \geq (p+1)$, *then the last statement is also valid when* $g_{n,T} \to \infty$.

Finally, we include here a central limit theorem for the kernel density estimator, when the process is positive recurrent but possibly nonstationary. It is derived using a second order limit theorem for occupation times of positive recurrent processes due to Tanaka [17].

Theorem 9. *Under Assumptions 1-3, if $T \to \infty$, $n \to \infty$, $\Delta_{n,T} \to 0$, $b_{n,T} \to 0$, and $\left(T\Delta_{n,T}\log(1/\Delta_{n,T})\right)^{1/2}/b_{n,T} = o_{a.s.}(1)$ then*

$$\sqrt{T}\left(f_{n,T}(x) - f(x)\right) \overset{d}{\to} N(0,\omega)$$

where

$$\omega = 4m\left(I\right) f^2\left(x\right) \int \left(1_{x \leq y} - F\left(y\right)\right)^2 p\left(y\right) dy.$$

The driving term in the previous result is the difference between the true local time scaled by total time and the stationary density. Under the conditions of the theorem, this term always dominates asymptotically the kernel estimation error for local time, the spatial discretization error and the time discretization error. Note that the limit theory for the estimation error is different with respect to analog results derived in a stationary discrete time context. The differences arise from two sources. The first is the discretization error terms that vanish asymptotically under the imposed conditions on the bandwidth. The second is that in continuous time the occupation times measure is absolutely continuous with respect to Lebesgue measure allowing a faster rate of convergence to the true stationary density.

5 Proofs.

5.1 Proof of Theorem 1.

This is a special case of Theorem 6.8 in Itô and McKean [9]. Let e be an arbitrary non-negative measure. Define

$$\overline{L}_X^e(T) = \int \overline{L}_X(T,a)\, e(a)\, da.$$

Denote by T_a the first passage or hitting time to $a \in I$:

$$T_a(\omega) = \inf(t \geq 0 : X_t = a).$$

Let $T^1 < T^2 < \ldots$ the succesive passage times to a via b, some $a, b \in I$. They are defined recursively by:

$$\begin{aligned} T^n &= T^{n-1} + \widehat{T}\left(\theta_{T^{n-1}} \circ \omega\right) \\ T^0 &= T_a(\omega), \end{aligned}$$

where θ is the shift operator and $\widehat{T}(\omega) = T_b + T_a\left(\theta_{T_b} \circ \omega\right)$ is an operator that finds the first passage to a via b. By the strong Markov property, the excursions $\{X_t : T^{n-1} \leq t \leq T^n\}$, $n \geq 1$ are independent and identical in law. Then so are the increments of $\overline{L}_X^e$

$$\overline{L}_X^e(T^n) - \overline{L}_X^e\left(T^{n-1}\right), \; n \geq 1.$$

using a decomposition of the Green function of the diffusion killed at an exit time (see proof of Theorem 14, below) it can be shown that

$$E_a\left[\overline{L}_X\left(\widehat{T}(\omega), \xi\right)\right] = 2\left(p(b) - p(a)\right) m(\xi), \; \xi \in I$$

invoking the strong law of large numbers, it is clear that

$$P_a\left[\lim_{n\to\infty} \frac{\overline{L}_X^e(T^n)}{n} = 2\left[p(b) - p(a)\right] \int e(x) m(dx)\right] = 1$$

and

$$P_a\left[\lim_{T\to\infty} \frac{\overline{L}_X^e(T)}{\min(n : T^n \geq T)} = 2\left[p(b) - p(a)\right] \int e(x) m(dx)\right] = 1$$

implying

$$P.\left[\lim_{T\to\infty}\frac{\overline{L}_X^{e_1}(T)}{\overline{L}_X^{e_2}(T)}=\frac{\int e_1(x)\,m(dx)}{\int e_2(x)\,m(dx)}\right]=1. \tag{8}$$

In particular

$$P.\left[\lim_{T\to\infty}\frac{\overline{L}_X^{\delta(a)}(T)}{\overline{L}_X^{\lambda}(T)}=\frac{m(a)}{M(I)}\right]=1$$

$$P.\left[\lim_{T\to\infty}\frac{\overline{L}_X^{\delta(a)}(T)}{\overline{L}_X^{\delta(b)}(T)}=\frac{m(a)}{m(b)}\right]=1$$

where $\delta(a)$ is the Dirac measure at a, and λ is Lebesgue measure; then note that $\overline{L}_X^{\delta(a)}(T)=\overline{L}_X(T,a)$ and $\overline{L}_X^{\lambda}(T)=T$. But since $\lim_{T\to\infty}\frac{\overline{L}_X(T,a)}{T}=\frac{m(a)}{M(I)}$ a.s., then $\lim_{T\to\infty}\frac{E.\left[\overline{L}_X(T,a)\right]}{T}=\frac{m(a)}{M(I)}$.

5.2 Proof of Theorem 2.

Let us begin by obtaining a spectral decomposition for the Green's function of possibly nonstationary diffusions. This decomposition will allow us to obtain rates for the Laplace transform of occupation times by applying a theorem of Kasahara, et.al. [12]. Then the result follows by applying the classic Karamata's Tauberian theorem, cf. Feller [6]. See also Zirbel [19].

Let $\mathcal{A}$ be the second order differential operator, that is $\mathcal{A}f=\mu f'+\frac{1}{2}\sigma^2 f''$, where f is a function in the domain of $\mathcal{A}$. It is enough for our purposes to deal only with $\mathcal{C}_b^{2+}$, the bounded, positive, continuous, twice differentiable functions. All these functions are in the domain of $\mathcal{A}$. Further, $\mathcal{A}$ is also valued in $\mathcal{C}_b^{2+}$.

As in Itô and McKean (1965) let us introduce two functions $\tilde{\psi}_\alpha$ and $\tilde{\varphi}_\alpha$ solving

$$\mathcal{A}u=\alpha u, \tag{9}$$

for some constant α such that $\tilde{\psi}_\alpha$ is strictly increasing and $\tilde{\varphi}_\alpha$ is strictly decreasing, with boundary conditions $\lim_{x\downarrow l}\tilde{\psi}_\alpha(x)=0$ and $\lim_{x\uparrow r}\tilde{\varphi}_\alpha(x)=0$. The solutions are then unique up to scale. These functions are linearly independent and all solutions of problem (9) can be expressed as their linear combinations.

Let us define the Wronskian

$$\begin{aligned}\tilde{W}_\alpha &= \tilde{\psi}_\alpha^+ (x) \, \tilde{\varphi}_\alpha (x) - \tilde{\psi}_\alpha (x) \, \tilde{\varphi}_\alpha^+ (x) \\ &= \tilde{\psi}_\alpha^- (x) \, \tilde{\varphi}_\alpha (x) - \tilde{\psi}_\alpha (x) \, \tilde{\varphi}_\alpha^- (x) \, ,\end{aligned}$$

where u^+ and u^- denote the right and left derivative of u with respect to the scale function p:

$$\begin{aligned}u^+ (x) &= \lim_{\varepsilon \downarrow 0} \frac{u (x + \varepsilon) - u (x)}{p (x + \varepsilon) - p (x)}, \\ u^- (x) &= \lim_{\varepsilon \downarrow 0} \frac{u (x) - u (x - \varepsilon)}{p (x) - p (x - \varepsilon)}.\end{aligned}$$

The Wronskian is constant as a function of x.

Green's function satisfies

$$G_\alpha (x, y) = \int_0^\infty e^{-\alpha s} p_s (x, y) \, ds,$$

where p is the transition density with respect to the speed measure. It possesses a decomposition in terms of $\tilde{\psi}_\alpha$ and $\tilde{\varphi}_\alpha$. In fact, Green's function is defined according to Itô and McKean (1965) as:

$$G_\alpha (x, y) = \begin{cases} \tilde{\psi}_\alpha (x) \, \tilde{\varphi}_\alpha (y) \, / \tilde{W}_\alpha, & x \leq y, \\ \tilde{\psi}_\alpha (y) \, \tilde{\varphi}_\alpha (x) \, / \tilde{W}_\alpha, & x \geq y. \end{cases}$$

This decomposition can be reformulated in terms of expected exponential hitting times, which are more intuitive and easier to analyze. Since

$$E_x \left[e^{-\alpha T_c} \right] = \begin{cases} \tilde{\psi}_\alpha (x) \, / \tilde{\psi}_\alpha (c) \, , & x \leq c, \\ \tilde{\varphi}_\alpha (x) \, / \tilde{\varphi}_\alpha (x) \, , & x \geq c, \end{cases}$$

defining

$$\begin{aligned}\psi_\alpha (x) &= \begin{cases} E_x \left[e^{-\alpha T_c} \right], & x \leq c, \\ 1/E_c \left[e^{-\alpha T_x} \right], & x > c, \end{cases} \\ \varphi_\alpha (x) &= \begin{cases} 1/E_c \left[e^{-\alpha T_x} \right], & x \leq c, \\ E_x \left[e^{-\alpha T_c} \right], & x > c, \end{cases}\end{aligned}$$

along with $W_\alpha = \tilde{W}_\alpha / \left(\tilde{\psi}_\alpha (c) \, \tilde{\varphi}_\alpha (c) \right)$, imply that

$$G_\alpha (x, y) = \psi_\alpha (x \wedge y) \, \varphi_\alpha (x \vee y) \, / W_\alpha,$$

We now characterize the behavior of Green's function as $\alpha \downarrow 0$. Since the functions ψ_α and φ_α solve (9), it is clear that as $\alpha \downarrow 0$, they become hitting time probabilities:

$$\begin{aligned}
\lim_{\alpha \downarrow 0} \psi_\alpha(x) &= \lim_{b \downarrow l} P_x[T_a < T_b],\\
\lim_{\alpha \downarrow 0} \varphi_\alpha(x) &= \lim_{b \uparrow r} P_x[T_a < T_b].
\end{aligned}$$

Then, by using well known formulas for expected hitting times, we can evaluate the limits explicitly, recalling that for recurrent diffusions, the scale function diverges at the boundaries:

$$\begin{aligned}
\lim_{\alpha \downarrow 0} \psi_\alpha(x) &= \frac{p(x) - p(l)}{p(r) - p(l)} = 1,\\
\lim_{\alpha \downarrow 0} \varphi_\alpha(x) &= \frac{p(x) - p(r)}{p(l) - p(r)} = 1.
\end{aligned}$$

Recurrence implies that the Wronskian converges to zero, $\lim_{\alpha \downarrow 0} W_\alpha = 0$, thus:

$$\lim_{\alpha \downarrow 0} W_\alpha = 0,\; x \le y$$

$$\lim_{\alpha \downarrow 0} W_\alpha = 0,\; y \le x.$$

It is clear now that the rate of divergence of the Green's function is completely determined by the Wronskian. The latter can be decomposed as

$$W(\alpha) = W_+(\alpha) + W_-(\alpha), \tag{10}$$

where

$$\begin{aligned}
W_+(\alpha)^{-1} &= \lim_{x \uparrow r} \frac{\psi_\alpha(x)}{\varphi_\alpha(x)},\\
W_-(\alpha)^{-1} &= -\lim_{x \downarrow l} \frac{\psi_\alpha(x)}{\varphi_\alpha(x)}.
\end{aligned}$$

The rates for W_+ and W_- can be fully characterized by the scale function and the speed measure. Theorem 2 in Kasahara, et.al. [12], relates the behavior at the boundaries of the speed measure to the convergence rates of W_+ and W_- for a diffusion process in the natural scale. The extension of this theorem for the general case is straightforward, see Zirbel

[19]. Then we can restate that theorem as follows: if h varies regularly with exponent $\delta \in (0,1)$, then $\lim_{a \uparrow r} h\left(p\left(a\right)m\left(c,a\right)\right)/p\left(a\right) = C_1$ if and only if $\lim_{\alpha \downarrow 0} W_+\left(\alpha\right)h\left(1/\alpha\right) = C_2$ and $\lim_{a \downarrow l} h\left(-p\left(a\right)m\left([a,c]\right)\right)/p\left(c\right) = C_3$ if and only if $\lim_{\alpha \downarrow 0} W_-\left(\alpha\right)h\left(1/\alpha\right) = C_4$, where $C_1,...,C_4$ are constants different from zero, $c \in I$. The h-functions obtained for each boundary are not necessarily identical, but they uniquely determine the rates for the Wronskian by using (10), and choosing the slowest one if they differ. Note also that if the Wronskian depends only on α, then h is necessarily regularly varying at infinity, see Darling and Kac [4], thus h exists and satisfy the stated conditions. In fact, the existence of h for all diffusions has been established by Kasahara [11].

The result follows by applying Karamata's Tauberian theorem, cf. Feller [6]. Suppose that the diffusion X is recurrent, then $E_\cdot\left[L_X\left(T,a\right)\right] = \int_0^T p_s\left(\cdot,a\right) ds$ is nondecreasing. Let $0 < c < \infty$ a constant. If $h\left(T\right)$ varies regularly at infinity with exponent $\delta \in (0,1)$, then $\int_0^T p_s\left(a,b\right)ds/h\left(T\right) \to c/\Gamma\left(1+\delta\right)$ as $T \to \infty$ if and only if $\int_0^\infty e^{-\alpha s}p_s\left(a,b\right)ds/h\left(1/\alpha\right) \to c$ as $\alpha \to \infty$ for all $a,b \in I$, where c is a constant that depends on b only. But by (8) the limit $c\left(b\right)/\Gamma\left(1+\delta\right)$ can be written as $D_X m\left(b\right)$, where D_X is a positive constant. Finally, the Tauberian theorem also states that the function $h\left(1/\alpha\right)$ found above is unique.

5.3 Proof of Theorem 3.

We prove convergence of an arbitrary element $s^k_{n,T}$, $k \in \{0,...,p\}$. First, note that

$$
\begin{aligned}
&\int_0^{\overline{T}} \frac{1}{h_{n,T}}\left(\frac{X_s-x}{h_{n,\overline{T}}}\right)^k K\left(\frac{X_s-x}{h_{n,\overline{T}}}\right)ds \\
&= \int_0^{\overline{T}} \frac{1}{h_{n,T}}\left(\frac{X_s-x}{h_{n,\overline{T}}}\right)^k K\left(\frac{X_s-x}{h_{n,T}}\right)\frac{d\langle X_s\rangle}{\sigma^2(X_s)} \\
&= \int \frac{1}{h_{n,T}}\left(\frac{a-x}{h_{n,\overline{T}}}\right)^k K\left(\frac{a-x}{h_{n,\overline{T}}}\right)\frac{L_X(\overline{T},a)}{\sigma^2(a)}da \\
&= \int u^k K\left(u\right)\frac{L_X(\overline{T},uh_{n,\overline{T}}+x)}{\sigma^2(uh_{n,\overline{T}}+x)}du \\
&\overset{a.s.}{\to} \int_{-\infty}^{+\infty} u^k K\left(u\right)du\frac{L_X(\overline{T},x)}{\sigma^2(x)}
\end{aligned}
$$

$$= \int u^k K(u)\, du \overline{L}_X(\overline{T}, x)$$

where we used the quadratic variation of the diffusion, the occupation time formula and the change of variable $u = (a-x)/h_{n,\overline{T}}$. Define

$$\kappa_{n,T} = \max_{i \leq n} \sup_{i\Delta_{n,T} \leq s \leq (i+1)\Delta_{n,T}} \left| X_s - X_{i\Delta_{n,T}} \right|$$

then by the regularity conditions on the kernel function

$$\begin{aligned}
&\left| \frac{\Delta_{n,\overline{T}}}{h_{n,\overline{T}}} \sum_{i=0}^{n-1} \left(\frac{X_{i\Delta_{n,T}} - x}{h_{n,\overline{T}}} \right)^k K\left(\frac{X_{i\Delta_{n,T}} - x}{h_{n,\overline{T}}} \right) \right. \\
&\qquad \left. - \int_0^{\overline{T}} \frac{1}{h_{n,T}} \left(\frac{X_s - x}{h_{n,\overline{T}}} \right)^k K\left(\frac{X_s - x}{h_{n,\overline{T}}} \right) \right| \\
&= \left| \frac{1}{h_{n,\overline{T}}} \sum_{i=0}^{n-1} \int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}} \left[\left(\frac{X_{i\Delta_{n,T}} - x}{h_{n,\overline{T}}} \right)^k K\left(\frac{X_{i\Delta_{n,T}} - x}{h_{n,\overline{T}}} \right) - \left(\frac{X_s - x}{h_{n,\overline{T}}} \right)^k \right. \right. \\
&\qquad \left. \left. \times K\left(\frac{X_s - x}{h_{n,\overline{T}}} \right) \right] ds + \frac{\Delta_{n,\overline{T}}}{h_{n,\overline{T}}} \left(\frac{X_{n\Delta_{n,T}} - x}{h_{n,\overline{T}}} \right)^k K\left(\frac{X_{n\Delta_{n,T}} - x}{h_{n,\overline{T}}} \right) \right| \\
&\leq \left| \frac{1}{h_{n,\overline{T}}} \sum_{i=0}^{n-1} \int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}} \left[\left(\frac{X_{i\Delta_{n,T}} - x}{h_{n,\overline{T}}} \right)^k K\left(\frac{X_{i\Delta_{n,T}} - x}{h_{n,\overline{T}}} \right) - \left(\frac{X_s - x}{h_{n,\overline{T}}} \right)^k \right. \right. \\
&\qquad \left. \left. \times K\left(\frac{X_s - x}{h_{n,\overline{T}}} \right) \right] ds \right| + \left| \frac{\Delta_{n,\overline{T}}}{h_{n,\overline{T}}} \left(\frac{X_{n\Delta_{n,T}} - x}{h_{n,\overline{T}}} \right)^k K\left(\frac{X_{n\Delta_{n,T}} - x}{h_{n,\overline{T}}} \right) \right| \\
&\leq \left| \frac{1}{h_{n,\overline{T}}} \frac{\kappa_{n,\overline{T}}}{h_{n,\overline{T}}} \sum_{i=0}^{n-1} \int_{i\Delta_{n,\overline{T}}}^{(i+1)\Delta_{n,\overline{T}}} D^k \left(\frac{X_s - x}{h_{n,\overline{T}}}, \frac{\kappa_{n,\overline{T}}}{h_{n,\overline{T}}} \right) ds \right| + O_{a.s.}\left(\frac{\Delta_{n,\overline{T}}}{h_{n,\overline{T}}} \right) \\
&= \left| \frac{1}{h_{n,\overline{T}}} \frac{\kappa_{n,\overline{T}}}{h_{n,\overline{T}}} \int_0^{\overline{T}} D^k \left(\frac{X_s - x}{h_{n,\overline{T}}}, \frac{\kappa_{n,\overline{T}}}{h_{n,\overline{T}}} \right) ds \right| + O_{a.s.}\left(\frac{\Delta_{n,\overline{T}}}{h_{n,\overline{T}}} \right) \\
&= \left| \frac{1}{h_{n,\overline{T}}} \frac{\kappa_{n,\overline{T}}}{h_{n,\overline{T}}} \int D^k \left(\frac{a - x}{h_{n,\overline{T}}}, \frac{\kappa_{n,\overline{T}}}{h_{n,\overline{T}}} \right) \overline{L}_X(\overline{T}, a) da \right| + O_{a.s.}\left(\frac{\Delta_{n,\overline{T}}}{h_{n,\overline{T}}} \right) \\
&= \left| \frac{\kappa_{n,\overline{T}}}{h_{n,\overline{T}}} \int D^k \left(u, \frac{\kappa_{n,\overline{T}}}{h_{n,\overline{T}}} \right) \overline{L}_X(\overline{T}, u h_{n,\overline{T}} + x) du \right| + O_{a.s.}\left(\frac{\Delta_{n,\overline{T}}}{h_{n,\overline{T}}} \right) \\
&= O_{a.s.}\left(\frac{\left(\Delta_{n,\overline{T}} \log(1/\Delta_{n,\overline{T}}) \right)^{1/2}}{h_{n,\overline{T}}} \overline{L}_X(\overline{T}, x) \right) + O_{a.s.}\left(\frac{\Delta_{n,\overline{T}}}{h_{n,\overline{T}}} \right)
\end{aligned}$$

by the occupation time formula and the change of variable $u = (a - x)/h_{n,\overline{T}}$. By assumption $\left(\Delta_{n,\overline{T}} \log(1/\Delta_{n,\overline{T}})\right)^{1/2} / h_{n,\overline{T}} \overset{a.s.}{\to} 0$, so $\Delta_{n,\overline{T}}/h_{n,\overline{T}} \overset{a.s.}{\to} 0$. This, along with

$$\lim\sup_{\Delta_{n,\overline{T}}\downarrow 0} \frac{\kappa_{n,\overline{T}}}{\left(\Delta_{n,\overline{T}} \log(1/\Delta_{n,\overline{T}})\right)^{1/2}} < \infty \text{ a.s.}$$

implies that

$$\begin{aligned}\lim_{\Delta_{n,\overline{T}}\downarrow 0, h_{n,\overline{T}}\downarrow 0} \frac{\kappa_{n,\overline{T}}}{h_{n,\overline{T}}} &\leq \lim\sup_{\Delta_{n,\overline{T}}\downarrow 0} \frac{\kappa_{n,\overline{T}}}{\left(\Delta_{n,\overline{T}} \log(1/\Delta_{n,\overline{T}})\right)^{1/2}} \\ &\times \lim_{\Delta_{n,\overline{T}}\downarrow 0, h_{n,\overline{T}}\downarrow 0} \frac{\left(\Delta_{n,\overline{T}} \log(1/\Delta_{n,\overline{T}})\right)^{1/2}}{h_{n,\overline{T}}}\end{aligned}$$

thus

$$\frac{\kappa_{n,\overline{T}}}{h_{n,\overline{T}}} = O_{a.s.}\left(\frac{\left(\Delta_{n,\overline{T}} \log(1/\Delta_{n,\overline{T}})\right)^{1/2}}{h_{n,\overline{T}}}\right)$$

verifying the result.

5.4 Proof of Corollary 1.

Direct from the proof of Theorem 1 and Bandi and Phillips [3].

5.5 Proof of formula (5).

Recall that

$$\begin{aligned}\sigma^2(X_{i\Delta_{n,T}}) &= \sum_{\nu=0}^{p} \frac{h_{n,T}^{\nu}}{\nu!} \sigma^{2(\nu)}(x) \left(\frac{X_{i\Delta_{n,T}} - x}{h_{n,T}}\right)^{\nu} \\ &+ \frac{h_{n,T}^{p+1}}{(p+1)!} \sigma^{2(p+1)}(\widetilde{x}_i) \left(\frac{X_{i\Delta_{n,T}} - x}{h_{n,T}}\right)^{p+1}\end{aligned}$$

where $\widetilde{x}_i \in \left(X_{i\Delta_{n,T}}, x\right)$. Then we can write

$$\mathbf{H}\left(\beta_{n,T} - \beta\right) = \mathbf{S}_{n,T}^{-1}\mathbf{t}_{n,T}^{*} - \mathbf{H}\left(\mathbf{X}_{n,T}'\mathbf{W}_{n,T}\mathbf{X}_{n,T}\right)^{-1}\mathbf{H}\widetilde{\mathbf{u}}_{n,T} ,$$

where

$$\widetilde{\mathbf{u}}_{n,T} = \begin{bmatrix} \widetilde{u}_{n,T}^{0} \\ \dots \\ \widetilde{u}_{n,T}^{p} \end{bmatrix} ,$$

$$\widetilde{u}_{n,T}^{k} = \sum_{i=0}^{n-1} \frac{\Delta_{n,T}}{h_{n,T}} \left(\frac{X_{i\Delta_{n,T}} - x}{h_{n,T}}\right)^k K\left(\frac{X_{i\Delta_{n,T}} - x}{h_{n,T}}\right) \frac{h_{n,T}^{p+1}}{(p+1)!} \sigma^{2(p+1)}(\widetilde{x}_i).$$

This term is bounded by

$$\begin{aligned}
&\sum_{i=0}^{n-1} \frac{\Delta_{n,T}}{h_{n,T}} \left(\frac{X_{i\Delta_{n,T}} - x}{h_{n,T}}\right)^k K\left(\frac{X_{i\Delta_{n,T}} - x}{h_{n,T}}\right) \frac{h_{n,T}^{p+1}}{(p+1)!} \sigma^{2(p+1)}(\tilde{x}_i) \\
\leq\ & \left| \sum_{i=0}^{n-1} \frac{\Delta_{n,T}}{h_{n,T}} \left(\frac{X_{i\Delta_{n,T}} - x}{h_{n,T}}\right)^k K\left(\frac{X_{i\Delta_{n,T}} - x}{h_{n,T}}\right) \frac{h_{n,T}^{p+1}}{(p+1)!} \sup_{z\in(X_{i\Delta_{n,T}},x)} \sigma^{2(p+1)}(z) \right. \\
& \left. - \int_0^T \frac{1}{h_{n,T}} \left(\frac{X_s - x}{h_{n,T}}\right)^k K\left(\frac{X_s - x}{h_{n,T}}\right) \frac{h_{n,T}^{p+1}}{(p+1)!} \sup_{z\in(X_s,x)} \sigma^{2(p+1)}(z) ds \right| \\
& + \left| \int_0^T \frac{1}{h_{n,T}} \left(\frac{X_s - x}{h_{n,T}}\right)^k K\left(\frac{X_s - x}{h_{n,T}}\right) \frac{h_{n,T}^{p+1}}{(p+1)!} \sup_{z\in(X_s,x)} \sigma^{2(p+1)}(z) ds \right| \\
\leq\ & \left| \frac{h_{n,T}^{p+1}}{(p+1)!} \sum_{i=0}^{n-1} \frac{1}{h_{n,T}} \int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}} \left(\frac{X_{i\Delta_{n,T}} - x}{h_{n,T}}\right)^k K\left(\frac{X_{i\Delta_{n,T}} - x}{h_{n,T}}\right) \right. \\
& \left. \left(\sup_{z\in(X_{i\Delta_{n,T}},x)} \sigma^{2(p+1)}(z) - \sup_{z\in(X_s,x)} \sigma^{2(p+1)}(z) \right) ds \right| \\
& + \left| \frac{h_{n,T}^{p+1}}{(p+1)!} \sum_{i=0}^{n-1} \frac{1}{h_{n,T}} \int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}} \left(\left(\frac{X_{i\Delta_{n,T}} - x}{h_{n,T}}\right)^k K\left(\frac{X_{i\Delta_{n,T}} - x}{h_{n,T}}\right) \right. \right. \\
& \left. \left. - \left(\frac{X_s - x}{h_{n,T}}\right)^k K\left(\frac{X_s - x}{h_{n,T}}\right) \right) \sup_{z\in(X_s,x)} \sigma^{2(p+1)}(z) ds \right| \\
& + \left| \int_0^T \frac{1}{h_{n,T}} \left(\frac{X_s - x}{h_{n,T}}\right)^k K\left(\frac{X_s - x}{h_{n,T}}\right) \frac{h_{n,T}^{p+1}}{(p+1)!} \sup_{z\in(X_s,x)} \sigma^{2(p+1)}(z) ds \right| \\
& + O_{a.s.}\left(\frac{\Delta_{n,T}}{h_{n,T}}\right) \\
=\ & O_{a.s.}\left(h_{n,T}^{p+1} \left(\Delta_{n,T} \log(1/\Delta_{n,T})\right)^{1/2} \overline{L}_X(T,x) \right) \\
& + O_{a.s.}\left(h_{n,T}^{p+1} \left(\Delta_{n,T} \log(1/\Delta_{n,T})\right)^{1/2} \overline{L}_X(T,x)/h_{n,T} \right) \\
& + O_{a.s.}\left(h_{n,T}^{p+1} \overline{L}_X(T,x) \right) \\
& + O_{a.s.}\left(\frac{\Delta_{n,T}}{h_{n,T}}\right)
\end{aligned}$$

Also from Theorem 1 we have

$$
\begin{aligned}
s_{n,T}^{k} &= O_{a.s.}(\overline{L}_X(T,x)) + O_{a.s.}\left((\Delta_{n,T}\log(1/\Delta_{n,T}))^{1/2}\,\overline{L}_X(T,x)/h_{n,T}\right) \\
&\quad + O_{a.s.}\left(\Delta_{n,T}/h_{n,T}\right),
\end{aligned}
$$

all $k \in \{0, ..., p\}$. Collecting results

$$
\mathbf{H}\left(\mathbf{X}_{n,T}'\mathbf{W}_{n,T}\mathbf{X}_{n,T}\right)^{-1}\mathbf{H}\widetilde{\mathbf{u}}_{n,T} = \mathbf{S}_{n,T}^{-1}\widetilde{\mathbf{u}}_{n,T} = O_{a.s.}\left(h_{n,T}^{p+1}\right)
$$

and formula (5) follows.

5.6 Proof of Theorem 4.

It is done for an arbitrary term $t_{n,T}^{*k}$, $k \in \{0, 1, ..., p\}$. Note that

$$
\begin{aligned}
&\left(X_{(i+1)\Delta_{n,T}} - X_{i\Delta_{n,T}}\right)^2 \\
&= \int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}} 2\left(X_s - X_{i\Delta_{n,T}}\right)\mu(X_s)ds \\
&\quad + \int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}} 2\left(X_s - X_{i\Delta_{n,T}}\right)\sigma(X_s)dB_s \\
&\quad + \int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}} \sigma^2(X_s)ds\ .
\end{aligned}
$$

Then $t_{n,T}^{*k}$ can be written as

$$
t_{n,T}^{*k} = a_{n,T}^{k} + b_{n,T}^{k} + c_{n,T}^{k}\ ,
$$

where

$$
\begin{aligned}
a_{n,T}^{k} &= \sum_{i=0}^{n-1}\frac{\Delta_{n,T}}{h_{n,T}}\left(\frac{X_{i\Delta_{n,T}} - x}{h_{n,T}}\right)^k K\left(\frac{X_{i\Delta_{n,T}} - x}{h_{n,T}}\right) \\
&\quad \times\left[\frac{1}{\Delta_{n,T}}\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}} 2\left(X_s - X_{i\Delta_{n,T}}\right)\mu(X_s)ds\right], \qquad (11)
\end{aligned}
$$

$$
\begin{aligned}
b_{n,T}^{k} &= \sum_{i=0}^{n-1}\frac{\Delta_{n,T}}{h_{n,T}}\left(\frac{X_{i\Delta_{n,T}} - x}{h_{n,T}}\right)^k K\left(\frac{X_{i\Delta_{n,T}} - x}{h_{n,T}}\right) \\
&\quad \times\left[\frac{1}{\Delta_{n,T}}\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}} 2\left(X_s - X_{i\Delta_{n,T}}\right)\sigma(X_s)dB_s\right], \qquad (12)
\end{aligned}
$$

$$
\begin{aligned}
c^k_{n,T} &= \sum_{i=0}^{n-1} \frac{\Delta_{n,T}}{h_{n,T}} \left(\frac{X_{i\Delta_{n,T}} - x}{h_{n,T}} \right)^k K \left(\frac{X_{i\Delta_{n,T}} - x}{h_{n,T}} \right) \\
&\quad \times \left[\frac{1}{\Delta_{n,T}} \int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}} \left(\sigma^2(X_s) - \sigma^2(X_{i\Delta_{n,T}}) \right) ds \right]. \qquad (13)
\end{aligned}
$$

The first term:

$$
\begin{aligned}
a^k_{n,T} &\leq 2\kappa_{n,T} \sum_{i=0}^{n-1} \frac{\Delta_{n,T}}{h_{n,T}} \left(\frac{X_{i\Delta_{n,T}} - x}{h_{n,T}} \right)^k \\
&\quad \times K \left(\frac{X_{i\Delta_{n,T}} - x}{h_{n,T}} \right) \frac{1}{\Delta_{n,T}} \int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}} \mu(X_s) ds \\
&\leq \Bigg| \Bigg(2\kappa_{n,T} \sum_{i=0}^{n-1} \frac{\Delta_{n,T}}{h_{n,T}} \left(\frac{X_{i\Delta_{n,T}} - x}{h_{n,T}} \right)^k \\
&\quad \times K \left(\frac{X_{i\Delta_{n,T}} - x}{h_{n,T}} \right) \frac{1}{\Delta_{n,T}} \int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}} \left(\mu(X_s) - \mu(X_{i\Delta_{n,T}}) \right) ds \Bigg) \\
&\quad + \Bigg(2\kappa_{n,T} \sum_{i=0}^{n-1} \frac{\Delta_{n,T}}{h_{n,T}} \left(\frac{X_{i\Delta_{n,T}} - x}{h_{n,T}} \right)^k \\
&\quad \times K \left(\frac{X_{i\Delta_{n,T}} - x}{h_{n,T}} \right) \frac{1}{\Delta_{n,T}} \int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}} \mu(X_{i\Delta_{n,T}}) ds \Bigg) \Bigg| \\
&\leq \Bigg| 2\kappa_{n,T} \sum_{i=0}^{n-1} \frac{\Delta_{n,T}}{h_{n,T}} \left(\frac{X_{i\Delta_{n,T}} - x}{h_{n,T}} \right)^k \\
&\quad \times K \left(\frac{X_{i\Delta_{n,T}} - x}{h_{n,T}} \right) \frac{1}{\Delta_{n,T}} \int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}} \left(\mu(X_s) - \mu(X_{i\Delta_{n,T}}) \right) ds \Bigg| \\
&\quad + \left| 2\kappa_{n,T} \sum_{i=0}^{n-1} \frac{\Delta_{n,T}}{h_{n,T}} \left(\frac{X_{i\Delta_{n,T}} - x}{h_{n,T}} \right)^k K \left(\frac{X_{i\Delta_{n,T}} - x}{h_{n,T}} \right) \mu(X_{i\Delta_{n,T}}) \right| \\
&= O_{a.s.} \left(\left(\Delta_{n,T} \log(1/\Delta_{n,T}) \right) \overline{L}_X(T,x) \right) \\
&\quad + O_{a.s.} \left(\left(\Delta_{n,T} \log(1/\Delta_{n,T}) \right)^{1/2} \overline{L}_X(T,x) \right) + O_{a.s.} \left(\Delta_{n,T}/h_{n,T} \right)
\end{aligned}
$$

due to the the boundedness of the drift on compact subsets and of its first derivative, and a local Lipschitz condition on the drift implied by Assumption

1. By a similar argument,

$$c^k_{n,T} = O_{a.s.}\left(\left(\Delta_{n,\overline{T}} \log(1/\Delta_{n,\overline{T}}) \right)^{1/2} \overline{L}_X(T,x) \right) + O_{a.s.}\left(\Delta_{n,T}/h_{n,T} \right) \ .$$

Since $\mathbf{S}^{-1}_{n,T} = O_{a.s.}(\overline{L}_X(T,x)^{-1})$, $\mathbf{S}^{-1}_{n,T}\mathbf{a}_{n,T} = O_{a.s.}\left(\Delta_{n,T}\log(1/\Delta_{n,T})\right)$ and $\mathbf{S}^{-1}_{n,T}$ $\mathbf{c}_{n,T} = O_{a.s.}\left(\left(\Delta_{n,T}\log(1/\Delta_{n,T}) \right)^{1/2} \right)$, where $\mathbf{a}_{n,T}$ and $\mathbf{c}_{n,T}$ are constructed from (11) and (13), respectively. To analyze $b^k_{n,T}$, we use a result in Theorem 5. We obtain below that $\sqrt{\frac{h_{n,T}\overline{L}_X(T,x)}{\Delta_{n,T}}}\mathbf{S}^{-1}_{n,T}b^k_{n,T}$ converges to a normal random variable. Then $\mathbf{S}^{-1}_{n,T}b^k_{n,T} \to 0$ almost surely if $\frac{h_{n,T}\overline{L}_X(T,x)}{\Delta_{n,T}} \to \infty$. But this is already implied by the condition $h_{n,T}/\sqrt{\Delta_{n,T}\log(1/\Delta_{n,T})} \to \infty$, since $\left(\Delta_{n,T}\log(1/\Delta_{n,T})\right)^{1/2}/\Delta_{n,T} \to \infty$ and $\overline{L}_X(T,x) \to \infty$.

5.7 Proof of Theorem 5.

We first derive the asymptotic distribution of $\mathbf{t}^*_{n,T}$. This vector can be decomposed as

$$\mathbf{t}^*_{n,T} = \mathbf{a}_{n,T} + \mathbf{b}_{n,T} + \mathbf{c}_{n,T}$$

where $\mathbf{a}_{n,T}$, $\mathbf{b}_{n,T}$ and $\mathbf{c}_{n,T}$ are constructed from (11)-(13). We study the destribution and the convergence rate of $\mathbf{b}_{n,T}$. Let $\mathbf{b}_{n,T} = \mathbf{B}_{n,T}(1)$, and write

$$U^k_{n,T}(x,r) = \sqrt{\frac{h_{n,T}}{\Delta_{n,T}}} B^k_{n,T}(r) \ ,$$

$$\begin{aligned} U^k_{n,T}(r) &= \frac{2}{\sqrt{h_{n,T}}} \sum_{i=0}^{[nr]-1} \left(\frac{X_{i\Delta_{n,T}} - x}{h_{n,T}} \right)^k K\left(\frac{X_{i\Delta_{n,T}} - x}{h_{n,T}} \right) \\ &\quad \times \frac{1}{\sqrt{\Delta_{n,T}}} \int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}} \left(X_s - X_{i\Delta_{n,T}} \right) \sigma(X_s) dB_s \ . \end{aligned}$$

$\mathbf{U}_{n,T}(r)$ is a continuous martingale whose quadratic variation process $\langle \mathbf{U}_{n,T} \rangle_r$ is given by

$$\begin{aligned} &\langle U^k_{n,T}, U^l_{n,T} \rangle_r \\ &= \frac{4}{h_{n,T}} \sum_{i=0}^{[nr]-1} \left(\frac{X_{i\Delta_{n,T}} - x}{h_{n,T}} \right)^{k+l} K^2\left(\frac{X_{i\Delta_{n,T}} - x}{h_{n,T}} \right) \end{aligned}$$

$$\times \frac{1}{\Delta_{n,T}} \int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}} \left(X_s - X_{i\Delta_{n,T}}\right)^2 \sigma^2(X_s) ds$$

$$\begin{aligned}
&= \frac{4}{h_{n,T}} \int_0^{rT} \left(\frac{X_s - x}{h_{n,T}}\right)^{k+l} K^2 \left(\frac{X_s - x}{h_{n,T}}\right) \sigma^4(X_s) ds \\
&\quad + O_{a.s.}\left(\frac{\overline{L}_X(rT,x)}{h_{n,\overline{T}}} \left(\Delta_{n,T} \log\left(\frac{1}{\Delta_{n,T}}\right)\right)^{1/2}\right) \\
&= \frac{4}{h_{n,T}} \int \left(\frac{a-x}{h_{n,T}}\right)^{k+l} K^2 \left(\frac{a-x}{h_{n,T}}\right) \sigma^4(a) \overline{L}_X(rT,a) da + o_{a.s.}(1) \\
&= 4 \int u^{k+l} K^2(u) \sigma^4(x + h_{n,T}u) \overline{L}_X(rT, x + h_{n,T}u) du + o_{a.s.}(1) \\
&\overset{a.s.}{\to} 4 \left(\int u^{k+l} K^2(u) du\right) \sigma^4(x) \overline{L}_X(rT,x),
\end{aligned}$$

following arguments similar to those used to show consistency. Let $W_{n,t} = \sum_{i=0}^{n-1} w_{(i+1)\Delta_{n,T}}$ and $w_{(i+1)\Delta_{n,T}} = (W_{(i+1)\Delta_{n,T}} - W_{i\Delta_{n,T}})$, where W_t is a Brownian Motion independent of B_t. Clearly, the covariation process $\langle U_{n,T}^k, W_{n,T}\rangle_r \overset{a.s.}{\to} 0$ for all $k = 0, ..., p$. $\mathbf{R}$ is a symmetric, positive definite matrix. Consequently there exists Λ such that $\Lambda \mathbf{R} \Lambda' = \mathbf{I}$, $\Lambda\Lambda' = \mathbf{R}^{-1}$. Define $\mathbf{M}_{n,T}(r) = \Lambda \mathbf{U}_{n,T}(r)$. $\mathbf{M}_{n,T}(r)$ is also a continuous martingale. From previous results,

$$\langle \mathbf{M}_{n,T}\rangle_r \overset{a.s.}{\to} 4\mathbf{I}\sigma^4(x)\overline{L}_X(rT,x) \ .$$

Since $\langle M_{n,T}^k, M_{n,T}^l\rangle_r \overset{a.s.}{\to} 0$, defining

$$\tau_{n,T}^k(r) = \inf\{s : \langle M_{n,T}^k, M_{n,T}^k\rangle_s > r\}$$

implies

$$\langle M_{n,T}^k, M_{n,T}^l\rangle_{\tau_{n,T}^k(r)} \to^p 0, \langle M_{n,T}^k, M_{n,T}^l\rangle_{\tau_{n,T}^l(r)} \overset{p}{\to} 0 \ ,$$

by the definition of time change. Also, $\langle M_{n,T}^k, W_{n,T}\rangle_{\tau_{n,T}^k(r)} \overset{p}{\to} 0$, as a consequence of $\langle M_{n,T}^k, W_{n,T}\rangle_r \overset{p}{\to} 0$. Applying the asymptotic Knight theorem (see Revuz and Yor [16], p.496):

$$(\mathbf{M}_{n,T}(\tau_{n,T}(r)), W_{n,T}(r)) = (\mathbf{V}_{n,T}(r), W_{n,T}(r)) \overset{d}{\to} (\mathbf{V}(r), W(r))$$

where the process $\mathbf{V}_{n,T}(r)$ is the Dambis, Dubins, Schwartz Brownian motion of the martingale $\mathbf{M}_{n,T}(r)$, and $(\mathbf{V}, W)$ is a $(p+3/2)$ −vector of independent Brownian motions. Hence,

$$M^k_{n,T}(r) \xrightarrow{d} V^k\left(4\sigma^4(x)\overline{L}_X(rT,x)\right),\ k = 0,1,...,p\ .$$

This implies

$$\begin{aligned}
&\sqrt{\overline{L}_X(rT,x)}\mathbf{S}^{-1}_{n,T}\mathbf{U}_{n,T}(1) \\
&= \sqrt{\frac{h_{n,T}\overline{L}_X(rT,x)}{\Delta_{n,T}}}\mathbf{S}^{-1}_{n,T}\mathbf{B}_{n,T}(1) \\
&\xrightarrow{d} \mathrm{N}\left(0, 4\mathbf{S}^{-1}\mathbf{R}\mathbf{S}^{-1}\sigma^4(x)\right)
\end{aligned}$$

Finally, if $h^{2p+3}_{n,T}\overline{L}_X(T,x)/\Delta_{n,T} = o_{a.s.}(1)$

$$\begin{aligned}
&\sqrt{\frac{h_{n,T}\overline{L}_X(T,x)}{\Delta_{n,T}}}\left(\mathbf{H}\left(\beta_{n,T}-\beta\right)\right) \\
&= \sqrt{\frac{h_{n,T}\overline{L}_X(T,x)}{\Delta_{n,T}}}\left(\mathbf{S}^{-1}_{n,T}\mathbf{t}^*_{n,T} + O_{a.s.}\left(h^{p+1}_{n,T}\right)\mathbf{1}^{p+1}\right) \\
&= \sqrt{\frac{h_{n,T}\overline{L}_X(T,x)}{\Delta_{n,T}}}\left(\mathbf{S}^{-1}_{n,T}\left(\mathbf{a}_{n,T}+\mathbf{b}_{n,T}+\mathbf{c}_{n,T}\right) + O_{a.s.}\left(h^{p+1}_{n,T}\right)\mathbf{1}^{p+1}\right) \\
&= \sqrt{\frac{h_{n,T}\overline{L}_X(T,x)}{\Delta_{n,T}}}\left(O_{a.s.}\left(\left(\Delta_{n,\overline{T}}\log(1/\Delta_{n,\overline{T}})\right)\right)\mathbf{1}^{p+1}\right. \\
&\quad + O_{a.s.}\left(\left(\Delta_{n,\overline{T}}\log(1/\Delta_{n,\overline{T}})\right)^{1/2}\right)\mathbf{1}^{p+1} \\
&\quad \left.+\ \mathbf{S}^{-1}_{n,T}\mathbf{b}_{n,T} + O_{a.s.}\left(h^{p+1}_{n,T}\right)\mathbf{1}^{p+1}\right) \\
&= \sqrt{\frac{h_{n,T}\overline{L}_X(T,x)}{\Delta_{n,T}}}\left(\mathbf{S}^{-1}_{n,T}\mathbf{b}_{n,T}\right) + o_{a.s.}(1)\mathbf{1}^{p+1} \\
&\xrightarrow{d} \mathrm{N}\left(0, 4\mathbf{S}^{-1}\mathbf{R}\mathbf{S}^{-1}\sigma^4(x)\right)\ .
\end{aligned}$$

5.8 Proof of Theorem 6.

It is done for an arbitrary term $u^{*k}_{n,T}$, $k \in \{0,1,...,p\}$. Note that

$$X_{(i+1)\Delta_{n,T}} - X_{i\Delta_{n,T}}$$

$$= \int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}} \mu(X_s)ds + \int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}} \sigma(X_s)dB_s$$

Then $u_{n,T}^{*k}$ can be written as

$$u_{n,T}^{*k} = d_{n,T}^k + e_{n,T}^k \ ,$$

where

$$\begin{aligned} d_{n,T}^k &= \sum_{i=0}^{n-1} \frac{\Delta_{n,T}}{g_{n,T}} \left(\frac{X_{i\Delta_{n,T}} - x}{g_{n,T}}\right)^k K\left(\frac{X_{i\Delta_{n,T}} - x}{g_{n,T}}\right) \\ &\times \left[\frac{1}{\Delta_{n,T}} \int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}} \left(\mu(X_s) - \mu\left(X_{i\Delta_{n,T}}\right)\right) ds\right], \end{aligned} \tag{14}$$

$$\begin{aligned} e_{n,T}^k &= \sum_{i=0}^{n-1} \frac{\Delta_{n,T}}{g_{n,T}} \left(\frac{X_{i\Delta_{n,T}} - x}{g_{n,T}}\right)^k K\left(\frac{X_{i\Delta_{n,T}} - x}{g_{n,T}}\right) \\ &\times \left[\frac{1}{\Delta_{n,T}} \int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}} \sigma(X_s)dB_s\right], \end{aligned} \tag{15}$$

The first term:

$$\begin{aligned} d_{n,T}^k &\leq 2C\kappa_{n,T} \sum_{i=0}^{n-1} \frac{\Delta_{n,T}}{g_{n,T}} \left(\frac{X_{i\Delta_{n,T}} - x}{g_{n,T}}\right)^k K\left(\frac{X_{i\Delta_{n,T}} - x}{g_{n,T}}\right) \\ &= O_{a.s.}\left(\left(\Delta_{n,\overline{T}} \log(1/\Delta_{n,\overline{T}})\right)^{1/2} \overline{L}_X(T,x)\right) \end{aligned}$$

due to a local Lipschitz condition on the drift implied by Assumption 1. Thus $\mathbf{S}_{n,T}^{-1}\mathbf{d}_{n,T} \overset{a.s.}{\to} 0$. The term $e_{n,T}^k$ is the martingale term and it is analyzed as usual.

5.9 Proof of Theorem 7.

We first derive the asymptotic distribution of $\mathbf{u}_{n,T}^*$. This vector can be decomposed as

$$\mathbf{u}_{n,T}^* = \mathbf{d}_{n,T} + \mathbf{e}_{n,T}$$

where $\mathbf{d}_{n,T}$ and $\mathbf{e}_{n,T}$ are constructed from (14) and (15). We first study the distribution and the convergence rate of $\mathbf{e}_{n,T}$. Let $\mathbf{e}_{n,T} = \mathbf{E}_{n,T}(1)$, and write

$$J^k_{n,T}(x,r) = \sqrt{g_{n,T}} E^k_{n,T}(r) \ ,$$

$$\begin{aligned} J^k_{n,T}(r) &= \frac{2}{\sqrt{g_{n,T}}} \sum_{i=0}^{[nr]-1} \left(\frac{X_{i\Delta_{n,T}} - x}{g_{n,T}}\right)^k K\left(\frac{X_{i\Delta_{n,T}} - x}{g_{n,T}}\right) \\ &\quad \times \int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}} \sigma(X_s) dB_s \ . \end{aligned}$$

$\mathbf{J}_{n,T}(r)$ is a continuous martingale whose quadratic variation process $\langle \mathbf{J}_{n,T} \rangle_r$ is given by

$$\begin{aligned} &\langle J^k_{n,T}, J^l_{n,T} \rangle_r \\ &= \frac{1}{g_{n,T}} \sum_{i=0}^{[nr]-1} \left(\frac{X_{i\Delta_{n,T}} - x}{g_{n,T}}\right)^{k+l} K^2\left(\frac{X_{i\Delta_{n,T}} - x}{g_{n,T}}\right) \\ &\quad \times \int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}} \sigma^2(X_s) ds \\ &= \frac{1}{g_{n,T}} \int_0^{rT} \left(\frac{X_s - x}{g_{n,T}}\right)^{k+l} K^2\left(\frac{X_s - x}{g_{n,T}}\right) \sigma^2(X_s) ds \\ &\quad + O_{a.s.}\left(\frac{\overline{L}_X(rT,x)}{g_{n,\overline{T}}} \left(\Delta_{n,T} \log\left(\frac{1}{\Delta_{n,T}}\right)\right)^{1/2}\right) \\ &= \frac{1}{g_{n,T}} \int \left(\frac{a-x}{g_{n,T}}\right)^{k+l} K^2\left(\frac{a-x}{g_{n,T}}\right) \sigma^2(a) \overline{L}_X(rT,a) da + o_{a.s.}(1) \\ &= \int u^{k+l} K^2(u)\, \sigma^2(x + g_{n,T}u) \overline{L}_X(rT, x + g_{n,T}u) du + o_{a.s.}(1) \\ &\overset{a.s.}{\to} \left(\int u^{k+l} K^2(u) du\right) \sigma^2(x) \overline{L}_X(rT,x) \end{aligned}$$

The quadratic covariation process $\langle J^k_{n,T}, W_{n,T} \rangle_r \overset{a.s.}{\to} 0$ for all $k = 0, ..., p$. Define $\mathbf{N}_{n,T}(r) = \Lambda \mathbf{J}_{n,T}(r)$. $\mathbf{N}_{n,T}(r)$ is clearly a continuous martingale. From results above,

$$\langle \mathbf{N}_{n,T} \rangle_r \overset{a.s.}{\to} \mathbf{I} \sigma^2(x) \overline{L}_X(rT,x) \ .$$

Defining $\rho^k_{n,T}(r) = \inf\{s : \langle N^k_{n,T}, N^k_{n,T}\rangle_s > r\}$ and verifying that the conditions for the asymptotic Knight theorem hold as in Theorem 3,

$$(\mathbf{N}_{n,T}(\tau_{n,T}(r)), W_{n,T}(r)) = (\mathbf{Z}_{n,T}(r), W_{n,T}(r)) \overset{d}{\to} (\mathbf{Z}(r), W(r))$$

where the process $\mathbf{Z}_{n,T}(r)$ is the Dambis, Dubins, Schwartz Brownian motion of the martingale $\mathbf{N}_{n,T}(r)$, and $(\mathbf{Z}, W)$ is a $(p+2)$ −vector of independent Brownian motions. Hence,

$$N^k_{n,T}(r) \overset{d}{\to} Z^k\left(\sigma^2(x)\overline{L}_X(rT,x)\right), \; k = 0, 1, ..., p \; .$$

This implies

$$\begin{aligned}
&\sqrt{\overline{L}_X(rT,x)}\mathbf{S}^{-1}_{n,T}\mathbf{J}_{n,T}(1) \\
&= \sqrt{g_{n,T}\overline{L}_X(rT,x)}\mathbf{S}^{-1}_{n,T}\mathbf{E}_{n,T}(1) \\
&\overset{d}{\to} \mathrm{N}\left(0, \mathbf{S}^{-1}\mathbf{R}\mathbf{S}^{-1}\sigma^2(x)\right)
\end{aligned}$$

Finally, if $g^{2p+3}_{n,T}\overline{L}_X(T,x) \to \infty$

$$\begin{aligned}
&\sqrt{g_{n,T}\overline{L}_X(T,x)}\left(\mathbf{H}\left(\alpha_{n,T} - \alpha\right)\right) \\
&= \sqrt{g_{n,T}\overline{L}_X(T,x)}\left(\mathbf{S}^{-1}_{n,T}\mathbf{u}^*_{n,T} + O_{a.s.}\left(g^{p+1}_{n,T}\right)\mathbf{1}^{p+1}\right) \\
&= \sqrt{g_{n,T}\overline{L}_X(T,x)}\left(\mathbf{S}^{-1}_{n,T}\left(\mathbf{d}_{n,T} + \mathbf{e}_{n,T}\right) + O_{a.s.}\left(g^{p+1}_{n,T}\right)\mathbf{1}^{p+1}\right) \\
&= \sqrt{g_{n,T}\overline{L}_X(T,x)}\Big(O_{a.s.}\left(\left(\Delta_{n,\overline{T}}\log(1/\Delta_{n,\overline{T}})\right)^{1/2}\right)\mathbf{1}^{p+1} \\
&\quad + \mathbf{S}^{-1}_{n,T}\mathbf{e}_{n,T} + O_{a.s.}\left(g^{p+1}_{n,T}\right)\mathbf{1}^{p+1}\Big) \\
&= \sqrt{g_{n,T}\overline{L}_X(T,x)}\left(\mathbf{S}^{-1}_{n,T}\mathbf{e}_{n,T}\right) + o_{a.s.}(1)\mathbf{1}^{p+1} \\
&\overset{d}{\to} \mathrm{N}\left(0, \mathbf{S}^{-1}\mathbf{R}\mathbf{S}^{-1}\sigma^2(x)\right) \; .
\end{aligned}$$

5.10 Proof of Theorem 8.

Decompose the estimation error as:

$$f_{n,T}(x) \;=\; \frac{1}{Tb_{n,T}}\int_0^T K\left(\frac{X_s - x}{b_{n,T}}\right)ds + O_{a.s.}\left(\frac{\left(\Delta_{n,T}\log(1/\Delta_{n,T})\right)^{1/2}}{b_{n,T}}\right)$$

$$
\begin{aligned}
&= \frac{\overline{L}_X(T,x)}{T} + O_{a.s.}\left(\frac{\sqrt{b_{n,T}\overline{L}_X(T,x)}}{T}\right) + O_{a.s.}\left(\sqrt{b_{n,T}}\right) \\
&\quad + O_{a.s.}\left(\frac{(\Delta_{n,T}\log(1/\Delta_{n,T}))^{1/2}}{b_{n,T}}\right) \\
&= f(x) + O_{a.s.}\left(\frac{1}{\sqrt{T}}\right) + O_{a.s.}\left(\frac{\sqrt{b_{n,T}\overline{L}_X(T,x)}}{T}\right) + O_{a.s.}\left(\sqrt{b_{n,T}}\right) \\
&\quad + O_{a.s.}\left(\frac{(\Delta_{n,T}\log(1/\Delta_{n,T}))^{1/2}}{b_{n,T}}\right)
\end{aligned}
$$

where the second line follows from Bandi and Phillips [3] and the third from Theorem 1 and Theorem 9, below. Under the stated conditions, it is easy to see that the assertion of the theorem holds.

5.11 Proof of Corollary 2.

Decompose the estimation error as:

$$
\begin{aligned}
f_{n,T}(x) &= \frac{1}{Tb_{n,T}}\int_0^T K\left(\frac{X_s - x}{b_{n,T}}\right) ds + O_{a.s.}\left(\frac{(\Delta_{n,T}\log(1/\Delta_{n,T}))^{1/2}}{b_{n,T}}\right) \\
&= \frac{\overline{L}_X(T,x)}{T} + O_{a.s.}\left(\frac{\sqrt{b_{n,T}\overline{L}_X(T,x)}}{T}\right) + O_{a.s.}\left(\sqrt{b_{n,T}}\right) \\
&\quad + O_{a.s.}\left(\frac{(\Delta_{n,T}\log(1/\Delta_{n,T}))^{1/2}}{b_{n,T}}\right) \\
&= \frac{m(x)}{m\,(I)} + o_{a.s.}\left(\frac{1}{T}\right) + O_{a.s.}\left(\frac{\sqrt{b_{n,T}\overline{L}_X(T,x)}}{T}\right) + O_{a.s.}\left(\sqrt{b_{n,T}}\right) \\
&\quad + O_{a.s.}\left(\frac{(\Delta_{n,T}\log(1/\Delta_{n,T}))^{1/2}}{b_{n,T}}\right) \\
&= o_{a.s.}\left(\frac{1}{T}\right) + O_{a.s.}\left(\frac{\sqrt{b_{n,T}\overline{L}_X(T,x)}}{T}\right) + O_{a.s.}\left(\sqrt{b_{n,T}}\right)
\end{aligned}
$$

$$+O_{a.s.}\left(\frac{(\Delta_{n,T}\log(1/\Delta_{n,T}))^{1/2}}{b_{n,T}}\right)$$

where the second line follows from Bandi and Phillips [3] and the third from Theorem 1. If Assumption 3 does not hold, $m(I)=\infty$, and provided all the asymptotic terms vanish, the Corollary is verified.

5.12 Proof of Corollaries 3-7.

They are a direct consequence of Theorems 3-7, and Theorems 1, 14:

$$\begin{aligned}\overline{L}_X(T,x) &= \overline{L}_X(T,x)-Tf(x)+Tf(x)\\ &\leq \left|\overline{L}_X(T,x)-Tf(x)\right|+|Tf(x)|\\ &= O_{a.s.}(\sqrt{T})+O_{a.s.}(T)\\ &= O_{a.s.}(T).\end{aligned}$$

5.13 Proof of Theorem 9.

Using the estimation error decomposition in Theorem 8, the stated assumptions and noting that the local time estimation error is always of smaller order of magnitude than the error between the density and the normalized local time, the result follows from analyzing the asymptotics of continuous additive functionals of diffusion processes of the form

$$\int_0^t \xi(X_s)\,ds$$

cf. Tanaka [17] with

$$\xi(y)=\frac{1}{\varepsilon}\left[1_{\{y\in(x,x+\varepsilon)\}}-m((x,x+\varepsilon))/m(I)\right]$$

and taking limits as $\varepsilon\downarrow 0$.

As in Theorem 1 let $T_{a,b}=\inf\{t\geq 0; X_t\notin(a,b)\}$.It is known that: (see Itô and McKean (1974))

$$\begin{aligned}E_x[T_{a,b}] &= -2\int_a^x (p(x)-p(y))\,m(dy) && (16)\\ &\quad +2\frac{p(x)-p(a)}{p(b)-p(a)}\int_a^b (p(b)-p(y))\,m(dy), && (17)\end{aligned}$$

for $a < x < b$. Then

$$
\begin{aligned}
P_x\left[X_{T_{a,b}} = a\right] &= \frac{p(b) - p(x)}{p(b) - p(a)} \\
P_x\left[X_{T_{a,b}} = b\right] &= \frac{p(x) - p(a)}{p(b) - p(a)},
\end{aligned}
$$

and by taking limits in (16), we obtain expected hitting times:

$$
\begin{aligned}
E_x[T_a] &= -2\int_a^x (p(x) - p(y))\, m(dy) + 2(p(x) - p(a))\, m((a,r)) \\
E_x[T_b] &= -2\int_x^b (p(y) - p(x))\, m(dy) + 2(p(b) - p(x))\, m((l,b)),
\end{aligned}
$$

note that if the speed measure is integrable, then the expected hitting times are finite. Next define the random time $T_{a;b} = T_b + \theta_{T_b} \circ T_a$. Then using the strong Markov property,

$$
\begin{aligned}
E_x[T_{u;x}] &= E_x[T_u] + E_x[\theta_{T_u} \circ T_x] \\
&= E_x[T_u] + E_u[T_x] \\
&= 2[p(u) - p(x)]\, m(I) < \infty,\ x < u < r.
\end{aligned}
$$

The finiteness of expected hitting times are also known as positive recurrence properties. Also

$$
\begin{aligned}
E_u\left[\int_0^{T_x} \xi(X_t)\, dt\right] &= -2\int_x^u (p(u) - p(y))\, \xi(y)\, m(dy) \\
&\quad +2(p(u) - p(x))\int_x^r \xi(y)\, m(y)\, dy \\
E_x\left[\int_0^{T_u} \xi(X_t)\, dt\right] &= -2\int_x^u (p(y) - p(x))\, \xi(y)\, m(dy) \\
&\quad +2(p(u) - p(x))\int_l^u \xi(y)\, m(y)\, dy,
\end{aligned}
$$

adding up

$$
E_x\left[\int_0^{T_u} \xi(X_s)\, ds\right] + E_u\left[\int_0^{T_x} \xi(X_s)\, ds\right] =
$$

$$\begin{aligned}
E_x\left[\int_0^{T_{u;x}} \xi(X_t)\,dt\right] &= E_x\left[\int_0^{T_u} \xi(X_s)\,ds\right] + E_x\left[\int_0^{T_x} \xi(\theta_{T_u}\circ X_s)\,ds\right] \\
&= E_x\left[\int_0^{T_u} \xi(X_s)\,ds\right] + E_u\left[\int_0^{T_x} \xi(X_s)\,ds\right] \\
&= 2\left[p(u)-p(x)\right]\int \xi(x)\,m(dx) = 0.
\end{aligned}$$

To obtain the variance, note that

$$\frac{d}{m(dx)}\left[\frac{d}{dp(x)}E_x\left[\int_0^{T_u} \xi(X_t)\,dt\right]^2\right] + 2\xi(x)\,E_x\left[\int_0^{T_u} \xi(X_t)\,dt\right] = 0$$

to obtain

$$\begin{aligned}
E_x\left[\int_0^{T_{u;x}} \xi(X_t)\,dt\right]^2 &= E_x\left[\int_0^{T_u} \xi(X_t)\,dt\right]^2 + E_u\left[\int_0^{T_x} \xi(X_t)\,dt\right]^2 \\
&\quad +2E_x\left[\int_0^{T_u} \xi(X_t)\,dt\right]E_u\left[\int_0^{T_x} \xi(X_t)\,dt\right] \\
&= 2\left(p(u)-p(x)\right)D(\xi,\varepsilon)
\end{aligned}$$

where

$$D(\xi,\varepsilon) = \frac{4}{m(I)}\int\left(\int^y \left(\xi(z,\varepsilon)\right)m(dz)\right)^2 p(y)\,dy,$$

after some rearranging. Note that

$$\lim_{\varepsilon\downarrow 0} D(\xi,\varepsilon) = 4m(I)\,f^2(x)\int\left(1_{x\le y}-F(y)\right)^2 p(y)\,dy.$$

The succesive increments at times $T_{u;x}$ are independent using the strong Markov property, and this implies, after letting $\varepsilon\downarrow 0$, that

$$\sqrt{T}\left[\frac{\overline{L}_X(T,x)}{T} - f(x)\right] = N(0,\omega)$$

with ω as in the statement of the Theorem.

References

[1] F.M. Bandi. Nonparametric fixed income pricing. Unpublished, Graduate School of Business, University of Chicago, 1999.

[2] F.M. Bandi and G. Moloche. On the functional estimation of multivariate diffusion processes. Unpublished, University of Chicago GSB and M.I.T., 2001.

[3] F.M. Bandi and P.C.B. Phillips. Fully nonparametric estimation of scalar diffusion models. Unpublished, Cowles Foundation for Research in Economics, Yale University, 2001.

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

[5] J. Fan and I. Gijbels. *Local Polynomial Modelling and Its Applications*. Chapman & Hall, 1996.

[6] W. Feller. *An introduction to probability theory and its applications. Vol. II.* John Wiley & Sons Inc., New York, second edition, 1971.

[7] D. Florens-Zmirou. On estimating the diffusion coefficient from discrete observations. *J. Appl. Probability*, 30:790–804, 1993.

[8] S.A. Geman. On a comon sense estimator for the drift of a diffusion. Unpublished, Brown University, 1979.

[9] K. Itô and H.P. McKean. *Diffusion processes and their sample paths.* Springer-Verlag, Berlin, 1974. Second printing, corrected.

[10] I. Karatzas and S. E. Shreve. *Brownian Motion and Stochastic Calculus.* Springer-Verlag, 1991.

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

[12] Y. Kasahara, S. Kotani, and H. Watanabe. On the Green functions of 1-dimensional diffusion processes. *Publ. Res. Inst. Math. Sci.*, 16:175–188, 1980.

[13] Y. A. Kutoyants. On density estimation by the observations of ergodic diffusion processes. In *Statistics and control of stochastic processes (Moscow, 1995/1996)*, pages 253–274. World Sci. Publishing, River Edge, NJ, 1997.

[14] G. Moloche. Kernel regression for non-stationary Harris-recurrent processes. Unpublished, M.I.T., 2001.

[15] P.C.B. Phillips and J.Y. Park. Nonstationary kernel regression and density estimation. Unpublished, Cowles Commision, Yale University, 1999.

[16] D. Revuz and M. Yor. *Continuous Martingales and Brownian Motion.* Springer-Verlag, 1994.

[17] H. Tanaka. Certain limit theorems concerning one-dimensional diffusion processes. *Mem. Fac. Sci. Kyusyu Univ. Ser. A. Math.*, 12:1–11, 1958.

[18] S. Yakowitz. Nonparametric density and regression estimation for Markov sequences without mixing assumptions. *J. Multivariate Anal.*, 30:124–136, 1989.

[19] C. L. Zirbel. Mean occupation times of continuous one-dimensional Markov processes. *Stochastic Process. Appl.*, 69:161–178, 1997.