Bandwidth Selection and Asymptotic Properties of Local Nonparametric Estimators in Possibly Nonstationary Continuous-Time Models*

Yacine Aït-Sahalia[†]

Department of Economics
Princeton University
and NBER

Joon Y. Park[‡]

Department of Economics
Indiana University
and Sungkyunkwan University

This Version: November 9, 2015

Abstract

We derive the asymptotic properties of nonparametric estimators of the drift and diffusion functions, and the local time, of a discretely sampled diffusion process that is possibly nonstationary. We provide complete two-dimensional asymptotics in both the time span and the sampling interval, allowing for a precise characterization of their distribution, as well as the determination of optimal bandwidths for these estimators.

Keywords: Kernel estimators; locally linear estimators; diffusions; local time; nonstationarity.

*We are grateful to Peter Robinson, the co-editor, and two anonymous referees for helpful comments. We also express our thanks to Minsoo Jeong, Jihyun Kim and Bin Wang for correcting many typographical errors in earlier versions of this paper.

[†]Princeton, NJ 08544-1021. E-mail: yacine@princeton.edu. Research supported by the NSF under grant SES-0850533.

[‡]Bloomington, IN 47405-7104. E-mail: joon@indiana.edu.

1. Introduction

This paper derives the asymptotic properties of nonparametric estimators of diffusion models, including their local time, drift and diffusion functions. We consider kernel, or locally constant, as well as locally linear estimators. Nonparametric methods in the context of discrete-time nonstationary time series have been analyzed by, e.g., Jeganathan (1995), Karlsen and Tjøstheim (2001), Park and Phillips (2001), Karlsen et al. (2007) and Schienle (2008); their properties, including their rates of convergence, often differ from those obtained in the stationary case (Robinson (1983)).

Discretely sampling from a continuous-time process introduces new issues, and in particular the fact that two separate parameters emerge to control the sample size and the resulting asymptotic theory. The sample is obtained from a diffusion process observed at equidistant intervals of length Δ over a time span T . Our asymptotics encompass the case where both $\Delta \rightarrow 0$ and $T \rightarrow \infty$. Therefore they are two dimensional, in contrast to more conventional asymptotics relying only on the sample size $n = T/\Delta$.

The two dimensional asymptotics are useful for several reasons when one is interested in analyzing nonstationary diffusion processes. First, they provide a single framework to unify the limit theories for the stationary and nonstationary diffusion models. Many models used in practice in financial applications (such as the interest rate models Vasicek (1977) and Cox et al. (1985)) are based on stationary processes, but many are not (for example the option pricing model of Black and Scholes (1973)). And often, the same parametric diffusion model can be stationary or not depending upon the values of its parameters, including displaying different forms of nonstationarity such as unit root or explosive behaviors. Nonstationarity in diffusion models can take multiple and interesting aspects. For instance, nonlinearities in the drift and diffusion coefficients influence temporal dependence in diffusion models, either in the stationary mixing case (see Chen et al. (2010)) or in the nonstationary case. Because diffusion models are fully characterized by their drift and diffusion functions, all the relevant properties of the model can be fully characterized by specific properties of these two functions, sometimes separately and sometimes jointly.

Second, compared to asymptotics where either Δ or T is fixed, and $n \rightarrow \infty$, nonstationarity is neither revealed nor in fact relevant if T is fixed, while samples with fixed Δ typically do not provide sufficient information to fully identify nonparametrically important components of the discretely sampled process, such as its diffusion function. The drift term can be consistently estimated only when we have $T \rightarrow \infty$, whereas it is sufficient to have $\Delta \rightarrow 0$ to identify the diffusion function.

Third, using two-dimensional asymptotics, we may analyze the performance of estimators under various sampling schemes corresponding for instance to sampling frequently over a short period of time compared to relatively sparse sampling over a long time span. Some sampling schemes may be more realistic than others in some practical applications, or they may lead to a better approximation to the small sample properties of the estimators.

Fourth, two-dimensional asymptotics allow for the determination of an optimal bandwidth for each esti-

mator under consideration. Asymptotic theories for various semi- and nonparametric estimators of different characteristics of a diffusion model have been developed by several authors, including Florens-Zmirou (1993), Brugière (1993), Genon-Catalot et al. (1992), Aït-Sahalia (1996), Jacod (1998), Jacod (2000), Jiang and Knight (1997), Corradi and White (1999), Hoffmann (1999), Bandi and Phillips (2003), Fan and Zhang (2003), Fan et al. (2003), Höpfner and Kutoyants (2003) and Renò (2008); see Cai and Hong (2003) and Fan (2005) for reviews.

Fifth, our asymptotic results are sufficiently fine to provide a characterization of the bias and variance terms for all these estimators. For example, we will see that, compared to kernel estimators, locally linear estimators eliminate the bias terms involving the second-order derivatives of the estimated functions, exactly as in the usual stationary case.

Our framework and approach are most closely related to those of Bandi and Phillips (2003) and Bandi and Phillips (2010), who establish the asymptotic properties of the kernel and locally polynomial estimators for the drift and diffusion functions of diffusion models, as well as their local time, in the nonstationary case. We extend their results in the following direction: our asymptotics allow the time span T , the sampling interval Δ and the bandwidth h to change simultaneously, after imposing appropriate conditions on how they may change. By contrast, the sequential asymptotics take limits sequentially in different orders of h , Δ and T . Besides estimators of the drift and diffusion functions, we also develop the asymptotic properties of the local time estimator, and the locally linear in addition to the kernel or locally constant estimators under this setting. Perhaps most importantly, the two-dimensional analysis in this paper yields the optimal bandwidth choices for all the nonparametric estimators considered. Bandwidth selection cannot be made on the basis of a sequential asymptotic analysis, since the optimal bandwidth h is generally obtained as a function of T and Δ , hence the need for the two-dimensional asymptotics we consider here.¹

Bandwidth selection in the context of nonparametric estimation for nonstationary continuous-time models is not a trivial endeavor. The nonstationary continuous-time case adds a layer of complications to both the stationary case and the discrete-time case. First, relative to the stationary case, where the optimal bandwidth is typically of the form cT^a where c and a are constant, the optimal bandwidth in the nonstationary case is now of the form $c\ell(T, x)^a$ where $\ell(T, x)$ is the local time on $[0, T]$ of the process at level x . The difficulty comes from the fact that $\ell(T, x)$ needs to be estimated, typically using nonparametric methods, themselves involving a bandwidth. Second, relative to the discrete-time case, where the sampling interval is irrelevant, the optimal bandwidth must be adapted to the time interval Δ at which the data are sampled since the asymptotics allow for the fact that $\Delta \rightarrow 0$. In the paper, we simply suggest to use consistent estimates of the optimal bandwidths that minimize the asymptotic mean squared errors.

The paper is organized as follows. Section 2 sets up the model and its properties, and introduces the assumptions. Section 3 concerns the nonparametric estimation of the local time and the kernel density estimator,

¹In fact, Bandi and Phillips (2010) (Section 3.2.1, p.162) note that that the “optimal bandwidth selection is technically very demanding in these models and represents an open field of research, no rigorous treatment being available at present (...)”

Section 4 the local estimation of the drift and diffusion functions of the process. Proofs are contained in Section 5.

2. The Model and Preliminaries

Consider the time-homogenous stochastic differential equation

$$dX_t = \mu(X_t) dt + \sigma(X_t) dW_t, \quad (2.1)$$

where μ and σ^2 are, respectively, the drift and diffusion functions. We let $D = (\underline{x}, \bar{x})$ denote the domain of the diffusion X . In general, $D = (-\infty, +\infty)$, but in many examples in finance, X is the price of an asset with limited liability (stock, foreign currency, bond, etc.) or a nominal interest rate, in which case $D = (0, +\infty)$. We start by making the following assumption.

Assumption 1. *We assume that (a) $\sigma^2(x) > 0$ for all x in D , and (b) $\mu(x)$ and $\sigma^2(x)$ are twice continuously differentiable for all x in D .*

We require that Assumption 1 hold through the paper. In particular, Assumption 1 ensures that the stochastic differential equation (2.1) has a weak solution which is unique. This is well known.

In the development of our theory, we suppose that the diffusion X is observed at intervals of length Δ over time $[0, T]$. The sample size n is, therefore, given by $n = T/\Delta$. In what follows, we let

$$X_\Delta, X_{2\Delta}, \dots, X_{n\Delta}$$

be observed. Here, we assume that the observation intervals are equally spaced. However, this is mainly for expositional simplicity. All of our subsequent results can easily be extended to more general cases where X are observed irregularly, i.e., at $\Delta_1, \Delta_1 + \Delta_2, \dots$ and so on, as long as the maximum of the observation intervals, $\max_{1 \leq i \leq n} \Delta_i$, decreases to zero and the observation intervals are either deterministic or, if stochastic, independent of the X process. For the development of our asymptotics, we assume that $\Delta \rightarrow 0$ sufficiently fast relative to T , which we allow to be fixed at $\bar{T}$, $T = \bar{T}$, or to increase up to infinity, $T \rightarrow \infty$. This is necessary to establish a unified asymptotic theory for stationary and nonstationary diffusions. In either case, we have $n \rightarrow \infty$.

Our subsequent theory heavily relies on the local time ℓ of X at an interior point x of D , which is defined by

$$\ell(T, x) = \lim_{\varepsilon \rightarrow 0} \frac{1}{2\varepsilon} \int_0^T 1\{|X_s - x| < \varepsilon\} ds,$$

at time T . Roughly speaking, $\ell(T, x)$ denotes the amount of time spent by X in a neighborhood of x for the time interval $[0, T]$. For this reason, it is often called the sojourn time. It is more common in the stochastic processes literature to define the local time to be the sojourn time measured by the quadratic variation of the underlying process X . However, our definition is more appropriate for the theory developed in this paper. It is

well known that the local time ℓ exists with a version continuous a.s. in both arguments for a general Markov process. See, e.g., Eisenbaum and Kaspi (2007). The local time ℓ allows us to have

$$\int_0^T f(X_s) ds = \int_{-\infty}^{\infty} f(x) \ell(T, x) dx \quad (2.2)$$

for any nonnegative measurable function f on R , which is referred to the occupation times formula. In what follows, we only consider $x \in D$ such that $\ell(T, x) > 0$.

Assumption 2. *We assume that (a) there exists $\varepsilon > 0$ such that $\bar{\ell}_\varepsilon(T, x) = O_p(\ell(T, x)^2)$, where $\bar{\ell}_\varepsilon(T, x) = \sup_{|x-y| \leq \varepsilon} \ell(T, y)$, and (b) $\ell(T, x) = O_p(\kappa_T)$ for some nonrandom sequence (κ_T) .*

The condition (a) in Assumption 2 controls the divergence rate of local time in the neighborhood of a spatial point. It is not stringent, and appears to be satisfied by a wide class of diffusion processes. Of course, it clearly holds if we let $T = \bar{T}$ fixed. The condition is satisfied by Brownian motion, in which case we have $\sup_{x \in R} \ell(T, x)/T^{1/2} \log \log T \rightarrow_{a.s.} 1$ as $T \rightarrow \infty$, due to the well known law of the iterated logarithm for Brownian local time. See, e.g., Theorem 2.1 (p.20) of Borodin (1989). The condition is introduced mainly to simplify our exposition, representing the orders of error terms solely as functions of $\ell(T, x)$. If we do not impose the condition, the orders of error terms in our subsequent results are determined by $\bar{\ell}_\varepsilon(T, x)$ as well as $\ell(T, x)$. The condition (b) of Assumption 2 is also fairly weak and holds for general diffusion processes. It is trivially satisfied if $T = \bar{T}$ fixed. In general, the asymptotic behavior of $\ell(T, x)$ as $T \rightarrow \infty$ is determined by the recurrence property of the diffusion X , as we will explain in detail below.

Under Assumption 1, it is straightforward to characterize the recurrence property of X . Following standard practice, we introduce the scale function s , whose derivative s' is given by

$$s'(x) = \exp \left[-2 \int^x \frac{\mu(y)}{\sigma^2(y)} dy \right],$$

on D . Then X becomes recurrent if and only if $s(\underline{x}) = -\infty$ and $s(\bar{x}) = \infty$. A process which is not recurrent is said to be transient. Moreover, we define the speed density of X , which is given by

$$m(x) = \frac{1}{(\sigma^2 s')(x)}$$

on D . A recurrent process X becomes positive recurrent if $\int_D m(x) dx < \infty$, and otherwise it becomes null recurrent. A positive recurrent process has a time invariant stationary marginal distribution and becomes stationary, if the process starts at its stationary distribution, whereas a null recurrent process or a transient process is nonstationary having no such time invariant distribution.

For a recurrent process, we have $\ell(T, x) \rightarrow \infty$ as $T \rightarrow \infty$ at each $x \in D$. In the case of positive recurrent diffusions, we have

$$\frac{\ell(T, x)}{T} \rightarrow_{a.s.} \pi(x) \quad (2.3)$$

as $T \rightarrow \infty$, as shown in, e.g., Theorem 6.3 in Bosq (1998), where

$$\pi(x) = \frac{m(x)}{\int_D m(y) dy} \quad (2.4)$$

is the time invariant stationary density. Moreover, for a wide class of null recurrent diffusion processes, $\ell(T, x)/T^r l(T)$ weakly converges as $T \rightarrow \infty$ to a nondegenerate limit distribution, and therefore, we have $\ell(T, x) = O_p(T^r l(T))$, for each x in D with some $r \in (0, 1]$ and slowly varying function l at infinity. In particular, we have $\ell(T, x) = O_p(T^{1/2})$ for Brownian motion. The reader is referred to Höpfner and Löcherbach (2003) for more details. Of course, $\ell(T, x)$ does not diverge as T for transient processes.

Assumption 3. *We assume that K is bounded, twice continuously differentiable and has support $[-1, 1]$, and that it has the properties $\int_{-\infty}^{\infty} K(x) dx = 1$ and $\int_{-\infty}^{\infty} xK(x) dx = 0$.*

The kernel conditions we introduced in Assumption 3 are standard, except for the boundedness of support. Though it does not seem to be essential, the boundedness assumption greatly simplifies the proofs of theorems in the paper.

In our theory, we require that $\Delta > 0$ be sufficiently small relative to the extremal bounds of various functional transforms of X over time interval $[0, T]$. Indeed, we define

$$T(f) = \max_{0 \leq t \leq T} |f(X_t)|$$

for a function f defined from the drift and diffusion functions μ and σ^2 of X . For several positive recurrent processes that are used commonly in economic and financial applications, the exact order of extremal process is known: see Borkovec and Klüppelberg (1998). For instance, the extremal processes of the Ornstein-Uhlenbeck process and the Feller's square root process are respectively of orders $O_p(\sqrt{\log T})$ and $O_p(\log T)$, and the extremal process of the general C.E.V. process is of order $O_p(T)$. Moreover, for general null recurrent processes in natural scale, i.e., $(s(X_t))$, the extremal process is known to be of order $O_p(T_r)$, where T_r is the T -inverse of $T^{r+2}l(T)$ for some $r > -1$ and slowly varying function l at infinity. Note that $T_r = o(T)$ for all $r > -1$. The reader is referred to Stone (1963) and Jeong and Park (2011) for details.

For the subsequent development of our asymptotics, we let $\mathcal{D}$ be the differential operator, and define

$$\mathcal{A} = \mu\mathcal{D} + \frac{1}{2}\sigma^2\mathcal{D}^2, \quad \mathcal{B} = \sigma\mathcal{D}.$$

If we define $f_A = \mathcal{A}f = \mu f' + \sigma^2 f''/2$ and $f_B = \mathcal{B}f = \sigma f'$ for a twice continuously differentiable function f , then it follows from Itô's formula that

$$f(X_t) - f(X_s) = \int_s^t f_A(X_u) du + \int_s^t f_B(X_u) dW_u$$

for any $0 \leq s \leq t$. In this representation, $\int_s^t f_A(X_u) du$ and $\int_s^t f_B(X_u) dW_u$ represent, respectively, the bounded variation and martingale parts of $f(X_t) - f(X_s)$. The representation will be used repetitively to derive the asymptotics in the paper. The operator $\mathcal{A}$ is called the infinitesimal generator of the diffusion X .

Finally, all our asymptotics in the paper are joint, not sequential, in bandwidth h , sampling interval Δ and time span T . Throughout the paper, we let $h \rightarrow 0$ and $\Delta \rightarrow 0$, while allowing T to be either fixed at $T = \bar{T}$ or $T \rightarrow \infty$, in such a way that they satisfy certain conditions. Our asymptotics jointly hold for any sequences

of h , Δ and T as long as they satisfy our conditions. Some of our asymptotic results are obtained without any additional conditions except that $h \rightarrow 0$ or $h \rightarrow 0$ jointly with $\Delta \rightarrow 0$, as in Theorem 1 and Lemmas 1-10 and 12. Others are established under various additional conditions for h , Δ and T such as those in Assumptions 4 and 5, and Lemmas 11 and 13. In the development of our asymptotics, we let $A = o_p(\ell(T, x))$ uniformly in T as $h \rightarrow 0$ or $h \rightarrow 0$ jointly with $\Delta \rightarrow 0$ if and only if we may write $A = a_h \ell(T, x)$ or $a_{h,\Delta} \ell(T, x)$ for all T with a sequence a_h of h such that $a_h \rightarrow_p 0$ as $h \rightarrow 0$ or with a sequence $a_{h,\Delta}$ of h and Δ such that $a_{h,\Delta} \rightarrow 0$ as $h \rightarrow 0$ and $\Delta \rightarrow 0$, and similarly, $B = O_p(\ell(T, x))$ uniformly in T as $h \rightarrow 0$ or $h \rightarrow 0$ jointly with $\Delta \rightarrow 0$ if and only if we may write $B = b_h \ell(T, x)$ or $b_{h,\Delta} \ell(T, x)$ for all T with a sequence b_h of h stochastically bounded or with a sequence $b_{h,\Delta}$ of h and Δ stochastically bounded. Moreover, we let $A = o_p(a_h \ell(T, x))$ and $B = O_p(b_h \ell(T, x))$ uniformly in T as $h \rightarrow 0$ if and only if $A/a_h = o_p(\ell(T, x))$ and $B/b_h = O_p(\ell(T, x))$ uniformly in T as $h \rightarrow 0$ for some sequences a_h and b_h of h , and similarly, let $A = o_p(a_{h,\Delta} \ell(T, x))$ and $B = O_p(b_{h,\Delta} \ell(T, x))$ uniformly in T as $h \rightarrow 0$ and $\Delta \rightarrow 0$ if and only if $A/a_{h,\Delta} = o_p(\ell(T, x))$ and $B/b_{h,\Delta} = O_p(\ell(T, x))$ uniformly in T as $h \rightarrow 0$ and $\Delta \rightarrow 0$, for some sequences $a_{h,\Delta}$ and $b_{h,\Delta}$ of h and Δ .

3. Asymptotics for the Local Time and Kernel Density Estimators

In this section, we consider the nonparametric inference on local time. Under stationarity, time invariant and stable marginal density can be obtained by taking T -limit to the averaged local time, as stated in (2.3). For nonstationary processes, local time remains to be random and path-dependent. For both cases, the local time plays a central role in the nonparametric inference on diffusions. We consider

$$\hat{\ell}(T, x) = \frac{\Delta}{h} \sum_{i=1}^n K\left(\frac{X_{i\Delta} - x}{h}\right),$$

where K is the kernel function and h is the bandwidth parameter, which is commonly used for the nonparametric estimation of the density with standardizing factor $1/nh$ in place of Δ/h . We also define

$$\tilde{\ell}(T, x) = \frac{1}{h} \int_0^T K\left(\frac{X_t - x}{h}\right) dt,$$

which is the continuous version of $\hat{\ell}(T, x)$. We call $\hat{\ell}(T, x)$ and $\tilde{\ell}(T, x)$ respectively the discrete and continuous sample kernel estimators for local time. Throughout the paper, we allow the bandwidth h to be random and given as a function of $\ell(T, x)$.

Theorem 1. *Let Assumptions 1, 2 and 3 hold. Then we have*

$$\hat{\ell}(T, x) = \tilde{\ell}(T, x) + O_p(h^{-2} \Delta \ell(T, x)) \quad (3.1)$$

for every $x \in D$ uniformly in T as $h \rightarrow 0$ and $\Delta \rightarrow 0$. Furthermore, conditional on $\ell(T, x)$, we have

$$\begin{aligned} \tilde{\ell}(T, x) =_d & \ell(T, x) + \iota(K_2)\tau(x)h^2\ell(T, x) + 2\iota(K_1^2)^{1/2}(1/\sigma)(x)(h\ell(T, x))^{1/2}Z \\ & + o_p(h^2\ell(T, x)) + o_p(h^{1/2}\ell(T, x)^{1/2}) \end{aligned} \quad (3.2)$$

for every $x \in D$, uniformly in T as $h \rightarrow 0$, where $\iota(K_2) = \int u^2 K(u) du$, $\iota(K_1^2) = \int u^2 K(u)^2 du$,

$$\tau(x) = \left(\frac{2\mu^2}{\sigma^4} - \frac{2\mu + 4\mu\sigma'}{\sigma^3} + \frac{\mu' - \sigma\sigma'' + 3\sigma'^2}{\sigma^2} \right) (x) \quad (3.3)$$

and Z is a standard normal random variate.

Remarks (a) The continuous sample kernel estimator $\tilde{\ell}(T, x)$ is consistent as long as $h \rightarrow_p 0$. For the consistency of the discrete sample kernel estimator $\hat{\ell}(T, x)$, however, we require an additional condition $h^{-2}\Delta\ell(T, x) \rightarrow_p 0$. If $T = \bar{T}$ fixed, the required condition reduces to $h^{-2}\Delta \rightarrow_p 0$, which corresponds to the condition $(nh^2)^{-1} \log n \rightarrow 0$ used by Florens-Zmirou (1993) to establish the strong consistency. Of course, the consistency holds under the same condition also for transient processes for which $\ell(T, x) < \infty$ for all T .

(b) We employ the notation (3.1) rather than the more standard

$$\frac{\hat{\ell}(T, x)}{\ell(T, x)} = \frac{\tilde{\ell}(T, x)}{\ell(T, x)} + O_p(h^{-2}\Delta) \quad (3.4)$$

because in the case where the terms present in the O_p term are more involved, as in (3.3) for example, then dividing by them throughout the equation to achieve form (3.4) makes the equations in that form more cumbersome to read.

(c) The error due to discrete sampling, which is of order $O_p(h^{-2}\Delta\ell(T, x))$, becomes unimportant for the local time inference when Δ is sufficiently small relative to h . In this case, the discrete and continuous sample kernel local time estimators, $\hat{\ell}(T, x)$ and $\tilde{\ell}(T, x)$, become asymptotically identical.

(d) The continuous sample kernel estimator $\tilde{\ell}(T, x)$ has two error terms, which are of orders respectively $O_p(h^2\ell(T, x))$ and $O_p(h^{1/2}\ell(T, x)^{1/2})$. Both terms, of course, disappear as h decreases down to zero. The first and second error terms can be regarded respectively as the bias and variance terms, conditional on $\ell(T, x)$. They play similar roles as the bias and variance terms in the standard analysis of kernel density estimator. The bias term dominates the variance term if h is chosen such that $h\ell(T, x)^{1/3} \rightarrow_p \infty$. The variance term, on the other hand, gets larger than the bias term if we set h smaller. As in the case of the usual kernel density estimation, the bias term can be made arbitrarily small if higher order kernels are used. In fact, we may show that the bias term can be made to be of order $O_p(h^k\ell(T, x))$ if the k -th order kernel is used. This can be easily seen in the proof of Theorem 1.

(e) Even though it has two error terms that we may interpret as the bias and variance terms, we may not define the optimal bandwidth parameter for the continuous sample kernel estimator $\tilde{\ell}(T, x)$. Both terms decrease monotonically down to zero as $h \rightarrow 0$, and there is no trade-off between bias and variance. In fact, the optimal bandwidth parameter minimizing the mean squared error of $\tilde{\ell}(T, x)$ is trivially given by $h^* = 0$.

(f) The choice of bandwidth for the discrete sample kernel estimator $\hat{\ell}(T, x)$ should consider Δ as well as $\ell(T, x)$, due to the discrete sampling error that is of order $O_p(h^{-2}\Delta\ell(T, x))$. First, we let

$$\Delta^3\ell(T, x)^4 \rightarrow_p 0. \quad (3.5)$$

In this case, the optimal bandwidth which balances off the discrete sampling error $\hat{\ell}(T, x) - \tilde{\ell}(T, x)$ and the bias and variance terms in $\tilde{\ell}(T, x)$ is given by

$$h^* = c \Delta^{2/5} \ell(T, x)^{1/5} \quad (3.6)$$

for some constant $c > 0$. Under condition (3.5), we have $h^* \ell(T, x)^{1/3} \rightarrow_p 0$ for the choice of bandwidth in (3.6) and the variance term dominates in $\tilde{\ell}(T, x)$. As can be easily seen, the bandwidth h^* make identical the orders of the discrete sampling error and the variance term in the continuous sample kernel estimator. Similarly, we may show that $h^* = c \Delta^{1/4}$ with some constant $c > 0$ balances off the error terms in $\hat{\ell}(T, x) - \tilde{\ell}(T, x)$ and $\tilde{\ell}(T, x)$, if $\Delta^3 \ell(T, x)^4 \rightarrow_p \infty$.

(g) For stationary diffusions,

$$\hat{\pi}(x) = \frac{\hat{\ell}(T, x)}{T} = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_{i\Delta}}{h}\right) \quad (3.7)$$

reduces to the usual kernel density estimator based on the sample $(X_{i\Delta})$. Our results in Theorem 1 may therefore be used to analyze $\hat{\pi}(x)$. For instance, we may easily deduce that

$$\hat{\pi}(x) = \frac{\ell(T, x)}{T} + O_p(h^{-2}\Delta) + O_p((h/T)^{1/2}) + O_p(h^2). \quad (3.8)$$

As discussed in (2.3), $\ell(T, x)/T \rightarrow_{a.s.} \pi(x)$ as $T \rightarrow \infty$, where π is the time invariant stationary density defined in (2.4). Therefore, $\hat{\pi}(x)$ becomes consistent for $\pi(x)$, if $h^{-2}\Delta \rightarrow 0$.

4. Nonparametric Estimation of the Drift and Diffusion Functions

Estimators of this type for diffusion processes have been studied in the stationary case by Stanton (1997) and Fan and Zhang (2003) and in the nonstationary case by Bandi and Phillips (2003).

4.1. Nonparametric Kernel Estimation of Drift Function

It follows from (2.1) that

$$\begin{aligned} X_{i\Delta} - X_{(i-1)\Delta} &= \Delta \mu(X_{(i-1)\Delta}) + \int_{(i-1)\Delta}^{i\Delta} \sigma(X_t) dW_t + \int_{(i-1)\Delta}^{i\Delta} [\mu(X_t) - \mu(X_{(i-1)\Delta})] dt \\ &\approx \Delta \mu(X_{(i-1)\Delta}) + \int_{(i-1)\Delta}^{i\Delta} \sigma(X_t) dW_t \end{aligned} \quad (4.1)$$

for small $\Delta > 0$. Therefore, we may estimate the drift function μ by the kernel regression of $(X_{i\Delta} - X_{(i-1)\Delta})/\Delta$ on $X_{(i-1)\Delta}$, $i = 1, \dots, n$, if $\Delta > 0$ is small. Below we consider the locally constant estimator and the locally linear estimator in sequel. To develop the asymptotics of the estimators, we need the following assumption.

Assumption 4. We assume $h \rightarrow_p 0$ and $\Delta \rightarrow 0$ such that (a) $h^{-4}\Delta = o_p(1)$, and (b) $\Delta^{1/2}T(\mu_A) = o_p(1)$, $\Delta^{1/2}T(\mu_B) = o_p((h\ell(T, x)^{1/2}))$ uniformly in T as $h \rightarrow 0$ and $\Delta \rightarrow 0$.

In Assumption 4, we assume neither $T \rightarrow \infty$ nor $\ell(T, x) \rightarrow_p \infty$. We only require that $h \rightarrow_p 0$ and that $\Delta \rightarrow 0$ sufficiently fast relative to h and $T(\mu_A)$ and $T(\mu_B)$. Of course, if $T = \bar{T}$ fixed, then the conditions in part (b) hold trivially as long as $\Delta \rightarrow 0$. In this case, the condition in part (a) reduces to $nh^4 \rightarrow \infty$.

4.1.1 Locally Constant Estimator

The locally constant estimator for $\mu(x)$ is given by

$$\hat{\mu}(x) = \frac{P_T(K, \mu)}{Q_T(K)},$$

where

$$P_T(K, \mu) = \frac{1}{h} \sum_{i=1}^n K\left(\frac{X_{(i-1)\Delta} - x}{h}\right) (X_{i\Delta} - X_{(i-1)\Delta}) \quad (4.2)$$

$$Q_T(K) = \frac{\Delta}{h} \sum_{i=1}^n K\left(\frac{X_{(i-1)\Delta} - x}{h}\right). \quad (4.3)$$

According to the representation (4.1), we write

$$\hat{\mu}(x) = \hat{\mu}_p(x) + \hat{\mu}_q(x) + \hat{\varepsilon}_\mu(x), \quad (4.4)$$

where

$$\hat{\mu}_p(x) = \mu(x) + \frac{N_T(K, \mu)}{Q_T(K)}, \quad \hat{\mu}_q(x) = \frac{M_T(K, \mu)}{Q_T(K)}$$

with

$$N_T(K, \mu) = \frac{\Delta}{h} \sum_{i=1}^n K\left(\frac{X_{(i-1)\Delta} - x}{h}\right) [\mu(X_{(i-1)\Delta}) - \mu(x)] \quad (4.5)$$

$$M_T(K, \mu) = \frac{1}{h} \sum_{i=1}^n K\left(\frac{X_{(i-1)\Delta} - x}{h}\right) \int_{(i-1)\Delta}^{i\Delta} \sigma(X_t) dW_t, \quad (4.6)$$

and

$$\hat{\varepsilon}_\mu(x) = \frac{R_T(K, \mu)}{Q_T(K)}$$

with

$$R_T(K, \mu) = \frac{1}{h} \sum_{i=1}^n K\left(\frac{X_{(i-1)\Delta} - x}{h}\right) \int_{(i-1)\Delta}^{i\Delta} [\mu(X_t) - \mu(X_{(i-1)\Delta})] dt. \quad (4.7)$$

We may view (4.4) as an approximate Doob decomposition. As is clearly seen, $\hat{\mu}_p(x)$ and $\hat{\mu}_q(x)$ signify the conditional mean and martingale terms, respectively, of the locally constant estimator $\hat{\mu}(x)$. Moreover, $\hat{\varepsilon}_\mu(x)$ represents the approximation error in the representation. As we show below, $\hat{\varepsilon}_\mu(x)$ becomes negligible asymptotically.

Theorem 2. *Let Assumptions 1, 2, 3 and 4 hold. Then we have*

$$\hat{\mu}_p(x) = \mu(x) + \frac{h^2}{2} \iota(K_2) \left(\mu''(x) + 2\mu'(x) \frac{m'(x)}{m(x)} \right) + o_p(h^2) + O_p(h^{3/2} \ell(T, x)^{-1/2})$$

uniformly in T as $h \rightarrow 0$ and $\Delta \rightarrow 0$, and

$$[h\ell(T, x)]^{1/2} \hat{\mu}_q(x) \rightarrow_d \sigma(x) \iota(K^2)^{1/2} Z,$$

where $\iota(K_2) = \int x^2 K(x) dx$, $\iota(K^2) = \int K^2(x) dx$ and Z is a standard normal random variate independent of $\ell(T, x)$. Moreover, we have

$$\hat{\varepsilon}_\mu(x) = o_p(h^2),$$

which becomes negligible asymptotically.

4.1.2 Locally Linear Estimator

The locally linear estimator for $\mu(x)$ is given by

$$\tilde{\mu}(x) = \frac{P_T(K, \mu)Q_T(K_2) - P_T(K_1, \mu)Q_T(K_1)}{Q_T(K)Q_T(K_2) - Q_T(K_1)^2},$$

where $P_T(K_1, \mu)$ is defined as $P_T(K, \mu)$ in (4.2), and $Q_T(K_1)$ and $Q_T(K_2)$ as $Q_T(K)$ in (4.3), only with K replaced by K_1 and K_2 , $K_1(x) = xK(x)$ and $K_2(x) = x^2K(x)$.

Similarly as in (4.4), we may decompose $\tilde{\mu}(x)$ as

$$\tilde{\mu}(x) = \tilde{\mu}_p(x) + \tilde{\mu}_q(x) + \tilde{\varepsilon}_\mu(x),$$

where $\tilde{\mu}_p(x)$ and $\tilde{\mu}_q(x)$ are respectively the conditional mean and martingale terms of $\tilde{\mu}(x)$, which are given by

$$\begin{aligned} \tilde{\mu}_p(x) &= \mu(x) + \frac{N_T(K, \mu)Q_T(K_2) - N_T(K_1, \mu)Q_T(K_1)}{Q_T(K)Q_T(K_2) - Q_T(K_1)^2} \\ \tilde{\mu}_q(x) &= \frac{M_T(K, \mu)Q_T(K_2) - M_T(K_1, \mu)Q_T(K_1)}{Q_T(K)Q_T(K_2) - Q_T(K_1)^2} \end{aligned}$$

with $N_T(K_1, \mu)$ and $M_T(K, \mu)$ defined as $N_T(K, \mu)$ and $M_T(K, \mu)$ in (4.5) and (4.6) respectively with K_1 , $K_1(x) = xK(x)$, in place of K , and $\tilde{\varepsilon}_\mu(x)$ is the approximation error term given by

$$\tilde{\varepsilon}_\mu(x) = \frac{R_T(K, \mu)Q_T(K_2) - R_T(K_1, \mu)Q_T(K_1)}{Q_T(K)Q_T(K_2) - Q_T(K_1)^2},$$

with $R_T(K_1, \mu)$ defined as $R_T(K, \mu)$ in (4.7) with the replacement of K by K_1 , $K_1(x) = xK(x)$.

Theorem 3. *Let Assumptions 1, 2, 3 and 4 hold. Then we have*

$$\tilde{\mu}_p(x) = \mu(x) + \frac{h^2}{2} \iota(K_2) \mu''(x) + o_p(h^2) + O_p(h^{3/2} \ell(T, x)^{-1/2})$$

uniformly in T as $h \rightarrow 0$ and $\Delta \rightarrow 0$, and

$$[h\ell(T, x)]^{1/2} \tilde{\mu}_q(x) \rightarrow_d \sigma(x) \iota(K^2)^{1/2} Z,$$

where $\iota(K_2) = \int x^2 K(x) dx$, $\iota(K^2) = \int K^2(x) dx$ and Z is a standard normal random variate independent of $\ell(T, x)$. Moreover, we have

$$\tilde{\varepsilon}_\mu(x) = o_p(h^2),$$

which becomes negligible asymptotically.

4.1.3 Remarks

(a) Note that $\hat{\mu}_q(x), \tilde{\mu}_q(x) \rightarrow_p \infty$ unless $\ell(T, x) \rightarrow_p \infty$. In particular, $\hat{\mu}(x)$ and $\tilde{\mu}(x)$ become consistent only if we have $T \rightarrow \infty$ and $\ell(T, x) \rightarrow_p \infty$. For instance, it becomes inconsistent if either $T = \bar{T}$ fixed or $\ell(T, x)$ is bounded, the underlying diffusion being transient.

(b) For the locally linear estimator $\tilde{\mu}(x)$, the bias term has the same expression as for the standard nonparametric regression model. The bias term of the locally constant estimator $\hat{\mu}(x)$ is also identical to that in the standard model if the underlying diffusion is stationary. Note that

$$\frac{m'(x)}{m(x)} = \frac{\pi'(x)}{\pi(x)}$$

in this case, as follows immediately from (2.4). If $\mu(x)$ is linear in x , then the leading bias term in the locally linear estimator disappears as in the standard nonparametric regression.

(c) Let $\ell(T, x) \rightarrow_p \infty$. In this case, the optimal bandwidth that minimizes the mean squared error of the leading terms is given respectively for the locally constant estimator $\hat{\mu}(x)$ and the locally linear estimator $\tilde{\mu}(x)$ by

$$\begin{aligned} h^*(\hat{\mu}) &= c(K) \sigma^{2/5}(x) \left(\mu''(x) + 2\mu'(x) \frac{m'(x)}{m(x)} \right)^{-2/5} \ell(T, x)^{-1/5} \\ h^*(\tilde{\mu}) &= c(K) \sigma^{2/5}(x) \mu''(x)^{-2/5} \ell(T, x)^{-1/5} \end{aligned}$$

with

$$c(K) = \frac{\iota(K^2)^{1/5}}{\iota(K_2)^{2/5}}. \quad (4.8)$$

In particular, h^* , $h^* = h^*(\hat{\mu})$ or $h^*(\tilde{\mu})$, is given as a function of local time by $h^* = c \ell(T, x)^{-1/5}$ for some constant c . For the stationary diffusions, we thus have $h^* = c T^{-1/5}$, as is in the standard case. For Brownian motions, on the other hand, we have $h^* = c T^{-1/10}$. In general, h^* is larger for a less recurrent process, which has a more diffuse spatial distribution. If we set

$$h = c \ell(T, x)^r$$

for some constant $c > 0$, the bias or variance term of both $\hat{\mu}(x)$ and $\tilde{\mu}(x)$ becomes dominant depending upon whether $r > -1/5$ or $r < -1/5$. Note that the optimal bandwidths $h^*(\hat{\mu})$ and $h^*(\tilde{\mu})$ can be estimated using an estimate for $\ell(T, x)$ and estimates for μ and σ^2 and their derivatives. Though we do not provide the details, it is well expected that our asymptotics are applicable also for the local constant or linear estimator with an estimated optimal bandwidth. Note that our asymptotics allow h to be given as a function of $\ell(T, x)$, and that Theorem 1 implies

$$\hat{\ell}(T, x) = \ell(T, x) \left[1 + O_p(h^{-2}\Delta) + O_p(h^2) + O_p(h^{1/2}\ell(T, x)^{-1/2}) \right], \quad (4.9)$$

from which we have $\hat{\ell}(T, x) = \ell(T, x)[1 + o_p(1)]$ as $h \rightarrow_p 0$ and $h^{-2}\Delta \rightarrow_p 0$.

(d) Assume $\ell(T, x) \rightarrow_p \infty$ and define

$$\lambda^2(h, T) = \frac{h\ell(T, x)}{\sigma^2(x)\iota(K^2)}.$$

If we set $h = c\ell(T, x)^r$ with $r < -1/5$ for some constant $c > 0$, then it follows that

$$\lambda(h, T) [\hat{\mu}(x) - \mu(x)], \lambda(h, T) [\tilde{\mu}(x) - \mu(x)] \rightarrow_d Z. \quad (4.10)$$

Since Z is independent of $\ell(T, x)$, we may deduce from (4.10) that both $\hat{\mu}(x)$ and $\tilde{\mu}(x)$ are distributed approximately as a normal mixture with mixing variate $\lambda^2(h, T)$. For $h = c\ell(T, x)^{-1/5}$, on the other hand, we have

$$\lambda(h, T) [\hat{\mu}(x) - (\mu + \mu_a + \mu_b)(x)], \lambda(h, T) [\tilde{\mu}(x) - (\mu + \mu_a)(x)] \rightarrow_d Z, \quad (4.11)$$

where $\mu_a(x) = (h^2/2)\iota(K_2)\mu''(x)$ and $\mu_b(x) = h^2\iota(K_2)[(m'/m)\mu'](x)$, and again the distributions of $\hat{\mu}(x)$ and $\tilde{\mu}(x)$ are a normal mixture with mixing variate $\lambda^2(h, T)$. Note that, due to (4.9), we may replace $\ell(T, x)$ in the definition of $\lambda(T, x)$ by $\hat{\ell}(T, x)$ without changing the limit distribution in (4.9) and (4.10) as long as $h \rightarrow_p 0$ and $h^{-2}\Delta \rightarrow_p 0$.

4.2. Nonparametric Kernel Estimation of the Diffusion Function

Using Itô's formula, we may easily deduce from (2.1) that

$$\begin{aligned} (X_{i\Delta} - X_{(i-1)\Delta})^2 &= \Delta\sigma^2(X_{(i-1)\Delta}) + 2 \int_{(i-1)\Delta}^{i\Delta} (X_t - X_{(i-1)\Delta})\sigma(X_t) dW_t \\ &\quad + 2 \int_{(i-1)\Delta}^{i\Delta} (X_t - X_{(i-1)\Delta})\mu(X_t) dt + \int_{(i-1)\Delta}^{i\Delta} [\sigma^2(X_t) - \sigma^2(X_{(i-1)\Delta})] dt \\ &\approx \Delta\sigma^2(X_{(i-1)\Delta}) + 2 \int_{(i-1)\Delta}^{i\Delta} (X_t - X_{(i-1)\Delta})\sigma(X_t) dW_t \end{aligned} \quad (4.12)$$

for small $\Delta > 0$. Therefore, we may estimate the diffusion function σ^2 by the kernel regression of $(X_{i\Delta} - X_{(i-1)\Delta})^2/\Delta$ on $X_{(i-1)\Delta}$, $i = 1, \dots, n$, if $\Delta > 0$ is small. Subsequently we consider each of the locally constant estimator and the locally linear estimator. To develop the asymptotics of the estimators, we need to introduce a new set of assumptions. In the following assumption, we define f_j on D by $f_j(x) = x^j f(x)$ for $j = 1, 2$ for f defined on D .

Assumption 5. We assume $h \rightarrow_p 0$ and $\Delta \rightarrow 0$ such that (a) $h^{-4}\Delta = o_p(1)$, (b) $\Delta^{1/2}T(\sigma_A^2) = o_p(1)$, $\Delta^{1/2}T(\sigma_B^2) = o_p((h\ell(T, x)^{1/2})$ uniformly in T as $h \rightarrow 0$ and $\Delta \rightarrow 0$, (c) $\Delta T(\mu_{jA}), \Delta T(\sigma_{jA}) = o_p(1)$, $\Delta T(\mu_{jB}), \Delta T(\sigma_{jB}) = o_p((h\ell(T, x))^{1/2})$ uniformly in T as $h \rightarrow 0$ and $\Delta \rightarrow 0$ for $j = 1, 2$, and (d) $\Delta T(\mu)T(\sigma^{2'})$, $\Delta^{1/2}T(\sigma)T(\sigma^{2'}) = o_p(1)$.

As in Assumption 4, we assume neither $T \rightarrow \infty$ nor $\ell(T, x) \rightarrow_p \infty$. We only require in Assumption 5 that $\Delta \rightarrow 0$ sufficiently fast relative to h and the extremal processes of X under various transforms given by the drift and diffusion functions. Clearly, if $T = \bar{T}$ fixed, then the conditions in parts (b), (c) and (d) are all trivially satisfied as long as $\Delta \rightarrow 0$. As discussed, the condition in part (a) becomes $nh^4 \rightarrow \infty$ in this case.

4.2.1 Locally Constant Estimator

The locally constant estimator for $\sigma^2(x)$ is given by

$$\hat{\sigma}^2(x) = \frac{P_T(K, \sigma^2)}{Q_T(K)},$$

where

$$P_T(K, \sigma^2) = \frac{1}{h} \sum_{i=1}^n K\left(\frac{X_{(i-1)\Delta} - x}{h}\right) (X_{i\Delta} - X_{(i-1)\Delta})^2 \quad (4.13)$$

and $Q_T(K)$ is defined in (4.3). Using (4.12), we write similarly as before

$$\hat{\sigma}^2(x) = \hat{\sigma}_p^2(x) + \hat{\sigma}_q^2(x) + \hat{\varepsilon}_{\sigma^2}(x),$$

where

$$\hat{\sigma}_p^2(x) = \sigma^2(x) + \frac{N_T(K, \sigma^2)}{Q_T(K)}, \quad \hat{\sigma}_q^2(x) = \frac{2M_T(K, \sigma^2)}{Q_T(K)}$$

with

$$N_T(K, \sigma^2) = \frac{\Delta}{h} \sum_{i=1}^n K\left(\frac{X_{(i-1)\Delta} - x}{h}\right) [\sigma^2(X_{(i-1)\Delta}) - \sigma^2(x)] \quad (4.14)$$

$$M_T(K, \sigma^2) = \frac{1}{h} \sum_{i=1}^n K\left(\frac{X_{(i-1)\Delta} - x}{h}\right) \int_{(i-1)\Delta}^{i\Delta} (X_t - X_{(i-1)\Delta}) \sigma(X_t) dW_t, \quad (4.15)$$

and

$$\hat{\varepsilon}_{\sigma^2}(x) = \frac{R_T(K, \sigma^2) + 2S_T(K)}{Q_T(K)}$$

with

$$R_T(K, \sigma^2) = \frac{1}{h} \sum_{i=1}^n K\left(\frac{X_{(i-1)\Delta} - x}{h}\right) \int_{(i-1)\Delta}^{i\Delta} [\sigma^2(X_t) - \sigma^2(X_{(i-1)\Delta})] dt \quad (4.16)$$

$$S_T(K) = \frac{1}{h} \sum_{i=1}^n K\left(\frac{X_{(i-1)\Delta} - x}{h}\right) \int_{(i-1)\Delta}^{i\Delta} (X_t - X_{(i-1)\Delta}) \mu(X_t) dt. \quad (4.17)$$

Just as in our decomposition for $\hat{\mu}(x)$ in (4.4), $\hat{\sigma}_p^2$ and $\hat{\sigma}_q^2$ denote the conditional mean and martingale terms of $\hat{\sigma}^2(x)$ respectively, and $\hat{\varepsilon}_{\sigma^2}(x)$ is the approximation error term which is of order smaller than the leading terms and becomes negligible asymptotically.

Theorem 4. *Let Assumptions 1, 2, 3 and 5 hold. Then we have*

$$\hat{\sigma}_p^2(x) = \sigma^2(x) + \frac{h^2}{2} \iota(K_2) \left(\sigma^{2''}(x) + 2\sigma^{2'}(x) \frac{m'(x)}{m(x)} \right) + o_p(h^2) + O_p(h^{3/2} \ell(T, x)^{-1/2})$$

uniformly in T as $h \rightarrow 0$ and $\Delta \rightarrow 0$, and

$$\left[\frac{h\ell(T, x)}{\Delta} \right]^{1/2} \hat{\sigma}_q^2 \rightarrow_d \sqrt{2}\sigma^2(x) \iota(K^2)^{1/2} Z,$$

where $\iota(K_2) = \int x^2 K(x) dx$, $\iota(K^2) = \int K^2(x) dx$ and Z is a standard normal random variate independent of $\ell(T, x)$. Moreover, we have

$$\hat{\varepsilon}_{\sigma^2}(x) = o_p(h^2),$$

which becomes negligible asymptotically.

4.2.2 Locally Linear Estimator

The locally linear estimator for $\sigma^2(x)$ is given by

$$\tilde{\sigma}^2(x) = \frac{P_T(K, \sigma^2)Q_T(K_2) - P_T(K_1, \sigma^2)Q_T(K_1)}{Q_T(K)Q_T(K_2) - Q_T(K_1)^2},$$

where Q_K and $P_T(K, \sigma^2)$ are defined respectively in (4.3) and (4.13), $Q_T(K_1)$ and $Q_T(K_2)$ are defined as $Q_T(K)$, and $P_T(K_1, \sigma^2)$ as $P_T(K, \sigma^2)$ with K substituted by K_1 , $K_1(x) = xK(x)$, and $K_2, K_2(x) = x^2K(x)$.

As before, we let

$$\tilde{\sigma}^2(x) = \tilde{\sigma}_p^2(x) + \tilde{\sigma}_q^2(x) + \tilde{\varepsilon}_{\sigma^2}(x),$$

where $\tilde{\sigma}_p^2(x)$ and $\tilde{\sigma}_q^2(x)$ are respectively the conditional mean and martingale terms of $\tilde{\sigma}^2(x)$, which are given by

$$\begin{aligned}\tilde{\sigma}_p^2(x) &= \sigma^2(x) + \frac{N_T(K, \sigma^2)Q_T(K_2) - N_T(K_1, \sigma^2)Q_T(K_1)}{Q_T(K)Q_T(K_2) - Q_T(K_1)^2} \\ \tilde{\sigma}_q^2(x) &= \frac{2[M_T(K, \sigma^2)Q_T(K_2) - M_T(K_1, \sigma^2)Q_T(K_1)]}{Q_T(K)Q_T(K_2) - Q_T(K_1)^2}\end{aligned}$$

with $N_T(K_1, \sigma^2)$ and $M_T(K_1, \sigma^2)$ defined respectively as $N_T(K, \sigma^2)$ and $M_T(K, \sigma^2)$ in (4.14) and (4.15), and $\tilde{\varepsilon}_{\sigma^2}(x)$ is the approximation error term given by

$$\tilde{\varepsilon}_{\sigma^2}(x) = \frac{[R_T(K, \sigma^2) + 2S_T(K)]Q_T(K_2) - [R_T(K_1, \sigma^2) + 2S_T(K_1)]Q_T(K_1)}{Q_T(K)Q_T(K_2) - Q_T(K_1)^2},$$

with $R_T(K_1, \sigma^2)$ and $S_T(K_1)$ are defined as $R_T(K, \sigma^2)$ and $S_T(K)$ in (4.16) and (4.17), respectively, using K_1 , $K_1(x) = xK(x)$, instead of K .

Theorem 5. *Let Assumptions 1, 2, 3 and 5 hold. Then we have*

$$\tilde{\sigma}_p^2(x) = \sigma^2(x) + \frac{h^2}{2}\iota(K_2)\sigma^2(x) + o_p(h^2) + O_p(h^{3/2}\ell(T, x)^{-1/2})$$

uniformly in T as $h \rightarrow 0$ and $\Delta \rightarrow 0$, and

$$\left[\frac{h\ell(T, x)}{\Delta} \right]^{1/2} \tilde{\sigma}_q^2 \rightarrow_d \sqrt{2}\sigma^2(x)\iota(K^2)^{1/2}Z,$$

where $\iota(K_2) = \int x^2 K(x)dx$, $\iota(K^2) = \int K^2(x)dx$ and Z is a standard normal random variate independent of $\ell(T, x)$. Moreover, we have

$$\tilde{\varepsilon}_{\sigma^2}(x) = o_p(h^2),$$

which becomes negligible asymptotically.

4.2.3 Remarks

(a) For the consistency of $\hat{\sigma}^2(x)$ and $\tilde{\sigma}^2(x)$, we only need $h \rightarrow_p 0$ and $\Delta \rightarrow 0$ sufficiently fast. In particular, we do not require $T \rightarrow \infty$ or $\ell(T, x) \rightarrow_p \infty$. Especially, $\sigma^2(x)$ may be consistently estimated even if we fix $T = \bar{T}$ or if the underlying diffusion is transient. We have $\hat{\sigma}_p^2(x), \tilde{\sigma}_p^2(x) \rightarrow_p \sigma^2(x)$ and $\hat{\varepsilon}_{\sigma^2}(x), \tilde{\varepsilon}_{\sigma^2}(x) \rightarrow_p 0$, as long as $h \rightarrow_p 0$. Moreover, we only need $\Delta \rightarrow 0$ to have $\hat{\sigma}_q^2(x), \tilde{\sigma}_q^2(x) \rightarrow_p 0$.

(b) The bias term of $\tilde{\sigma}^2(x)$ is identical to that in the standard nonparametric regression model. If the underlying diffusion is stationary, the locally constant estimator $\hat{\sigma}^2(x)$ also has the same bias term as in the standard model, as noted earlier for $\hat{\mu}(x)$. Again, the leading bias term in $\tilde{\sigma}^2(x)$ vanishes if and only if $\sigma^2(x)$ is linear in x .

(c) Let $\ell(T, x)/\Delta \rightarrow_p \infty$. The optimal bandwidth that minimizes the mean squared error of the leading terms is given by

$$\begin{aligned} h^*(\hat{\sigma}^2) &= \sqrt[5]{2} c(K) \sigma^{4/5}(x) \left(\sigma^{2''}(x) + 2\sigma^{2'}(x) \frac{m'(x)}{m(x)} \right)^{-2/5} \Delta^{1/5} \ell(T, x)^{-1/5} \\ h^*(\tilde{\sigma}^2) &= \sqrt[5]{2} c(K) \sigma^{4/5}(x) \sigma^{2''}(x)^{-2/5} \Delta^{1/5} \ell(T, x)^{-1/5} \end{aligned}$$

respectively for the locally constant estimator $\hat{\sigma}^2(x)$ and the locally linear estimator $\tilde{\sigma}^2(x)$ of $\sigma^2(x)$, where $c(K)$ is a constant introduced earlier in (4.8). In what follows, we let $h^* = h^*(\hat{\sigma}^2)$ or $h^*(\tilde{\sigma}^2)$ for notational brevity. For the stationary diffusion, we have $h^* = c(\Delta/T)^{1/5} = cn^{1/5}$ with some constant $c > 0$. Therefore, our optimal bandwidth in this case has the usual rate given in the conventional asymptotic analysis of the standard nonparametric regression, based on the single parameter n , the sample size. In general, the optimal bandwidth is not given by the sample size alone. It gets smaller as Δ becomes small and $\ell(T, x)$ becomes large, i.e., the observations collected more frequently and the underlying process is more recurrent. If $\ell(T, x)$ is bounded as is the case of $T = \bar{T}$ being fixed or the underlying diffusion being transient, then the optimal bandwidth is given by $h^* = c\Delta^{1/5}$ with some constant $c > 0$. Note that this choice of the bandwidth parameter h satisfies the part (a) of Assumption 5. If we set

$$h = c[\ell(T, x)/\Delta]^r$$

for some constant $c > 0$, the bias or variance term of both $\hat{\sigma}^2(x)$ and $\tilde{\sigma}^2(x)$ dominates depending upon whether $r > -1/5$ or $r < -1/5$. Finally, note that the optimal bandwidths $h^*(\hat{\sigma}^2)$ and $h^*(\tilde{\sigma}^2)$ can be estimated using an estimate for $\ell(T, x)$ and estimates for μ and σ^2 and their derivatives. As discussed earlier, our asymptotics are generally applicable also for the local constant or linear estimator with an estimated optimal bandwidth.

(d) Assume $\ell(T, x)/\Delta \rightarrow_p \infty$ and define

$$\lambda^2(h, \Delta, T) = \frac{h\ell(T, x)}{2\sigma^4(x)\iota(K^2)\Delta}.$$

If $h = c[\ell(T, x)/\Delta]^r$ with $r < -1/5$ for some constant $c > 0$, then we have

$$\lambda(h, \Delta, T) [\hat{\sigma}^2(x) - \sigma^2(x)], \lambda(h, \Delta, T) [\tilde{\sigma}^2(x) - \sigma^2(x)] \rightarrow_d Z. \quad (4.18)$$

Moreover, if $h = c[\ell(T, x)/\Delta]^{-1/5}$, then it follows that

$$\lambda(h, \Delta, T) [\hat{\sigma}^2(x) - (\sigma^2 + \sigma_a^2 + \sigma_b^2)(x)], \lambda(h, \Delta, T) [\tilde{\sigma}^2(x) - (\sigma^2 + \sigma_a^2)(x)] \rightarrow_d Z, \quad (4.19)$$

where $\sigma_a^2(x) = (h^2/2)\iota(K_2)\sigma^{2''}(x)$ and $\sigma_b^2(x) = h^2\iota(K_2)[(m'/m)\sigma^{2'}](x)$, similarly as for the kernel estimators for $\mu(x)$. Since Z is independent of $\ell(T, x)$, we may also deduce from (4.18) and (4.19) that both $\hat{\sigma}^2(x)$ and

$\tilde{\sigma}^2(x)$ are distributed approximately as a normal mixture with mixing variate $\lambda^2(h, \Delta, T)$. Finally, the limit distribution in (4.18) and (4.19) do not change if we replace $\ell(T, x)$ in the definition of $\lambda(h, \Delta, T)$ by $\hat{\ell}(T, x)$.

(e) If we set $T = \bar{T}$ fixed, our results for $\hat{\sigma}^2(x)$ in Theorem 4 is comparable with those obtained by Florens-Zmirou (1993). However, he assumes that $nh^4 \rightarrow 0$ and $nh^3 \rightarrow 0$, respectively, to obtain the consistency and limit distribution of $\hat{\sigma}^2(x)$ (in Proposition 4 and Theorem 1, p.800). Under his assumptions, the variance term dominates the bias term, and this is why the bias term does not appear in his asymptotics. As discussed above, the optimal bandwidth in this case is given by $h = cn^{1/5}$ with some constant $c > 0$, for which his assumptions are not met.

4.3. Asymptotic Independence

The nonparametric kernel estimators for the drift and diffusion functions become independent asymptotically under conditions that we introduce in earlier assumptions. This is shown in the following theorem.

Theorem 6. *Let Assumptions 1, 2, 3, 4 and 5 hold. Then, conditional on $\ell(T, x)$, $\hat{\mu}(x)$ and $\tilde{\mu}(x)$ are asymptotically independent of $\hat{\sigma}^2(x)$ and $\tilde{\sigma}^2(x)$.*

The joint asymptotics of the nonparametric kernel estimators of the drift function μ and the diffusion function σ^2 are therefore completely characterized by Theorems 2, 3, 4, 5 and 6.

5. Proofs

5.1. Preliminary Lemmas

Throughout this section, we let

$$\ell[T, x] = \sigma^2(x)\ell(T, x).$$

Note that $\ell[\cdot, \cdot]$ is the sojourn time measured by the quadratic variation $[X]$ of X . In the semimartingale literature $\ell[\cdot, \cdot]$ is more commonly referred to as the local time, than $\ell(\cdot, \cdot)$. With $\ell[\cdot, \cdot]$, we have an occupation times formula

$$\int_0^T f(X_t) d[X]_t = \int_{-\infty}^{\infty} f(x) \ell[T, x] dx$$

for any nonnegative f on R , in place of our earlier one in (2.2). In what follows, we let f be bounded and have support $[-1, 1]$, as we assume for the kernel function in Assumption 3.

Lemma 1. *Assume that g is continuous on D . Then we have*

$$\begin{aligned} \frac{1}{h} \int_0^T f\left(\frac{X_t - x}{h}\right) g(X_t) d[X]_t &= \int_{-\infty}^{\infty} f(u) g(x + hu) \ell[T, x + hu] du \\ &= \iota(f) g(x) \ell[T, x] [1 + o_p(1)] \end{aligned}$$

uniformly in T as $h \rightarrow 0$, where $\iota(f) = \int f(x) dx$.

Proof. By the successive applications of occupation times formula and change of variables in integral, we have

$$\begin{aligned} \frac{1}{h} \int_0^T f\left(\frac{X_t - x}{h}\right) g(X_t) d[X]_t &= \frac{1}{h} \int_{-\infty}^{\infty} f\left(\frac{u - x}{h}\right) g(u) \ell[T, x] du \\ &= \int_{-\infty}^{\infty} f(u) g(x + hu) \ell[T, x + hu] du. \end{aligned} \quad (5.1)$$

However, we have

$$\begin{aligned} &\int_{-\infty}^{\infty} f(u) g(x + hu) \ell[T, x + hu] du - g(x) \ell[T, x] \int_{-\infty}^{\infty} f(u) du \\ &= \int_{-\infty}^{\infty} f(u) g(x + hu) \left(\ell[T, x + hu] - \ell[T, x] \right) du + \ell[T, x] \int_{-\infty}^{\infty} f(u) \left(g(x + hu) - g(x) \right) du \\ &= \int_{-\infty}^{\infty} f(u) g(x + hu) \left(\ell[T, x + hu] - \ell[T, x] \right) du + o_{a.s.}(\ell[T, x]) \end{aligned} \quad (5.2)$$

uniformly in T as $h \rightarrow 0$. Moreover, it follows from Theorem 1.1 of Eisenbaum and Kaspi (2007) and condition (a) of Assumption 2 that

$$\sup_{|u| \leq 1} \left| \ell[T, x + hu] - \ell[T, x] \right| \leq \delta_h \bar{\ell}_\varepsilon[T, x]^{1/2} = o_p(\ell[T, x])$$

uniformly in T as $h \rightarrow 0$, where δ_h is a sequence of h such that $\delta_h \rightarrow 0$ as $h \rightarrow 0$, from which we may easily deduce that

$$\int_{-\infty}^{\infty} f(u) g(x + hu) \ell[T, x + hu] du = \iota(f) g(x) \ell[T, x] + o_p(\ell[T, x]) \quad (5.3)$$

uniformly in T as $h \rightarrow 0$. The stated result follows immediately from (5.1), (5.2) and (5.3). $\square$

Lemma 2. Assume that (a) $\int f(x) dx = 0$ and (b) $\nu = 2\mu/\sigma^2$ is continuous on D . Then we have

$$\int_{-\infty}^{\infty} f(u) \ell[T, x + hu] du = \iota_1(f) \nu(x) h \ell[T, x] + o_p(h \ell[T, x]) + O_p\left((h \ell[T, x])^{1/2}\right)$$

uniformly in T as $h \rightarrow 0$, where $\iota_1(f) = \int x f(x) dx$.

Proof. We define

$$f_1(x) = 1\{x \geq 0\} \int_x^{\infty} f(u) du - 1\{x < 0\} \int_{-\infty}^x f(u) du. \quad (5.4)$$

It is easy to see that f_1 is a negative anti-derivative of f , i.e., $f'_1 = -f$. Note that we have

$$\iota_1(f) = \int_{-\infty}^{\infty} x f(x) dx = \int_{-\infty}^{\infty} f_1(x) dx \quad (5.5)$$

under condition (a). Moreover, f_1 is square integrable.

We first note that

$$\int_{-\infty}^{\infty} f(u) \ell[T, x + hu] du = \int_{-\infty}^{\infty} f(u) \left(\ell[T, x + hu] - \ell[T, x] \right) du, \quad (5.6)$$

which follows from the condition $\int f(x) dx = 0$ in (a), and subsequently establish that

$$\int_{-\infty}^{\infty} f(u) \left(\ell[T, x + hu] - \ell[T, x] \right) du = 2 \int_0^T f_1\left(\frac{X_t - x}{h}\right) dX_t + O_{a.s.}(h), \quad (5.7)$$

uniformly in T as $h \rightarrow 0$. To show (5.7), we let $u > 0$ be fixed and apply the result in Exercise 1.28 (p.226) of Revuz and Yor (1994) with the function $1[x, x + hu]$ and the stopping time T . It follows that

$$\begin{aligned}\ell[T, x + hu] - \ell[T, x] &= 2 \int_0^T 1\{x \leq X_t \leq x + hu\} dX_t + O_{a.s.}(hu) \\ &= 2 \int_0^T 1\left\{0 \leq \frac{X_t - x}{h} \leq u\right\} dX_t + O_{a.s.}(hu)\end{aligned}\quad (5.8)$$

uniformly for all $u \in R$. Moreover, we have

$$\begin{aligned}\int_{-\infty}^{\infty} f(u) \left(\int_0^T 1\left\{0 \leq \frac{X_t - x}{h} \leq u\right\} dX_t \right) du \\ = \int_0^T \left(\int_{-\infty}^{\infty} f(u) 1\left\{0 \leq \frac{X_t - x}{h} \leq u\right\} du \right) dX_t\end{aligned}\quad (5.9)$$

by the Fubini's theorem for stochastic integrals [see e.g., Problem 6.12 (p.209) in Karatzas and Shreve (1991)].

Finally, note that

$$\int_{-\infty}^{\infty} f(u) 1\{0 \leq x \leq u\} du = \int_x^{\infty} f(u) du \quad (5.10)$$

for $x \geq 0$. Now the result in (5.7) follows from (5.8), (5.9) and (5.10). The proof for the case of $u < 0$ is entirely analogous. In this case, we write

$$\begin{aligned}\ell[T, x] - \ell[T, x + hu] &= 2 \int_0^T 1\{x + hu \leq X_t \leq x\} dX_t + O_{a.s.}(hu) \\ &= 2 \int_0^T 1\left\{u \leq \frac{X_t - x}{h} \leq 0\right\} dX_t + O_{a.s.}(hu),\end{aligned}$$

and note that

$$\int_{-\infty}^{\infty} f(u) 1\{u \leq x \leq 0\} du = \int_{-\infty}^x f(u) du$$

for $x < 0$.

Now we define a continuous semimartingale

$$Y_T = 2 \int_0^T f_1\left(\frac{X_t - x}{h}\right) dX_t,$$

which was introduced in the right hand side of (5.7), and decompose it as

$$Y_T = 2 \int_0^T f_1\left(\frac{X_t - x}{h}\right) \sigma(X_t) dW_t + 2 \int_0^T f_1\left(\frac{X_t - x}{h}\right) \mu(X_t) dt = M_T + N_T. \quad (5.11)$$

The martingale component M_T in (5.11) has quadratic variation $[M]_T$, which is given by

$$4 \int_0^T f_1^2\left(\frac{X_t - x}{h}\right) d[X]_t = 4h\ell[T, x] \left(\int_{-\infty}^{\infty} f_1^2(u) du + o_p(1) \right) \quad (5.12)$$

uniformly in T as $h \rightarrow 0$, due to Lemma 1. Therefore, we may readily deduce that

$$M_T = O_p\left((h\ell[T, x])^{1/2}\right) \quad (5.13)$$

uniformly in T as $h \rightarrow 0$ from the representation of M_T as the DDS Brownian motion with time change $[M]_T$.

Finally, the bounded variation component N_T in (5.11) becomes

$$\int_0^T f_1\left(\frac{X_t - x}{h}\right) \nu(X_t) d[X]_t = h\nu(x)\ell[T, x] \left(\int_{-\infty}^{\infty} f_1(u) du + o_p(1) \right)$$

uniformly in T as $h \rightarrow 0$, due to Lemma 1 and condition (b), from which and (5.5) it follows immediately that

$$N_T = \iota_1(f)\nu(x)h\ell[T, x] + o_p(h\ell[T, x]) \quad (5.14)$$

uniformly in T as $h \rightarrow 0$. The stated result can now be easily obtained from (5.6), (5.7), (5.11), (5.13) and (5.14). $\square$

Lemma 3. Assume that (a) $\int xf(x) dx = 0$, (b) $\nu = 2\mu/\sigma^2$ is continuously differentiable on D , and (c) $\sigma^2 > 0$ and σ^2 is continuous on D . Then we have

$$\begin{aligned} \int_{-\infty}^{\infty} f(u) \left(\ell[T, x + hu] - \ell[T, x] \right) du = & \frac{\iota_2(f)}{2} (\nu' + \nu^2)(x) h^2 \ell[T, x] + 2\iota(f_1^2)^{1/2} (h\ell[T, x])^{1/2} Z \\ & + o_p(h^2 \ell[T, x]) + o_p((h\ell[T, x])^{1/2}), \end{aligned}$$

conditional on $\ell[T, x]$, uniformly in T as $h \rightarrow 0$, where Z is a standard normal random variate, $\iota_2(f) = \int x^2 f(x) dx$ and $\iota(f_1^2) = \int f_1^2(x) dx$ with f_1 defined in (5.4).

Proof. Let f_1 be defined as in (5.4) from f , and let f_2 be defined from f_1 in the same way. Note that we have

$$\int_{-\infty}^{\infty} xf(x) dx = \int_{-\infty}^{\infty} f_1(x) dx = 0 \quad (5.15)$$

and

$$\iota_2(f) = \int_{-\infty}^{\infty} x^2 f(x) dx = 2 \int_{-\infty}^{\infty} xf_1(x) dx = 2 \int_{-\infty}^{\infty} f_2(x) dx \quad (5.16)$$

under condition in (a).

Let M_T and N_T be defined as in (5.11). Following the proof of Lemma 2, we may deduce that

$$\int_{-\infty}^{\infty} f(u) (\ell[T, x + hu] - \ell[T, x]) du = M_T + N_T + O_{a.s.}(h). \quad (5.17)$$

We will first show that

$$\frac{M_T}{2\iota(f_1^2)^{1/2} (h\ell[T, x])^{1/2}} =_d Z + o_p(1) \quad (5.18)$$

uniformly in T as $h \rightarrow 0$, where Z is a standard normal random variate independent of $\ell[T, x]$ for all $T > 0$. For simplicity, we momentarily assume that h is a nonrandom sequence of T , and let $m_T = 4\iota(f_1^2)h\kappa_T$. We write $M_T = V([M]_T)$, where V is the DDS Brownian motion of M , for which the reader is referred to, e.g., Theorem V.1.6 (p.173) of Revuz and Yor (1994). Moreover, we define $V_T(\cdot) = m_T^{-1/2}V(m_T\cdot)$, so that we may write

$$\frac{M_T}{\sqrt{m_T}} = V_T\left(\frac{[M]_T}{m_T}\right) \quad (5.19)$$

for each $T > 0$. By the well known scaling property of Brownian motion, V_T is standard Brownian motion for all T .

We may easily deduce from Lemma 1 that

$$[W, M]_T = o_p(h\ell[T, x])$$

uniformly in T as $h \rightarrow 0$ under condition (c), since $\iota(f_1) = 0$, and it follows immediately that

$$\frac{[W, M]_T}{[M]_T} \rightarrow_p 0, \quad \frac{[W, M]_T}{[W]_T} = \frac{[W, M]_T}{T} \rightarrow_p 0. \quad (5.20)$$

Therefore, we may apply Theorem XIII.2.3 and Corollary XIII.2.4 (pp.496-497) of Revuz and Yor (1994) to establish

$$(V_T, W) \rightarrow_d (V_\infty, W),$$

where V_∞ is a standard Brownian motion independent of W . The conclusion of Corollary XIII.2.4 of Revuz and Yor (1994) holds under (5.20), though a stronger condition is required.

Now we use the almost sure representation of weakly convergence sequences in, e.g., Theorem IV.13 (p.71) of Pollard (1984), and define distributionally equivalent copies of (V_T) and V_∞ on a common probability space with W such that

$$V_T \rightarrow_{a.s.} V_\infty \quad (5.21)$$

as $T \rightarrow \infty$. We do not distinguish (V_T) and V_∞ from their distributionally equivalent copies, since we are only interested in their distributions. Note that we have

$$\frac{[M]_T}{m_T} = \frac{\ell[T, x]}{m_T} + o_p(1)$$

as $T \rightarrow \infty$, due in particular to condition (b) of Assumption 2, Therefore, we may deduce from (5.19) and (5.21) that

$$\frac{M_T}{\sqrt{m_T}} = V_\infty \left(\frac{\ell[T, x]}{m_T} \right) + o_p(1)$$

as $h \rightarrow 0$ and $T \rightarrow \infty$, from which (5.18) follows immediately, since V_∞ is a standard Brownian motion that is independent of W and therefore of also X and $\ell[T, x]$.

It is clear that our proof of (5.18) can be easily modified to allow for h given as a function of $\ell[T, x]$. In fact, to deal with such a general h , we only need to redefine m_T accordingly, so that $h\ell[T, x]/m_T = O_p(1)$ as $T \rightarrow \infty$. With this simple modification, the rest of the proof will go through as it is. For a positive recurrent diffusion X , we expect $h\ell[T, x]/m_T$ to converge in probability under general regularity condition, in which case we may directly apply Corollary 3.2 (p.233) of van Zanten (2000) to obtain (5.18).

Next we analyze N_T in (5.17). Similarly as in the proof of Lemma 2, we may deduce

$$\begin{aligned} N_T &= h \int_{-\infty}^{\infty} f_1(u) \nu(x + hu) \ell[T, x + hu] du \\ &= h\nu(x) \int_{-\infty}^{\infty} f_1(u) \ell[T, x + hu] du + h^2 \nu'(x) \int_{-\infty}^{\infty} u f_1(u) \ell[T, x + hu] du + o_p(h^2 \ell[T, x]) \\ &= h\nu(x) \int_{-\infty}^{\infty} f_1(u) \ell[T, x + hu] du + \frac{\iota_2(f)}{2} \nu'(x) h^2 \ell[T, x] + o_p(h^2 \ell[T, x]) \end{aligned} \quad (5.22)$$

uniformly in T as $h \rightarrow 0$, since ν is continuously differentiable by condition (b) and $\iota_2(f) = 2 \int x f_1(x) dx$ as shown in (5.16). Moreover, f_1 satisfies condition (a) in Lemma 2 with $2\iota_1(f_1) = \iota_2(f)$, as shown in (5.15) and (5.16). Therefore, it follows from Lemma 2 that

$$\int_{-\infty}^{\infty} f_1(u) \ell[T, x + hu] du = \frac{\iota_2(f)}{2} \nu(x) h \ell[T, x] + o_p(h \ell[T, x]) + O_p\left((h \ell[T, x])^{1/2}\right) \quad (5.23)$$

uniformly in T as $h \rightarrow 0$. Consequently, we have from (5.22) and (5.23) that

$$N_T = \frac{\iota_2(f)}{2} (\nu' + \nu^2)(x) h^2 \ell[T, x] + o_p(h^2 \ell[T, x]) + O_p\left(h^{3/2} \ell[T, x]^{1/2}\right) \quad (5.24)$$

uniformly in T as $h \rightarrow 0$. The stated result now follows readily from (5.17), (5.18) and (5.24). The proof is therefore complete. $\square$

Lemma 4. Assume that (a) $\int f(x) dx = 0$, (b) $\nu = 2\mu/\sigma^2$ is continuous on D , and (c) $\sigma^2 > 0$ and σ^2 is continuously differentiable on D . Then we have

$$\int_{-\infty}^{\infty} f(u) \ell(T, x + hu) du = \iota_1(f) [m'/m](x) h \ell(T, x) + o_p(h \ell(T, x)) + O_p\left((h \ell(T, x))^{1/2}\right)$$

uniformly in T as $h \rightarrow 0$, where $\iota_1(f) = \int x f(x) dx$.

Proof. Note that

$$\begin{aligned} \int_{-\infty}^{\infty} f(u) \ell(T, x + hu) du &= \int_{-\infty}^{\infty} f(u) (1/\sigma^2)(x + hu) \ell[T, x + hu] du \\ &= A_T + B_T + o_p(h \ell(T, x)) \end{aligned} \quad (5.25)$$

uniformly in T as $h \rightarrow 0$, where

$$\begin{aligned} A_T &= (1/\sigma^2)(x) \int_{-\infty}^{\infty} f(u) \ell[T, x + hu] du \\ B_T &= (1/\sigma^2)'(x) h \int_{-\infty}^{\infty} u f(u) \ell[T, x + hu] du, \end{aligned}$$

due to Lemma 1 and condition (c). We have

$$\begin{aligned} A_T &= (1/\sigma^2)(x) \left(\iota_1(f) \nu(x) h \ell[T, x] + o_p(h \ell(T, x)) + O_p\left((h \ell(T, x))^{1/2}\right) \right) \\ &= \iota_1(f) (2\mu/\sigma^2)(x) h \ell(T, x) + o_p(h \ell(T, x)) + O_p\left((h \ell(T, x))^{1/2}\right) \end{aligned} \quad (5.26)$$

uniformly in T as $h \rightarrow 0$, due to Lemma 2, which holds under conditions (a) and (b). On the other hand,

$$\begin{aligned} B_T &= (1/\sigma^2)'(x) h \left(\iota_1(f) \ell[T, x] \right) + o_p(h \ell(T, x)) \\ &= -\iota_1(f) (2\sigma'/\sigma)(x) h \ell(T, x) + o_p(h \ell(T, x)) \end{aligned} \quad (5.27)$$

uniformly in T as $h \rightarrow 0$, due again to Lemma 1. Note that

$$\frac{m'(x)}{m(x)} = (-\log s(x) - \log \sigma^2(x))' = \frac{2\mu(x)}{\sigma^2(x)} - \frac{2(\sigma\sigma')(x)}{\sigma^2(x)}$$

to obtain the stated result from (5.25), (5.26) and (5.27). $\square$

Lemma 5. Assume that (a) $\int xf(x) dx = 0$, (b) $\nu = 2\mu/\sigma^2$ is continuously differentiable on D , and (c) $\sigma^2 > 0$ and σ^2 is twice continuously differentiable on D . Then we have

$$\begin{aligned} \int_{-\infty}^{\infty} f(u) \left(\ell(T, x + hu) - \ell(T, x) \right) du = & \iota_2(f) \tau(x) h^2 \ell(T, x) + 2\iota(f_1^2)^{1/2} (1/\sigma)(x) (h\ell(T, x))^{1/2} Z \\ & + o_p(h^2 \ell(T, x)) + o_p((h\ell(T, x))^{1/2}) \end{aligned}$$

conditional on $\ell(T, x)$, uniformly in T as $h \rightarrow 0$, where Z is a standard normal random variate, $\iota_2(f) = \int x^2 f(x) dx$, $\iota(f_1^2) = \int f_1^2(x) dx$ with f_1 defined in (5.4), and $\tau(x) = [2\mu^2/\sigma^4 - (2\mu + 4\mu\sigma')/\sigma^3 + (\mu' - \sigma\sigma'' + 3\sigma'^2)/\sigma^2](x)$.

Proof. As in the proof of Lemma 4, we may write

$$\begin{aligned} \int_{-\infty}^{\infty} f(u) \ell(T, x + hu) du &= \int_{-\infty}^{\infty} f(u) (1/\sigma^2)(x + hu) \ell[T, x + hu] du \\ &= A_T + B_T + C_T + o_p(h^2 \ell(T, x)) \end{aligned} \quad (5.28)$$

uniformly in T as $h \rightarrow 0$, where

$$\begin{aligned} A_T &= (1/\sigma^2)(x) \int_{-\infty}^{\infty} f(u) \ell[T, x + hu] du \\ B_T &= (1/\sigma^2)'(x) h \int_{-\infty}^{\infty} u f(u) \ell[T, x + hu] du \\ C_T &= \frac{(1/\sigma^2)''(x)}{2} h^2 \int_{-\infty}^{\infty} u^2 f(u) \ell[T, x + hu] du, \end{aligned}$$

due to Lemma 1 and condition (c).

Furthermore, we have

$$\begin{aligned} A_T = & \iota(f) \ell(T, x) + \iota_2(f) [(2\mu^2 - 2\mu\sigma\sigma' + \mu'\sigma^2)/\sigma^4](x) h^2 \ell(T, x) + 2\iota(f_1^2)^{1/2} (1/\sigma)(x) (h\ell(T, x))^{1/2} Z \\ & + o_p((h\ell(T, x))^{1/2}) + o_p(h^2 \ell(T, x)), \end{aligned} \quad (5.29)$$

conditional on $\ell(T, x)$, uniformly in T as $h \rightarrow 0$, due to Lemma 3. Likewise, it follows from Lemma 2 that

$$B_T = -\iota_2(f) (4\mu\sigma'/\sigma^3)(x) h^2 \ell(T, x) + o_p(h^2 \ell(T, x)) \quad (5.30)$$

uniformly in T as $h \rightarrow 0$. Note that the conditions in Lemmas 2 and 3 are satisfied under conditions (a), (b) and (c). Finally, we have

$$C_T = \iota_2(f) [(3\sigma'^2 - \sigma\sigma'')/\sigma^2](x) h^2 \ell(T, x) + o_p(h^2 \ell(T, x)) \quad (5.31)$$

uniformly in T as $h \rightarrow 0$ from Lemma 1. The stated result now follows immediately from (5.29), (5.30) and (5.31) and the expansion (5.28). $\square$

Lemma 6. *We assume that g is continuous on D . Then we have*

$$\begin{aligned} \frac{1}{h} \int_0^T f\left(\frac{X_t - x}{h}\right) g(X_t) dt &= \int_{-\infty}^{\infty} f(u) g(x + hu) \ell(T, x + hu) du \\ &= \iota(f) g(x) \ell(T, x) + o_p(\ell(T, x)) \end{aligned}$$

uniformly in T as $h \rightarrow 0$, where $\iota(f) = \int f(x) dx$.

Proof. The proof is essentially identical to that of Lemma 1, and the details are therefore omitted. $\square$

Lemma 7. *Assume that g is continuously differentiable on D . Then we have*

$$\frac{1}{h^2} \int_0^T f\left(\frac{X_t - x}{h}\right) [g(X_t) - g(x)] dt = \iota_1(f) g'(x) \ell(T, x) + o_p(\ell(T, x))$$

uniformly in T as $h \rightarrow 0$, where $\iota_1(f) = \int x f(x) dx$.

Proof. We write

$$\frac{1}{h} \int_0^T f\left(\frac{X_t - x}{h}\right) [g(X_t) - g(x)] dt = \int_{-\infty}^{\infty} f(u) [g(x + hu) - g(x)] \ell(T, x + hu) du$$

using the occupation times formula and change of variables. Since g is continuously differentiable, we have

$$\begin{aligned} \int_{-\infty}^{\infty} f(u) [g(x + hu) - g(x)] \ell(T, x + hu) du &= g'(x) h \int_{-\infty}^{\infty} u f(u) \ell(T, x + hu) du + o_p(h \ell(T, x)) \\ &= \iota_1(f) g'(x) h \ell(T, x) + o_p(h \ell(T, x)) \end{aligned}$$

uniformly in T as $h \rightarrow 0$, due to Lemma 6. This completes the proof. $\square$

Lemma 8. *Assume that (a) $\int f(x) dx = 0$, (b) $\nu = 2\mu/\sigma^2$ is continuous on D , and (c) $\sigma^2 > 0$ and σ^2 is continuously differentiable on D . Moreover, we assume that g is continuously differentiable on D . Then we have*

$$\frac{1}{h^2} \int_0^T f\left(\frac{X_t - x}{h}\right) g(X_t) dt = \iota_1(f) \left(g'(x) + g(x) \frac{m'(x)}{m(x)} \right) \ell(T, x) + o_p(\ell(T, x)) + O_p\left(h^{-1/2} \ell(T, x)^{1/2}\right)$$

uniformly in T as $h \rightarrow 0$, where $\iota_1(f) = \int x f(x) dx$.

Proof. As before, we write

$$\frac{1}{h} \int_0^T f\left(\frac{X_t - x}{h}\right) g(X_t) dt = \int_{-\infty}^{\infty} f(u) g(x + hu) \ell(T, x + hu) du$$

using the occupation times formula and change of variables. Since g is continuously differentiable, it follows that

$$\begin{aligned} &\int_{-\infty}^{\infty} f(u) g(x + hu) \ell(T, x + hu) du \\ &= g(x) \int_{-\infty}^{\infty} f(u) \ell(T, x + hu) du + g'(x) h \int_{-\infty}^{\infty} u f(u) \ell(T, x + hu) du + o_p(h \ell(T, x)) \\ &= A_T + B_T + o_p(h \ell(T, x)) \end{aligned} \tag{5.32}$$

uniformly in T as $h \rightarrow 0$, due to Lemma 6. However, we have

$$A_T = \iota_1(f)g(x)[m'/m](x)h\ell(T, x) + o_p(h\ell(T, x)) + O_p\left(h^{-1/2}\ell(T, x)^{1/2}\right) \quad (5.33)$$

uniformly in T as $h \rightarrow 0$ from Lemma 4, and

$$B_T = \iota_1(f)g'(x)h\ell(T, x) + o_p(h\ell(T, x)) \quad (5.34)$$

uniformly in T as $h \rightarrow 0$, due to Lemma 6. The stated result now follows immediately from (5.32), (5.33) and (5.34). $\square$

Lemma 9. Assume that (a) $\int xf(x)dx = 0$, (b) $\nu = 2\mu/\sigma^2$ is continuous on D , and (c) $\sigma^2 > 0$ and σ^2 is continuously differentiable on D . Moreover, we assume that g is twice continuously differentiable on D . Then we have

$$\begin{aligned} \frac{1}{h^3} \int_0^T f\left(\frac{X_t - x}{h}\right) [g(X_t) - g(x)] dt &= \frac{\iota_2(f)}{2} \left(g''(x) + 2g'(x) \frac{m'(x)}{m(x)} \right) \ell(T, x) \\ &\quad + o_p(\ell(T, x)) + O_p\left(h^{-1/2}\ell(T, x)^{1/2}\right) \end{aligned}$$

uniformly in T as $h \rightarrow 0$, where $\iota_2(f) = \int x^2 f(x) dx$.

Proof. Again, we write

$$\begin{aligned} \frac{1}{h} \int_0^T f\left(\frac{X_t - x}{h}\right) [g(X_t) - g(x)] dt &= \frac{1}{h} \int_{-\infty}^{\infty} f\left(\frac{u - x}{h}\right) [g(u) - g(x)] \ell(T, u) du \\ &= \int_{-\infty}^{\infty} f(u) [g(x + hu) - g(x)] \ell(T, x + hu) du. \end{aligned}$$

Since g is continuously differentiable, it follows that

$$\begin{aligned} &\int_{-\infty}^{\infty} f(u) [g(x + hu) - g(x)] \ell(T, x + hu) du \\ &= g'(x)h \int_{-\infty}^{\infty} u f(u) \ell(T, x + hu) du + \frac{g''(x)}{2} h^2 \int_{-\infty}^{\infty} u^2 f(u) \ell(T, x + hu) du + o_p(h^2 \ell(T, x)) \\ &= A_T + B_T + o_p(h^2 \ell(T, x)) \end{aligned}$$

uniformly in T as $h \rightarrow 0$, due to Lemma 6. However, we have

$$\begin{aligned} A_T &= \iota_2(f)g'(x)[m'/m](x)h^2\ell(T, x) + o_p(h^2\ell(T, x)) + O_p\left(h^{3/2}\ell(T, x)^{1/2}\right) \\ B_T &= \frac{\iota_2(f)}{2}g''(x)h^2\ell(T, x) + o_p(h^2\ell(T, x)) \end{aligned}$$

uniformly in T as $h \rightarrow 0$, similarly as in (5.33) and (5.34), from which the stated result can be readily obtained. $\square$

Lemma 10. Let (a) f be twice continuously differentiable, and let (b) μ and σ^2 are continuous on D . Then we have

$$\frac{\Delta}{h} \sum_{i=1}^n f\left(\frac{X_{(i-1)\Delta} - x}{h}\right) = \frac{1}{h} \int_0^T f\left(\frac{X_t - x}{h}\right) dt + O_p\left(\frac{\Delta}{h^2}\ell(T, x)\right)$$

uniformly in T as $h \rightarrow 0$ and $\Delta \rightarrow 0$.

Proof. We write

$$\frac{\Delta}{h} \sum_{i=1}^n f\left(\frac{X_{(i-1)\Delta} - x}{h}\right) = \frac{1}{h} \int_0^T f\left(\frac{X_t - x}{h}\right) dt + R_T,$$

where

$$R_T = -\frac{1}{h} \sum_{i=1}^n \int_{(i-1)\Delta}^{i\Delta} \left[f\left(\frac{X_t - x}{h}\right) - f\left(\frac{X_{(i-1)\Delta} - x}{h}\right) \right] dt.$$

We may apply Itô's formula to deduce

$$\begin{aligned} & f\left(\frac{X_t - x}{h}\right) - f\left(\frac{X_{(i-1)\Delta} - x}{h}\right) \\ &= \frac{1}{h} \int_{(i-1)\Delta}^t f'\left(\frac{X_s - x}{h}\right) dX_s + \frac{1}{2h^2} \int_{(i-1)\Delta}^t f''\left(\frac{X_s - x}{h}\right) d[X]_s \\ &= \frac{1}{h} \int_{(i-1)\Delta}^t \mu(X_s) f'\left(\frac{X_s - x}{h}\right) ds + \frac{1}{2h^2} \int_{(i-1)\Delta}^t \sigma^2(X_s) f''\left(\frac{X_s - x}{h}\right) ds \\ &\quad + \frac{1}{h} \int_{(i-1)\Delta}^t \sigma(X_s) f'\left(\frac{X_s - x}{h}\right) dW_s \end{aligned}$$

for $t \in [(i-1)\Delta, i\Delta]$.

Consequently, we may write

$$R_T = -(A_T + B_T + C_T),$$

where

$$\begin{aligned} A_T &= \frac{1}{h^2} \sum_{i=1}^n \int_{(i-1)\Delta}^{i\Delta} \int_{(i-1)\Delta}^t \mu(X_s) f'\left(\frac{X_s - x}{h}\right) ds dt = \frac{1}{h^2} \sum_{i=1}^n \int_{(i-1)\Delta}^{i\Delta} (i\Delta - t) \mu(X_t) f'\left(\frac{X_t - x}{h}\right) dt \\ B_T &= \frac{1}{2h^3} \sum_{i=1}^n \int_{(i-1)\Delta}^{i\Delta} \int_{(i-1)\Delta}^t \sigma^2(X_s) f''\left(\frac{X_s - x}{h}\right) ds dt = \frac{1}{2h^3} \sum_{i=1}^n \int_{(i-1)\Delta}^{i\Delta} (i\Delta - t) \sigma^2(X_t) f''\left(\frac{X_t - x}{h}\right) dt \\ C_T &= \frac{1}{h^2} \sum_{i=1}^n \int_{(i-1)\Delta}^{i\Delta} \int_{(i-1)\Delta}^t \sigma(X_s) f'\left(\frac{X_s - x}{h}\right) dW_s dt = \frac{1}{h^2} \sum_{i=1}^n \int_{(i-1)\Delta}^{i\Delta} (i\Delta - t) \sigma(X_t) f'\left(\frac{X_t - x}{h}\right) dW_t. \end{aligned}$$

Clearly, A_T is bounded by

$$h^{-2} \Delta \int_0^T |\mu(X_t)| \left| f'\left(\frac{X_t - x}{h}\right) \right| dt = h^{-1} \Delta |\mu(x)| \ell(T, x) \left(\int_{-\infty}^{\infty} |f'(u)| du + o_p(1) \right) \quad (5.35)$$

uniformly in T as $h \rightarrow 0$ and $\Delta \rightarrow 0$, due to Lemma 6, since μ is continuous and $\int |f'(x)| dx < \infty$. We may similarly deduce that B_T is bounded by

$$h^{-3} \Delta \int_0^T \sigma^2(X_t) \left| f''\left(\frac{X_t - x}{h}\right) \right| dt = h^{-2} \Delta \sigma^2(x) \ell(T, x) \left(\int_{-\infty}^{\infty} |f''(u)| du + o_p(1) \right) \quad (5.36)$$

uniformly in T as $h \rightarrow 0$ and $\Delta \rightarrow 0$, since σ is continuous and $\int |f''(x)| dx < \infty$.

For C_T , we consider a continuous martingale M defined by

$$M_t = \frac{1}{h^2} \left[\sum_{j=1}^{i-1} \int_{(j-1)\Delta}^{j\Delta} (j\Delta - s) \sigma(X_s) f'\left(\frac{X_s - x}{h}\right) dW_s + \int_{(i-1)\Delta}^t (i\Delta - s) \sigma(X_s) f'\left(\frac{X_s - x}{h}\right) dW_s \right]$$

for $t \in [(i-1)\Delta, i\Delta]$, $i = 1, 2, \dots$, so that $M_T = C_T$. We may deduce from Lemma 6 that

$$[M]_T = h^{-4} \Delta^2 \int_0^T \sigma^2(X_t) \left| f'^2 \left(\frac{X_t - x}{h} \right) \right| dt = h^{-3} \Delta^2 \sigma^2(x) \ell(T, x) \left(\int_{-\infty}^{\infty} |f'^2(u)| du + o_p(1) \right)$$

uniformly in T as $h \rightarrow 0$ and $\Delta \rightarrow 0$, and it follows that

$$C_T = O_p \left(h^{-3/2} \Delta \ell(T, x)^{1/2} \right), \quad (5.37)$$

due to the martingale representation theorem in, e.g., Theorem V.1.6 (p.173) of Revuz and Yor (1994). The stated result now follows immediately from (5.35), (5.36) and (5.37). $\square$

Lemma 11. *Let the assumptions in Lemma 10 hold. Moreover, we assume that g is twice continuously differentiable on D , and that (a) $h^{-4} \Delta \rightarrow_p 0$, (b) $\Delta^{1/2} T(g_A) \rightarrow_p 0$ and (c) $\Delta^{1/2} T(g_B) = o_p((h\ell(T, x))^{1/2})$ uniformly in T as $h \rightarrow 0$ and $\Delta \rightarrow 0$. Then we have*

$$\frac{1}{h} \sum_{i=1}^n f \left(\frac{X_{(i-1)\Delta} - x}{h} \right) \int_{(i-1)\Delta}^{i\Delta} [g(X_t) - g(X_{(i-1)\Delta})] dt = o_p(h^2 \ell(T, x)),$$

and

$$\frac{1}{\sqrt{h\Delta}} \sum_{i=1}^n f \left(\frac{X_{(i-1)\Delta} - x}{h} \right) \int_{(i-1)\Delta}^{i\Delta} [g(X_t) - g(X_{(i-1)\Delta})] dt = o_p(h^{1/2} \ell(T, x))$$

uniformly in T as $h \rightarrow 0$ and $\Delta \rightarrow 0$.

Proof. It follows from Itô's formula that

$$\begin{aligned} \int_{(i-1)\Delta}^{i\Delta} [g(X_t) - g(X_{(i-1)\Delta})] dt &= \int_{(i-1)\Delta}^{i\Delta} dt \left[\int_{(i-1)\Delta}^t g_A(X_s) ds + \int_{(i-1)\Delta}^t g_B(X_s) dW_s \right] \\ &= \int_{(i-1)\Delta}^{i\Delta} (i\Delta - t) g_A(X_t) dt + \int_{(i-1)\Delta}^{i\Delta} (i\Delta - t) g_B(X_t) dW_t. \end{aligned}$$

Therefore, we have

$$\begin{aligned} &\sum_{i=1}^n f \left(\frac{X_{(i-1)\Delta} - x}{h} \right) \int_{(i-1)\Delta}^{i\Delta} [g(X_t) - g(X_{(i-1)\Delta})] dt \\ &= \sum_{i=1}^n f \left(\frac{X_{(i-1)\Delta} - x}{h} \right) \int_{(i-1)\Delta}^{i\Delta} (i\Delta - t) g_A(X_t) dt + \sum_{i=1}^n f \left(\frac{X_{(i-1)\Delta} - x}{h} \right) \int_{(i-1)\Delta}^{i\Delta} (i\Delta - t) g_B(X_t) dW_t \quad (5.38) \end{aligned}$$

each term of which will be looked at separately below.

For the first term in (5.38), we have

$$\begin{aligned} &\left| \sum_{i=1}^n f \left(\frac{X_{(i-1)\Delta} - x}{h} \right) \int_{(i-1)\Delta}^{i\Delta} (i\Delta - t) g_A(X_t) dt \right| \\ &\leq (h\Delta) \frac{\Delta}{h} \sum_{i=1}^n |f| \left(\frac{X_{(i-1)\Delta} - x}{h} \right) \sup_{t \in [0, T]} |g_A(X_t)| = O_p(h\Delta \ell(T, x) T(g_A)) \quad (5.39) \end{aligned}$$

uniformly in T as $h \rightarrow 0$ and $\Delta \rightarrow 0$, due to Lemma 10 and conditions (a) and (b). For the second term in (5.38), we define a continuous martingale M as

$$M_t = \sum_{j=1}^{i-1} f \left(\frac{X_{(j-1)\Delta} - x}{h} \right) \int_{(j-1)\Delta}^{j\Delta} (j\Delta - t) g_B(X_s) dW_s + f \left(\frac{X_{(i-1)\Delta} - x}{h} \right) \int_{(i-1)\Delta}^t (i\Delta - s) g_B(X_s) dW_s$$

for $t \in [(i-1)\Delta, i\Delta]$, $i = 1, 2, \dots$, similarly as in the proof of Lemma 10. It follows that

$$M_T = \sum_{i=1}^n f\left(\frac{X_{(i-1)\Delta} - x}{h}\right) \int_{(i-1)\Delta}^{i\Delta} (i\Delta - t) g_B(X_t) dW_t$$

and

$$\begin{aligned} [M]_T &= \left| \sum_{i=1}^n f^2\left(\frac{X_{(i-1)\Delta} - x}{h}\right) \int_{(i-1)\Delta}^{i\Delta} (i\Delta - t)^2 g_B^2(X_t) dt \right| \\ &\leq (h\Delta^2) \frac{\Delta}{h} \sum_{i=1}^n f^2\left(\frac{X_{(i-1)\Delta} - x}{h}\right) \sup_{t \in [0, T]} |g_B^2(X_t)| = O_p(h\Delta^2 \ell(T, x) T(g_B^2)) \end{aligned} \quad (5.40)$$

from Lemma 10, uniformly in T as $h \rightarrow 0$ and $\Delta \rightarrow 0$, due to conditions (a) and (c).

Now we have from (5.38), (5.39) and (5.40) that

$$\begin{aligned} &\sum_{i=1}^n f\left(\frac{X_{(i-1)\Delta} - x}{h}\right) \int_{(i-1)\Delta}^{i\Delta} [g(X_t) - g(X_{(i-1)\Delta})] dt \\ &= O_p(h\Delta \ell(T, x) T(g_A)) + O_p(h^{1/2} \Delta \ell(T, x)^{1/2} T(g_B)) \end{aligned} \quad (5.41)$$

uniformly in T as $h \rightarrow 0$ and $\Delta \rightarrow 0$. To obtain the first part, we note that

$$\Delta \ell(T, x) T(g_A), \quad h^{-1/2} \Delta \ell(T, x)^{1/2} T(g_B) = o_p(h^2 \ell(T, x))$$

uniformly in T as $h \rightarrow 0$ and $\Delta \rightarrow 0$ under conditions (a), (b) and (c). For the second part, we similarly note that

$$h^{1/2} \Delta^{1/2} \ell(T, x) T(g_A), \quad \Delta^{1/2} \ell(T, x)^{1/2} T(g_B) = o_p(h^{1/2} \ell(T, x))$$

uniformly in T as $h \rightarrow 0$ and $\Delta \rightarrow 0$ under conditions (a), (b) and (c). The stated results therefore follow immediately from (5.41), and the proof is complete. $\square$

Lemma 12. *Let the assumptions in Lemma 10 hold. Moreover, we assume that g is twice continuously differentiable on D . Then we have*

$$\frac{\Delta}{h} \sum_{i=1}^n f\left(\frac{X_{(i-1)\Delta} - x}{h}\right) g(X_{(i-1)\Delta}) = \frac{1}{h} \int_0^T f\left(\frac{X_t - x}{h}\right) g(X_t) dt + o_p(h^{-2} \Delta \ell(T, x))$$

uniformly in T as $h \rightarrow 0$ and $\Delta \rightarrow 0$.

Proof. Let

$$\frac{\Delta}{h} \sum_{i=1}^n f\left(\frac{X_{(i-1)\Delta} - x}{h}\right) g(X_{(i-1)\Delta}) = \frac{1}{h} \int_0^T f\left(\frac{X_t - x}{h}\right) g(X_t) dt + R_T, \quad (5.42)$$

where

$$R_T = -\frac{1}{h} \sum_{i=1}^n \int_{(i-1)\Delta}^{i\Delta} \left[f\left(\frac{X_t - x}{h}\right) g(X_t) - f\left(\frac{X_{(i-1)\Delta} - x}{h}\right) g(X_{(i-1)\Delta}) \right] dt.$$

We write

$$R_T = -(A_T + B_T + C_T),$$

where

$$\begin{aligned}
A_T &= \frac{1}{h} \sum_{i=1}^n \int_{(i-1)\Delta}^{i\Delta} \int_{(i-1)\Delta}^t \mu(X_s) \left[\frac{1}{h} f'_{x,h}(X_s) g(X_s) + f_{x,h}(X_s) g'(X_s) \right] ds dt \\
&= \frac{1}{h} \sum_{i=1}^n \int_{(i-1)\Delta}^{i\Delta} (i\Delta - t) \mu(X_t) \left[\frac{1}{h} f'_{x,h}(X_t) g(X_t) + f_{x,h}(X_t) g'(X_t) \right] dt \\
B_T &= \frac{1}{2h} \sum_{i=1}^n \int_{(i-1)\Delta}^{i\Delta} \int_{(i-1)\Delta}^t \sigma^2(X_s) \left[\frac{1}{h^2} f''_{x,h}(X_s) g(X_s) + \frac{2}{h} f'_{x,h}(X_s) g'(X_s) + f_{x,h}(X_s) g''(X_s) \right] ds dt \\
&= \frac{1}{2h} \sum_{i=1}^n \int_{(i-1)\Delta}^{i\Delta} (i\Delta - t) \sigma^2(X_t) \left[\frac{1}{h^2} f''_{x,h}(X_t) g(X_t) + \frac{2}{h} f'_{x,h}(X_t) g'(X_t) + f_{x,h}(X_t) g''(X_t) \right] dt \\
C_T &= \frac{1}{h} \sum_{i=1}^n \int_{(i-1)\Delta}^{i\Delta} \int_{(i-1)\Delta}^t \sigma(X_s) \left[\frac{1}{h} f'_{x,h}(X_s) g(X_s) + f_{x,h}(X_s) g'(X_s) \right] dW_s dt \\
&= \frac{1}{h} \sum_{i=1}^n \int_{(i-1)\Delta}^{i\Delta} (i\Delta - t) \sigma(X_t) \left[\frac{1}{h} f'_{x,h}(X_t) g(X_t) + f_{x,h}(X_t) g'(X_t) \right] dW_t
\end{aligned}$$

with

$$f_{x,h}(X_t) = f\left(\frac{X_t - x}{h}\right)$$

and $f'_{x,h}(X_t)$ and $f''_{x,h}(X_t)$ defined similarly from f' and f'' .

Similarly as in the proof of Lemma 10, we may readily deduce that

$$\begin{aligned}
A_T &= O_p(h^{-1} \Delta \ell(T, x)) \\
B_T &= O_p(h^{-2} \Delta \ell(T, x)) \\
C_T &= O_p(h^{-3/2} \Delta \ell(T, x)^{1/2})
\end{aligned}$$

uniformly in T as $h \rightarrow 0$ and $\Delta \rightarrow 0$, from which the stated result follows immediately. $\square$

Lemma 13. *Let the assumptions in Lemma 10 hold. Moreover, we assume that g is twice continuously differentiable on D , and that (a) $h^{-4} \Delta \rightarrow_p 0$, (b) $\Delta^{1/2} T(g_{jA}) \rightarrow_p 0$ and (c) $\Delta^{1/2} T(g_{jB}) = o_p((h \ell(T, x))^{1/2})$ uniformly in T as $h \rightarrow 0$ and $\Delta \rightarrow 0$ for $j = 0, 1$, where $g_j(x) = x^j g(x)$. Then we have*

$$\frac{1}{h} \sum_{i=1}^n f\left(\frac{X_{(i-1)\Delta} - x}{h}\right) \int_{(i-1)\Delta}^{i\Delta} (X_t - X_{(i-1)\Delta}) g(X_t) dt = o_p(h^2 \ell(T, x))$$

and

$$\frac{1}{\sqrt{h\Delta}} \sum_{i=1}^n f\left(\frac{X_{(i-1)\Delta} - x}{h}\right) \int_{(i-1)\Delta}^{i\Delta} (X_t - X_{(i-1)\Delta}) g(X_t) dt = o_p(h^{1/2} \ell(T, x))$$

uniformly in T as $h \rightarrow 0$ and $\Delta \rightarrow 0$.

Proof. Write

$$\begin{aligned}
(X_t - X_{(i-1)\Delta}) g(X_t) &= [g_1(X_t) - g_1(X_{(i-1)\Delta})] - X_{(i-1)\Delta} [g(X_t) - g(X_{(i-1)\Delta})] \\
&= [g_1(X_t) - g_1(X_{(i-1)\Delta})] - \left[x + h \left(\frac{X_{(i-1)\Delta} - x}{h} \right) \right] [g(X_t) - g(X_{(i-1)\Delta})],
\end{aligned}$$

so that we have

$$\begin{aligned}
& \frac{1}{h} \sum_{i=1}^n f\left(\frac{X_{(i-1)\Delta} - x}{h}\right) \int_{(i-1)\Delta}^{i\Delta} (X_t - X_{(i-1)\Delta}) g(X_t) dt \\
&= \frac{1}{h} \sum_{i=1}^n f\left(\frac{X_{(i-1)\Delta} - x}{h}\right) \int_{(i-1)\Delta}^{i\Delta} [g_1(X_t) - g_1(X_{(i-1)\Delta})] dt \\
&\quad - \frac{1}{h} \sum_{i=1}^n (xf + hf_1) \left(\frac{X_{(i-1)\Delta} - x}{h}\right) \int_{(i-1)\Delta}^{i\Delta} [g(X_t) - g(X_{(i-1)\Delta})] dt,
\end{aligned}$$

where $f_1(x) = xf(x)$. The stated result now follows immediately from Lemma 11. $\square$

Lemma 14. *Let the assumptions in Lemma 10 hold. Moreover, we assume that g is twice continuously differentiable on D , and that (a) $h^{-2}\Delta \rightarrow_p 0$, (b) $\Delta T(g_A) \rightarrow_p 0$ and (c) $\Delta T(g_B) = o_p((h\ell(T, x))^{1/2})$ uniformly in T as $h \rightarrow 0$ and $\Delta \rightarrow 0$. Then we have*

$$\frac{1}{h\Delta} \sum_{i=1}^n f\left(\frac{X_{(i-1)\Delta} - x}{h}\right) \int_{(i-1)\Delta}^{i\Delta} (i\Delta - t)[g(X_t) - g(X_{(i-1)\Delta})] dt = o_p(\ell(T, x))$$

uniformly in T as $h \rightarrow 0$ and $\Delta \rightarrow 0$.

Proof. The proof is similar to that of Lemma 11. We have from Itô's formula

$$\begin{aligned}
& \int_{(i-1)\Delta}^{i\Delta} (i\Delta - t)[g(X_t) - g(X_{(i-1)\Delta})] dt \\
&= \int_{(i-1)\Delta}^{i\Delta} dt(i\Delta - t) \left[\int_{(i-1)\Delta}^t g_A(X_s) ds + \int_{(i-1)\Delta}^t g_B(X_s) dW_s \right] \\
&= \frac{1}{2} \left[\int_{(i-1)\Delta}^{i\Delta} (i\Delta - t)^2 g_A(X_t) dt + \int_{(i-1)\Delta}^{i\Delta} (i\Delta - t)^2 g_B(X_t) dW_t \right].
\end{aligned}$$

Consequently, it follows that

$$\frac{1}{h\Delta} \sum_{i=1}^n f\left(\frac{X_{(i-1)\Delta} - x}{h}\right) \int_{(i-1)\Delta}^{i\Delta} (i\Delta - t)[g(X_t) - g(X_{(i-1)\Delta})] dt = \frac{M_T + N_T}{2}, \quad (5.43)$$

where

$$\begin{aligned}
N_T &= \frac{1}{h\Delta} \sum_{i=1}^n f\left(\frac{X_{(i-1)\Delta} - x}{h}\right) \int_{(i-1)\Delta}^{i\Delta} (i\Delta - t)^2 g_A(X_t) dt \\
M_T &= \frac{1}{h\Delta} \sum_{i=1}^n f\left(\frac{X_{(i-1)\Delta} - x}{h}\right) \int_{(i-1)\Delta}^{i\Delta} (i\Delta - t)^2 g_B(X_t) dW_t.
\end{aligned}$$

As in the proof of Lemma 11, we let M be a continuous martingale.

We have

$$|N_T| \leq \Delta \frac{\Delta}{h} \sum_{i=1}^n |f|\left(\frac{X_{(i-1)\Delta} - x}{h}\right) \sup_{t \in [0, T]} |g_A(X_t)| = O_p(\Delta \ell(T, x) T(g_A)) = o_p(\ell(T, x)) \quad (5.44)$$

uniformly in T as $h \rightarrow 0$ and $\Delta \rightarrow 0$, due to Lemma 10 and conditions (a) and (b). Moreover, we have

$$\begin{aligned} [M]_T &= \frac{1}{h^2 \Delta^2} \sum_{i=1}^n f^2 \left(\frac{X_{(i-1)\Delta} - x}{h} \right) \int_{(i-1)\Delta}^{i\Delta} (i\Delta - t)^4 g_B^2(X_t) dt \\ &\leq \left(\frac{\Delta^2}{h} \right) \frac{\Delta}{h} \sum_{i=1}^n f^2 \left(\frac{X_{(i-1)\Delta} - x}{h} \right) \sup_{t \in [0, T]} |g_B^2(X_t)| \\ &= O_p(h^{-1} \Delta^2 \ell(T, x) T(g_B^2)) = o_p(\ell^2(T, x)) \end{aligned}$$

uniformly in T as $h \rightarrow 0$ and $\Delta \rightarrow 0$, due to Lemma 10 and conditions (a) and (c), and it follows that

$$M_T = o_p(\ell(T, x)) \quad (5.45)$$

uniformly in T as $h \rightarrow 0$ and $\Delta \rightarrow 0$. The stated result now follows directly from (5.43), (5.44) and (5.45). $\square$

Lemma 15. *Let the assumptions in Lemma 10 hold. Moreover, we assume that g is twice continuously differentiable on D , and that (a) $h^{-2} \Delta \rightarrow_p 0$, (b) $\Delta T(g_{jA}) \rightarrow_p 0$ and (c) $\Delta T(g_{jB}) = o_p((h\ell(T, x))^{1/2})$ uniformly in T as $h \rightarrow 0$ and $\Delta \rightarrow 0$ for $j = 0, 1$ with $g_j(x) = x^j g(x)$. Then we have*

$$\frac{1}{h\Delta} \sum_{i=1}^n f \left(\frac{X_{(i-1)\Delta} - x}{h} \right) \int_{(i-1)\Delta}^{i\Delta} (i\Delta - t)(X_t - X_{(i-1)\Delta}) g(X_t) dt = o_p(\ell(T, x))$$

uniformly in T as $h \rightarrow 0$ and $\Delta \rightarrow 0$.

Proof. The stated result may easily be derived as in Lemma 14, following the proof of Lemma 13. The details are therefore omitted. $\square$

5.2. Proofs of the Theorems

5.2.1 Proof of Theorem 1

Proof. The first part is a direct consequence of Lemma 10 with $f = K$. Notice that

$$\frac{\Delta}{h} \sum_{i=1}^n K \left(\frac{X_{i\Delta} - x}{h} \right) = \frac{\Delta}{h} \sum_{i=1}^n K \left(\frac{X_{(i-1)\Delta} - x}{h} \right) + O(\Delta h^{-1}) \text{ a.s.},$$

due to the boundedness of K . To prove the first part, we apply the occupation times formula and change of variables successively to get

$$\begin{aligned} \frac{1}{h} \int_0^T K \left(\frac{X_t - x}{h} \right) dt &= \frac{1}{h} \int_{-\infty}^{\infty} K \left(\frac{u - x}{h} \right) \ell(T, u) du \\ &= \int_{-\infty}^{\infty} K(u) \ell(T, x + hu) du. \end{aligned}$$

The stated result now follows immediately from Lemma 5 with $f = K$ upon noticing $\int K(x) dx = 1$. $\square$

5.2.2 Proof of Theorem 2

Proof. By the successive applications of Lemmas 10 and 6 with $f = K$ and $g = 1$, we have

$$\begin{aligned} Q_T(K) &= \frac{1}{h} \int_0^T K\left(\frac{X_t - x}{h}\right) dt + O_p\left(\frac{\Delta}{h^2} \ell(T, x)\right) \\ &= \ell(T, x) + o_p(\ell(T, x)) + O_p\left(\frac{\Delta}{h^2} \ell(T, x)\right) \end{aligned} \quad (5.46)$$

uniformly in T as $h \rightarrow 0$ and $\Delta \rightarrow 0$. Moreover, we may set $f = K$ and $g = \mu$ and apply Lemmas 12 and 9 to deduce that

$$\begin{aligned} N_T(K, \mu) &= \frac{1}{h} \int_0^T K\left(\frac{X_t - x}{h}\right) [\mu(X_t) - \mu(x)] dt + O_p\left(\frac{\Delta}{h^2} \ell(T, x)\right) \\ &= \frac{h^2}{2} \iota(K_2) \left(\mu''(x) + 2\mu'(x) \frac{m'(x)}{m(x)} \right) \ell(T, x) \\ &\quad + o_p(h^2 \ell(T, x)) + O_p\left(h^{3/2} \ell(T, x)^{1/2}\right) + O_p\left(\frac{\Delta}{h^2} \ell(T, x)\right) \end{aligned} \quad (5.47)$$

uniformly in T as $h \rightarrow 0$ and $\Delta \rightarrow 0$. Therefore, it follows immediately from (5.46) and (5.47) that

$$\hat{\mu}_p(x) - \mu(x) = \frac{N_T(K, \mu)}{Q_T(K)} = \frac{h^2}{2} \left(\mu''(x) + 2\mu'(x) \frac{m'(x)}{m(x)} \right) \iota_2(K) + o_p(h^2) + O_p\left(h^{3/2} \ell(T, x)^{-1/2}\right)$$

uniformly in T as $h \rightarrow 0$ and $\Delta \rightarrow 0$, under condition (a) of Assumption 4. This establishes the first part for $\hat{\mu}_p(x)$.

To prove the second part, we define M to be a continuous martingale such that

$$M_T = \frac{1}{\sqrt{h}} \sum_{i=1}^n K\left(\frac{X_{(i-1)\Delta} - x}{h}\right) \int_{(i-1)\Delta}^{i\Delta} \sigma(X_t) dW_t,$$

as in the proof of Lemma 11. Note that $M_T = \sqrt{h} M_T(K, \mu)$. We may easily deduce that

$$\begin{aligned} [M]_T &= \frac{1}{h} \sum_{i=1}^n K^2\left(\frac{X_{(i-1)\Delta} - x}{h}\right) \int_{(i-1)\Delta}^{i\Delta} \sigma^2(X_t) dt \\ &= \frac{\Delta}{h} \sum_{i=1}^n K^2\left(\frac{X_{(i-1)\Delta} - x}{h}\right) \sigma^2(X_{(i-1)\Delta}) \\ &\quad + \frac{1}{h} \sum_{i=1}^n K^2\left(\frac{X_{(i-1)\Delta} - x}{h}\right) \int_{(i-1)\Delta}^{i\Delta} [\sigma^2(X_t) - \sigma^2(X_{(i-1)\Delta})] dt \\ &= \sigma^2(x) \iota(K^2) \ell(T, x) [1 + o_p(1)] \end{aligned} \quad (5.48)$$

uniformly in T as $h \rightarrow 0$ and $\Delta \rightarrow 0$, since

$$\begin{aligned} \frac{\Delta}{h} \sum_{i=1}^n K^2\left(\frac{X_{(i-1)\Delta} - x}{h}\right) \sigma^2(X_{(i-1)\Delta}) &= \frac{1}{h} \int_0^T K^2\left(\frac{X_t - x}{h}\right) \sigma^2(X_t) dt + o_p(h^2 \ell(T, x)) \\ &= \sigma^2(x) \iota(K^2) \ell(T, x) + o_p(\ell(T, x)) + o_p(h^2 \ell(T, x)) \end{aligned}$$

by the successive applications of Lemma 12 under conditions (a) and (b) of Assumption 4 and Lemma 6 with $f = K^2$ and $g = \sigma^2$, and

$$\frac{1}{h} \sum_{i=1}^n K^2 \left(\frac{X_{(i-1)\Delta} - x}{h} \right) \int_{(i-1)\Delta}^{i\Delta} [\sigma^2(X_t) - \sigma^2(X_{(i-1)\Delta})] dt = o_p(h^2 \ell(T, x))$$

uniformly in T as $h \rightarrow 0$ and $\Delta \rightarrow 0$, due to Lemma 11 with $f = K^2$ and $g = \sigma^2$.

Moreover, we have

$$\begin{aligned} [W, M]_T &= \frac{1}{\sqrt{h}} \sum_{i=1}^n K \left(\frac{X_{(i-1)\Delta} - x}{h} \right) \int_{(i-1)\Delta}^{i\Delta} \sigma(X_s) ds \\ &= \frac{\Delta}{\sqrt{h}} \sum_{i=1}^n K \left(\frac{X_{(i-1)\Delta} - x}{h} \right) \sigma(X_{(i-1)\Delta}) \\ &\quad + \frac{1}{\sqrt{h}} \sum_{i=1}^n K \left(\frac{X_{(i-1)\Delta} - x}{h} \right) \int_{(i-1)\Delta}^{i\Delta} [\sigma(X_s) - \sigma(X_{(i-1)\Delta})] ds \\ &= \sqrt{h} \left[\frac{\Delta}{h} \sum_{i=1}^n K \left(\frac{X_{(i-1)\Delta} - x}{h} \right) \sigma(X_{(i-1)\Delta}) \right. \\ &\quad \left. + \frac{1}{h} \sum_{i=1}^n K \left(\frac{X_{(i-1)\Delta} - x}{h} \right) \int_{(i-1)\Delta}^{i\Delta} [\sigma(X_s) - \sigma(X_{(i-1)\Delta})] ds \right] \\ &= \sqrt{h} \sigma(x) \ell(T, x) [1 + o_p(1)] \end{aligned} \tag{5.49}$$

uniformly in T as $h \rightarrow 0$ and $\Delta \rightarrow 0$, similarly for $[M]$ above. Consequently, it follows that

$$\frac{[W, M]_T}{[M]_T} = \frac{\sqrt{h} \sigma(x) \ell(T, x) [1 + o_p(1)]}{\sigma^2(x) \iota(K^2) \ell(T, x) [1 + o_p(1)]} = o_p(h^{1/2}) \tag{5.50}$$

and

$$\frac{[W, M]_T}{[W]_T} = \frac{\sqrt{h} \sigma(x) \ell(T, x) [1 + o_p(1)]}{T} = o_p(h^{1/2} T^{-1}) \tag{5.51}$$

uniformly in T as $h \rightarrow 0$ and $\Delta \rightarrow 0$, due to (5.48) and (5.49).

As in the proof of Lemma 3, we may deduce from (5.48) and (5.50) and (5.51) that

$$\ell(T, x)^{-1/2} M_T \rightarrow_d \sigma(x) \iota(K^2)^{1/2} Z, \tag{5.52}$$

where Z is a standard normal random variate independent of $\ell(T, x)$. Since we define $M_T = \sqrt{h} M_T(K, \mu)$ and we have

$$[h \ell(T, x)]^{1/2} \hat{\mu}_q(x) = \frac{\ell(T, x)^{1/2}}{Q_T(K)} \sqrt{h} M_T(K, \mu) = \ell(T, x)^{-1/2} M_T [1 + o_p(1)]$$

uniformly in T as $h \rightarrow 0$ and $\Delta \rightarrow 0$ due to (5.46), the second part for $\hat{\mu}_q(x)$ follows immediately from (5.52).

Finally, for the third part for $\hat{\varepsilon}_\mu(x)$, we simply note that

$$R(K, \mu) = o_p(h^2 \ell(T, x)) \tag{5.53}$$

uniformly in T as $h \rightarrow 0$ and $\Delta \rightarrow 0$ under conditions (a) and (b) of Assumption 4, which is due to Lemma 11 with $g = \mu$. Therefore, the stated result follows immediately from (5.46). The proof is therefore complete. $\square$

5.2.3 Proof of Theorem 3

Proof. Similarly as in (5.46), we have

$$Q_T(K_2) = \iota(K_2)\ell(T, x) + o_p(\ell(T, x)) + O_p\left(\frac{\Delta}{h^2}\ell(T, x)\right) \quad (5.54)$$

uniformly in T as $h \rightarrow 0$ and $\Delta \rightarrow 0$. Moreover, it follows from Lemma 8 with $f = K_1$ and $g = 1$ that

$$\begin{aligned} Q_T(K_1) &= \frac{1}{h} \int_0^T K_1\left(\frac{X_t - x}{h}\right) dt + O_p\left(\frac{\Delta}{h^2}\ell(T, x)\right) \\ &= \iota(K_2) \frac{m'(x)}{m(x)} h\ell(T, x) + o_p(h\ell(T, x)) \end{aligned} \quad (5.55)$$

uniformly in T as $h \rightarrow 0$ and $\Delta \rightarrow 0$. Therefore, we have

$$Q_T(K)Q_T(K_2) - Q_T(K_1)^2 = \iota(K_2)\ell(T, x)^2 + o_p(\ell(T, x)^2) \quad (5.56)$$

uniformly in T as $h \rightarrow 0$ and $\Delta \rightarrow 0$, due to (5.46), (5.54) and (5.55).

We may apply Lemmas 12 and 7 successively with $f = K_1$ and $g = \mu$, and deduce that

$$\begin{aligned} N_T(K_1, \mu) &= \frac{1}{h} \int_0^T K_1\left(\frac{X_t - x}{h}\right) [\mu(X_t) - \mu(x)] dt + O_p\left(\frac{\Delta}{h^2}\ell(T, x)\right) \\ &= \iota(K_2)\mu'(x)h\ell(T, x) + o_p(h\ell(T, x)) \end{aligned} \quad (5.57)$$

uniformly in T as $h \rightarrow 0$ and $\Delta \rightarrow 0$. Moreover, it follows from (5.47), (5.54), (5.55) and (5.57) that

$$\begin{aligned} &N_T(K, \mu)Q_T(K_2) - N_T(K_1, \mu)Q_T(K_1) \\ &= \left[\frac{\iota(K_2)}{2} \left(\mu''(x) + 2\mu'(x) \frac{m'(x)}{m(x)} \right) h^2\ell(T, x) + o_p(h^2\ell(T, x)) \right] \left[\iota(K_2)\ell(T, x) + o_p(\ell(T, x)) \right] \\ &\quad - \left[\iota(K_2)\mu'(x)h\ell(T, x) + o_p(h\ell(T, x)) \right] \left[\iota(K_2) \frac{m'(x)}{m(x)} h\ell(T, x) + o_p(h\ell(T, x)) \right] \\ &= \frac{\iota(K_2)^2}{2} \mu''(x) h^2\ell(T, x)^2 + o_p(h^2\ell(T, x)^2) + O_p\left((h\ell(T, x))^{3/2}\right) \end{aligned} \quad (5.58)$$

uniformly in T as $h \rightarrow 0$ and $\Delta \rightarrow 0$. Consequently, we have

$$\begin{aligned} \tilde{\mu}_p(x) - \mu(x) &= \frac{N_T(K, \mu)Q_T(K_2) - N_T(K_1, \mu)Q_T(K_1)}{Q_T(K)Q_T(K_2) - Q_T(K_1)^2} \\ &= \frac{h^2}{2} \iota(K_2)\mu''(x) + o_p(h^2) + O_p\left(h^{3/2}\ell(T, x)^{-1/2}\right) \end{aligned}$$

from (5.56) and (5.58), uniformly in T as $h \rightarrow 0$ and $\Delta \rightarrow 0$, as was to be shown for the first part for $\tilde{\mu}_p(x)$.

For the second part, we note that

$$M_T(K_1, \mu) = O_p\left(h^{-1/2}\ell(T, x)^{1/2}\right)$$

uniformly in T as $h \rightarrow 0$ and $\Delta \rightarrow 0$, which follows exactly as in the proof of Theorem 2 for $M_T(K, \mu)$. As a result, it follows from (5.55) that

$$M_T(K_1, \mu)Q_T(K_1) = O_p(h^{-1/2}\ell(T, x)^{1/2})O_p(h\ell(T, x)) = O_p\left(h^{1/2}\ell(T, x)^{3/2}\right) \quad (5.59)$$

uniformly in T as $h \rightarrow 0$ and $\Delta \rightarrow 0$. Therefore, we have

$$\begin{aligned}
& [h\ell(T, x)]^{1/2} \tilde{\mu}_q(x) \\
&= [h\ell(T, x)]^{1/2} \frac{M_T(K, \mu)Q_T(K_2) - M_T(K_1, \mu)Q_T(K_1)}{Q_T(K)Q_T(K_2) - Q_T(K_1)^2} \\
&= [h\ell(T, x)]^{1/2} \left(\frac{\iota(K_2)\ell(T, x)[1 + o_p(1)]}{\iota(K_2)\ell(T, x)^2[1 + o_p(1)]} M_T(K, \mu) + \frac{O_p(h^{1/2}\ell(T, x)^{3/2})}{O_p(\ell(T, x)^2)} \right) \\
&= \frac{\sqrt{h}M_T(K, \mu)}{\ell(T, x)^{1/2}} [1 + o_p(1)] + O_p(h) \\
&= \frac{\sqrt{h}M_T(K, \mu)}{\ell(T, x)^{1/2}} + o_p(1)
\end{aligned}$$

uniformly in T as $h \rightarrow 0$ and $\Delta \rightarrow 0$, due to (5.54), (5.56) and (5.59). As shown in the proof of Theorem 2, we have

$$\frac{\sqrt{h}M_T(K, \mu)}{\ell(T, x)^{1/2}} \rightarrow_d \sigma(x)\iota(K^2)^{1/2}Z,$$

where Z is a standard normal random variate independent of $\ell(T, x)$. The proof of the second part for $\tilde{\mu}_q(x)$ is therefore complete.

Finally, we have from Lemma 11

$$R_T(K_1, \mu) = o_p(h^2\ell(T, x)) \quad (5.60)$$

uniformly in T as $h \rightarrow 0$ and $\Delta \rightarrow 0$, under conditions (a) and (b) of Assumption 4. It therefore follows from (5.53), (5.54), (5.55) and (5.60) that

$$R_T(K, \mu)Q_T(K_2) - R_T(K_1, \mu)Q_T(K_1) = o_p(h^2\ell(T, x)^2) + o_p(h^3\ell(T, x)^2)$$

uniformly in T as $h \rightarrow 0$ and $\Delta \rightarrow 0$, and it follows from (5.56) that

$$\tilde{\varepsilon}_\mu(x) = \frac{R_T(K, \mu)Q_T(K_2) - R_T(K_1, \mu)Q_T(K_1)}{Q_T(K)Q_T(K_2) - Q_T(K_1)^2} = o_p(h^2),$$

as was to be shown. $\square$

5.2.4 Proof of Theorem 4

Proof. The proof of $\hat{\sigma}_p^2(x)$ is completely analogous to that of $\hat{\mu}_p(x)$ in the proof of Theorem 2. The details are therefore omitted. Note that we have the same conditions in (a) and (b) of Assumption 5 for $\sigma^2(x)$ as those introduced in Theorem 2 for $\mu(x)$. To establish the second part for $\hat{\sigma}_q^2(x)$, we first define a continuous martingale M such that

$$M_T = \sqrt{\frac{2}{h\Delta}} \sum_{i=1}^n K\left(\frac{X_{(i-1)\Delta} - x}{h}\right) \int_{(i-1)\Delta}^{i\Delta} (X_t - X_{(i-1)\Delta}) \sigma(X_t) dW_t,$$

whose quadratic variation $[M]$ at T is given by

$$[M]_T = \frac{2}{h\Delta} \sum_{i=1}^n K^2\left(\frac{X_{(i-1)\Delta} - x}{h}\right) \int_{(i-1)\Delta}^{i\Delta} (X_t - X_{(i-1)\Delta})^2 \sigma^2(X_t) dt.$$

We write

$$\begin{aligned}
& \int_{(i-1)\Delta}^{i\Delta} (X_t - X_{(i-1)\Delta})^2 \sigma^2(X_t) dt \\
&= \sigma^2(X_{(i-1)\Delta}) \int_{(i-1)\Delta}^{i\Delta} (X_t - X_{(i-1)\Delta})^2 dt \\
&+ \int_{(i-1)\Delta}^{i\Delta} (X_t - X_{(i-1)\Delta})^2 [\sigma^2(X_t) - \sigma^2(X_{(i-1)\Delta})] dt,
\end{aligned} \tag{5.61}$$

and use Itô's formula to deduce that

$$\begin{aligned}
\int_{(i-1)\Delta}^{i\Delta} (X_t - X_{(i-1)\Delta})^2 dt &= \frac{\Delta^2}{2} \sigma^2(X_{(i-1)\Delta}) \\
&+ 2 \int_{(i-1)\Delta}^{i\Delta} (i\Delta - t)(X_t - X_{(i-1)\Delta}) \mu(X_t) dt \\
&+ \int_{(i-1)\Delta}^{i\Delta} (i\Delta - t) [\sigma^2(X_t) - \sigma^2(X_{(i-1)\Delta})] dt \\
&+ 2 \int_{(i-1)\Delta}^{i\Delta} (i\Delta - t)(X_t - X_{(i-1)\Delta}) \sigma(X_t) dW_t.
\end{aligned} \tag{5.62}$$

Then it follows from (5.61) and (5.62) that

$$[M]_T = A_T + B_T + C_T + D_T + E_T, \tag{5.63}$$

where

$$\begin{aligned}
A_T &= \frac{\Delta}{h} \sum_{i=1}^n K^2 \left(\frac{X_{(i-1)\Delta} - x}{h} \right) \sigma^4(X_{(i-1)\Delta}) \\
B_T &= \frac{4}{h\Delta} \sum_{i=1}^n K^2 \left(\frac{X_{(i-1)\Delta} - x}{h} \right) \sigma^2(X_{(i-1)\Delta}) \int_{(i-1)\Delta}^{i\Delta} (i\Delta - t)(X_t - X_{(i-1)\Delta}) \mu(X_t) dt \\
C_T &= \frac{2}{h\Delta} \sum_{i=1}^n K^2 \left(\frac{X_{(i-1)\Delta} - x}{h} \right) \sigma^2(X_{(i-1)\Delta}) \int_{(i-1)\Delta}^{i\Delta} (i\Delta - t) [\sigma^2(X_t) - \sigma^2(X_{(i-1)\Delta})] dt \\
D_T &= \frac{4}{h\Delta} \sum_{i=1}^n K^2 \left(\frac{X_{(i-1)\Delta} - x}{h} \right) \sigma^2(X_{(i-1)\Delta}) \int_{(i-1)\Delta}^{i\Delta} (i\Delta - t)(X_t - X_{(i-1)\Delta}) \sigma(X_t) dW_t \\
E_T &= \frac{2}{h\Delta} \sum_{i=1}^n K^2 \left(\frac{X_{(i-1)\Delta} - x}{h} \right) \int_{(i-1)\Delta}^{i\Delta} (X_t - X_{(i-1)\Delta})^2 [\sigma^2(X_t) - \sigma^2(X_{(i-1)\Delta})] dt,
\end{aligned}$$

which we consider below in sequel.

For the first term, we apply Lemmas 12 and 6 successively with $f = K^2$ and $g = \sigma^4$ and have

$$\begin{aligned}
A_T &= \frac{1}{h} \int_0^T K^2 \left(\frac{X_t - x}{h} \right) \sigma^4(X_t) dt + O_p \left(\frac{\Delta}{h^2} \ell(T, x) \right) \\
&= \sigma^4(x) \iota(K^2) \ell(T, x) [1 + o_p(1)]
\end{aligned} \tag{5.64}$$

uniformly in T as $h \rightarrow 0$ and $\Delta \rightarrow 0$, under conditions (a) and (b) of Assumption 5. This is the leading term, and all other terms become asymptotically negligible, as we will show below. For the second and fourth terms, we have

$$B_T, D_T = o_p(\ell(T, x)) \tag{5.65}$$

uniformly in T as $h \rightarrow 0$ and $\Delta \rightarrow 0$, under condition (c) of Assumption 5, due to Lemma 15 with $f = K^2$ and $g = \mu, \sigma$, respectively for B_T and D_T . Similarly, we may deduce from Lemma 14 that

$$C_T = o_p(\ell(T, x)) \quad (5.66)$$

uniformly in T as $h \rightarrow 0$ and $\Delta \rightarrow 0$, under condition (b) of Assumption 5. Note that Lemmas 14 and 15 continue to hold under the presence of an additional term $\sigma^2(X_{(i-1)\Delta})$, since

$$\sigma^2(X_{(i-1)\Delta}) = \sigma^2\left(x + h \left(\frac{X_{(i-1)\Delta} - x}{h}\right)\right)$$

and σ^2 is continuous.

For the last term, we note that

$$|\sigma^2(X_t) - \sigma^2(X_{(i-1)\Delta})| \leq T(\sigma^{2'})|X_t - X_{(i-1)\Delta}|,$$

and that

$$|X_t - X_{(i-1)\Delta}| \leq \left| \int_{(i-1)\Delta}^t \mu(X_s) ds \right| + \left| \int_{(i-1)\Delta}^t \sigma(X_s) dW_s \right|$$

for all $1 \leq i \leq n$ and $t \in \Delta_i$, where $\Delta_i = [(i-1)\Delta, i\Delta]$. We have

$$\left| \int_{(i-1)\Delta}^t \mu(X_s) ds \right| \leq \Delta T(\mu), \quad \left| \int_{(i-1)\Delta}^t \sigma^2(X_s) ds \right| \leq \Delta T(\sigma^2)$$

for all $1 \leq i \leq n$ and $t \in \Delta_i$. Moreover, by the martingale representation theorem, we have

$$\begin{aligned} \int_{(i-1)\Delta}^t \sigma(X_s) dW_s &= V\left(\int_0^t \sigma(X_s)^2 ds\right) - V\left(\int_0^{(i-1)\Delta} \sigma(X_s)^2 ds\right) \\ &= O_p\left((\Delta T(\sigma^2))^{1/2}\right) \end{aligned}$$

uniformly for all $1 \leq i \leq n$ and $t \in \Delta_i$, where V is a standard Brownian motion. Consequently, we have

$$|\sigma^2(X_t) - \sigma^2(X_{(i-1)\Delta})| \leq O_p(\Delta T(\mu)T(\sigma^{2'})) + O_p(\Delta^{1/2}T(\sigma)T(\sigma^{2'})) = o_p(1) \quad (5.67)$$

uniformly for all $1 \leq i \leq n$ and $t \in \Delta_i$, under condition (d) of Assumption 5.

It follows from (5.67) that

$$E = o_p\left(\frac{1}{h\Delta} \sum_{i=1}^n K^2\left(\frac{X_{(i-1)\Delta} - x}{h}\right) \int_{(i-1)\Delta}^{i\Delta} (X_t - X_{(i-1)\Delta})^2 dt\right).$$

However, we may write

$$\begin{aligned}
& \frac{1}{h\Delta} \sum_{i=1}^n K^2 \left(\frac{X_{(i-1)\Delta} - x}{h} \right) \int_{(i-1)\Delta}^{i\Delta} (X_t - X_{(i-1)\Delta})^2 dt \\
&= \frac{\Delta}{2h} \sum_{i=1}^n K^2 \left(\frac{X_{(i-1)\Delta} - x}{h} \right) \sigma^2(X_{(i-1)\Delta}) \\
&\quad + \frac{2}{h\Delta} \sum_{i=1}^n K^2 \left(\frac{X_{(i-1)\Delta} - x}{h} \right) \int_{(i-1)\Delta}^{i\Delta} (i\Delta - t)(X_t - X_{(i-1)\Delta}) \mu(X_t) dt \\
&\quad + \frac{1}{h\Delta} \sum_{i=1}^n K^2 \left(\frac{X_{(i-1)\Delta} - x}{h} \right) \int_{(i-1)\Delta}^{i\Delta} (i\Delta - t)[\sigma^2(X_t) - \sigma^2(X_{(i-1)\Delta})] dt \\
&\quad + \frac{2}{h\Delta} \sum_{i=1}^n K^2 \left(\frac{X_{(i-1)\Delta} - x}{h} \right) \int_{(i-1)\Delta}^{i\Delta} (i\Delta - t)(X_t - X_{(i-1)\Delta}) \sigma(X_t) dW_t,
\end{aligned}$$

and show that it is of order $O_p(\ell(T, x))$ uniformly in T as $h \rightarrow 0$ and $\Delta \rightarrow 0$ by applying Lemmas 12, 6, 14 and 15 successively as in (5.64), (5.65) and (5.66). Therefore, we have

$$E_T = o_p(\ell(T, x))$$

uniformly in T as $h \rightarrow 0$ and $\Delta \rightarrow 0$, which, together with (5.64), (5.65) and (5.66), yields

$$[M]_T = \sigma^4(x) \iota(K^2) \ell(T, x) [1 + o_p(1)] \quad (5.68)$$

from (5.63), uniformly in T as $h \rightarrow 0$ and $\Delta \rightarrow 0$.

Moreover, we have

$$[W, M]_T = \sqrt{\frac{2}{h\Delta}} \sum_{i=1}^n K \left(\frac{X_{(i-1)\Delta} - x}{h} \right) \int_{(i-1)\Delta}^{i\Delta} (X_t - X_{(i-1)\Delta}) \sigma(X_t) dt,$$

and it follows from the application of Lemma 13 with $f = K$ and $g = \sigma$ that

$$[W, M]_T = o_p \left(h^{1/2} \ell(T, x) \right) \quad (5.69)$$

uniformly in T as $h \rightarrow 0$ and $\Delta \rightarrow 0$, under condition (c) of Assumption 5. Therefore, we may deduce from (5.68) and (5.69) that

$$\frac{[W, M]_T}{[M]_T} = o_p(h^{1/2}),$$

and

$$\frac{[W, M]_T}{[W]_T} = o_p(h^{1/2} T^{-1}).$$

Consequently, we may deduce similarly as in the proof of Theorem 2 that

$$\ell(T, x)^{-1/2} M_T \rightarrow_d \sigma^2(x) \iota(K^2)^{1/2} Z, \quad (5.70)$$

where Z is a standard normal random variate independent of $\ell(T, x)$.

To obtain the stated result for $\hat{\sigma}_q^2(x)$, we simply note that

$$\begin{aligned} \left[\frac{h\ell(T, x)}{\Delta} \right]^{1/2} \hat{\sigma}_q^2(x) &= \left[\frac{h\ell(T, x)}{\Delta} \right]^{1/2} \frac{2M_T(K, \sigma^2)}{Q_T(K)} \\ &= \frac{\sqrt{2}\ell(T, x)^{1/2}}{\ell(T, x)[1 + o_p(1)]} \sqrt{\frac{2h}{\Delta}} M_T(K, \sigma^2) \\ &= \frac{\sqrt{2}M_T}{\ell(T, x)^{1/2}[1 + o_p(1)]} \rightarrow_d \sqrt{2}\sigma^2(x)\iota(K^2)^{1/2}Z, \end{aligned}$$

as was to be shown.

For $\hat{\varepsilon}_{\sigma^2}(x)$, we apply Lemma 11 with $f = K$ and $g = \sigma^2$ to have

$$R_T(K, \sigma^2) = o_p(h^2\ell(T, x)) \quad (5.71)$$

uniformly in T as $h \rightarrow 0$ and $\Delta \rightarrow 0$, under conditions (a) and (b) of Assumption 5, and Lemma 13 with $f = K$ and $g = \mu$ to have

$$S_T(K) = o_p(h^2\ell(T, x)) \quad (5.72)$$

uniformly in T as $h \rightarrow 0$ and $\Delta \rightarrow 0$, under conditions (a) and (c) of Assumption 5. Now it follows from (5.46) that

$$\hat{\varepsilon}_{\sigma^2}(x) = \frac{R_T(K, \sigma^2) + 2S_T(K)}{Q_T(K)} = o_p(h^2),$$

and the proof is complete. $\square$

5.2.5 Proof of Theorem 5

Proof. The proof of $\tilde{\sigma}_p^2(x)$ is essentially the same as that of $\tilde{\mu}_p(x)$ in the proof of Theorem 3, if we replace μ with σ^2 . For $\tilde{\sigma}_q^2(x)$, we note that

$$M_T(K_1, \sigma^2) = O_p(h^{-1/2}\Delta^{1/2}\ell(T, x)^{1/2})$$

uniformly in T as $h \rightarrow 0$ and $\Delta \rightarrow 0$, which can be shown similarly as $M_T(K, \sigma^2)$ in the proof of Theorem 4. Therefore, we may deduce from (5.55) that

$$\begin{aligned} M_T(K_1, \sigma^2)Q_T(K_1) &= O_p(h^{-1/2}\Delta^{1/2}\ell(T, x)^{1/2})O_p(h\ell(T, x)) \\ &= O_p\left((h\Delta)^{1/2}\ell(T, x)^{3/2}\right) \end{aligned} \quad (5.73)$$

uniformly in T as $h \rightarrow 0$ and $\Delta \rightarrow 0$. Consequently, we have

$$\begin{aligned}
& \left[\frac{h\ell(T, x)}{\Delta} \right]^{1/2} \tilde{\sigma}_q^2(x) \\
&= 2 \left[\frac{h\ell(T, x)}{\Delta} \right]^{1/2} \frac{M_T(K, \sigma^2)Q_T(K_2) - M_T(K_1, \sigma^2)Q_T(K_1)}{Q_T(K)Q_T(K_2) - Q_T(K_1)^2} \\
&= 2 \left[\frac{h\ell(T, x)}{\Delta} \right]^{1/2} \left(\frac{\iota(K_2)\ell(T, x)[1 + o_p(1)]}{\iota(K_2)\ell(T, x)^2[1 + o_p(1)]} M_T(K, \sigma^2) + \frac{O_p((h\Delta)^{1/2}\ell(T, x)^{3/2})}{O_p(\ell(T, x)^2)} \right) \\
&= \sqrt{2} \sqrt{\frac{2h}{\Delta}} \frac{M_T(K, \sigma^2)}{\ell(T, x)^{1/2}} [1 + o_p(1)] + O_p(h) \\
&= \sqrt{2} \sqrt{\frac{2h}{\Delta}} \frac{M_T(K, \sigma^2)}{\ell(T, x)^{1/2}} + o_p(1),
\end{aligned}$$

due to (5.54), (5.56) and (5.73). As shown in the proof of Theorem 4, we have

$$\sqrt{\frac{2h}{\Delta}} \frac{M_T(K, \sigma^2)}{\ell(T, x)^{1/2}} \rightarrow_d \sigma^2(x) \iota(K^2)^{1/2} Z,$$

where Z is a standard normal random variate independent of $\ell(T, x)$. This was to be shown for the second part.

Finally, it follows from Lemma 11

$$R_T(K_1, \sigma^2) = o_p(h^2\ell(T, x)) \quad (5.74)$$

uniformly in T as $h \rightarrow 0$ and $\Delta \rightarrow 0$, under conditions (a) and (b) of Assumption 5, with $f = K_1$ and $g = \sigma^2$, analogously as (5.71). Therefore, we may deduce from (5.71), (5.54), (5.55), (5.72) and (5.74) that

$$\begin{aligned}
& [R_T(K, \sigma^2) + 2S_T(K)]Q_T(K_2) - [R_T(K_1, \sigma^2) + 2S_T(K_1)]Q_T(K_1) \\
&= o_p(h^2\ell(T, x)^2) + o_p(h^3\ell(T, x)^2)
\end{aligned}$$

uniformly in T as $h \rightarrow 0$ and $\Delta \rightarrow 0$, from which and (5.56) we have

$$\tilde{\varepsilon}_{\sigma^2}(x) = \frac{[R_T(K, \sigma^2) + 2S_T(K)]Q_T(K_2) - [R_T(K_1, \sigma^2) + 2S_T(K_1)]Q_T(K_1)}{Q_T(K)Q_T(K_2) - Q_T(K_1)^2} = o_p(h^2),$$

and the proof is complete. $\square$

5.2.6 Proof of Theorem 6

Proof. Define two continuous martingales M^a and M^b such that

$$\begin{aligned}
M_T^a &= \frac{1}{\sqrt{h}} \sum_{i=1}^n K\left(\frac{X_{(i-1)\Delta} - x}{h}\right) \int_{(i-1)\Delta}^{i\Delta} \sigma(X_t) dW_t \\
M_T^b &= \sqrt{\frac{2}{h\Delta}} \sum_{i=1}^n K\left(\frac{X_{(i-1)\Delta} - x}{h}\right) \int_{(i-1)\Delta}^{i\Delta} (X_t - X_{(i-1)\Delta}) \sigma(X_t) dW_t,
\end{aligned}$$

which are introduced respectively in the proofs of Theorems 2 and 4. It is shown in the proofs of Theorems 2 and 4 that

$$\ell(T, x)^{-1/2} M_T^a \rightarrow_d Z_a \quad (5.75)$$

$$\ell(T, x)^{-1/2} M_T^b \rightarrow_d Z_b, \quad (5.76)$$

where Z_a and Z_b are normal random variates that are independent of $\ell(T, x)$.

It is straightforward to deduce that

$$\begin{aligned} [M^a, M^b]_T &= \sqrt{\frac{2}{\Delta}} \frac{1}{h} \sum_{i=1}^n K^2 \left(\frac{X_{(i-1)\Delta} - x}{h} \right) \int_{(i-1)\Delta}^{i\Delta} (X_t - X_{(i-1)\Delta}) \sigma^2(X_t) dt \\ &= O(\Delta^{-1/2}) o_p(h^2 \ell(T, x)) = o_p \left(h^2 \Delta^{-1/2} \ell(T, x) \right) \end{aligned} \quad (5.77)$$

uniformly in T as $h \rightarrow 0$ and $\Delta \rightarrow 0$, due to Lemma 13. However, we have

$$[M^a]_T, [M^b]_T = O_p(\ell(T, x)) \quad (5.78)$$

uniformly in T as $h \rightarrow 0$ and $\Delta \rightarrow 0$, as shown in the proofs of Theorems 2 and 4. We may therefore deduce from (5.77) and (5.78) that

$$\begin{aligned} \frac{[M^a, M^b]_T}{[M^a]_T} &= o_p(h^2 \Delta^{-1/2}) = o_p(1) \\ \frac{[M^a, M^b]_T}{[M^b]_T} &= o_p(h^2 \Delta^{-1/2}) = o_p(1) \end{aligned}$$

under condition (a) of Assumption 4 or 5. Now we may show as in the proof of Lemma 3 that the limit random variates Z_a and Z_b in (5.75) and (5.76) are independent of each other, from which the asymptotic independence of $\hat{\mu}(x)$ and $\tilde{\mu}(x)$ with $\hat{\sigma}^2(x)$ and $\tilde{\sigma}^2(x)$ readily follow. The proof is therefore complete. $\square$

References

- Aït-Sahalia, Y., 1996. Nonparametric pricing of interest rate derivative securities. *Econometrica* 64, 527–560.
- Bandi, F. M., Phillips, P. C. B., 2003. Fully nonparametric estimation of scalar diffusion models. *Econometrica* 71, 241–283.
- Bandi, F. M., Phillips, P. C. B., 2010. Nonstationary continuous-time processes. In: Aït-Sahalia, Y., Hansen, L. P. (Eds.), *Handbook of Financial Econometrics*. North-Holland, Amsterdam, The Netherlands, pp. 139–201.
- Black, F., Scholes, M., 1973. The pricing of options and corporate liabilities. *Journal of Political Economy* 81, 637–654.
- Borkovec, M., Klüppelberg, C., 1998. Extremal behaviour of diffusion models in finance. *Extremes* 1, 47–80.
- Borodin, A. N., 1989. Brownian local time. *Russian Mathematical Surveys* 44, 1–51.
- Bosq, D., 1998. *Nonparametric Statistics for Stochastic Processes*, Lecture Notes in Statistics 110, 2nd Edition. Springer, New York.

- Brugière, P., 1993. Théorème de limite centrale pour un estimateur nonparamétrique de la variance d'un processus de diffusion multidimensionnelle. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques* 29, 357–389.
- Cai, Z., Hong, Y., 2003. Nonparametric methods in continuous-time finance: A selective review. In: Akritas, M. G., Politis, D. N. (Eds.), *Recent Advances and Trends in Nonparametric Statistics*. Elsevier, The Netherlands, pp. 283–302.
- Chen, X., Hansen, L. P., Carrasco, M., 2010. Nonlinearity and temporal dependence. *Journal of Econometrics* 155, 155–169.
- Corradi, V., White, H., 1999. Specification tests for the variance of a diffusion. *Journal of Time Series Analysis* 20, 253–270.
- Cox, J. C., Ingersoll, J. E., Ross, S. A., 1985. A theory of the term structure of interest rates. *Econometrica* 53, 385–408.
- Eisenbaum, N., Kaspi, H., 2007. On the continuity of local times of Borel right Markov processes. *Annals of Probability* 35, 915–934.
- Fan, J., 2005. A selective overview of nonparametric methods in financial econometrics (with discussion). *Statistical Science* 20, 317–337.
- Fan, J., Jiang, J., Zhang, C., Zhou, Z., 2003. Time-dependent diffusion models for term structure dynamics and the stock price volatility. *Statistica Sinica* 13, 965–992.
- Fan, J., Zhang, C., 2003. A reexamination of diffusion estimations with applications to financial model validation. *Journal of American Statistical Association* 13, 965–992.
- Florens-Zmirou, D., 1993. On estimating the diffusion coefficient from discrete observations. *Journal of Applied Probability* 30, 790–804.
- Genon-Catalot, V., Larédo, C., Picard, D., 1992. Non-parametric estimation of the diffusion coefficient by wavelets methods. *Scandinavian Journal of Statistics* 19, 317–335.
- Hoffmann, M., 1999. Adaptive estimation in diffusion processes. *Stochastic Processes and Their Applications* 79, 135–163.
- Höpfner, R., Kutoyants, Y. A., 2003. On a problem of statistical inference in null recurrent diffusions. *Statistical Inference for Stochastic Processes* 6, 25–42.
- Höpfner, R., Löcherbach, E., 2003. Limit Theorems for Null Recurrent Markov Processes. *Memoirs of the American Mathematical Society* 768, Providence, RI.

- Jacod, J., 1998. Rates of convergence to the local time of a diffusion. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques* 34, 505–544.
- Jacod, J., 2000. Non-parametric kernel estimation of the coefficient of a diffusion. *Scandinavian Journal of Statistics* 27, 83–96.
- Jeganathan, P., 1995. Some aspects of asymptotic theory with applications to time series models. *Econometric Theory* 11, 818–887.
- Jeong, M., Park, J. Y., 2011. Asymptotic theory of maximum likelihood estimator for diffusion model. Tech. rep., Indiana University.
- Jiang, G. J., Knight, J., 1997. A nonparametric approach to the estimation of diffusion processes – with an application to a short-term interest rate model. *Econometric Theory* 13, 615–645.
- Karatzas, I., Shreve, S. E., 1991. *Brownian Motion and Stochastic Calculus*. Springer-Verlag.
- Karlsen, H. A., Myklebust, T., Tjøstheim, D., 2007. Nonparametric estimation in a nonlinear cointegration type model. *Annals of Statistics* 35, 252–299.
- Karlsen, H. A., Tjøstheim, D., 2001. Nonparametric estimation in null recurrent time series. *Annals of Statistics* 29, 372–416.
- Park, J. Y., Phillips, P. C., 2001. Nonlinear regressions with integrated time series. *Econometrica* 69, 117–161.
- Pollard, D., 1984. *Convergence of Stochastic Processes*. Springer-Verlag.
- Renò, R., 2008. Nonparametric estimation of the diffusion coefficient of stochastic volatility models. *Econometric Theory* 24, 1174–1206.
- Revuz, D., Yor, M., 1994. *Continuous Martingales and Brownian Motion*, 2nd Edition. Springer-Verlag.
- Robinson, P. M., 1983. Nonparametric estimators for time series. *Journal of Time Series Analysis* 4, 185–207.
- Schienle, M., 2008. Nonparametric nonstationary regression. Tech. rep., University of Mannheim.
- Stanton, R., 1997. A nonparametric model of term structure dynamics and the market price of interest rate risk. *The Journal of Finance* 52, 1973–2002.
- Stone, C., 1963. Limit theorems for random walks, birth and death processes, and diffusion processes. *Illinois Journal of Mathematics* 7, 638–660.
- van Zanten, H., 2000. A multivariate central limit theorem for continuous local martingales. *Statistics & Probability Letters* 3, 229–235.
- Vasicek, O., 1977. An equilibrium characterization of the term structure. *Journal of Financial Economics* 5, 177–188.