Abstract

Essays in the Econometrics of Continuous-Time Finance

Federico M. Bandi

1999

This dissertation is devoted to the study and empirical implementation of new methods in the econometrics of continuous-time finance. The first chapter is concerned with the nonparametric estimation of the drift and diffusion function of general continuous-time homogeneous stochastic differential equations. Minimal requirements are placed on the data generating mechanism, allowing for both stationary and nonstationary systems, and the available data is assumed to be a set of discrete sample observations. Econometric estimation proceeds by constructing refined sample analogues of unknown drift and diffusion function. Cross-restrictions on the functional forms are not imposed, nor is the existence of a time-invariant marginal data density and, in consequence, the new approach is robust against deviations from stationarity. We prove consistency of the point estimates and pointwise weak convergence to mixtures of normal laws, where the mixture variates depend on the chronological local time of the underlying semimartingale, that is on the amount of time spent by the process in the spatial vicinity of each point.

The second chapter focuses on the application of this new method to a well-known problem in empirical finance, namely the estimation of the short-term interest rate dynamics in a continuous-time framework. The approach to data analysis is twofold. First, a descriptive analysis of the time series is conducted using econometric estimates of the local time, which is treated as a spatial density function, along lines pioneered in Phillips (1998). Spatial densities (and various functionals of them, such as spatial hazard rates) are newly developed descriptive

tools for data analysis that are applicable when the series is nonstationary or, more strictly, when stationarity cannot be guaranteed [c.f. Phillips (1998) and Phillips and Park (1998)]. Second, nonparametric estimates of the drift and diffusion function and associated confidence intervals are obtained for the interest rate process.

The third chapter of this thesis discusses the finite sample performance of fully nonparametric estimators of the drift and diffusion function of general, potentially nonlinear and homogeneous stochastic differential equations. We compare the estimators in the first chapter to those suggested in recent papers by Jiang and Knight (1997) and Stanton (1997). Theoretical justification for the different functional approaches is based on specific assumptions on the limit theory and the underlying process. The stringency of these assumptions in finite sample is investigated by evaluating the performance of the estimators in the presence of various simulated underlying processes.

Essays in the Econometrics of Continuous-Time Finance

A Dissertation
Presented to the Faculty of the Graduate School
of
Yale University
in Candidacy for the Degree of
Doctor of Philosophy

by

Federico M. Bandi

Dissertation Director: Peter C. B. Phillips

November 1999

UMI Number: 9954370

Bell & Howell Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

Contents

List of Tables

List of Figures

List of Figures

ACKNOWLEDGMENTS

I am deeply grateful to the members of my dissertation committee, Peter C. B. Phillips (chair), Nagpurnanand R. Prabhala and Christopher Sims. Special thanks go to Peter for the continuous and invaluable support and for the many hours we spent together. It has been a personal pleasure to work with him and a fantastic learning experience. His teaching will always be a great source of inspiration for me. I thank Nagpurnanand and Chris for encouragement and numerous suggestions. I recall a New Year's day that Nagpurnanand spent with me discussing my work and my academic job market. His words on that day made it easier for me to face the first stage of my career.

I am also indebted to colleagues and friends who spent their precious time working with me, reading my papers and trying to understand them at seminars. In particular, I thank Anders Karlsson, Jerry Mc Donald, Thong Nguyen and Kostya Tyurin. I also wish to thank seminar participants at Berkeley, Bocconi University, Chicago, Duke, Ente Einaudi, Lehman Brothers, London Business School, New York University, Princeton, Rochester, Yale, Universita' di Bologna, Universita' di Napoli Federico II and the 1999 WFA meetings at Santa Monica.

Finally, I appreciated the financial support of Mediocredito Centrale and the Alfred P. Sloan Foundation.

A Francesca e ai miei genitori, Silvana e Massimo

Part I

Econometric Estimation of Diffusion Models [with Peter C. B. Phillips]

1. Introduction

Many popular models in economics and finance, like those for pricing derivative securities, involve diffusion processes formulated in continuous-time as stochastic differential equations. These processes have been used to model options prices, the term structure of interest rates, exchange rates, and foreign currency interest rates, *inter alia*. A recent introduction to some of these applications is given in Baxter and Rennie (1996). Stochastic differential equations have also been used to model macroeconomic aggregates like consumption and investment, and systems of such equations have been used for many years to model economic activity at the national level, as described in Bergstrom (1988). In all these applications, statistical estimation involves the use of discrete data. It is then necessary to identify and estimate with discretely sampled observations the parameters and functionals of a process that is defined in continuous time.

The stochastic differential equation that defines a diffusion process, like X_t in (2.1) below, involves two components. These components measure the conditional drift, $\mu(X_t)$, and the conditional variation, $\sigma^2(X_t)$, of the process in the vicinity of each point visited by X_t. The most general approach to estimating stochastic differential equations is to avoid any functional form specification for the drift and the diffusion term. In some cases, attention may focus on one of the functions and it is then of interest to estimate it in the context of the other function being treated as a nuisance parameter. A substantial simplification to the estimation problem is obtained by the commonly made assumption of stationarity. Indeed, under stationarity and provided suitable regularity conditions are met, the marginal density of the process is fully characterized by the two functions of interest [e.g. see Karatzas and Shreve (1991) and Karlin and Taylor (1981)]. This fact justifies some estimation methods that have appeared recently in the literature which exploit the restrictions imposed on the drift and diffusion function by virtue of the existence of a time-invariant density of the process [see, in particular, Aït-Sahalia (1996a,b) and Jiang and Knight (1997)]. Notwithstanding the advantages of assuming stationarity, it would appear that, for many of the empirical applications mentioned in the preceding paragraph at least, it would be more appropriate to allow for martingale and other possible forms of nonstationary behavior in the process. In such cases, it becomes necessary to achieve identification without resorting to cross restrictions delivered from the existence of a time-invariant density and transitional density, and estimation and inference must be performed when such restrictions cannot be imposed, namely when the process is nonstationary. Of course, there may also be interest in testing either local or more general martingale behavior in the process.

The aim of this chapter is to construct a nonparametric estimation method for diffusion models without imposing a stationarity assumption. Our approach is simply a refined sample analog method, which builds local estimates of the drift and diffusion components from the local behavior of the process at each spatial point that the process visits. We assume that the process is discretely sampled, but we explore the limit theory of the proposed estimators as the sample frequency increases [i.e. as the interval between observations tends to zero, as in Florens-Zmirou (1993), Jacod (1997) and Jiang and Knight (1997)] and also as the total time span of observation lengthens. In technical terms this amounts to both *infill* and *long span* asymptotics. We give conditions for almost sure convergence of the proposed sample analog estimators to the theoretical functions and provide a limit distribution theory for the general case. The asymptotic distributions of the estimates are mixed normal and the mixture variates can be expressed in terms of the *chronological local time* [see Phillips and Park (1998)] of the underlying process, a random quantity that measures in chronological time units the amount of time the process spends in the vicinity of each spatial point. Our results also enable us to comment on the fixed time span situation. We confirm earlier findings that the diffusion term can be consistently estimated over a fixed time span [as in Florens-Zmirou (1993) and Jacod (1997), for example]. We also confirm that, in general, the drift term can not be identified nonparametrically on a fixed interval without cross-restrictions, no matter how frequently the data is sampled [c.f. Merton (1973), Aït-Sahalia (1996a) and the discussion in Part II]. Despite this, by letting the time span increase to infinity, the theoretical drift term can be recovered in the limit, provided the process continues to repeat itself, that is provided the process is recurrent. Geman (1979) utilized the same property but assumed the availability of a continuous record of observations. To our knowledge, our drift estimator is the first fully nonparametric estimator which permits identification of the drift function by use of discretely sampled data, without relying on cross-restrictions based on the existence of a time-invariant marginal density. It is therefore robust against deviations from stationarity.

This chapter is presented as follows. Section 2 lays out the model and objects of interest. Section 3 gives some useful theoretical preliminaries. Section 4 contains a description of the methodology. Section 5 presents the main results and Section 6 concludes. Section 7 provides proofs and technicalities. Notation is laid out in Section 8.

2. The Model, Assumptions and Objects of Interest

The model we consider is the autonomous stochastic differential equation

$$dX_t = \mu(X_t)dt + \sigma(X_t)dB_t, \tag{2.1}$$

with initial condition $X_0 = \overline{X}$ and where B_t is a standard Brownian motion defined on the filtered probability space $(\Omega, \Im^B, (\Im_t^B)_{t\geq 0}, P)$. The initial condition $\overline{X} \in L^2$ and is taken to be independent of $\{B_t : t \geq 0\}$. We define the left-continuous filtration

$$\overline{\Im}_t := \sigma(\overline{X}) \vee \Im_t^B = \sigma(\overline{X}, B_s; 0 \leq s \leq t) \qquad 0 \leq t < \infty$$

and the collection of null sets

$$\aleph := \{N \subseteq \Omega; \exists G \in \overline{\Im}_\infty \textit{ with } N \subseteq G \textit{ and } P(G) = 0\}.$$

We create the *augmented* filtration

$$\tilde{\Im}_t^X := \sigma(\overline{\Im}_t \cup \aleph) \qquad 0 \leq t < \infty.$$

The following conditions will be used in the study of (2.1). They will assure the existence and pathwise uniqueness of a nonexplosive solution to (2.1) that is adapted to the *augmented* filtration $\{\tilde{\Im}_t^X\}$.

2.1 Assumption

(A) *$\mu(\cdot)$ and $\sigma(\cdot)$ are time-homogeneous, $\mathfrak{B}$-measurable functions on $\mathfrak{D} = (l, u)$ with $-\infty \leq l < u \leq \infty$ where $\mathfrak{B}$ is the σ-field generated by Borel sets on $\mathfrak{D}$. Both functions are at least once continuously differentiable. Hence, they satisfy local Lipschitz and growth conditions. Thus, for every compact subset $J = [1/H, H]$ with $H > 0$ of the range of the process, there exist constants C_1 and C_2 such that, for all x and y in J,*

$$|\mu(x) - \mu(y)| + |\sigma(x) - \sigma(y)| \leq C_1|x - y|,$$

and

$$|\mu(x)| + |\sigma(x)| \leq C_2\{1 + |x|\}.$$

(B) $\sigma^2(\cdot) > 0$ *on* $\mathfrak{D}$.

(C) [Feller's (1952) necessary and sufficient condition for nonexplosion]. *We define* $V(\alpha)$ *as*

$$\int_0^\alpha S'(y)\{\int_0^y \left[\frac{2}{S'(x)\sigma^2(x)}\right] dx\}dy$$

where $S'(x)$ *is the first derivative of the natural scale measure,*

$$S(\alpha) = \int_0^\alpha \exp\{\int_0^y \left[-\frac{2\mu(x)}{\sigma^2(x)}\right] dx\}dy.$$

We require $V(\alpha)$ *to diverge at the boundaries of* $\mathfrak{D}$, *i.e.*

$$\lim_{\alpha \to l^+} V(\alpha) = \lim_{\alpha \to u^-} V(\alpha) = \infty.$$

Assumption (A) is sufficient for pathwise uniqueness of the solution to (2.1) [c.f. Karatzas and Shreve (1991, Theorem 5.2.5, page 287)]. Assumptions (A) and (B) assure the existence of a unique strong solution up to an explosion time [c.f. Karatzas and Shreve (1991, Theorem 5.5.15, page 341 and Corollary 5.3.23, page 310)]. Assumption (C) guarantees that neither l nor u are attained in finite time [c.f. Karatzas and Shreve (1991, Theorem 5.5.29, page 348)]; and the same condition is necessary and sufficient for recurrence, meaning that, for each $c \in (l, u)$, there exist a sequence of times $\{t_i\}$ increasing to infinity such that $X_{t_i} = c$ for each i, almost surely.

2.2 Remark Global Lipschitz and growth conditions are typically assumed to guarantee existence and uniqueness of a strong solution to (2.1) [c.f. Karatzas and Shreve (1991, Theorem 5.2.9, page 289), for example]. We do not impose these conditions here because, as Aït-Sahalia (1996a,b) points out, they fail to be satisfied for interesting models in economics and finance.

2.3 Remark Geman (1979) requires the natural scale measure $S(\alpha)$ to diverge to ∞ as $\alpha \to u$, and to $-\infty$ as $\alpha \to l$. Notice that this condition is only sufficient for nonexplosion and recurrence. Feller's (1952) condition based on the function $V(\alpha)$ is necessary and sufficient. The following implications are easily derived [c.f. Karatzas and Shreve (1991, Problem 5.5.27, page 348)]:

$$S(l^+) = -\infty \Rightarrow V(l^+) = \infty$$

and

$$S(u^-) = \infty \Rightarrow V(u^-) = \infty.$$

Thus, under conditions (A), (B) and (C), the stochastic differential equation has a strong solution X_t that is unique, recurrent and continuous in $t \in [0, T]$. X_t satisfies

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

a.s., with $\int_0^T E[X_t^2]dt < \infty$.

The objects of econometric interest are the drift and diffusion terms in (2.1). These functions have the following conditional moment definitions:

$$\begin{aligned}
\mu(x) &= \lim_{h\to 0} E\{\frac{X_{t+h} - X_t}{h}|X_t = x\} = \\
&= \lim_{h\to 0} \frac{1}{h}\int_{[|X_{t+h}-X_t|<\epsilon]} (X_{t+h} - X_t)dP(X_{t+h}|X_t = x), \qquad (2.2) \\
\sigma^2(x) &= \lim_{h\to 0} E\left\{\frac{[X_{t+h} - X_t]^2}{h}|X_t = x\right\} = \\
&= \lim_{h\to 0} \frac{1}{h}\int_{[|X_{t+h}-X_t|<\epsilon]} (X_{t+h} - X_t)^2 dP(X_{t+h}|X_t = x). \qquad (2.3)
\end{aligned}$$

Also,

$$\lim_{h\to 0} h^{-1}P(|X_{t+h} - X_t| \geq \epsilon|X_t = x) = 0.$$

Loosely speaking, (2.2) and (2.3) can be interpreted as representing the "instantaneous" conditional mean and the "instantaneous" conditional variance of the process when $X_t = x$. More precisely, (2.2) describes the conditional expected rate of change of the process for infinitesimal time changes, whereas (2.3) gives the conditional rate of change of volatility at x.

3. Local Time Preliminaries

In what follows we introduce some preliminary theory regarding the local or *sojourn* time of a semimartingale (SMG). All of what is needed below is contained in standard treatments like those of Protter (1990) and Revuz and Yor (1991). Continuous-time stochastic differential equations like (2.1) have solutions that are semimartingales and hence the theory comes within the ambit of SMG analysis.

The local time of a continuous SMG M is defined as follows:

3.1 Definition (The Tanaka Formula) *For any real number a, there exists a non-decreasing continuous process $L_M(.,a)$ called the local time of M at a, such that*

$$\begin{aligned} |M_t - a| &= |M_0 - a| + \int_0^t sgn(M_s - a)dM_s + L_M(t,a), \\ (M_t - a)^+ &= (M_0 - a)^+ + \int_0^t \mathbf{1}_{\{M_s > a\}} dM_s + \frac{1}{2} L_M(t,a), \\ (M_t - a)^- &= (M_0 - a)^- - \int_0^t \mathbf{1}_{\{M_s \leq a\}} dM_s + \frac{1}{2} L_M(t,a). \end{aligned}$$

In particular, $|M_t - a|$, $(M_t - a)^+$ and $(M_t - a)^-$ are semimartingales.

3.2 Lemma (Continuity of Martingale local time) *For any continuous SMG M, there exists a version of the local time such that $(t,a) \mapsto L_M(t,a)$ is a.s. continuous in both t and a. Moreover, it can be chosen so that $a \mapsto L_M(t,a)$ is Hölder continuous of order k for every $k < 1/2$ uniformly in t on every compact interval.*

3.3 Lemma (The occupation time formula) *Let M be a continuous SMG with quadratic variation process $[M]_s$ and let L^a be the local time at a. Then,*

$$\int_0^t f(M_s, s)d[M]_s = \int_{-\infty}^{+\infty} da \int_0^t f(a,s)dL_M(s,a)$$

for every Borel function f. If f is homogeneous, then the expression simplifies to

$$\int_0^t f(M_s)d[M]_s = \int_{-\infty}^{+\infty} f(a)L_M(t,a)da. \tag{3.1}$$

3.4 Lemma *If M is a continuous SMG then, almost surely*

$$L_M(t,a) = \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \int_0^t \mathbf{1}_{[a,\, a+\varepsilon[}(M_s)d[M]_s \ \forall a, t. \tag{3.2}$$

If M is a continuous local martingale then, almost surely

$$L_M(t,a) = \lim_{\varepsilon\to 0}\frac{1}{2\varepsilon}\int_0^t 1_{]a-\varepsilon,\ a+\varepsilon[}(M_s)d[M]_s\ \forall a,t. \tag{3.3}$$

The process $L_M(t,a)$ is called the *local time* of M at the point a over the time interval $[0,t]$. It is measured in units of the quadratic variation process and gives the amount of time that the process spends in the vicinity of a. The *chronological local time* [from Phillips and Park (1998)] is a standardized version of the conventional local time that is defined in terms of pure time units. It can be easily derived in the Brownian motion case. From (3.3), the local time of a standard Brownian motion W is

$$L_W(t,a) = \lim_{\varepsilon\to 0}\frac{1}{2\varepsilon}\int_0^t 1_{(|W_s-a|<\varepsilon)}ds \text{ a.s. } \forall a,t.$$

Now, consider the Brownian motion $B = \sigma W$ with variance σ^2. We can write, as in Phillips and Park (1998),

$$L_B(t,a) = \lim_{\varepsilon\to 0}\frac{1}{2\varepsilon}\int_0^t 1_{(|B_s-a|<\varepsilon)}\sigma^2 ds = \sigma L_W(t,\frac{a}{\sigma}) \text{ a.s. } \forall a,t.$$

Since the quadratic variation of Brownian motion is deterministic, the chronological local time can be obtained as a scaled version of the conventional sojourn time as

$$\overline{L}_B(t,a) = \lim_{\varepsilon\to 0}\frac{1}{2\varepsilon}\int_0^t 1_{(|B_s-a|<\varepsilon)}ds = \sigma^{-2}L_B(t,a) \text{ a.s. } \forall a,t. \tag{3.4}$$

Equation (3.4) clarifies the sense in which $\overline{L}_B(t,a)$ measures the amount of time (out of t) that the process spends in the neighborhood of a generic spatial point a.

It turns out that a similar expression can be defined for more general processes such as those driven by stochastic differential equations like (2.1). In this case, the measure $d[X]_s$ is random and equal to $\sigma^2(X_s)ds$. Hence, given the limit operation, a natural way to define the *chronological local time* of a process like (2.1) is by

$$\overline{L}_X(t,a) = \frac{1}{\sigma^2(a)}\lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\int_0^t 1_{[a,\ a+\varepsilon[}(X_s)\sigma^2(X_s)ds = \frac{1}{\sigma^2(a)}L_X(t,a) \text{ a.s. } \forall a,t. \tag{3.5}$$

This is the notion of local time that we will use extensively in what follows.

The following result generalizes to diffusion processes the limit theory for Brownian local time [see Yor (1983), Revuz and Yor (1994) and Phillips and Park (1998)]. This result will be useful in the development of our limit theory.

3.5 Lemma (Limit theory for the local time of a diffusion) *Let X satisfy the properties in Section 2. Let r and a be fixed real numbers and treat $\{L_X(t, r+\frac{a}{\lambda}) - L_X(t,r)\}$ as a double indexed stochastic process in (t, a). Then, as $\lambda \to \infty$*

$$\frac{1}{2}\sqrt{\lambda}\{L_X(t, r+\frac{a}{\lambda}) - L_X(t,r)\} \xrightarrow{d} \mathfrak{B}(L_X(t,r), a)$$

where $\mathfrak{B}(t,a)$ is a standard Brownian sheet independent of X.

4. Econometric Estimation

Assume we observe the process X_t at $\{t = t_1, t_2, .., t_n\}$ in the time interval $[0, T]$, with $T \geq T_0 > 0$, where T_0 is a positive constant. Further assume that the observations are equispaced. Then, $\{X_t = X_{\Delta_{n,T}}, X_{2\Delta_{n,T}}, X_{3\Delta_{n,T}}, ..., X_{n\Delta_{n,T}}\}$ are n observations on the process X_t at $\{t_1 = \Delta_{n,T}, t_2 = 2\Delta_{n,T}, t_3 = 3\Delta_{n,T}, ..., t_n = n\Delta_{n,T}\}$ where $\Delta_{n,T} = T/n$.

We want the number of sampled points (n) to increase as the time span (T) lengthens. We also want the frequency of observation to increase with n. Thus, we will explore the limit theory of the proposed estimators as $n \to \infty$, $T \to \infty$ and $\Delta_{n,T} = T/n \to 0$. We will also comment on the fixed T case where $T = \overline{T}$.

We propose the following estimators for (2.2) and (2.3).

$$\begin{aligned}\hat{\sigma}^2_{(n,T)}(x) &= \frac{\sum_{i=1}^n \mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})(\frac{1}{m_{n,T}(i\Delta_{n,T})\Delta_{n,T}}\sum_{j=0}^{m_{n,T}(i\Delta_{n,T})-1}[X_{t(i\Delta_{n,T})_j+\Delta_{n,T}} - X_{t(i\Delta_{n,T})_j}]^2)}{\sum_{i=1}^n \mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})} \\ &:= \frac{\sum_{i=1}^n \mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})\tilde{\sigma}^2_{n,T}(X_{i\Delta_{n,T}})}{\sum_{i=1}^n \mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})}. \qquad (4.1)\end{aligned}$$

$$\begin{aligned}\hat{\mu}_{(n,T)}(x) &= \frac{\sum_{i=1}^n \mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})(\frac{1}{m_{n,T}(i\Delta_{n,T})\Delta_{n,T}}\sum_{j=0}^{m_{n,T}(i\Delta_{n,T})-1}[X_{t(i\Delta_{n,T})_j+\Delta_{n,T}} - X_{t(i\Delta_{n,T})_j}])}{\sum_{i=1}^n \mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})} \\ &:= \frac{\sum_{i=1}^n \mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})\tilde{\mu}_{n,T}(X_{i\Delta_{n,T}})}{\sum_{i=1}^n \mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})}. \qquad (4.2)\end{aligned}$$

In the above formulae, $\{t(i\Delta_{n,T})_j\}$ is a sequence of random times defined in the following manner:

$$t(i\Delta_{n,T})_0 = \inf\{t \geq 0 : |X_t - X_{i\Delta_{n,T}}| \leq \varepsilon_{n,T}\},$$

and

$$t(i\Delta_{n,T})_{j+1} = \inf\{t \geq t(i\Delta_{n,T})_j + \Delta_{n,T} : |X_t - X_{i\Delta_{n,T}}| \leq \varepsilon_{n,T}\}.$$

The number $m_{n,T}(i\Delta_{n,T}) \leq n$ counts the stopping times associated with the value $X_{i\Delta_{n,T}}$ and is defined as

$$m_{n,T}(i\Delta_{n,T}) = \sum_{j=1}^{n} \mathbf{1}_{\left[\left|X_{j\Delta_{n,T}} - X_{i\Delta_{n,T}}\right| \leq \varepsilon_{n,T}\right]},$$

where $\mathbf{1}_A$ denotes the indicator of A. The quantity $\varepsilon_{n,T}$ is a bandwidth-like parameter that is taken to depend on the time span and on the sample size. We call this parameter the spatial bandwidth. As usual, the random time $t(i\Delta_{n,T})$ is defined on Ω and takes values on $[0,\infty)$. Further, $\{t(i\Delta_{n,T}) < t^*\} \in \Im^X_{t+^*}$ where $\Im^X_{t+^*}$ is a right-continuous filtration defined as $\bigcap_{u>t^*} \Im^X_u$.

The kernel function $\mathbf{K}(\cdot)$ that appears in (4.1) and (4.2) is assumed to satisfy the following condition.

4.1 Assumption *The kernel* $\mathbf{K}(.)$ *is a continuous differentiable, symmetric and nonnegative function whose derivative* $\mathbf{K}'$ *is absolutely integrable and for which*

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

and

$$\int_{-\infty}^{\infty} s^q \mathbf{K}(s)ds < \infty,$$

for some $q \geq 1$.

4.2 Heuristics of the Estimation Procedure The method hinges on the simultaneous operation of *infill* and *long span* asymptotics. The intuition underlying the construction of (4.1) and (4.2) is fairly clear. By using observations over a lengthening time span as well as of increasing frequency we aim to "reconstruct" as well as possible the path of the process in terms of the key objects of interest, the drift and diffusion functions, which vary over the path. The idea is twofold.

First, the use of local averaging and stopping times in the algorithm is designed to replicate as well as possible the instantaneous features of the actual functions. Notice, in fact, that the components $\tilde{\sigma}^2_{n,T}(X_{i\Delta_{n,T}})$ and $\tilde{\mu}_{n,T}(X_{i\Delta_{n,T}})$ in (4.1) and (4.2) are defined as empirical analogs to the true functions for all i. Further, the estimates $\tilde{\sigma}^2_{n,T}(X_{i\Delta_{n,T}})$ and $\tilde{\mu}_{n,T}(X_{i\Delta_{n,T}})$ are consistent for $\sigma^2(X_{i\Delta_{n,T}})$ and $\mu(X_{i\Delta_{n,T}})$ as the random quantity

$m_{n,T}(i\Delta_{n,T})$ goes to infinity $\forall i$. Under suitable conditions on the bandwidths, $m_{n,T}(i\Delta_{n,T})$ diverges to infinity almost surely when $T \to \infty$. In particular, given appropriate choices of the smoothing sequences, divergence occurs when the process X_t is recurrent, as it is under Assumption 2.1. In this case, the process almost surely hits any point in its range an infinite number of times, i.e. $P_x\{X_t \text{ hits } z \text{ at a sequence of times increasing to } \infty\} = 1$, $\forall x, z$ (here x represents possible initializations of the process X_t).

Second, we apply standard nonparametric smoothing to recover the two functions of interest from the crude estimates $\tilde{\sigma}^2_{n,T}(X_{i\Delta_{n,T}})$ and $\tilde{\mu}_{n,T}(X_{i\Delta_{n,T}})$ calculated at the sample points.

5. Main Results

5A. Some Preliminary Theory

We start with the following preliminary result. Throughout, we will assume that Assumptions 2.1 and 4.1 hold.

5.1 Theorem (Almost sure convergence to the chronological local time) *Given $n \to \infty$, T fixed ($= \overline{T}$) and $h_{n,\overline{T}} \to 0$ (as $n \to \infty$) in such a way that $\frac{1}{h_{n,\overline{T}}}(\Delta_{n,\overline{T}})^{\alpha} = O(1)$ for some $\alpha \in (0, \frac{1}{2})$, the estimator $\frac{\Delta_{n,\overline{T}}}{h_{n,\overline{T}}} \sum_{i=1}^{n} \mathbf{K}(\frac{X_{i\Delta_{n,\overline{T}}} - x}{h_{n,\overline{T}}})$ converges to $\overline{L}_X(\overline{T}, x)$ a.s.*

5.2 Remark Theorem (5.1) is general enough to be applicable to *transient* processes. The following corollary illustrates the difference between the two cases when we let T go to infinity.

5.3 Corollary *If $T \to \infty$ with n but $\frac{T}{n} = \Delta_{n,T} \to 0$ and $h_{n,T} \to 0$ (as $n \to \infty$) in such a way that $\frac{\overline{L}_X(T,x)}{h_{n,T}}(\Delta_{n,T})^{\alpha} = O_{a.s.}(1)$ for some $\alpha \in (0, \frac{1}{2})$, and $\frac{\Delta_{n,T}}{h_{n,T}} \to 0$, then*

$$\frac{\Delta_{n,T}}{h_{n,T}} \sum_{i=1}^{n} \mathbf{K}(\frac{X_{i\Delta_{n,T}} - x}{h_{n,T}}) \overset{a.s.}{\to} \overline{L}_X(\sup\{t : X_t = x\}, x).$$

Further, if the process is recurrent, then $\overline{L}_X(\sup\{t : X_t = x\}, x) = \infty$ a.s.

5.4 Remark In applications it is often conventional to normalize T to 1. This implies that the admissible bandwidth $h_{n,T}$ is proportional to n^{-k} with $k \in (0, \frac{1}{2})$.

5B. Function Estimation of the Diffusion

We next develop the asymptotic theory for the diffusion estimator (4.1).

5.5 Theorem (Almost sure convergence of the diffusion estimator) *Given* $n \to \infty$, $T \to \infty$ *and* $h_{n,T} \to 0$ *(as* $n, T \to \infty$*) such that* $\frac{\Delta_{n,T}}{h_{n,T}} \to 0$ *and* $\frac{\overline{L}_X(T,x)}{h_{n,T}}(\Delta_{n,T})^{\alpha} = O_{a.s.}(1)$ *for some* $\alpha \in (0, \frac{1}{2})$, *and provided* $\varepsilon_{n,T} \to 0$ *(as* $n, T \to \infty$*) such that* $\frac{\Delta_{n,T}}{\varepsilon_{n,T}} \to 0$ *and* $\frac{\overline{L}_X(T,x)}{\varepsilon_{n,T}}(\Delta_{n,T})^{\beta} = O_{a.s.}(1)$ *for some* $\beta \in (0, \frac{1}{2})$, *the estimator*

$$\frac{\sum_{i=1}^{n} \mathbf{K}\left(\frac{X_{i\Delta_{n,T}} - x}{h_{n,T}}\right)\tilde{\sigma}^2_{n,T}(X_{i\Delta_{n,T}})}{\sum_{i=1}^{n} \mathbf{K}\left(\frac{X_{i\Delta_{n,T}} - x}{h_{n,T}}\right)} \overset{a.s.}{\to} \sigma^2(x),$$

where

$$\tilde{\sigma}^2_{n,T}(X_{i\Delta_{n,T}}) = \frac{1}{m_{n,T}(i\Delta_{n,T})\Delta_{n,T}} \sum_{j=0}^{m_{n,T}(i\Delta_{n,T})-1} [X_{t(i\Delta_{n,T})_j + \Delta_{n,T}} - X_{t(i\Delta_{n,T})_j}]^2,$$

and where the sequence of stopping times $\{t(i\Delta_{n,T})_j\}$ $j = 1, 2, ..$ *satisfies*

$$t(i\Delta_{n,T})_0 = \inf\{t \geq 0 : |X_t - X_{i\Delta_{n,T}}| \leq \varepsilon_{n,T}\},$$

and

$$t(i\Delta_{n,T})_{j+1} = \inf\{t \geq t(i\Delta_{n,T})_j + \Delta_{n,T} : |X_t - X_{i\Delta_{n,T}}| \leq \varepsilon_{n,T}\},$$

for all i.

5.6 Theorem (limiting distribution of the diffusion estimator) *Assume* $n \to \infty$, $T \to \infty$, $h_{n,T} \to 0$ *(as* $n, T \to \infty$*) such that* $\frac{\Delta_{n,T}}{h_{n,T}} \to 0$ *and* $\frac{\overline{L}_X(T,x)}{h_{n,T}}(\Delta_{n,T})^{\alpha} = O_{a.s.}(1)$ *for some* $\alpha \in (0, \frac{1}{2})$. *Also, assume* $\varepsilon_{n,T} \to 0$ *(as* $n, T \to \infty$*) such that* $\frac{\Delta_{n,T}}{\varepsilon_{n,T}} \to 0$, *and* $\frac{\overline{L}_X(T,x)}{\varepsilon_{n,T}}(\Delta_{n,T})^{\beta} = O_{a.s.}(1)$ *for some* $\beta \in (0, \frac{1}{2})$.

If $h_{n,T} = o(\varepsilon_{n,T})$, $\varepsilon_{n,T}\overline{L}_X(T,x) \overset{a.s.}{\to} 0$ *and* $\frac{\varepsilon^2_{n,T}}{\sqrt{\Delta_{n,T}}} \to 0$, *then the asymptotic distribution of the diffusion function estimator is driven by a 'martingale' effect and has the form*

$$\sqrt{\frac{\varepsilon_{n,T}\overline{L}_X(T,x)}{\Delta_{n,T}}}\{\hat{\sigma}^2_{(n,T)}(x) - \sigma^2(x)\} \overset{d}{\to} N\left(0, 2\sigma^4(x)\right). \tag{5.1}$$

If $h_{n,T} = o(\varepsilon_{n,T})$, $\frac{\sqrt{\overline{L}_X(T,x)\Delta_{n,T}}}{\varepsilon^{3/2}_{n,T}} \overset{a.s.}{\to} 0$ *and* $\frac{\varepsilon^2_{n,T}}{\sqrt{\Delta_{n,T}}} \to \infty$, *then the asymptotic distribution of the diffusion function estimator is driven by a 'bias' effect and has the form*

$$\frac{\sqrt{\overline{L}_X(T,x)}}{\varepsilon_{n,T}^{3/2}}\{\widehat{\sigma}^2_{(n,T)}(x)-\sigma^2(x)\}\xrightarrow{d} N\left(0,4\varphi^*\left(\sigma'(x)\right)^2\right), \tag{5.2}$$

where $\varphi^* = -2\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}|a-b|ab\left(\frac{1}{2}\mathbf{1}_{\{|a|\leq 1\}}\right)\left(\frac{1}{2}\mathbf{1}_{\{|b|\leq 1\}}\right)dadb = 0.2666.$

If $h_{n,T}=O(\varepsilon_{n,T})$ *with* $h_{n,T}/\varepsilon_{n,T}\to\phi>0$, $\varepsilon_{n,T}\overline{L}_X(T,x)\overset{a.s.}{\to}0$ *and* $\frac{\varepsilon_{n,T}^2}{\sqrt{\Delta_{n,T}}}\to 0$, *then the asymptotic distribution of the diffusion function estimator is driven by a 'martingale' effect and is of the form*

$$\sqrt{\frac{\varepsilon_{n,T}\overline{L}_X(T,x)}{\Delta_{n,T}}}\{\widehat{\sigma}^2_{(n,T)}(x)-\sigma^2(x)\}\xrightarrow{d} N\left(0,2\theta_\phi\sigma^4(x)\right). \tag{5.3}$$

where $\theta_\phi=\frac{1}{2}\int_{-\infty}^{\infty}\int_{(z-1)/\phi}^{(z+1)/\phi}\int_{(z-1)/\phi}^{(z+1)/\phi}\mathbf{K}(a)\mathbf{K}(e)dzdade.$

If $h_{n,T}=O(\varepsilon_{n,T})$ *with* $h_{n,T}/\varepsilon_{n,T}\to\phi>0$, $\frac{\sqrt{\overline{L}_X(T,x)\Delta_{n,T}}}{\varepsilon_{n,T}^{3/2}}\overset{a.s.}{\to}0$ *and* $\frac{\varepsilon_{n,T}^2}{\sqrt{\Delta_{n,T}}}\to\infty$, *then the asymptotic distribution of the diffusion function estimator is driven by a 'bias' effect and is of the form*

$$\frac{\sqrt{\overline{L}_X(T,x)}}{\varepsilon_{n,T}^{3/2}}\{\widehat{\sigma}^2_{(n,T)}(x)-\sigma^2(x)\}\xrightarrow{d} N\left(0,4\left(\vartheta_\phi+\phi^3\varphi+\eta_\phi\right)\left(\sigma'(x)\right)^2\right), \tag{5.4}$$

where

$$\vartheta_\phi=\frac{1}{8}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{\infty}^{\infty}\mathbf{K}(s)\mathbf{K}(c)(a-\phi s)\mathbf{1}_{\{|a-\phi s|\leq 1\}}(b-\phi s)\mathbf{1}_{\{|b-\phi s|\leq 1\}}(a-\phi c)\times$$
$$\times\mathbf{1}_{\{|a-\phi c|\leq 1\}}(b-\phi c)\mathbf{1}_{\{|b-\phi c|\leq 1\}}\mathbf{1}_{(0\leq v\leq a)}\mathbf{1}_{(0\leq v\leq b)}dadbdcdsdv,$$

$$\varphi=-\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}2|a-b|ab\mathbf{K}(a)\mathbf{K}(b)dadb,$$

and

$$\eta_\phi=4\phi^2\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}u\left(a-\phi c\right)\mathbf{K}(u)\mathbf{K}(c)\mathbf{1}_{\{|a-\phi c|\leq 1\}}\mathbf{1}_{(0\leq b\leq u)}\mathbf{1}_{(0\leq b\phi\leq a)}dbdudadc.$$

5.7 Remarks The statement of Theorem 5.6 uses the terms 'bias' effect and 'martingale' effect to refer to the principal terms that govern the asymptotic distribution. These effects are revealed in the proof of the theorem. The essential factor governing the magnitude of the two effects is the relation of the observation rate, $\Delta_{n,T}$, of the process to the spatial bandwidth parameter, $\varepsilon_{n,T}$. If $\Delta_{n,T}$ is small relative to $\varepsilon_{n,T}$, so that $\varepsilon_{n,T}^2/\sqrt{\Delta_{n,T}}\to\infty$, then the bias effect dominates the asymptotics. In contrast to conventional nonparametric regression situations (Härdle, 1990), the bias effect turns out to be random, as it is in the nonstationary autoregressive case studied in Phillips and Park (1998). If the spatial bandwidth $\varepsilon_{n,T}$ is small relative to the observation interval and $\varepsilon^2_{n,T}/\sqrt{\Delta_{n,T}} \to 0$, the bias effects are eliminated asymptotically and the martingale effect governs the limit theory.

Theorems 5.5 and 5.6 state the a.s. consistency and convergence in distribution of the diffusion function estimator as we enlarge the time span and increase the frequency of observations. It is easy to see that if we fix T ($T = \overline{T}$ with $\Delta_{n,\overline{T}} \to 0$), the previous results do not change: that is, the diffusion function estimator is still consistent and distributed in the limit as a mixed Gaussian distribution with mixing variate depending on the chronological local time of the underlying diffusion process. Assume $h_{n,T} = o(\varepsilon_{n,T})$. When $\overline{L}_X(\overline{T}, x) = O_p(1)$, that is when $T = \overline{T}$, the asymptotic distribution can be written as

$$\sqrt{n\varepsilon_{n,\overline{T}}}\{\widehat{\sigma}^2_{(n,\overline{T})}(x) - \sigma^2(x)\} \xrightarrow{d} MN\left(0, 2\frac{\sigma^4(x)}{(\overline{L}_X(\overline{T},x)/\overline{T})}\right), \tag{5.5}$$

under the following conditions on $h_{n,\overline{T}}$ and $\varepsilon_{n,\overline{T}}$:

$$\varepsilon_{n,\overline{T}} \propto n^{-k_1} \text{ with } k_1 \in \left(\frac{1}{4}, \frac{1}{2}\right)$$

and

$$h_{n,\overline{T}} \propto n^{-k_2} \text{ with } k_2 \in \left(0, \frac{1}{2}\right).$$

On the other hand, if

$$\varepsilon_{n,\overline{T}} \propto n^{-k_1} \text{ with } k_1 \in \left(0, \frac{1}{4}\right)$$

and

$$h_{n,\overline{T}} \propto n^{-k_2} \text{ with } k_2 \in \left(0, \frac{1}{2}\right)$$

then the 'bias' term dominates and

$$\frac{1}{\varepsilon^{3/2}_{n,\overline{T}}}\{\widehat{\sigma}^2_{(n,\overline{T})}(x) - \sigma^2(x)\} \xrightarrow{d} MN\left(0, 4\varphi^* \frac{\left(\sigma'(x)\right)^2}{(\overline{L}_X(\overline{T},x))}\right) \tag{5.6}$$

with $\varphi^* = -2\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |a-b| ab \left(\frac{1}{2}\mathbf{1}_{\{|a|\le 1\}}\right)\left(\frac{1}{2}\mathbf{1}_{\{|b|\le 1\}}\right) da db = 0.2666$.

5.8 Relation to Florens-Zmirou (1993) There is an important similarity between (5.5) and the limiting distribution obtained in Florens-Zmirou (1993). It is useful to recall her results before commenting further.

5.9 Theorem (Florens-Zmirou (1993)) *Assume we observe X_t at $\{t = t_1, t_2, ..., t_n\}$ in the time interval $[0, \overline{T}]$ where $\overline{T}$ can be normalized to 1. Also, the data is equispaced. Consequently, $\{X_t = X_{\Delta_n}, X_{2\Delta_n}, X_{3\Delta_n}, ..., X_{n\Delta_n}\}$ are n observations at points $\{t_1 = \Delta_n, t_2 = 2\Delta_n, ..., t_n = \Delta_n\}$, where $\Delta_n = 1/n$. The estimator*

$$\widehat{\sigma}^2_{(n)}(x) = \frac{1}{\Delta_n} \frac{\sum_{i=1}^{n-1} \mathbf{1}_{\{|X_{i/n} - x| \leq h_n\}} [X_{(i+1)/n} - X_{i/n}]^2}{\sum_{i=1}^{n-1} \mathbf{1}_{\{|X_{i/n} - x| \leq h_n\}}} \xrightarrow{L_2} \sigma^2(x)$$

provided the sequence h_n is such that $nh_n \to \infty$ and $nh_n^4 \to 0$. Further, if $nh_n^3 \to 0$, then

$$\sqrt{nh_n}\{\widehat{\sigma}^2_{(n)}(x) - \sigma^2(x)\} \xrightarrow{d} MN\left(0, 2\frac{\sigma^4(x)}{(L_X(1, x))}\right)$$

where $\overline{L}_X(1, x)$ is the local time of the process.

What Florens-Zmirou calls local time $(L_X(., x))$ is what we refer to here as the *chronological local time* $(\overline{L}_X(., x))$ of the process.

Provided $nh_n^4 \to 0$, the bias term disappears asymptotically and the limiting distribution is the normal distribution to which the 'martingale' term converges. It is not surprising that the limiting distribution in Florens-Zmirou (1993) resembles the limiting distribution of the estimator proposed here for choices of $\varepsilon_{n,\overline{T}}$ and $h_{n,\overline{T}}$ that make the bias term negligible [and provided $h_{n,\overline{T}} = o(\varepsilon_{n,\overline{T}})$]. Note, in fact, that in the fixed T case the estimator that we suggest here can be interpreted as a convoluted version of Florens-Zmirou's estimator. In particular, it can be written as a weighted average of estimates obtained using Florens-Zmirou's method. In effect, $\tilde{\sigma}^2_{n,\overline{T}}(X_{i\Delta_{n,\overline{T}}})$ can be rearranged as follows $\forall i$,

$$\begin{aligned} \tilde{\sigma}^2_{n,\overline{T}}(X_{i\Delta_{n,\overline{T}}}) &= \frac{1}{m_{n,\overline{T}}(i\Delta_{n,\overline{T}})\Delta_{n,\overline{T}}} \sum_{j=0}^{m_{n,\overline{T}}(i\Delta_{n,\overline{T}})-1} [X_{t(i\Delta_{n,\overline{T}})_j + \Delta_{n,\overline{T}}} - X_{t(i\Delta_{n,\overline{T}})_j}]^2 \\ &= \frac{1}{\Delta_{n,\overline{T}}} \frac{\sum_{j=1}^{n-1} \mathbf{1}_{\{|X_{j\Delta_{n,\overline{T}}} - X_{i\Delta_{n,\overline{T}}}| \leq \varepsilon_{n,\overline{T}}\}} [X_{(j+1)\Delta_{n,\overline{T}}} - X_{j\Delta_{n,\overline{T}}}]^2}{\sum_{j=1}^{n-1} \mathbf{1}_{\{|X_{j\Delta_{n,\overline{T}}} - X_{i\Delta_{n,\overline{T}}}| \leq \varepsilon_{n,\overline{T}}\}}}. \end{aligned}$$

It is easy to prove that when $nh_n^4 \to \infty$ Florens-Zmirou's estimator is still consistent but, in the same manner as our own limit theory, the 'bias' term drives the asymptotic distribution, namely

$$\frac{1}{h_n^{3/2}}\{\widehat{\sigma}^2_{(n)}(x) - \sigma^2(x)\} \xrightarrow{d} MN\left(0, 4\varphi^* \frac{\left(\sigma'(x)\right)^2}{(L_X(1, x))}\right)$$

where $\varphi^* = -2\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |a - b| ab \left(\frac{1}{2}\mathbf{1}_{\{|a| \leq 1\}}\right)\left(\frac{1}{2}\mathbf{1}_{\{|b| \leq 1\}}\right) da db = 0.2666.$

5.10 Remark When we let $T \to \infty$, the normalizations in (5.1), (5.2), (5.3) and (5.4) are random because of the presence of the local time factor $\overline{L}_X(T,x)^{\frac{1}{2}}$. In general, therefore, the rate of convergence will be path dependent and will depend, in particular, on the sample trajectory of the conditional variance function. The precise rate of convergence in (5.1), (5.2), (5.3) and (5.4) depends on the asymptotic divergence characteristics of the chronological local time $\overline{L}_X(T,x)$ of the process $\{X_t; t \geq 0\}$. When X_t is a standard Brownian motion (i.e., $\mu(X) = 0$, and $\sigma(X) = 1$), then $\overline{L}_X(T,x) = \overline{L}_W(T,x) = T^{\frac{1}{2}} L_W(1, x/T^{\frac{1}{2}}) = O_{a.s.}(T^{\frac{1}{2}})$. In this case, which is explored further below, the convergence rates of $\widehat{\sigma}^2_{(n,T)}(x)$ are $\sqrt{\frac{\varepsilon_{n,T} T^{\frac{1}{2}}}{\Delta_{n,T}}}$ and $\sqrt{\frac{T^{\frac{1}{2}}}{\varepsilon^3_{n,T}}}$ in (5.1)-(5.3) and (5.2)-(5.4) respectively, and are not path dependent.

5C. Function Estimation of the Drift

We now turn to the analysis of the drift function.

5.11 Theorem (Almost sure convergence to the drift term) *Given $n \to \infty$, $T \to \infty$ and $h_{n,T} \to 0$ (as $n, T \to \infty$) such that $\frac{\Delta_{n,T}}{h_{n,T}} \to 0$ and $\frac{\overline{L}_X(T,x)}{h_{n,T}}(\Delta_{n,T})^{\alpha} = O_{a.s.}(1)$ for some $\alpha \in (0, \frac{1}{2})$, and provided $\varepsilon_{n,T} \to 0$ (as $n, T \to \infty$) such that $\frac{\Delta_{n,T}}{\varepsilon_{n,T}} \to 0$, $\frac{\overline{L}_X(T,x)}{\varepsilon_{n,T}}(\Delta_{n,T})^{\beta} = O_{a.s.}(1)$ for some $\beta \in (0, \frac{1}{2})$ and $\varepsilon_{n,T}\overline{L}_X(T,x) \to \infty$,* the estimator

$$\frac{\sum_{i=1}^{n} \mathbf{K}\left(\frac{X_{i\Delta_{n,T}} - x}{h_{n,T}}\right) \widetilde{\mu}_{n,T}(X_{i\Delta_{n,T}})}{\sum_{i=1}^{n} \mathbf{K}\left(\frac{X_{i\Delta_{n,T}} - x}{h_{n,T}}\right)} \stackrel{a.s.}{\to} \mu(x),$$

with

$$\widetilde{\mu}_{n,T}(X_{i\Delta_{n,T}}) = \frac{1}{m_{n,T}(i\Delta_{n,T})\Delta_{n,T}} \sum_{j=0}^{m_{n,T}(i\Delta_{n,T})-1} [X_{t(i\Delta_{n,T})_j + \Delta_{n,T}} - X_{t(i\Delta_{n,T})_j}],$$

where the sequence of stopping times $\{t(i\Delta_{n,T})_j\}$ $j = 1, 2, ..$ satisfies

$$t(i\Delta_{n,T})_0 = \inf\{t \geq 0 : |X_t - X_{i\Delta_{n,T}}| \leq \varepsilon_{n,T}\},$$

and

$$t(i\Delta_{n,T})_{j+1} = \inf\{t \geq t(i\Delta_{n,T})_j + \Delta_{n,T} : |X_t - X_{i\Delta_{n,T}}| \leq \varepsilon_{n,T}\},$$

for all i.

5.12 Theorem *Given $n \to \infty$, $T \to \infty$, $h_{n,T} \to 0$ (as $n, T \to \infty$) such that $\frac{\Delta_{n,T}}{h_{n,T}} \to 0$ and $\frac{\overline{L}_X(T,x)}{h_{n,T}}(\Delta_{n,T})^{\alpha} = O_{a.s.}(1)$ for some $\alpha \in (0, \frac{1}{2})$, and provided $\varepsilon_{n,T} \to 0$ (as $n, T \to \infty$) such that $\frac{\Delta_{n,T}}{\varepsilon_{n,T}} \to 0$, $\frac{\overline{L}_X(T,x)}{\varepsilon_{n,T}}(\Delta_{n,T})^{\beta} = O_{a.s.}(1)$ for some $\beta \in (0, \frac{1}{2})$ and $\varepsilon_{n,T}\overline{L}_X(T,x) \overset{a.s.}{\to} \infty$, the asymptotic distribution of the drift function estimator is of the form*

$$\sqrt{\varepsilon_{n,T}\overline{L}_X(T,x)}\{\widehat{\mu}_{(n,T)}(x) - \mu(x)\} \overset{d}{\to} N\left(0, \frac{1}{2}\sigma^2(x)\right) \tag{5.7}$$

if $h_{n,T} = o(\varepsilon_{n,T})$. If $h_{n,T} = O(\varepsilon_{n,T})$ with $h_{n,T}/\varepsilon_{n,T} \to \phi > 0$, then

$$\sqrt{\varepsilon_{n,T}\overline{L}_X(T,x)}\{\widehat{\mu}_{(n,T)}(x) - \mu(x)\} \overset{d}{\to} N\left(0, \frac{1}{2}\theta_{\phi}\sigma^2(x)\right)$$

where $\theta_{\phi} = \frac{1}{2}\int_{-\infty}^{\infty}\int_{(z-1)/\phi}^{(z+1)/\phi}\int_{(z-1)/\phi}^{(z+1)/\phi}\mathbf{K}(a)\mathbf{K}(e)dzdade$.

5.13 Remark (the fixed T case) If we fix the time span T the drift function cannot be identified. In particular, the drift estimator would diverge at a speed equal to $\frac{1}{\sqrt{\varepsilon_{n,T}}}$ [c.f. Theorem 10.1 in Part II]. However, if we do not constrain the time span to be fixed, by virtue of recurrence, there are repeated visits to every level over time and this opens up the possibility of recovering the true function by using a single trajectory of the process over a long time, through a combination of *infill* and *long* span asymptotics.

Since the local dynamics of the underlying continuous process reflect more of the features of the diffusion function than those of the drift, only the diffusion function estimator can be meaningfully defined over a fixed time span of observations [c.f. Geman (1979), Merton (1973), Aït-Sahalia (1996a), *inter alia*, and the discussion in Part II].

5.14 Remark Due to the very slow rate of convergence of the variance term in the estimation error decomposition for the drift, the bias term never plays a role in the limit. The rate of convergence is $\sqrt{\varepsilon_{n,T}\overline{L}_X(T,x)}$. This rate cannot be defined in closed form apart from few specific cases [see Remark 5.16]. Technically, it is achieved by choosing a slowly decaying sequence $\varepsilon_{n,T}$ so that $\varepsilon_{n,T}\overline{L}_X(T,x) \overset{a.s.}{\to} \infty$ [see Remark 5.16].

5.15 Remark The drift estimator converges at a substantially slower rate than the diffusion function estimator. This is due to the smaller order of magnitude of the infinitesimal conditional volatility of the process.

5.16 Remark It is useful to consider the limit theory in the Brownian motion case in more detail. This is an interesting special case as it allows us to obtain "closed-form" conditions for the bandwidth $\varepsilon_{n,T}$ and shows the applicabilty of our method to general processes of the type described in Section 2. Further, when the underlying process is Brownian motion, it is possible to illustrate the rate of convergence (as $T \to \infty$ with n) in a more precise fashion. Consider the Brownian motion $B = \sigma W$ with variance σ^2. The main assumption that ensures consistent estimation of the drift estimator is

$$\varepsilon_{n,T}\overline{L}_B(T,x) \overset{a.s.}{\to} \infty$$

for all $x \in \Re$. By using properties of Brownian local time, as indicated in Remark 5.10, we can write

$$\overline{L}_B(T,x) = T^{1/2}L_W(1,\frac{a}{T^{1/2}})$$

where $a = \frac{x}{\sigma}$. Hence,

$$\varepsilon_{n,T}\overline{L}_B(T,x) = \varepsilon_{n,T}T^{1/2}L_W(1,\frac{a}{T^{1/2}}) = \varepsilon_{n,T}T^{1/2}O_p(1)$$

and

$$\varepsilon_{n,T}T^{1/2} \to \infty \Rightarrow \varepsilon_{n,T}\overline{L}_B(T,x) \overset{a.s.}{\to} \infty$$

We want the window width $\varepsilon_{n,T}$ to tend to zero at a slow rate, and T needs to outweight $\varepsilon^2_{n,T}$. This condition does not pose difficulties and is consistent with the other requirements on the bandwidths. The more general case can be easily accommodated by assuming a slowly decaying $\varepsilon_{n,T}$ capable of taking into account plausible rates of divergence of $\overline{L}_X(T,x)$ to infinity in order to guarantee that $\varepsilon_{n,T}\overline{L}_X(T,x) \overset{a.s.}{\to} \infty$.

The rate of convergence in the Brownian motion case is easy to determine. For example, we can express (5.7) as

$$T^{1/4}\varepsilon^{1/2}_{n,T}\{\widehat{\mu}_{(n,T)}(x)\} \overset{d}{\to} (L_W(1,0))^{-\frac{1}{2}}N\left(0,\frac{1}{2}\sigma^2(x)\right) = MN\left(0,\frac{1}{2}\frac{\sigma^2(x)}{L_W(1,0)}\right). \tag{5.8}$$

The local time at the origin plays an important role in determining the asymptotic variance, $\frac{1}{2}\sigma^2(x)/L_W(1,0)$, that appears in this mixed normal limiting distribution.

5.17 Remark $\tilde{\mu}_{(n,T)}(X_{i\Delta_{n,T}})$ and $\tilde{\sigma}^2_{(n,T)}(X_{i\Delta_{n,T}})$ in (4.1) and (4.2) are defined as follows:

$$\tilde{\mu}_{(n,T)}(X_{i\Delta_{n,T}}) = \frac{1}{m_{n,T}(i\Delta_{n,T})\Delta_{n,T}} \sum_{j=0}^{m_{n,T}(i\Delta_{n,T})-1} [X_{t(i\Delta_{n,T})_j+\Delta_{n,T}} - X_{t(i\Delta_{n,T})_j}], \quad (5.9)$$

and

$$\tilde{\sigma}^2_{(n,T)}(X_{i\Delta_{n,T}}) = \frac{1}{m_{n,T}(i\Delta_{n,T})\Delta_{n,T}} \sum_{j=0}^{m_{n,T}(i\Delta_{n,T})-1} [X_{t(i\Delta_{n,T})_j+\Delta_{n,T}} - X_{t(i\Delta_{n,T})_j}]^2. \quad (5.10)$$

where the sequence of stopping times $\{t(i\Delta_{n,T})_j\}$ $j = 1, 2, ..$ satisfies

$$t(i\Delta_{n,T})_0 = \inf\{t \geq 0 : |X_t - X_{i\Delta_{n,T}}| \leq \varepsilon_{n,T}\},$$

and

$$t(i\Delta_{n,T})_{j+1} = \inf\{t \geq t(i\Delta_{n,T})_j + \Delta_{n,T} : |X_t - X_{i\Delta_{n,T}}| \leq \varepsilon_{n,T}\},$$

for all i. We know that formulae (5.9) and (5.10) can be rewritten using indicator kernels, namely

$$\tilde{\mu}_{(n,T)}(X_{i\Delta_{n,T}}) = \frac{1}{\Delta_{n,T}} \frac{\sum_{j=1}^{n-1} \mathbf{1}_{\{|X_{j\Delta_{n,T}} - X_{i\Delta_{n,T}}| \leq \varepsilon_{n,T}\}}[X_{(j+1)\Delta_{n,T}} - X_{j\Delta_{n,T}}]}{\sum_{j=1}^{n} \mathbf{1}_{\{|X_{j\Delta_{n,T}} - X_{i\Delta_{n,T}}| \leq \varepsilon_{n,T}\}}},$$

and

$$\tilde{\sigma}^2_{(n,T)}(X_{i\Delta_{n,T}}) = \frac{1}{\Delta_{n,T}} \frac{\sum_{j=1}^{n-1} \mathbf{1}_{\{|X_{j\Delta_{n,T}} - X_{i\Delta_{n,T}}| \leq \varepsilon_{n,T}\}}[X_{(j+1)\Delta_{n,T}} - X_{j\Delta_{n,T}}]^2}{\sum_{j=1}^{n} \mathbf{1}_{\{|X_{j\Delta_{n,T}} - X_{i\Delta_{n,T}}| \leq \varepsilon_{n,T}\}}}.$$

If a generic smooth kernel is used in place of indicator functions, then the constants of proportionality in the asymptotic variances need to be modified accordingly. For example, the factor of proportionality would be equal to $\int_{-\infty}^{\infty} \mathbf{K}^2(s)ds$, rather than $\frac{1}{2}$, in the case of the drift provided $h_{n,T} = o(\varepsilon_{n,T})$. As for the diffusion function estimator, the factor in (5.1), for instance, would be equal to $4\int_{-\infty}^{\infty} \mathbf{K}^2(s)ds$, rather than 2. An appropriate choice of the kernel could bring about an improvement in efficiency. As an example, in the case of the Gaussian kernel $\int_{-\infty}^{\infty} \mathbf{K}^2(s)ds = \frac{1}{2\sqrt{\pi}} < \frac{1}{2}$.

5.18 Remark The estimators presented and discussed in this chapter are sample analogues to the true theoretical functions. They are written as weighted averages based on convoluted smoothing functions. Our asymptotic results readily apply to weighted averages based on simple kernels. In this case, by virtue of the generality of our formulations, only straightforward modifications to the theory outlined here are needed.

6. Conclusion

This chapter shows how to identify and consistently estimate both the drift and diffusion term of a general homogeneous stochastic differential equation under broad assumptions on the data generating process. The methods can, in principle, be extended to multi-equation specifications although important difficulties associated with the curse of dimensionality arise in that case in the estimation of local time.

The methods presented here are also useful in assessing the asymptotic behavior of functionals of homogeneous diffusions. A typical example that is important in financial applications is the price of derivative securities. In this case, the limit theory that is obtained here for the drift estimator is ideally suited for exploring the limit behavior of functional estimators of fixed-income securities prices. The reason is that the value of these securities depends on the drift of the underlying process even under the no-arbitrage restrictions imposed by martingale pricing. In fixed-income pricing the underlying process is generally a short-term interest rate process. The next chapter applies the methodology suggested above to the analysis of the spot interest rate dynamics in continuous time.

7. Proofs

7.1. Proof of Lemma 3.2

See Revuz and Yor (1991), Corollary 1.8, page 217.

7.2. Proof of Lemma 3.3

See Revuz and Yor (1991), Exercise 1.15, page 222.

7.3. Proof of Lemma 3.4

This is a straightforward consequence of the occupation time formula [c.f. Lemma 3.3] and the right continuity in a of $L_X(t,a)$ (See Revuz and Yor (1991), Corollary 1.9, page 218).

7.4. Proof of Lemma 3.5

The first part of the proof follows Yor (1983). Start by considering a simple application of *Tanaka formula* [c.f Definition 3.1], namely

$$X_t^+ = X_0^+ + \int_0^t 1_{(X_s>0)} dX_s + \frac{1}{2} L_X(t,0),$$

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

Subtract the second expression from the first expression, giving

$$\begin{aligned} & X_t^+ - (X_t - a)^+ \\ = \; & X_0^+ - (X_0 - a)^+ + \int_0^t \mathbf{1}_{(0 \le X_s \le a)} dX_s + \frac{1}{2}(L_X(t,0) - L_X(t,a)). \end{aligned}$$

Equivalently, we can write

$$\begin{aligned} & X_t^+ - (X_t - a/\lambda)^+ \\ = \; & X_0^+ - (X_0 - a/\lambda)^+ + \int_0^t \mathbf{1}_{(0 \le X_s \le a/\lambda)} dX_s + \frac{1}{2}(L_X(t,0) - L_X(t,a/\lambda)). \end{aligned}$$

Now, multiply through by $\sqrt{\lambda}$. This gives,

$$\begin{aligned} & \sqrt{\lambda}(X_t^+ - (X_t - a/\lambda)^+) \\ = \; & \sqrt{\lambda}(X_0^+ - (X_0 - a/\lambda)^+) + \sqrt{\lambda} \int_0^t \mathbf{1}_{(0 \le X_s \le a/\lambda)} dX_s + \\ & + \frac{1}{2}\sqrt{\lambda}(L_X(t,0) - L_X(t,a/\lambda)). \end{aligned}$$

Apparently,

$$\sqrt{\lambda}|X_t^+ - (X_t - a/\lambda)^+| + \sqrt{\lambda}|X_0^+ - (X_0 - a/\lambda)^+| \le 2\frac{a}{\sqrt{\lambda}}.$$

Hence, the asymptotic distribution of $\frac{1}{2}\sqrt{\lambda}(L_X(t,0) - L_X(t,a/\lambda))$ is determined by the term $\sqrt{\lambda}\int_0^t \mathbf{1}_{(0 \le X_s \le a/\lambda)} dX_s$ as $\lambda \to \infty$. Further,

$$\sqrt{\lambda}\int_0^t \mathbf{1}_{(0 \le X_s \le a/\lambda)} dX_s = \sqrt{\lambda}\int_0^t \mathbf{1}_{(0 \le X_s \le a/\lambda)} \mu(X_s) ds + \sqrt{\lambda}\int_0^t \mathbf{1}_{(0 \le X_s \le a/\lambda)} \sigma(X_s) dB_s. \quad (7.1)$$

Now notice that $\sqrt{\lambda}\int_0^t \mathbf{1}_{(0 \le X_s \le a/\lambda)} \mu(X_s) ds \overset{a.s.}{\to} 0$ as $\lambda \to \infty$. In fact, by the occupation time formula [c.f. Lemma 3.3] we can write

$$\begin{aligned} & \sqrt{\lambda}\int_0^t \mathbf{1}_{(0 \le X_s \le a/\lambda)} \mu(X_s) ds \\ = \; & \sqrt{\lambda}\int_{-\infty}^{\infty} \mathbf{1}_{(0 \le b \le a/\lambda)} \frac{\mu(b)}{\sigma^2(b)} L_X(t,b) db \\ = \; & \sqrt{\lambda}\int_{-\infty}^{\infty} \mathbf{1}_{(0 \le \lambda b \le a)} \frac{\mu(b)}{\sigma^2(b)} L_X(t,b) db, \end{aligned}$$

and, setting $\lambda b = c$, this becomes

$$\frac{1}{\sqrt{\lambda}} \int_{-\infty}^{\infty} 1_{(0 \le c \le a)} \frac{\mu(c/\lambda)}{\sigma^2(c/\lambda)} L_X(t, c/\lambda) dc.$$

By the properties of the local time [in particular, the map $a \to L_X(t, a)$ is a.s. continuous and has compact support - c.f. Lemma 3.2] and the dominated convergence theorem, it follows that

$$\int_{-\infty}^{\infty} 1_{(0 \le c \le a)} \frac{\mu(c/\lambda)}{\sigma^2(c/\lambda)} L_X(t, c/\lambda) dc \overset{a.s.}{\to} a \frac{\mu(0)}{\sigma^2(0)} L(t, 0),$$

as $\lambda \to \infty$. In consequence.

$$\frac{1}{\sqrt{\lambda}} \int_{-\infty}^{\infty} 1_{(0 \le c \le a)} \frac{\mu(c/\lambda)}{\sigma^2(c/\lambda)} L_X(t, c/\lambda) dc \overset{a.s.}{\to} 0.$$

This, in turn, implies that the asymptotic behavior of (7.1) is determined by $\sqrt{\lambda} \int_0^t 1_{(0 \le X_s \le a/\lambda)} \sigma(X_s) dB_s$. Now define

$$M^\lambda(t) := \sqrt{\lambda} \int_0^t 1_{(0 \le X_s \le a/\lambda)} \sigma(X_s) dB_s.$$

M^λ is a continuous martingale with quadratic variation process $\{[M^\lambda]_t : t \ge 0\}$ given by

$$\lambda \int_0^t 1_{(0 \le X_s \le a/\lambda)} \sigma^2(X_s) ds.$$

Again, by the occupation time formula, the properties of the local time and dominated convergence, we get

$$\begin{aligned}
[M\lambda]_t &= \lambda \int_0^t 1_{(0 \le X_s \le a/\lambda)} \sigma^2(X_s) ds \\
&= \lambda \int_{-\infty}^{\infty} 1_{(0 \le \lambda b \le a)} L_X(t, b) db \\
&= \int_{-\infty}^{\infty} 1_{(0 \le c \le a)} L_X(t, c/\lambda) dc \\
&\overset{a.s.}{\to} a L_X(t, 0).
\end{aligned}$$

Setting

$$T_t^\lambda = \inf\{s : [M^\lambda]_s > t\},$$

$\tilde{B}_t = M^\lambda_{T_t^\lambda}$ is a Brownian motion and $M_t^\lambda = \tilde{B}_{[M^\lambda]_t}$. In fact, $\tilde{B}_t$ is the so-called *Dambis, Dubins-Schwarz* (DDS, henceforth) Brownian motion of M_t^λ [c.f. Revuz and Yor (1994, Theorem 1.6, page 173 and, for an asymptotic version, Theorem 2.3, page 496)]. It follows that

$$
\begin{aligned}
M_t^\lambda &= \sqrt{\lambda}\int_0^t \mathbf{1}_{(0\le\lambda X_s\le a)}\sigma(X_s)dB_s \\
&\underset{\lambda\to\infty}{\overset{d}{\to}} \widetilde{B}_{aL_X(t,0)} \\
&\overset{d}{=} \sqrt{a}\widetilde{B}_{L_X(t,0)} \\
&\overset{d}{=} \mathfrak{B}_{(L_X(t,0),a)},
\end{aligned}
$$

where $L_X(t,x) = \lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\int_0^t 1_{[x,\, x+\varepsilon[}\sigma^2(X_s)ds$ a.s. $\forall x, t$ and $\mathfrak{B}$ is a standard Brownian sheet.

So far, we have proved convergence of the marginals of a generic family $\mathfrak{P}_\lambda$ of probability measures to corresponding marginal limit distributions. It is easy to verify the compactness of $\mathfrak{P}_\lambda$. The proof follows standard arguments and is omitted here for brevity [see Billingsley (1968)]. Weak convergence then follows. In particular, as $\lambda\to\infty$, the process (indexed by $(t,a)\in\Re_+^2$)

$$(X_t \; ; \; L_X(t,a) \; ; \; \frac{\sqrt{\lambda}}{2}\{L_X(t,\frac{a}{\lambda}) - L_X(t,0)\})$$

converges weakly to

$$(X_t \; ; \; L_X(t,a) \; ; \; \mathfrak{B}(L_X(t,0),a),$$

where $(\mathfrak{B}(s,a) \; ; \; (s,a)\in\Re_+^2)$ is a standard Brownian sheet independent of X [For the independence property, see Revuz and Yor (1994, Exercise 2.12, Chapter XIII)]. Then, a simple generalization of the previous finding to spatial location $r\neq 0$ gives

$$\frac{1}{2}\sqrt{\lambda}\{L_X(t,r+\frac{a}{\lambda}) - L_X(t,r)\} \overset{d}{\to} \mathfrak{B}(L_X(t,r),a),$$

as $\lambda\to\infty$, and this proves the stated result.

7.5. Proof of Theorem 5.1

First, consider the quantity

$$\int_0^{\overline{T}} \frac{1}{h_{n,\overline{T}}}\mathbf{K}(\frac{X_s - x}{h_{n,\overline{T}}})ds.$$

From the occupation time formula, it is easy to see that

$$\int_0^{\overline{T}} \frac{1}{h_{n,\overline{T}}}\mathbf{K}(\frac{X_s - x}{h_{n,\overline{T}}})\frac{d[X]_s}{\sigma^2(X_s)} = \int_{-\infty}^{\infty} \frac{1}{h_{n,\overline{T}}}\mathbf{K}(\frac{a-x}{h_{n,\overline{T}}})\frac{1}{\sigma^2(a)}L_X(\overline{T},a)da$$

Consider the transformation $a \to q$ where $q = (a-x)/h_{n,\overline{T}}$. The previous expression becomes

$$\int_{-\infty}^{\infty} \mathbf{K}(q)\frac{1}{\sigma^2(h_{n,\overline{T}}q+x)}L_X(\overline{T}, h_{n,\overline{T}}q + x)dq.$$

As in the proof of Lemma 3.5, notice that, for a fixed t, the map $a \mapsto \frac{1}{\sigma^2(a)}L_t^a$ is a.s. continuous with compact support by the properties of the diffusion function and the local time. We know, in fact, that for any continuous SMG M, there exists a version of the local time such that the map $(a,t) \mapsto L_t^a$ is a.s. continuous in t and a [Lemma 3.2]. Hence, as $n \to \infty$ and $h_{n,\overline{T}} \to 0$, and since $\int_{-\infty}^{\infty}\mathbf{K}(q)dq = 1$, by dominated convergence, we have

$$\begin{aligned}&\int_{-\infty}^{\infty} \mathbf{K}(q)\frac{1}{\sigma^2(h_{n,\overline{T}}q+x)}L_X(\overline{T}, h_{n,\overline{T}}q + x)dq \overset{a.s.}{\to} \frac{1}{\sigma^2(x)}L_X(\overline{T},x).\\ = \ & \overline{L}_X(\overline{T},x).\end{aligned}$$

Thus, to prove the stated result we just have to prove that

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

under the stated conditions. This is equivalent to proving that

$$\begin{aligned}&\frac{1}{h_{n,\overline{T}}}\sum_{i=0}^{n-1}\int_{i\overline{T}/n}^{(i+1)\overline{T}/n}[\mathbf{K}(\frac{X_{i\Delta_{n,\overline{T}}}-x}{h_{n,\overline{T}}}) - \mathbf{K}(\frac{X_s-x}{h_{n,\overline{T}}})]ds\\ &-\frac{\Delta_{n,\overline{T}}}{h_{n,\overline{T}}}\mathbf{K}(\frac{X_0-x}{h_{n,\overline{T}}}) + \frac{\Delta_{n,\overline{T}}}{h_{n,\overline{T}}}\mathbf{K}(\frac{X_{n\Delta_{n,\overline{T}}}-x}{h_{n,\overline{T}}}) \overset{a.s.}{\to} 0. \qquad (7.2)\end{aligned}$$

But, the left side of the previous expression is bounded by

$$\begin{aligned}&\left|\frac{1}{h_{n,\overline{T}}}\sum_{i=0}^{n-1}\int_{i\overline{T}/n}^{(i+1)\overline{T}/n}[\mathbf{K}(\frac{X_{i\Delta_{n,\overline{T}}}-x}{h_{n,\overline{T}}}) - \mathbf{K}(\frac{X_s-x}{h_{n,\overline{T}}})]ds\right|\\ &+\frac{\Delta_{n,\overline{T}}}{h_{n,\overline{T}}}\left|\mathbf{K}(\frac{X_0-x}{h_{n,\overline{T}}})\right| + \frac{\Delta_{n,\overline{T}}}{h_{n,\overline{T}}}\left|\mathbf{K}(\frac{X_{n\Delta_{n,\overline{T}}}-x}{h_{n,\overline{T}}})\right|\\ \leq \ & \frac{1}{h_{n,\overline{T}}}|\sum_{i=0}^{n-1}\int_{i\overline{T}/n}^{(i+1)\overline{T}/n}\mathbf{K}'(\frac{\tilde{X}_{is}-x}{h_{n,\overline{T}}})\left(\frac{X_s - X_{i\Delta_{n,\overline{T}}}}{h_{n,\overline{T}}}\right)ds| + 2C_3\frac{\Delta_{n,\overline{T}}}{h_{n,\overline{T}}}\\ \leq \ & \left(\frac{1}{h_{n,\overline{T}}}\right)\frac{1}{h_{n,\overline{T}}}\sum_{i=0}^{n-1}\int_{i\overline{T}/n}^{(i+1)\overline{T}/n}\left|\mathbf{K}'(\frac{\tilde{X}_{is}-x}{h_{n,\overline{T}}})\right|\left|\left(X_s - X_{i\Delta_{n,\overline{T}}}\right)\right|ds + 2C_3\frac{\Delta_{n,\overline{T}}}{h_{n,\overline{T}}} \qquad (7.3)\end{aligned}$$

where $\tilde{X}_{is}$ is on the line segment connecting X_s and $X_{i\Delta_{n,\overline{T}}}$. Define

$$\kappa_{n,\overline{T}} = \max_{i\leq n}\sup_{i\Delta_{n,\overline{T}}\leq s\leq (i+1)\Delta_{n,\overline{T}}}|X_s - X_{i\Delta_{n,\overline{T}}}|. \qquad (7.4)$$

By the *Hölder property* for continuous SMGs [e.g. Revuz and Yor (1994, Exercise 1.20, Chapter V)]

$$\mathfrak{P}\left(\left[t \geq 0 : \limsup_{\varepsilon \to 0} \frac{|X_{t+\varepsilon} - X_t|}{\varepsilon^\alpha} > 0\right]\right) = 0 \quad a.s., \tag{7.5}$$

where $\mathfrak{P}$ is Lebesgue measure on $\Re_+$ and (7.5) holds for every $\alpha < \frac{1}{2}$. In turn, (7.5) implies that

$$\frac{\kappa_{n,\overline{T}}}{\left(\Delta_{n,\overline{T}}\right)^\alpha} = o_{a.s.}(1) \tag{7.6}$$

for every $\alpha < \frac{1}{2}$. Hence, if $h_{n,\overline{T}}$ is such that $\frac{1}{h_{n,\overline{T}}}(\Delta_{n,\overline{T}})^\alpha = O(1)$ for some $\alpha \in (0, \frac{1}{2})$, then

$$\frac{\kappa_{n,\overline{T}}}{h_{n,\overline{T}}} = \frac{\kappa_{n,\overline{T}}}{\left(\Delta_{n,\overline{T}}\right)^\alpha} \frac{(\Delta_{n,\overline{T}})^\alpha}{h_{n,\overline{T}}} = o_{a.s.}(1) \tag{7.7}$$

as $n \to \infty$. In view of (7.7) we have

$$\mathbf{K}'\left(\frac{\tilde{X}_{is} - x}{h_{n,\overline{T}}}\right) = \mathbf{K}'\left(\frac{X_s - x}{h_{n,\overline{T}}} + o_{a.s}(1)\right), \tag{7.8}$$

uniformly over $i = 1, ..., n$. It follows from (7.4) and (7.8) that (7.3) is bounded by

$$\begin{aligned}
&\left(\frac{\kappa_{n,\overline{T}}}{h_{n,\overline{T}}}\right) \frac{1}{h_{n,\overline{T}}} \sum_{i=0}^{n-1} \int_{i\overline{T}/n}^{(i+1)\overline{T}/n} \left|\mathbf{K}'\left(\frac{X_s - x}{h_{n,\overline{T}}} + o_{a.s}(1)\right)\right| ds + 2C_3 \frac{\Delta_{n,\overline{T}}}{h_{n,\overline{T}}} \\
\leq\ &\left(\frac{\kappa_{n,\overline{T}}}{h_{n,\overline{T}}}\right) \frac{1}{h_{n,\overline{T}}} \int_0^{\overline{T}} \left|\mathbf{K}'\left(\frac{X_s - x}{h_{n,\overline{T}}} + o_{a.s}(1)\right)\right| ds + 2C_3 \frac{\Delta_{n,\overline{T}}}{h_{n,\overline{T}}} \\
=\ &\left(\frac{\kappa_{n,\overline{T}}}{h_{n,\overline{T}}}\right) \frac{1}{h_{n,\overline{T}}} \int_{-\infty}^{\infty} \left|\mathbf{K}'\left(\frac{p - x}{h_{n,\overline{T}}} + o_{a.s}(1)\right)\right| \overline{L}_X\left(\overline{T}, p\right) dp + 2C_3 \frac{\Delta_{n,\overline{T}}}{h_{n,\overline{T}}} \\
=\ &\left(\frac{\kappa_{n,\overline{T}}}{h_{n,\overline{T}}}\right) \int_{-\infty}^{\infty} \left|\mathbf{K}'(q + o_{a.s}(1))\right| \overline{L}_X\left(\overline{T}, q h_{n,\overline{T}} + x\right) dq + 2C_3 \frac{\Delta_{n,\overline{T}}}{h_{n,\overline{T}}} \\
\leq\ &C_4 \left(\frac{\kappa_{n,\overline{T}}}{h_{n,\overline{T}}}\right) O_{a.s}\left(\overline{L}_X(\overline{T}, x)\right) + 2C_3 \frac{\Delta_{n,\overline{T}}}{h_{n,\overline{T}}},
\end{aligned}$$

for some constant C_4, by virtue of the integrability of $\mathbf{K}'$ and the continuity of $\overline{L}_X$. Since $\frac{\Delta_{n,\overline{T}}}{h_{n,\overline{T}}} \to 0$ and $\frac{\kappa_{n,\overline{T}}}{h_{n,\overline{T}}} \to 0$ by (7.7), the bound goes to zero as $n \to \infty$. This last observation establishes (7.2) and thereby proves the stated result.

7.6. Proof of Corollary 5.2

If $T \to \infty$ and $\frac{T}{n} = \Delta_{n,T} \to 0$, then $\frac{\Delta_{n,T}}{h_{n,T}} \sum_{i=1}^{n} \mathbf{K}(\frac{X_{i\Delta_{n,T}} - x}{h_{n,T}})$ converges to $\overline{L}_X(\infty, x)$ provided $h_{n,T} \to 0$ (as $n \to \infty$) in such a way that $\frac{\overline{L}_X(T,x)}{h_{n,T}}(\Delta_{n,T})^\alpha = O_{a.s.}(1)$ for some $\alpha \in (0, \frac{1}{2})$ and

$\frac{\Delta_{n,T}}{h_{n,T}} \to 0$. But $\overline{L}_X(\infty, x) = \overline{L}_X(\sup\{t : X_t = x\}, x)$ a.s. [Revuz and Yor (1994, Proposition 1.3, Remark 2, page 214)]. And, if the process is recurrent, then $\overline{L}_X((\sup\{t : X_t = x\}), x) = \infty$ a.s.

7.7. Proof of Theorem 5.5

We start by considering the expression

$$\frac{\frac{\Delta_{n,T}}{h_{n,T}} \sum_{i=1}^{n} \mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})(\tilde{\sigma}^2_{n,T}(X_{i\Delta_{n,T}}) - \sigma^2(X_{i\Delta_{n,T}}))}{\frac{\Delta_{n,T}}{h_{n,T}} \sum_{i=1}^{n} \mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})} \tag{7.9}$$

$$+\frac{\frac{\Delta_{n,T}}{h_{n,T}} \sum_{i=1}^{n} \mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})\sigma^2(X_{i\Delta_{n,T}})}{\frac{\Delta_{n,T}}{h_{n,T}} \sum_{i=1}^{n} \mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})}. \tag{7.10}$$

First, we examine (7.10). We want to prove that for some $\varepsilon > 0$

$$\frac{\frac{\Delta_{n,T}}{h_{n,T}} \sum_{i=1}^{n} \mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})\sigma^2(X_{i\Delta_{n,T}})}{\frac{\Delta_{n,T}}{h_{n,T}} \sum_{i=1}^{n} \mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})} \tag{7.11}$$

$$= \frac{\int_0^T \frac{1}{h_{n,T}}\mathbf{K}(\frac{X_s-x}{h_{n,T}})\sigma^2(X_s)ds + o_{a.s.}\left(\frac{\overline{L}_X(T,x)}{h_{n,T}}(\Delta_{n,T})^{1/2-\varepsilon}\right) + O_{a.s.}\left(\frac{\Delta_{n,T}}{h_{n,T}}\right)}{\int_0^T \frac{1}{h_{n,T}}\mathbf{K}(\frac{X_s-x}{h_{n,T}})ds + o_{a.s.}\left(\frac{\overline{L}_X(T,x)}{h_{n,T}}(\Delta_{n,T})^{1/2-\varepsilon}\right) + O_{a.s.}\left(\frac{\Delta_{n,T}}{h_{n,T}}\right)}. \tag{7.12}$$

For the term in the denominator we can utilize the argument used in Theorem 5.1 and Corollary 5.2. As for the numerator, we look at the quantity

$$\frac{\Delta_{n,T}}{h_{n,T}} \sum_{i=1}^{n} \mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})\sigma^2(X_{i\Delta_{n,T}}) - \int_0^T \frac{1}{h_{n,T}}\mathbf{K}(\frac{X_s-x}{h_{n,T}})\sigma^2(X_s)ds \tag{7.13}$$

Given the properties of $\mathbf{K}(\cdot)$, the assumptions on $\sigma(\cdot)$, and proceeding as in the proof of Theorem 5.1, (7.13) is seen to be bounded as follows

$$\begin{aligned}
&\frac{1}{h_{n,T}}\left|\sum_{i=0}^{n-1}\int_{iT/n}^{(i+1)T/n}[\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})\sigma^2(X_{i\Delta_{n,T}}) - \mathbf{K}(\frac{X_s-x}{h_{n,T}})\sigma^2(X_s)]ds\right| \\
&+\frac{\Delta_{n,T}}{h_{n,T}}\mathbf{K}(\frac{X_0-x}{h_{n,T}})\sigma^2(X_0) + \frac{\Delta_{n,T}}{h_{n,T}}\mathbf{K}(\frac{X_{n\Delta_{n,T}}-x}{h_{n,T}})\sigma^2(X_{n\Delta_{n,T}}) \\
\leq\ &\frac{1}{h_{n,T}}\left|\sum_{i=0}^{n-1}\int_{iT/n}^{(i+1)T/n}[\mathbf{K}(\frac{X_s-x}{h_{n,T}})\sigma^2(X_{i\Delta_{n,T}}) - \mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})\sigma^2(X_{i\Delta_{n,T}})]ds\right| \\
&+\frac{1}{h_{n,T}}\left|\sum_{i=0}^{n-1}\int_{iT/n}^{(i+1)T/n}[\mathbf{K}(\frac{X_s-x}{h_{n,T}})\sigma^2(X_s) - \mathbf{K}(\frac{X_s-x}{h_{n,T}})\sigma^2(X_{i\Delta_{n,T}})]ds\right| + 2C_3 O_{a.s.}(\frac{\Delta_{n,T}}{h_{n,T}})
\end{aligned}$$

$$
\begin{aligned}
&\le \frac{1}{h_{n,T}}\sum_{i=0}^{n-1}\int_{iT/n}^{(i+1)T/n}\left|\mathbf{K}'(\frac{\tilde{X}_{is}-x}{h_{n,T}})\right|\left|\left(\frac{X_s-X_{i\Delta_{n,T}}}{h_{n,T}}\right)\right|\sigma^2(X_{i\Delta_{n,T}}) \\
&+\frac{1}{h_{n,\overline{T}}}\left|\sum_{i=0}^{n-1}\int_{iT/n}^{(i+1)T/n}\left[\mathbf{K}(\frac{X_s-x}{h_{n,T}})\left(\sigma^2(X_s)-\sigma^2(X_{i\Delta_{n,T}})\right)\right]ds\right|+2C_3O_{a.s.}(\frac{\Delta_{n,T}}{h_{n,T}}) \\
&\le \left(\frac{\kappa_{n,T}}{h_{n,T}}\right)\frac{1}{h_{n,T}}\int_0^T\left|\mathbf{K}'\left(\frac{X_s-x}{h_{n,T}}+o_{a.s.}(1)\right)\right|\sigma^2\left(X_s+o_{a.s.}(1)\right)ds \\
&+\kappa_{n,T}O_{a.s.}\left(\overline{L}_X(T,x)\right)+2C_3O_{a.s.}(\frac{\Delta_{n,T}}{h_{n,T}}) \\
&\le C_4\left(\frac{\kappa_{n,T}}{h_{n,T}}\right)O_{a.s}\left(L_X(T,x)\right)+\kappa_{n,T}O_{a.s.}\left(\overline{L}_X(T,x)\right)+2C_3O_{a.s.}(\frac{\Delta_{n,T}}{h_{n,T}})
\end{aligned}
$$

where $\kappa_{n,T}=\max_{i\le n}\sup_{i\Delta_{n,\overline{T}}\le s\le(i+1)\Delta_{n,\overline{T}}}|X_s-X_{i\Delta_{n,\overline{T}}}|$ as before. Under the stated conditions as $n,T\to\infty$, that is $h_{n,T}\to 0$ and

$$
\frac{\overline{L}_X(T,x)}{h_{n,T}}(\Delta_{n,T})^{\alpha}=O_{a.s.}(1) \tag{7.14}
$$

for some $\alpha\in(0,\frac{1}{2})$ and $\frac{\Delta_{n,T}}{h_{n,T}}\to 0$, the three terms are negligible in the limit and formula (7.11) holds for some $\varepsilon>0$ such that $\alpha\le\frac{1}{2}-\varepsilon$. Next, we prove that

$$
\begin{aligned}
&\frac{\int_0^T\frac{1}{h_{n,T}}\mathbf{K}\left(\frac{X_s-x}{h_{n,T}}\right)\sigma^2(X_s)ds+o_{a.s.}\left(\frac{\overline{L}_X(T,x)}{h_{n,T}}(\Delta_{n,T})^{1/2-\epsilon}\right)+O_{a.s.}\left(\frac{\Delta_{n,T}}{h_{n,T}}\right)}{\int_0^T\frac{1}{h_{n,T}}\mathbf{K}\left(\frac{X_s-x}{h_{n,T}}\right)ds+o_{a.s.}\left(\frac{\overline{L}_X(T,x)}{h_{n,T}}(\Delta_{n,T})^{1/2-\epsilon}\right)+O_{a.s.}\left(\frac{\Delta_{n,T}}{h_{n,T}}\right)} \\
&= \frac{\sigma^2(x)\overline{L}_X(T,x)+o_{a.s.}(1)}{\overline{L}_X(T,x)+o_{a.s.}(1)}\overset{a.s.}{\to}\sigma^2(x).
\end{aligned} \tag{7.15}
$$

By virtue of the occupation time formula, Lemma 3.5 and the fact that $h_{n,T}\to 0$, we have

$$
\begin{aligned}
\int_0^T\frac{1}{h_{n,T}}\mathbf{K}\left(\frac{X_s-x}{h_{n,T}}\right)ds &= \frac{1}{h_{n,T}}\int_{-\infty}^{\infty}\mathbf{K}\left(\frac{p-x}{h_{n,T}}\right)\overline{L}_X(T,p)\,dp \\
&= \int_{-\infty}^{\infty}\mathbf{K}(q)\,\overline{L}_X(T,qh_{n,T}+x)\,dq \\
&= \overline{L}_X(T,x)+o_{a.s.}(1),
\end{aligned}
$$

giving the required result for the denominator of (7.15). It remains to verify that

$$
\int_0^T\frac{1}{h_{n,T}}\mathbf{K}(\frac{X_s-x}{h_{n,T}})\sigma^2(X_s)ds=\sigma^2(x)\overline{L}_X(T,x)+o_{a.s.}(1). \tag{7.16}
$$

Using the occupation time formula again, we have

$$
\begin{aligned}
\int_0^T\frac{1}{h_{n,T}}\mathbf{K}\left(\frac{X_s-x}{h_{n,T}}\right)\sigma^2(X_s)ds &= \int_0^T\frac{1}{h_{n,T}}\mathbf{K}\left(\frac{X_s-x}{h_{n,T}}\right)d[X]_s \\
&= \int_{-\infty}^{\infty}\frac{1}{h_{n,T}}\mathbf{K}\left(\frac{p-x}{h_{n,T}}\right)L_X(T,p)dp
\end{aligned}
$$

$$
\begin{aligned}
&= \int_{-\infty}^{\infty} \mathbf{K}(q)\, L_X(T, h_{n,T}q + x)dq \\
&= L_X(T,x) + o_{a.s.}(1) \\
&= \sigma^2(x)\overline{L}_X(T,x) + o_{a.s.}(1),
\end{aligned}
$$

establishing (7.16) as required, and then (7.15) follows. We now turn to the analysis of (7.9). It is sufficient to prove that

$$
\frac{1}{m_{n,T}(i\Delta_{n,T})} \sum_{j=0}^{m_{n,T}(i\Delta_{n,T})-1} \frac{[X_{t(i\Delta_{n,T})_j+\Delta_{n,T}} - X_{t(i\Delta_{n,T})_j}]^2}{\Delta_{n,T}} - \sigma^2(X_{i\Delta_{n,T}}) = o_{a.s.}(1). \quad (7.17)
$$

in order to verify the stated result. By stochastic differentiation we have

$$
dX_s^2 = 2X_s dX_s + \sigma^2(X_s)\, ds = 2X_s\mu(X_s)\, ds + 2X_s\sigma(X_s)\, dB_s + \sigma^2(X_s)\, ds,
$$

so that

$$
\begin{aligned}
& X^2_{t(i\Delta_{n,T})_j+\Delta_{n,T}} - X^2_{t(i\Delta_{n,T})_j} \\
= \; & 2\int_{t(i\Delta_{n,T})_j}^{t(i\Delta_{n,T})_j+\Delta_{n,T}} X_s\mu(X_s)ds + 2\int_{t(i\Delta_{n,T})_j}^{t(i\Delta_{n,T})_j+\Delta_{n,T}} X_s\sigma(X_s)dB_s + \int_{t(i\Delta_{n,T})_j}^{t(i\Delta_{n,T})_j+\Delta_{n,T}} \sigma^2(X_s)ds.
\end{aligned}
$$

Further

$$
X_{t(i\Delta_{n,T})_j+\Delta_{n,T}} = X_{t(i\Delta_{n,T})_j} + \int_{t(i\Delta_{n,T})_j}^{t(i\Delta_{n,T})_j+\Delta_{n,T}} \mu(X_s)ds + \int_{t(i\Delta_{n,T})_j}^{t(i\Delta_{n,T})_j+\Delta_{n,T}} \sigma(X_s)dB_s, \quad (7.18)
$$

so that

$$
\begin{aligned}
& \left[X_{t(i\Delta_{n,T})_j+\Delta_{n,T}} - X_{t(i\Delta_{n,T})_j}\right]^2 \\
= \; & X^2_{t(i\Delta_{n,T})_j+\Delta_{n,T}} - 2X_{t(i\Delta_{n,T})_j+\Delta_{n,T}}X_{t(i\Delta_{n,T})_j} + X^2_{t(i\Delta_{n,T})_j} \\
= \; & X^2_{t(i\Delta_{n,T})_j+\Delta_{n,T}} - X^2_{t(i\Delta_{n,T})_j} - 2X_{t(i\Delta_{n,T})_j}\left[\int_{t(i\Delta_{n,T})_j}^{t(i\Delta_{n,T})_j+\Delta_{n,T}} \mu(X_s)ds + \int_{t(i\Delta_{n,T})_j}^{t(i\Delta_{n,T})_j+\Delta_{n,T}} \sigma(X_s)dB_s\right] \\
= \; & \int_{t(i\Delta_{n,T})_j}^{t(i\Delta_{n,T})_j+\Delta_{n,T}} 2\left(X_s - X_{t(i\Delta_{n,T})}\right)\mu(X_s)ds + \int_{t(i\Delta_{n,T})_j}^{t(i\Delta_{n,T})_j+\Delta_{n,T}} 2\left(X_s - X_{t(i\Delta_{n,T})}\right)\sigma(X_s)dB_s \\
& + \int_{t(i\Delta_{n,T})_j}^{t(i\Delta_{n,T})_j+\Delta_{n,T}} \sigma^2(X_s)ds.
\end{aligned}
$$

Then

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

It follows that

$$
\begin{aligned}
&\frac{1}{m_{n,T}(i\Delta_{n,T})}\sum_{j=0}^{m_{n,T}(i\Delta_{n,T})-1}\frac{\left[X_{t(i\Delta_{n,T})_j+\Delta_{n,T}}-X_{t(i\Delta_{n,T})_j}\right]^2}{\Delta_{n,T}}-\sigma^2(X_{i\Delta_{n,T}})\\
=\;&\frac{1}{m_{n,T}(i\Delta_{n,T})\Delta_{n,T}}\sum_{j=0}^{m_{n,T}(i\Delta_{n,T})-1}\int_{t(i\Delta_{n,T})_j}^{t(i\Delta_{n,T})_j+\Delta_{n,T}}2\left(X_s-X_{t(i\Delta_{n,T})}\right)\mu(X_s)ds\\
&+\frac{1}{m_{n,T}(i\Delta_{n,T})\Delta_{n,T}}\sum_{j=0}^{m_{n,T}(i\Delta_{n,T})-1}\int_{t(i\Delta_{n,T})_j}^{t(i\Delta_{n,T})_j+\Delta_{n,T}}2\left(X_s-X_{t(i\Delta_{n,T})}\right)\sigma(X_s)dB_s\\
&+\frac{1}{m_{n,T}(i\Delta_{n,T})\Delta_{n,T}}\sum_{j=0}^{m_{n,T}(i\Delta_{n,T})-1}\int_{t(i\Delta_{n,T})_j}^{t(i\Delta_{n,T})_j+\Delta_{n,T}}\left[\sigma^2(X_s)-\sigma^2(X_{i\Delta_{n,T}})\right]ds\\
=\;&\frac{1}{m_{n,T}(i\Delta_{n,T})\Delta_{n,T}}\{\sum_{j=0}^{m_{n,T}(i\Delta_{n,T})-1}\int_{t(i\Delta_{n,T})_j}^{t(i\Delta_{n,T})_j+\Delta_{n,T}}\left[\sigma^2(X_s)-\sigma^2(X_{t(i\Delta_{n,T})_j})\right]ds\\
&+\sum_{j=0}^{m_{n,T}(i\Delta_{n,T})-1}\int_{t(i\Delta_{n,T})_j}^{t(i\Delta_{n,T})_j+\Delta_{n,T}}\left[\sigma^2(X_{t(i\Delta_{n,T})_j})-\sigma^2(X_{i\Delta_{n,T}})\right]ds\}\\
&+\frac{1}{m_{n,T}(i\Delta_{n,T})\Delta_{n,T}}\sum_{j=0}^{m_{n,T}(i\Delta_{n,T})-1}\int_{t(i\Delta_{n,T})_j}^{t(i\Delta_{n,T})_j+\Delta_{n,T}}2\left(X_s-X_{t(i\Delta_{n,T})}\right)\sigma(X_s)dB_s\\
&+C_5O_{a.s.}(\kappa_{n,T})\\
=\;&C_6O_{a.s.}(\kappa_{n,T})+C_7O_{a.s.}(\epsilon_{n,T})\\
&+O_{a.s.}\left(\frac{\kappa_{n,T}}{m_{n,T}(i\Delta_{n,T})\Delta_{n,T}}\sum_{j=0}^{m_{n,T}(i\Delta_{n,T})-1}\int_{t(i\Delta_{n,T})_j}^{t(i\Delta_{n,T})_j+\Delta_{n,T}}\sigma(X_s)dB_s\right). \qquad (7.19)
\end{aligned}
$$

It remains to determine the order of the last term of (7.19).

Define $y_{t(i\Delta_{n,T})_j+\Delta_{n,T}}=\int_{t(i\Delta_{n,T})_j}^{t(i\Delta_{n,T})_j+\Delta_{n,T}}\sigma(X_s)dB_s$, which is measurable with respect to $\Im_{t(i\Delta_{n,T})_j+\Delta_{n,T}}$, where $\Im_{t(i\Delta_{n,T})_j+\Delta_{n,T}}=\{A\in\Im : A\{t(i\Delta_{n,T})_j+\Delta_{n,T}\leq t^*\}\in\Im_{t^+}\forall t\geq 0\}$ for all $j\leq m_{n,T}$. Further,

$$E\left(y_{t(i\Delta_{n,T})_j+\Delta_{n,T}}\right)=0,$$

and, by the Ito isometry,

$$\theta_{t(i\Delta_{n,T})_j+\Delta_{n,T}}=var\left(y_{t(i\Delta_{n,T})_j+\Delta_{n,T}}\right)=E\left(\int_{t(i\Delta_{n,T})_j}^{t(i\Delta_{n,T})_j+\Delta_{n,T}}\sigma^2(X_s)ds\right)<\infty,$$

for all $j\leq m_{n,T}$. So, $\left(y_{t(i\Delta_{n,T})_j+\Delta_{n,T}},\Im_{t(i\Delta_{n,T})_j+\Delta_{n,T}}\right)$ is a martingale difference sequence with zero mean and variance $\theta_{t(i\Delta_{n,T})_j+\Delta_{n,T}}$. Invoking a strong law of large numbers for martingale differences [e.g. Hall and Heyde (1980, Theorem 2.19, page 36)], we have

$$\frac{1}{m_{n,T}(i\Delta_{n,T})}\sum_{j=0}^{m_{n,T}(i\Delta_{n,T})-1} y_{t(i\Delta_{n,T})_j+\Delta_{n,T}}$$

$$= \frac{1}{m_{n,T}(i\Delta_{n,T})}\sum_{j=0}^{m_{n,T}(i\Delta_{n,T})-1}\int_{t(i\Delta_{n,T})_j}^{t(i\Delta_{n,T})_j+\Delta_{n,T}} \sigma(X_s)dB_s \overset{a.s.}{\to} 0 \text{ as } n,T\to\infty,$$

as $m_{n,T}\to\infty$ $(\forall i)$. We now explore the rate of convergence. Consider,

$$\frac{1}{m_{n,T}(i\Delta_{n,T})\Delta_{n,T}}\sum_{j=0}^{m_{n,T}(i\Delta_{n,T})-1}\int_{t(i\Delta_{n,T})_j}^{t(i\Delta_{n,T})_j+\Delta_{n,T}} \sigma(X_s)dB_s$$

$$= \frac{1}{\Delta_{n,T}}\frac{\frac{\Delta_{n,T}}{\varepsilon_{n,T}}\sum_{j=1}^{n-1}\mathbf{1}_{\{|X_{j\Delta_{n,T}}-X_{i\Delta_{n,T}}|\le\varepsilon_{n,T}\}}\int_{j\Delta_{n,T}}^{(j+1)\Delta_{n,T}}\sigma(X_s)dB_s}{\frac{\Delta_{n,T}}{\varepsilon_{n,T}}\sum_{j=1}^{n}\mathbf{1}_{\{|X_{j\Delta_{n,T}}-X_{i\Delta_{n,T}}|\le\varepsilon_{n,T}\}}}$$

$$= \frac{\frac{1}{2\varepsilon_{n,T}}\sum_{j=1}^{n-1}\mathbf{1}_{\{|X_{j\Delta_{n,T}}-X_{i\Delta_{n,T}}|\le\varepsilon_{n,T}\}}\int_{j\Delta_{n,T}}^{(j+1)\Delta_{n,T}}\sigma(X_s)dB_s}{\frac{\Delta_{n,T}}{2\varepsilon_{n,T}}\sum_{j=1}^{n}\mathbf{1}_{\{|X_{j\Delta_{n,T}}-X_{i\Delta_{n,T}}|\le\varepsilon_{n,T}\}}}.$$

First, analyze the numerator of this expression. Write,

$$\mathbf{U}_{n,T}^{X_{i\Delta_{n,T}}}(r) = \sqrt{\varepsilon_{n,T}}\left(\frac{1}{2\varepsilon_{n,T}}\sum_{j=1}^{[nr]-1}\mathbf{1}_{\{|X_{j\Delta_{n,T}}-X_{i\Delta_{n,T}}|\le\varepsilon_{n,T}\}}\int_{j\Delta_{n,T}}^{(j+1)\Delta_{n,T}}\sigma(X_s)dB_s\right)$$

$$= \frac{1}{2\sqrt{\varepsilon_{n,T}}}\sum_{j=1}^{[nr]-1}\mathbf{1}_{\{|X_{j\Delta_{n,T}}-X_{i\Delta_{n,T}}|\le\varepsilon_{n,T}\}}\int_{j\Delta_{n,T}}^{(j+1)\Delta_{n,T}}\sigma(X_s)dB_s.$$

$\mathbf{U}_{n,T}^{X_{i\Delta_{n,T}}}$ is a continuous martingale whose quadratic variation process $[\mathbf{U}_{n,T}]_r$ is

$$[\mathbf{U}_{n,T}^{X_{i\Delta_{n,T}}}]_r = \frac{1}{4\varepsilon_{n,T}}\sum_{j=1}^{[nr]-1}\mathbf{1}_{\{|X_{j\Delta_{n,T}}-X_{i\Delta_{n,T}}|\le\varepsilon_{n,T}\}}\int_{j\Delta_{n,T}}^{(j+1)\Delta_{n,T}}\sigma^2(X_s)ds$$

$$= \frac{1}{4\varepsilon_{n,T}}\sum_{j=1}^{[nr]-1}\mathbf{1}_{\{|X_{j\Delta_{n,T}}-X_{i\Delta_{n,T}}|\le\varepsilon_{n,T}\}}\sigma^2(X_{j\Delta_{n,T}}+o_{a.s.}(1))\Delta_{n,T}$$

$$= \frac{1}{4\varepsilon_{n,T}}\int_0^{rT}\mathbf{1}_{\{|X_s-X_{i\Delta_{n,T}}|\le\varepsilon_{n,T}\}}\sigma^2(X_s+o_{a.s.}(1))ds + o_{a.s.}(1)$$

$$= \frac{1}{2}L(rT,X_{i\Delta_{n,T}}) + o_{a.s.}(1)$$

$$= \frac{1}{2}\sigma^2(X_{i\Delta_{n,T}})\overline{L}(rT,X_{i\Delta_{n,T}}) + o_{a.s.}(1),$$

by virtue of (3.5). Now, as in Theorem 3.4 in Phillips and Ploberger (1996), expanding the probability as needed, we have

$$\left(\mathbf{U}_{n,T}^{X_{i\Delta_{n,T}}}\right)^2/[\mathbf{U}_{n,T}^{X_{i\Delta_{n,T}}}]_1 = O_{a.s.}(1),$$

and then it follows that

$$\sqrt{\overline{L}(T, X_{i\Delta_{n,T}})\varepsilon_{n,T}} \left(\frac{\frac{1}{\Delta_{n,T}} \frac{\Delta_{n,T}}{2\varepsilon_{n,T}} \sum_{j=1}^{[nr]-1} \mathbf{1}_{\{|X_{j\Delta_{n,T}} - X_{i\Delta_{n,T}}| \leq \varepsilon_{n,T}\}} \int_{j\Delta_{n,T}}^{(j+1)\Delta_{n,T}} \sigma(X_s) dB_s}{\frac{\Delta_{n,T}}{2\varepsilon_{n,T}} \sum_{j=1}^{n-1} \mathbf{1}_{\{|X_{j\Delta_{n,T}} - X_{i\Delta_{n,T}}| \leq \varepsilon_{n,T}\}}} \right)$$
$$= O_{a.s.}(1).$$

This result implies that the bound (7.19) becomes

$$C_8(\kappa_{n,T} + \varepsilon_{n,T}) + C_9 \kappa_{n,T} O_{a.s.} \left(\sqrt{\frac{1}{\overline{L}(T, X_{i\Delta_{n,T}})\varepsilon_{n,T}}} \right)$$
$$= o_{a.s}(1) + o(1) + o_{a.s.} \left(\frac{\Delta_{n,T}^{1/2-\delta}}{\sqrt{\overline{L}(T, X_{i\Delta_{n,T}})\varepsilon_{n,T}}} \right) \overset{a.s}{\to} 0$$

since $\frac{\Delta_{n,T}^{1/2-\delta}}{\sqrt{\overline{L}(T, X_{i\Delta_{n,T}})\varepsilon_{n,T}}} \overset{a.s.}{\to} 0$ by assumption and in view of (7.6). This proves the stated result.

7.8. Proof of Theorem 5.6

We write the estimation error in two components as follows:

$$\frac{\frac{\Delta_{n,T}}{h_{n,T}} \sum_{i=1}^{n} \mathbf{K}(\frac{X_{i\Delta_{n,T}} - x}{h_{n,T}}) \tilde{\sigma}_{n,T}(X_{i\Delta_{n,T}})}{\frac{\Delta_{n,T}}{h_{n,T}} \sum_{i=1}^{n} \mathbf{K}(\frac{X_{i\Delta_{n,T}} - x}{h_{n,T}})} - \sigma^2(x)$$
$$= \underbrace{\frac{\frac{\Delta_{n,T}}{h_{n,T}} \sum_{i=1}^{n} \mathbf{K}(\frac{X_{i\Delta_{n,T}} - x}{h_{n,T}}) \tilde{\sigma}_{n,T}^2(X_{i\Delta_{n,T}})}{\frac{\Delta_{n,T}}{h_{n,T}} \sum_{i=1}^{n} \mathbf{K}(\frac{X_{i\Delta_{n,T}} - x}{h_{n,T}})} - \frac{\frac{\Delta_{n,T}}{h_{n,T}} \sum_{i=1}^{n} \mathbf{K}(\frac{X_{i\Delta_{n,T}} - x}{h_{n,T}}) \sigma^2(X_{i\Delta_{n,T}})}{\frac{\Delta_{n,T}}{h_{n,T}} \sum_{i=1}^{n} \mathbf{K}(\frac{X_{i\Delta_{n,T}} - x}{h_{n,T}})}}_{\text{term } V}$$
$$+ \underbrace{\frac{\frac{\Delta_{n,T}}{h_{n,T}} \sum_{i=1}^{n} \mathbf{K}(\frac{X_{i\Delta_{n,T}} - x}{h_{n,T}}) \sigma^2(X_{i\Delta_{n,T}})}{\frac{\Delta_{n,T}}{h_{n,T}} \sum_{i=1}^{n} \mathbf{K}(\frac{X_{i\Delta_{n,T}} - x}{h_{n,T}})} - \frac{\sigma^2(x) \frac{\Delta_{n,T}}{h_{n,T}} \sum_{i=1}^{n} \mathbf{K}(\frac{X_{i\Delta_{n,T}} - x}{h_{n,T}})}{\frac{\Delta_{n,T}}{h_{n,T}} \sum_{i=1}^{n} \mathbf{K}(\frac{X_{i\Delta_{n,T}} - x}{h_{n,T}})}}_{\text{term } B}$$
$$= \text{term } V + \text{term } B.$$

Roughly speaking this is a decomposition into a bias term B and second effect, V. We start with the bias term B. Combining the two fractions constituting B, the numerator of the term is

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

By the mean-value theorem, the occupation time formula, and using the same approach as in the proof of (7.11) above, we find that

$$\begin{aligned}
&\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})2\sigma(x_i^*)\sigma'(x_i^*)(X_{i\Delta_{n,T}}-x)\\
=&\ \frac{1}{h_{n,T}}\int_0^T\mathbf{K}\left(\frac{X_s-x}{h_{n,T}}+o_{a.s.}(1)\right)2\sigma\left(f(X_s,x)+o_{a.s.}(1)\right)\times\\
&\times\sigma'\left(f(X_s,x)+o_{a.s.}(1)\right)(X_s-x+o_{a.s.}(1))ds+o_{a.s.}\left(\overline{L}_X(T,x)(\Delta_{n,T})^{1/2-\varepsilon}\right)\\
=&\ \frac{1}{h_{n,T}}\int_{-\infty}^{\infty}\mathbf{K}(\frac{a-x}{h_{n,T}})2\sigma\left(f(a,x)\right)\sigma'\left(f(a,x)\right)(a-x)\overline{L}_X(T,a)da\\
&+o_{a.s.}\left(\overline{L}_X(T,x)(\Delta_{n,T})^{1/2-\varepsilon}\right),
\end{aligned}$$

for some $\varepsilon>0$, and where $x_i^*=f(X_{i\Delta_{n,T}},x)\in[X_{i\Delta_{n,T}},x]\ \forall i$. If we multiply by $\frac{1}{h_{n,T}}$, then the first term becomes

$$\begin{aligned}
&\frac{1}{h_{n,T}}\left(\frac{1}{h_{n,T}}\int_{-\infty}^{\infty}\mathbf{K}(\frac{a-x}{h_{n,T}})2\sigma\left(f(a,x)\right)\sigma'\left(f(a,x)\right)(a-x)\overline{L}_X(T,a)da\right)\\
=&\ \frac{1}{h_{n,T}}\int_{-\infty}^{\infty}\mathbf{K}(\frac{a-x}{h_{n,T}})2\sigma\left(f(a,x)\right)\sigma'\left(f(a,x)\right)\left(\frac{a-x}{h_{n,T}}\right)\overline{L}_X(T,a)da\\
=&\ \int_{-\infty}^{\infty}c\mathbf{K}(c)2\sigma(x)\sigma'(x)\overline{L}_X(T,x+h_{n,T}c)dc+o_{a.s.}(1)\\
=&\ \int_{-\infty}^{\infty}c\mathbf{K}(c)2\left(\frac{\sigma'(x)}{\sigma(x)}\right)L_X(T,x+h_{n,T}c)dc+o_{a.s.}(1)\\
=&\ \int_{-\infty}^{\infty}c\mathbf{K}(c)2\left(\frac{\sigma'(x)}{\sigma(x)}\right)\left(L_X(T,x+h_{n,T}c)-L_X(T,x)\right)dc\ +o_{a.s.}(1),
\end{aligned}\tag{7.20}$$

since $\int_{-\infty}^{\infty}c\mathbf{K}(c)dc=0$. By Lemma 3.5 and neglecting the term of smaller order of magnitude, (7.20) has the following limiting form as functional of a Brownian sheet $\mathfrak{B}$:

$$\begin{aligned}
&\int_{-\infty}^{\infty}c\mathbf{K}(c)4\left(\frac{\sigma'(x)}{\sigma(x)}\right)\frac{1}{2\sqrt{h_{n,T}}}\left(L_X(T,x+h_{n,T}c)-L_X(T,x)\right)dc\\
=&\ 4\left(\frac{\sigma'(x)}{\sigma(x)}\right)\int_{-\infty}^{\infty}c\mathbf{K}(c)\frac{1}{2\sqrt{h_{n,T}}}\left(L_X(T,x+h_{n,T}c)-L_X(T,x)\right)dc\\
\overset{d}{\to}&\ 4\left(\frac{\sigma'(x)}{\sigma(x)}\right)\int_{-\infty}^{\infty}c\mathbf{K}(c)\mathfrak{B}(L_X(T,x),c)dc\\
\overset{d}{=}&\ 4\left(\frac{\sigma'(x)}{\sigma(x)}\right)\sqrt{L_X(T,x)}\int_{-\infty}^{\infty}c\mathbf{K}(c)\mathfrak{B}(1,c)dc.
\end{aligned}\tag{7.21}$$

Now, define $G(u) = \int_{-\infty}^{u} c\mathbf{K}(c)dc$, and integrate $\int_{-\infty}^{\infty} c\mathbf{K}(c)\mathfrak{B}(1,c)dc$ by parts, giving

$$\begin{aligned}
&\int_{-\infty}^{\infty} c\mathbf{K}(c)\mathfrak{B}(1,c)dc \\
=\ & G(c)\mathfrak{B}(1,c)|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} G(c)d\mathfrak{B}(1,c) \\
=\ & -\int_{-\infty}^{\infty} G(c)d\mathfrak{B}(1,c) \\
\overset{d}{=}\ & \mathbf{B}(\int_{-\infty}^{\infty} G(c)^2 dc) \\
\overset{d}{=}\ & \mathbf{B}(\varphi/4) \qquad (7.22)
\end{aligned}$$

where $\varphi = -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} 2|a-b|ab\mathbf{K}(a)\mathbf{K}(b)dadb$. In consequence,

$$\begin{aligned}
&\frac{1}{(h_{n,T})^{3/2}}\left(\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})\left(\sigma^2(X_{i\Delta_{n,T}})-\sigma^2(x)\right)\right) \\
&\overset{d}{\to} \mathbf{B}\left(4\varphi\left(\sigma'(x)\right)^2\overline{L}_X(T,x)\right),
\end{aligned}$$

where $\mathbf{B}$ is a standard Brownian motion independent of $\overline{L}_X(T,x)$ and φ is a constant of proportionality equal to $2\langle f,f\rangle$ where $\langle f,f\rangle$ is called the "energy" of the function $f(s) = s\mathbf{K}(s)$, i.e. $\langle f,f\rangle = -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}|a-b|ab\mathbf{K}(a)\mathbf{K}(b)dadb$ - see Revuz and Yor (1991, Lemma 2.7, Chapter XIII). In turn,

$$\begin{aligned}
&\frac{\sqrt{\overline{L}_X(T,x)}}{(h_{n,T})^{3/2}}\left(\frac{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})\left(\sigma^2(X_{i\Delta_{n,T}})-\sigma^2(x)\right)}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})}\right) \\
&\overset{d}{\to} \mathbf{B}\left(4\varphi\left(\sigma'(x)\right)^2\right) \\
&\overset{d}{=} N\left(0, 4\varphi\left(\sigma'(x)\right)^2\right).
\end{aligned}$$

Next consider term V :

$$V = \frac{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})\left(\tilde{\sigma}^2_{n,T}(X_{i\Delta_{n,T}})-\sigma^2(X_{i\Delta_{n,T}})\right)}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})}.$$

The numerator can be written as

$$\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})\left(\tilde{\sigma}^2_{n,T}(X_{i\Delta_{n,T}})-\sigma^2(X_{i\Delta_{n,T}})\right) =$$

$$\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})\frac{\sum_{j=1}^{n-1}\mathbf{1}_{\{|X_{j\Delta_{n,T}}-X_{i\Delta_{n,T}}|\leq\varepsilon_{n,T}\}}\frac{n}{T}\left[\left(X_{(j+1)T/n}-X_{jT/n}\right)^2-\sigma^2(X_{iT/n})\Delta_{n,T}\right]}{\sum_{j=1}^{n}\mathbf{1}_{\{|X_{j\Delta_{n,T}}-X_{i\Delta_{n,T}}|\leq\varepsilon_{n,T}\}}}.$$

By Ito's lemma [see the proof of Theorem 5.5]

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

Hence,

$$\begin{aligned}&\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})\frac{\sum_{j=1}^{n-1}\mathbf{1}_{\{|X_{j\Delta_{n,T}}-X_{i\Delta_{n,T}}|\leq\varepsilon_{n,T}\}}\frac{n}{T}\left[\left(X_{(j+1)T/n}-X_{jT/n}\right)^2-\sigma^2(X_{iT/n})\Delta_{n,T}\right]}{\sum_{j=1}^{n}\mathbf{1}_{\{|X_{j\Delta_{n,T}}-X_{i\Delta_{n,T}}|\leq\varepsilon_{n,T}\}}}\\ =\ &\underbrace{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})\frac{\sum_{j=1}^{n-1}\mathbf{1}_{\{|X_{j\Delta_{n,T}}-X_{i\Delta_{n,T}}|\leq\varepsilon_{n,T}\}}\frac{n}{T}\left[\int_{jT/n}^{(j+1)T/n}\left(\sigma^2(X_s)-\sigma^2(X_{iT/n})\right)ds\right]}{\sum_{j=1}^{n}\mathbf{1}_{\{|X_{j\Delta_{n,T}}-X_{i\Delta_{n,T}}|\leq\varepsilon_{n,T}\}}}}_{(\mathbf{A}_{n,T})}\\ &+\underbrace{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})\frac{\sum_{j=1}^{n-1}\mathbf{1}_{\{|X_{j\Delta_{n,T}}-X_{i\Delta_{n,T}}|\leq\varepsilon_{n,T}\}}\frac{n}{T}\left[\int_{jT/n}^{(j+1)T/n}2\left(X_s-X_{jT/n}\right)\sigma(X_s)dB_s\right]}{\sum_{j=1}^{n}\mathbf{1}_{\{|X_{j\Delta_{n,T}}-X_{i\Delta_{n,T}}|\leq\varepsilon_{n,T}\}}}}_{(\mathbf{B}_{n,T}(1))}\\ &+\underbrace{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})\frac{\sum_{j=1}^{n-1}\mathbf{1}_{\{|X_{j\Delta_{n,T}}-X_{i\Delta_{n,T}}|\leq\varepsilon_{n,T}\}}\frac{n}{T}\left[\int_{jT/n}^{(j+1)T/n}2\left(X_s-X_{jT/n}\right)\mu(X_s)ds\right]}{\sum_{j=1}^{n}\mathbf{1}_{\{|X_{j\Delta_{n,T}}-X_{i\Delta_{n,T}}|\leq\varepsilon_{n,T}\}}}}_{(\mathbf{C}_{n,T})}\\ =\ &\mathbf{A}_{n,T}+\mathbf{B}_{n,T}(1)+\mathbf{C}_{n,T}.\end{aligned}\tag{7.23}$$

These three terms comprise an additional bias effect, $\mathbf{A}_{n,T}$, a martingale effect, $\mathbf{B}_{n,T}(1)$, and a residual effect, $\mathbf{C}_{n,T}$. As we shall see, depending on the bandwidth choices, either $\mathbf{A}_{n,T}$ or $\mathbf{B}_{n,T}$ may dominate. First, examine $\sqrt{\frac{\varepsilon_{n,T}}{\Delta_{n,T}}}\mathbf{B}_{n,T}(r)$, which takes the form

$$\begin{aligned}&\sqrt{\frac{\varepsilon_{n,T}}{\Delta_{n,T}}}\mathbf{B}_{n,T}(r)\\ =\ &\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{[nr]}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})\times\\ &\times\frac{\frac{1}{2\sqrt{\varepsilon_{n,T}}}\sum_{j=1}^{n-1}\mathbf{1}_{\{|X_{j\Delta_{n,T}}-X_{i\Delta_{n,T}}|\leq\varepsilon_{n,T}\}}\sqrt{\frac{1}{\Delta_{n,T}}}\left[\int_{jT/n}^{(j+1)T/n}2\left(X_s-X_{jT/n}\right)\sigma(X_s)dB_s\right]}{\frac{\Delta_{n,T}}{2\varepsilon_{n,T}}\sum_{j=1}^{n}\mathbf{1}_{\{|X_{j\Delta_{n,T}}-X_{i\Delta_{n,T}}|\leq\varepsilon_{n,T}\}}}.\end{aligned}$$

The martingale $\mathbf{B}_{n,T}(r)$ has quadratic variation process which can be analysed as follows, using the same approach as that employed in deriving (7.11),

$$
\begin{aligned}
&[\mathbf{B}_{n,T}]_r \\
= & \frac{1}{4}\left(\frac{\Delta_{n,T}}{\sqrt{\varepsilon_{n,T}}h_{n,T}}\right)^2 \sum_{i=1}^{[nr]}\sum_{k=1}^{[nr]} \mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})\mathbf{K}(\frac{X_{k\Delta_{n,T}}-x}{h_{n,T}}) \times \\
& \times \frac{\sum_{j=1}^{n-1} \mathbf{1}_{\{|X_{j\Delta_{n,T}}-X_{i\Delta_{n,T}}|\leq\varepsilon_{n,T}\}} \mathbf{1}_{\{|X_{j\Delta_{n,T}}-X_{k\Delta_{n,T}}|\leq\varepsilon_{n,T}\}} \frac{1}{\Delta_{n,T}}\left[\int_{jT/n}^{(j+1)T/n} 4\left(X_s - X_{jT/n}\right)^2 \sigma^2(X_s)ds\right]}{\left(\frac{\Delta_{n,T}}{2\varepsilon_{n,T}}\sum_{j=1}^n \mathbf{1}_{\{|X_{j\Delta_{n,T}}-X_{i\Delta_{n,T}}|\leq\varepsilon_{n,T}\}}\right)\left(\frac{\Delta_{n,T}}{2\varepsilon_{n,T}}\sum_{j=1}^n \mathbf{1}_{\{|X_{j\Delta_{n,T}}-X_{k\Delta_{n,T}}|\leq\varepsilon_{n,T}\}}\right)} \\
= & \left(\frac{1}{h_{n,T}}\right)^2 \int_0^{[Tr]} ds \int_0^{[Tr]} du \mathbf{K}(\frac{X_s-x}{h_{n,T}})\mathbf{K}(\frac{X_u-x}{h_{n,T}}) \times \\
& \times \frac{\frac{1}{\varepsilon_{n,T}}\int_0^T db \mathbf{1}_{\{|X_b-X_s|\leq\varepsilon_{n,T}\}}\mathbf{1}_{\{|X_b-X_u|\leq\varepsilon_{n,T}\}}\sigma^4(X_b+o_{a.s.}(1))}{\left(\frac{1}{2\varepsilon_{n,T}}\int_0^T \mathbf{1}_{\{|X_b-X_s|\leq\varepsilon_{n,T}\}}db + o_{a.s}(1)\right)\left(\frac{1}{2\varepsilon_{n,T}}\int_0^T \mathbf{1}_{\{|X_b-X_u|\leq\varepsilon_{n,T}\}}db + o_{a.s}(1)\right)} \\
& + o_{a.s}(1) \\
= & \left(\frac{1}{h_{n,T}}\right)^2 \int_{-\infty}^{\infty}\int_{-\infty}^{+\infty} dsdu \mathbf{K}(\frac{s-x}{h_{n,T}})\mathbf{K}(\frac{u-x}{h_{n,T}}) \times \\
& \times \frac{\frac{1}{\varepsilon_{n,T}}\int_{-\infty}^{\infty} db \mathbf{1}_{\{|b-s|\leq\varepsilon_{n,T}\}}\mathbf{1}_{\{|b-u|\leq\varepsilon_{n,T}\}}\sigma^4(b)\overline{L}_X(T,b)\overline{L}_X(rT,s)\overline{L}_X(rT,u)}{\left(\frac{1}{2\varepsilon_{n,T}}\int_{-\infty}^{\infty} \mathbf{1}_{\{|b-s|\leq\varepsilon_{n,T}\}}\overline{L}_X(T,b)db + o_{a.s}(1)\right)\left(\frac{1}{2\varepsilon_{n,T}}\int_{-\infty}^{\infty} \mathbf{1}_{\{|b-u|\leq\varepsilon_{n,T}\}}\overline{L}_X(T,b)db + o_{a.s}(1)\right)} \\
& + o_{a.s}(1).
\end{aligned} \tag{7.24}
$$

Let

$$\frac{s-x}{h_{n,T}} = a \text{ and } \frac{u-x}{h_{n,T}} = e.$$

Then, (7.22) is

$$
\begin{aligned}
& \int_{-\infty}^{\infty}\int_{-\infty}^{+\infty} dade\mathbf{K}(a)\mathbf{K}(e) \times \\
& \times \frac{\frac{1}{\varepsilon_{n,T}}\int_{-\infty}^{\infty} db \mathbf{1}_{\{|b-x-ah_{n,T}|\leq\varepsilon_{n,T}\}}\mathbf{1}_{\{|b-x-eh_{n,T}|\leq\varepsilon_{n,T}\}}\sigma^4(b)\overline{L}_X(T,b)\overline{L}_X(rT,x+ah_{n,T})\overline{L}_X(rT,x+eh_{n,T})}{\left(\frac{1}{2\varepsilon_{n,T}}\int_{-\infty}^{\infty} \mathbf{1}_{\{|b-x-ah_{n,T}|\leq\varepsilon_{n,T}\}}\overline{L}_X(T,b)db\right)\left(\frac{1}{2\varepsilon_{n,T}}\int_{-\infty}^{\infty} \mathbf{1}_{\{|b-x-eh_{n,T}|\leq\varepsilon_{n,T}\}}\overline{L}_X(T,b)db\right)} \\
& + o_{a.s}(1),
\end{aligned}
$$

and setting

$$\frac{b-x}{\varepsilon_{n,T}} = z$$

this becomes

$$\int_{-\infty}^{\infty}\int_{-\infty}^{+\infty} dade\mathbf{K}(a)\mathbf{K}(e) \times$$
$$\times \frac{\int_{-\infty}^{\infty} dz \mathbf{1}_{\{|z-a\frac{h_{n,T}}{\varepsilon_{n,T}}|\leq 1\}}\mathbf{1}_{\{|z-e\frac{h_{n,T}}{\varepsilon_{n,T}}|\leq 1\}}\sigma^4(x)\overline{L}_X(T,x+z\varepsilon_{n,T})\overline{L}_X(rT,x+ah_{n,T})\overline{L}_X(rT,x+eh_{n,T})}{\left(\frac{1}{2}\int_{-\infty}^{\infty}\mathbf{1}_{\{|z-a\frac{h_{n,T}}{\varepsilon_{n,T}}|\leq 1\}}\overline{L}_X(T,x+z\varepsilon_{n,T})dz\right)\left(\frac{1}{2}\int_{-\infty}^{\infty}\mathbf{1}_{\{|z-e\frac{h_{n,T}}{\varepsilon_{n,T}}|\leq 1\}}\overline{L}_X(T,x+z\varepsilon_{n,T})dz\right)}$$
$$+o_{a.s}(1).$$

Now, if $h_{n,T} = o(\varepsilon_{n,T})$, then

$$[\mathbf{B}_{n,T}]_r \overset{a.s.}{\to} 2\sigma^4(x)\frac{\left(\overline{L}_X(rT,x)\right)^2}{\overline{L}_X(T,x)}.$$

If $h_{n,T} = O(\varepsilon_{n,T})$ with $h_{n,T}/\varepsilon_{n,T} \to \phi > 0$, then

$$[\mathbf{B}_{n,T}]_r \overset{a.s.}{\to} 2\theta_\phi\sigma^4(x)\frac{\left(\overline{L}_X(rT,x)\right)^2}{\overline{L}_X(T,x)}$$

where

$$\begin{aligned}
\theta_\phi &= \int_{-\infty}^{\infty}\int_{-\infty}^{+\infty} dade\mathbf{K}(a)\mathbf{K}(e)\left(\frac{\frac{1}{2}\int_{-\infty}^{\infty}\mathbf{1}_{\{|z-\phi a|\leq 1\}}\mathbf{1}_{\{|z-\phi e|\leq 1\}}dz}{\left(\frac{1}{2}\int_{-\infty}^{\infty}\mathbf{1}_{\{|z-\phi a|\leq 1\}}dz\right)\left(\frac{1}{2}\int_{-\infty}^{\infty}\mathbf{1}_{\{|z-\phi e|\leq 1\}}dz\right)}\right) \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{+\infty} dade\mathbf{K}(a)\mathbf{K}(e)\frac{1}{2}\int_{-\infty}^{\infty} dz\mathbf{1}_{\{|z-\phi a|\leq 1\}}\mathbf{1}_{\{|z-\phi e|\leq 1\}} \\
&= \frac{1}{2}\int_{-\infty}^{\infty}\int_{-\infty}^{+\infty}\int_{-\infty}^{\infty} dadedz\mathbf{K}(a)\mathbf{K}(e)\mathbf{1}_{\{|z-\phi a|\leq 1\}}\mathbf{1}_{\{|z-\phi e|\leq 1\}} \\
&= \frac{1}{2}\int_{-\infty}^{\infty}\int_{(z-1)/\phi}^{(z+1)/\phi}\int_{(z-1)/\phi}^{(z+1)/\phi} dadedz\mathbf{K}(a)\mathbf{K}(e).
\end{aligned}$$

By earlier arguments [e.g., the proofs of Lemma 3.5 and Theorem 5.2] this implies that

$$\sqrt{\frac{\varepsilon_{n,T}\overline{L}_X(T,x)}{\Delta_{n,T}}}\left(\frac{\mathbf{B}_{n,T}(1)}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})}\right) \overset{d}{\to} N\left(0,2\sigma^4(x)\right) \qquad (7.25)$$

if $h_{n,T} = o(\varepsilon_{n,T})$, and

$$\sqrt{\frac{\varepsilon_{n,T}\overline{L}_X(T,x)}{\Delta_{n,T}}}\left(\frac{\mathbf{B}_{n,T}(1)}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})}\right) \overset{d}{\to} N\left(0,2\theta_\phi\sigma^4(x)\right) \qquad (7.26)$$

if $h_{n,T} = O(\varepsilon_{n,T})$ and $h_{n,T}/\varepsilon_{n,T} \to \phi > 0$. Next, examine $\mathbf{A}_{n,T}$:

$$
\begin{aligned}
&\mathbf{A}_{n,T} \\
= & \frac{\Delta_{n,T}}{h_{n,T}} \sum_{i=1}^{n} \mathbf{K}\left(\frac{X_{i\Delta_{n,T}} - x}{h_{n,T}}\right) \frac{\sum_{j=1}^{n-1} \mathbf{1}_{\{|X_{j\Delta_{n,T}} - X_{i\Delta_{n,T}}| \le \varepsilon_{n,T}\}} \frac{n}{T} \left[\int_{jT/n}^{(j+1)T/n} \left(\sigma^2(X_s) - \sigma^2(X_{iT/n})\right) ds\right]}{\sum_{j=1}^{n} \mathbf{1}_{\{|X_{j\Delta_{n,T}} - X_{i\Delta_{n,T}}| \le \varepsilon_{n,T}\}}} \\
= & \frac{\Delta_{n,T}}{h_{n,T}} \sum_{i=1}^{n} \mathbf{K}\left(\frac{X_{i\Delta_{n,T}} - x}{h_{n,T}}\right) \frac{\sum_{j=1}^{n-1} \mathbf{1}_{\{|X_{j\Delta_{n,T}} - X_{i\Delta_{n,T}}| \le \varepsilon_{n,T}\}} \frac{n}{T} \left[\int_{jT/n}^{(j+1)T/n} \left(\sigma^2(X_s) - \sigma^2(X_{jT/n})\right) ds\right]}{\sum_{j=1}^{n} \mathbf{1}_{\{|X_{j\Delta_{n,T}} - X_{i\Delta_{n,T}}| \le \varepsilon_{n,T}\}}} \\
& + \frac{\Delta_{n,T}}{h_{n,T}} \sum_{i=1}^{n} \mathbf{K}\left(\frac{X_{i\Delta_{n,T}} - x}{h_{n,T}}\right) \frac{\frac{\Delta_{n,T}}{\varepsilon_{n,T}} \sum_{j=1}^{n-1} \mathbf{1}_{\{|X_{j\Delta_{n,T}} - X_{i\Delta_{n,T}}| \le \varepsilon_{n,T}\}} \left(\sigma^2(X_{jT/n}) - \sigma^2(X_{iT/n})\right)}{\frac{\Delta_{n,T}}{\varepsilon_{n,T}} \sum_{j=1}^{n} \mathbf{1}_{\{|X_{j\Delta_{n,T}} - X_{i\Delta_{n,T}}| \le \varepsilon_{n,T}\}}} \\
= & C_{10} O_{a.s.}(\kappa_{n,T}) \frac{\Delta_{n,T}}{h_{n,T}} \sum_{i=1}^{n} \mathbf{K}\left(\frac{X_{i\Delta_{n,T}} - x}{h_{n,T}}\right) + \\
& + \frac{\Delta_{n,T}}{h_{n,T}} \sum_{i=1}^{n} \mathbf{K}\left(\frac{X_{i\Delta_{n,T}} - x}{h_{n,T}}\right) \frac{\frac{\Delta_{n,T}}{\varepsilon_{n,T}} \sum_{j=1}^{n-1} \mathbf{1}_{\{|X_{j\Delta_{n,T}} - X_{i\Delta_{n,T}}| \le \varepsilon_{n,T}\}} 2\sigma(x_{ij}^*)\sigma'(x_{ij}^*) \left(X_{jT/n} - X_{iT/n}\right)}{\frac{\Delta_{n,T}}{\varepsilon_{n,T}} \sum_{j=1}^{n} \mathbf{1}_{\{|X_{j\Delta_{n,T}} - X_{i\Delta_{n,T}}| \le \varepsilon_{n,T}\}}}
\end{aligned}
$$

where $\kappa_{n,T}$ has its usual meaning and $x_{ij}^* = f\left(X_{jT/n}, X_{iT/n}\right) \in [X_{jT/n}, X_{iT/n}]$. Then, proceeding as in the derivation of (7.11), we find that

$$
\begin{aligned}
&\mathbf{A}_{n,T} \\
= & \frac{1}{h_{n,T}} \int_0^T ds \mathbf{K}\left(\frac{X_s - x}{h_{n,T}}\right) \frac{\frac{1}{2\varepsilon_{n,T}} \int_0^T du \mathbf{1}_{\{|X_u - X_s| \le \varepsilon_{n,T}\}} 2\sigma\left(f(X_u, X_s)\right) \sigma'\left(f(X_u, X_s)\right) (X_u - X_s)}{\frac{1}{2\varepsilon_{n,T}} \int_0^T \mathbf{1}_{\{|X_u - X_s| \le \varepsilon_{n,T}\}} du} \\
& + o_{a.s}(1) \\
= & \frac{1}{h_{n,T}} \int_{-\infty}^{\infty} ds \mathbf{K}\left(\frac{s - x}{h_{n,T}}\right) \frac{\frac{1}{2\varepsilon_{n,T}} \int_{-\infty}^{\infty} du \mathbf{1}_{\{|u - s| \le \varepsilon_{n,T}\}} 2\sigma\left(f(u, s)\right) \sigma'\left(f(u, s)\right) (u - s)}{\frac{1}{2\varepsilon_{n,T}} \int_{-\infty}^{\infty} \mathbf{1}_{\{|u - s| \le \varepsilon_{n,T}\}} \overline{L}_X(T, u) du} \overline{L}_X(T, u) \overline{L}_X(T, s) \\
& + o_{a.s}(1).
\end{aligned}
$$

Let

$$
\frac{s - x}{h_{n,T}} = c \text{ and } \frac{u - x}{\varepsilon_{n,T}} = a
$$

then

$$
\begin{aligned}
&\frac{1}{\varepsilon_{n,T}} \mathbf{A}_{n,T} \\
= & \int_{-\infty}^{\infty} \mathbf{K}(c) \frac{\frac{1}{2\varepsilon_{n,T}^2} \int_{-\infty}^{\infty} \mathbf{1}_{\{|u - x - ch_{n,T}| \le \varepsilon_{n,T}\}} 2\sigma\left(f(u, x + ch_{n,T})\right) \sigma'\left(f(u, x + ch_{n,T})\right) (u - x - ch_{n,T})}{\frac{1}{2\varepsilon_{n,T}} \int_{-\infty}^{\infty} \mathbf{1}_{\{|u - x - ch_{n,T}| \le \varepsilon_{n,T}\}} \overline{L}_X(T, u) du} \times \\
& \times \overline{L}_X(T, u) \overline{L}_X(T, x + ch_{n,T}) du dc
\end{aligned}
$$

$$
\begin{aligned}
&= \int_{-\infty}^{\infty} \mathbf{K}(c) \frac{\int_{-\infty}^{\infty} \mathbf{1}_{\{|a-c\frac{h_{n,T}}{\varepsilon_{n,T}}|\le 1\}} \sigma\left(f(x+a\varepsilon_{n,T}, x+ch_{n,T})\right) \sigma'\left(f(x+a\varepsilon_{n,T}, x+ch_{n,T})\right)\left(a - c\frac{h_{n,T}}{\varepsilon_{n,T}}\right)}{\frac{1}{2}\int_{-\infty}^{\infty} \mathbf{1}_{\{|a-c\frac{h_{n,T}}{\varepsilon_{n,T}}|\le 1\}} \overline{L}_X(T, x+a\varepsilon_{n,T})da} \times \\
&\quad \times \overline{L}_X(T, x+a\varepsilon_{n,T})\overline{L}_X(T, x+ch_{n,T})dadc + o_{a.s}(1) \\
&= \int_{-\infty}^{\infty} dc\mathbf{K}(c) \frac{\int_{-\infty}^{\infty} da \mathbf{1}_{\{|a-c\frac{h_{n,T}}{\varepsilon_{n,T}}|\le 1\}} \sigma(x)\sigma'(x)\left(a - c\frac{h_{n,T}}{\varepsilon_{n,T}}\right)}{\overline{L}_X(T,x) + o_{a.s.}(1)} \left(\overline{L}_X(T, x+a\varepsilon_{n,T}) - \overline{L}_X(T,x)\right)\overline{L}_X(T,x) \\
&\quad + o_{a.s}(1).
\end{aligned}
$$

because $\int_{-\infty}^{\infty} \mathbf{1}_{\{|g|\le 1\}} g dg = 0$. Hence, if $h_{n,T} = o(\varepsilon_{n,T})$, using Lemma 3.5 and proceeding as in (7.21) and (7.22) above, we find that

$$
\frac{1}{\varepsilon_{n,T}^{3/2}} \mathbf{A}_{n,T} \xrightarrow{d} N\left(0, 4\varphi^* \left(\sigma'(x)\right)^2 \overline{L}_X(T,x)\right).
$$

where

$$
\begin{aligned}
\varphi^* &= -2\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |a-b| ab \left(\frac{1}{2}\mathbf{1}_{\{|a|\le 1\}}\right)\left(\frac{1}{2}\mathbf{1}_{\{|b|\le 1\}}\right) dadb \\
&= -\frac{1}{2}\int_{-1}^{1}\int_{-1}^{1} |a-b|\, abdadb = 0.2666,
\end{aligned}
$$

by direct calculation. Then,

$$
\frac{\sqrt{\overline{L}_X(T,x)}}{\varepsilon_{n,T}^{3/2}} \left(\frac{\mathbf{A}_{n,T}}{\frac{\Delta_{n,T}}{h_{n,T}} \sum_{i=1}^{n} \mathbf{K}(\frac{X_{i\Delta_{n,T}} - x}{h_{n,T}})} \right) \xrightarrow{d} N\left(0, 4\varphi^* \left(\sigma'(x)\right)^2\right). \tag{7.27}
$$

Next, if $h_{n,T} = O(\varepsilon_{n,T})$ and $h_{n,T}/\varepsilon_{n,T} \to \phi > 0$, then

$$
\begin{aligned}
&\int_{-\infty}^{\infty} dc\mathbf{K}(c) \frac{\frac{1}{\sqrt{\varepsilon_{n,T}}}\int_{-\infty}^{\infty} da \mathbf{1}_{\{|a-c\frac{h_{n,T}}{\varepsilon_{n,T}}|\le 1\}} \sigma(x)\sigma'(x)\left(a - c\frac{h_{n,T}}{\varepsilon_{n,T}}\right)}{\overline{L}_X(T,x) + o_{a.s.}(1)} \times \\
&\times \left(\overline{L}_X(T, x+a\varepsilon_{n,T}) - \overline{L}_X(T,x)\right)\overline{L}_X(T,x) \\
&= 2\left(\frac{\sigma'(x)}{\sigma(x)}\right)\int_{-\infty}^{\infty} dc\mathbf{K}(c)\left(\frac{1}{2}\int_{-\infty}^{\infty} da \mathbf{1}_{\{|a-\phi c|\le 1\}}\right)(a-\phi c)\frac{1}{\sqrt{\varepsilon_{n,T}}}\left(L_X(T, x+a\varepsilon_{n,T}) - L_X(T,x)\right) \\
&+ o_{a.s}(1) \\
&\xrightarrow{d} 4\left(\frac{\sigma'(x)}{\sigma(x)}\right)\int_{-\infty}^{\infty} dc\mathbf{K}(c)\left(\frac{1}{2}\int_{-\infty}^{\infty} da \mathbf{1}_{\{|a-\phi c|\le 1\}}\right)(a-\phi c)\,\mathfrak{B}(1,a)
\end{aligned}
$$

Now, define $G_\phi(u) = \int_{-\infty}^{u}\int_{-\infty}^{\infty} \mathbf{K}(c)(a-\phi c)\frac{1}{2}\mathbf{1}_{\{|a-\phi c|\le 1\}} dcda$. We proceed as in (7.22) above and integrate

$$
\int_{-\infty}^{\infty} dc\mathbf{K}(c)\left(\frac{1}{2}\int_{-\infty}^{\infty} da \mathbf{1}_{\{|a-\phi c|\le 1\}}\right)(a-\phi c)\,\mathfrak{B}(1,a)
$$

by parts to obtain

$$\begin{aligned}
&\int_{-\infty}^{\infty} dc \left(\int_{-\infty}^{\infty} da \mathbf{K}(c) \frac{1}{2} \mathbf{1}_{\{|a-\phi c| \leq 1\}} (a - \phi c) \right) \mathfrak{B}(1, a) \\
&= G_\phi(a) \mathfrak{B}(1,a) |_{-\infty}^{\infty} - \int_{-\infty}^{\infty} G_\phi(a) d\mathfrak{B}(1,a) \\
&= -\int_{-\infty}^{\infty} G_\phi(a) d\mathfrak{B}(1,a) \\
&\stackrel{d}{=} \mathbf{B}\left(\int_{-\infty}^{\infty} G_\phi(a)^2 da\right) \\
&\stackrel{d}{=} \mathbf{B}(\vartheta_\phi / 4),
\end{aligned}$$

where

$$\vartheta_\phi = \frac{1}{8} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{\infty}^{\infty} \mathbf{K}(s) \mathbf{K}(c) (a - \phi s) \mathbf{1}_{\{|a-\phi s| \leq 1\}} (b - \phi s) \mathbf{1}_{\{|b-\phi s| \leq 1\}} (a - \phi c) \times$$
$$\times \mathbf{1}_{\{|a-\phi c| \leq 1\}} (b - \phi c) \mathbf{1}_{\{|b-\phi c| \leq 1\}} \mathbf{1}_{(0 \leq v \leq a)} \mathbf{1}_{(0 \leq v \leq b)} da db dc ds dv.$$

Hence,

$$\frac{1}{\varepsilon_{n,T}^{3/2}} \mathbf{A}_{n,T} \stackrel{d}{\to} N\left(0, 4\vartheta_\phi \left(\sigma'(x)\right)^2 \overline{L}_X(T,x)\right),$$

and, in consequence

$$\frac{\sqrt{\overline{L}_X(T,x)}}{\varepsilon_{n,T}^{3/2}} \left(\frac{\mathbf{A}_{n,T}}{\frac{\Delta_{n,T}}{h_{n,T}} \sum_{i=1}^{n} \mathbf{K}\left(\frac{X_{i\Delta_{n,T}} - x}{h_{n,T}}\right)} \right) \stackrel{d}{\to} N\left(0, 4\vartheta_\phi \left(\sigma'(x)\right)^2\right). \tag{7.28}$$

As for $\frac{\mathbf{C}_{n,T}}{\frac{\Delta_{n,T}}{h_{n,T}} \sum_{i=1}^{n} \mathbf{K}\left(\frac{X_{i\Delta_{n,T}} - x}{h_{n,T}}\right)}$, this term can be bounded as follows,

$$\begin{aligned}
&\frac{\frac{\Delta_{n,T}}{h_{n,T}} \sum_{i=1}^{n} \mathbf{K}\left(\frac{X_{i\Delta_{n,T}} - x}{h_{n,T}}\right) \frac{\sum_{j=1}^{n-1} \mathbf{1}_{\{|X_{j\Delta_{n,T}} - X_{i\Delta_{n,T}}| \leq \varepsilon_{n,T}\}} \frac{n}{T} \left[\int_{jT/n}^{(j+1)T/n} (X_s - X_{jT/n}) \mu(X_s) ds \right]}{\sum_{j=1}^{n} \mathbf{1}_{\{|X_{j\Delta_{n,T}} - X_{i\Delta_{n,T}}| \leq \varepsilon_{n,T}\}}}}{\frac{\Delta_{n,T}}{h_{n,T}} \sum_{i=1}^{n} \mathbf{K}\left(\frac{X_{i\Delta_{n,T}} - x}{h_{n,T}}\right)} \\
&= o_{a.s.}\left(\sqrt{\frac{\Delta_{n,T}}{\varepsilon_{n,T} \overline{L}_X(T,x)}} \right).
\end{aligned}$$

Then, defining the overall estimation error as

$$E = B + \frac{\mathbf{A}_{n,T}}{\frac{\Delta_{n,T}}{h_{n,T}} \sum_{i=1}^{n} \mathbf{K}\left(\frac{X_{i\Delta_{n,T}} - x}{h_{n,T}}\right)} + \frac{\mathbf{C}_{n,T}}{\frac{\Delta_{n,T}}{h_{n,T}} \sum_{i=1}^{n} \mathbf{K}\left(\frac{X_{i\Delta_{n,T}} - x}{h_{n,T}}\right)} + \frac{\mathbf{B}_{n,T}(1)}{\frac{\Delta_{n,T}}{h_{n,T}} \sum_{i=1}^{n} \mathbf{K}\left(\frac{X_{i\Delta_{n,T}} - x}{h_{n,T}}\right)}$$

and scaling by $\sqrt{\frac{\varepsilon_{n,T} \overline{L}_X(T,x)}{\Delta_{n,T}}}$ we have

$$\sqrt{\frac{\varepsilon_{n,T}\overline{L}_X(T,x)}{\Delta_{n,T}}}\left(\frac{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})\widetilde{\sigma}_{n,T}(X_{i\Delta_{n,T}})}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})}-\sigma^2(x)\right)$$
$$= \sqrt{\frac{\varepsilon_{n,T}\overline{L}_X(T,x)}{\Delta_{n,T}}}\left\{O_p\left(\frac{h_{n,T}^{3/2}}{\sqrt{\overline{L}(T,x)}}\right)+o_{a.s.}\left((\Delta_{n,T})^{1/2-\varepsilon}\right)\right.$$
$$\left.+o_{a.s.}\left(\sqrt{\frac{\Delta_{n,T}}{\varepsilon_{n,T}\overline{L}_X(T,x)}}\right)+O_p\left(\frac{\varepsilon_{n,T}^{3/2}}{\sqrt{\overline{L}_X(T,x)}}\right)+\frac{\mathbf{B}_{n,T}(1)}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})}\right\}$$
$$= O_p\left(\frac{\varepsilon_{n,T}^{1/2}h_{n,T}^{3/2}}{\sqrt{\Delta_{n,T}}}\right)+o_{a.s.}(1)+o_{a.s.}\left(\sqrt{\varepsilon_{n,T}\overline{L}_X(T,x)}\right)+O_p\left(\frac{\varepsilon_{n,T}^{2}}{\sqrt{\Delta_{n,T}}}\right)$$
$$+\sqrt{\frac{\varepsilon_{n,T}\overline{L}_X(T,x)}{\Delta_{n,T}}}\left(\frac{\mathbf{B}_{n,T}(1)}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})}\right)$$
$$\xrightarrow{d} N\left(0,2\sigma^4(x)\right)$$

from (7.25), for choices of $\varepsilon_{n,T}$ such that $\frac{\varepsilon_{n,T}^2}{\sqrt{\Delta_{n,T}}}\to 0$, $\sqrt{\varepsilon_{n,T}\overline{L}_X(T,x)}\xrightarrow{a.s.}0$ and $h_{n,T}=o(\varepsilon_{n,T})$. If $\frac{\varepsilon_{n,T}^2}{\sqrt{\Delta_{n,T}}}\to 0$, $\varepsilon_{n,T}\overline{L}_X(T,x)\xrightarrow{a.s.}0$ and $h_{n,T}=O(\varepsilon_{n,T})$ with $h_{n,T}/\varepsilon_{n,T}\to\phi>0$, then

$$\sqrt{\frac{\varepsilon_{n,T}\overline{L}_X(T,x)}{\Delta_{n,T}}}\left(\frac{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})\widetilde{\sigma}_{n,T}(X_{i\Delta_{n,T}})}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})}-\sigma^2(x)\right)$$
$$\xrightarrow{d} N\left(0,2\theta_\phi\sigma^4(x)\right)$$

from (7.26), where $\theta_\phi=\frac{1}{2}\int_{-\infty}^{\infty}\int_{(z-1)/\phi}^{(z+1)/\phi}\int_{(z-1)/\phi}^{(z+1)/\phi}\mathbf{K}(a)\mathbf{K}(e)dzdade$. Finally, provided that $\frac{\varepsilon_{n,T}^2}{\sqrt{\Delta_{n,T}}}\to\infty$, $\sqrt{\overline{L}_X(T,x)}\frac{(\Delta_{n,T})^{1/2-\epsilon}}{\varepsilon_{n,T}^{3/2}}\xrightarrow{a.s.}0$ and $h_{n,T}=o(\varepsilon_{n,T})$, the $\mathbf{A}_{n,T}$ term dominates, leading to

$$\frac{\sqrt{\overline{L}_X(T,x)}}{\varepsilon_{n,T}^{3/2}}\left\{O_p\left(\sqrt{\frac{\Delta_{n,T}}{\varepsilon_{n,T}\overline{L}_X(T,x)}}\right)+O_p\left(\frac{h_{n,T}^{3/2}}{\sqrt{\overline{L}_X(T,x)}}\right)+o_{a.s.}\left((\Delta_{n,T})^{1/2-\varepsilon}\right)\right.$$
$$\left.+\frac{\mathbf{A}_{n,T}}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})}\right\}\xrightarrow{d} N\left(0,4\varphi^*\left(\sigma'(x)\right)^2\right)$$

from (7.27), where $\varphi^*=-2\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}|a-b|ab\left(\frac{1}{2}\mathbf{1}_{\{|a|\le 1\}}\right)\left(\frac{1}{2}\mathbf{1}_{\{|b|\le 1\}}\right)dadb=0.2666$. Under the same conditions, but when $h_{n,T}=O(\varepsilon_{n,T})$ with $h_{n,T}/\varepsilon_{n,T}\to\phi>0$, we have

$$\frac{\sqrt{\overline{L}_X(T,x)}}{\varepsilon_{n,T}^{3/2}}\left\{O_p\left(\sqrt{\frac{\Delta_{n,T}}{\varepsilon_{n,T}\overline{L}(T,x)}}\right)+\frac{B(1)}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})}+o_{a.s.}\left((\Delta_{n,T})^{1/2-\epsilon}\right)\right.$$

$$\left.+\frac{\mathbf{A}_{n,T}}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})}\right\}\xrightarrow{d}N\left(0,4\left(\vartheta_\phi+\phi^3\varphi+\eta_\phi\right)\left(\sigma'(x)\right)^2\right)$$

from (7.28) and (7.22), where

$$\vartheta_\phi=\tfrac{1}{8}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{\infty}^{\infty}\mathbf{K}(s)\mathbf{K}(c)(a-\phi s)\mathbf{1}_{\{|a-\phi s|\leq 1\}}(b-\phi s)\mathbf{1}_{\{|b-\phi s|\leq 1\}}(a-\phi c)\times$$
$$\times\mathbf{1}_{\{|a-\phi c|\leq 1\}}(b-\phi c)\mathbf{1}_{\{|b-\phi c|\leq 1\}}\mathbf{1}_{(0\leq v\leq a)}\mathbf{1}_{(0\leq v\leq b)}dadbdcdsdv$$

$$\varphi=-\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}2|a-b|ab\mathbf{K}(a)\mathbf{K}(b)dadb$$

and

$$\eta_\phi=4\phi^2\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}u\left(a-\phi c\right)\mathbf{K}(u)\mathbf{K}(c)\mathbf{1}_{\{|a-\phi c|\leq 1\}}\mathbf{1}_{(0\leq b\leq u)}\mathbf{1}_{(0\leq b\phi\leq a)}dbdudadc.$$

Notice that η_ϕ is the constant (for a given ϕ) of proportionality in the asymptotic covariance between

$$\frac{1}{\varepsilon_{n,T}^{3/2}}\left(\frac{B(1)}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})}\right)=\frac{B^*(1)}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})}$$

and

$$\frac{1}{\varepsilon_{n,T}^{3/2}}\left(\frac{\mathbf{A}_{n,T}(1)}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})}\right)=\frac{\mathbf{A}^*_{n,T}(1)}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})}.$$

The scalar η_ϕ can be obtained as follows. Consider the process

$$\begin{aligned}B^*(r) &= \frac{1}{\varepsilon_{n,T}^{3/2}}\left(\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{[nr]}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})\left(\sigma^2(X_{i\Delta_{n,T}})-\sigma^2(x)\right)\right)\\ &= 4\phi^{3/2}\left(\frac{\sigma'(x)}{\sigma(x)}\right)\int_{-\infty}^{\infty}u\mathbf{K}(u)\frac{1}{2\sqrt{h_{n,T}}}\left(L_X(rT,x+h_{n,T}u)-L_X(rT,x)\right)du\ +o_{a.s.}(1)\end{aligned}$$

from (7.20). This quantity is distributed asymptotically [c.f. proof of Lemma 3.5] as a martingale, viz.

$$4\left(\frac{\sigma'(x)}{\sigma(x)}\right)\int_{-\infty}^{\infty}u\mathbf{K}(u)\frac{1}{2\sqrt{h_{n,T}}}\left(\int_0^{rT}\mathbf{1}_{(x\leq X_s\leq x+h_{n,T}u)}\sigma(X_s)dB_s\right)du. \qquad (7.29)$$

Now consider the process

$$
\begin{aligned}
& \mathbf{A}^{*}_{n,T}(r) \\
= & \left(\sum_{i=1}^{[nr]} \mathbf{K}\left(\frac{X_{i\Delta_{n,T}} - x}{h_{n,T}}\right) \frac{\frac{1}{h_{n,T}\varepsilon_{n,T}^{3/2}} \sum_{j=1}^{n-1} \mathbf{1}_{\{|X_{j\Delta_{n,T}} - X_{i\Delta_{n,T}}| \leq \varepsilon_{n,T}\}} \left[\int_{jT/n}^{(j+1)T/n} \left(\sigma^2(X_s) - \sigma^2(X_{iT/n})\right) ds \right]}{\sum_{j=1}^{n} \mathbf{1}_{\{|X_{j\Delta_{n,T}} - X_{i\Delta_{n,T}}| \leq \varepsilon_{n,T}\}}} \right) \\
= & \left(\frac{\sigma'(x)}{\sigma(x)}\right) \int_{-\infty}^{\infty} dc \mathbf{K}(c) \left(\frac{1}{2} \int_{-\infty}^{\infty} da \mathbf{1}_{\{|a-\phi c| \leq 1\}} \right) (a - \phi c) \frac{4}{2\sqrt{\varepsilon_{n,T}}} \times \\
& \times \left(L_X(T, x + a\varepsilon_{n,T}) - L_X(T, x) \right) + o_{a.s}(1).
\end{aligned}
$$

Again, from Lemma 3.5, the asymptotic behavior of this object is driven by the following martingale term,

$$
\left(\frac{\sigma'(x)}{\sigma(x)}\right) \int_{-\infty}^{\infty} dc \mathbf{K}(c) \left(\frac{1}{2} \int_{-\infty}^{\infty} da \mathbf{1}_{\{|a-\phi c| \leq 1\}} \right) (a - \phi c) \frac{4}{\sqrt{\varepsilon_{n,T}}} \left(\int_0^{rT} \mathbf{1}_{(x \leq X_s \leq x + \varepsilon_{n,T} a)} \sigma(X_s) dB_s \right). \tag{7.30}
$$

The covariance process between (7.29) and (7.30) can then be expressed as follows,

$$
\begin{aligned}
& [\mathbf{A}^{*}_{n,T}, B^{*}]_r \\
= & \frac{16\phi^{3/2}}{\sqrt{\varepsilon_{n,T}}\sqrt{h_{n,T}}} \left(\frac{\sigma'(x)}{\sigma(x)}\right)^2 \times \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathbf{K}(u)\mathbf{K}(c) \left(\frac{1}{2} \int_{-\infty}^{\infty} \mathbf{1}_{\{|a-\phi c| \leq 1\}} \right) \\
& \times u (a - \phi c) \left(\int_0^{rT} \mathbf{1}_{(x \leq X_s \leq x + h_{n,T} u)} \mathbf{1}_{(x \leq X_s \leq x + \varepsilon_{n,T} a)} \sigma^2(X_s) ds \right) dudcda \\
= & \frac{16\phi^{3/2}}{\sqrt{\varepsilon_{n,T}}\sqrt{h_{n,T}}} \left(\frac{\sigma'(x)}{\sigma(x)}\right)^2 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathbf{K}(u)\mathbf{K}(c) \left(\frac{1}{2} \int_{-\infty}^{\infty} \mathbf{1}_{\{|a-\phi c| \leq 1\}} \right) \\
& \times u (a - \phi c) \left(\int_{-\infty}^{\infty} \mathbf{1}_{(x \leq s \leq x + h_{n,T} u)} \mathbf{1}_{(x \leq s \leq x + \varepsilon_{n,T} a)} L_X(rT, s) ds \right) dudcda \\
= & \frac{16\phi^{3/2}}{\sqrt{\varepsilon_{n,T}}\sqrt{h_{n,T}}} \left(\frac{\sigma'(x)}{\sigma(x)}\right)^2 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathbf{K}(u)\mathbf{K}(c) \left(\frac{1}{2} \int_{-\infty}^{\infty} \mathbf{1}_{\{|a-\phi c| \leq 1\}} \right) \\
& \times u (a - \phi c) \left(\int_{-\infty}^{\infty} \mathbf{1}_{(0 \leq \frac{s-x}{h_{n,T}} \leq u)} \mathbf{1}_{(0 \leq \frac{s-x}{\varepsilon_{n,T}} \leq a)} L_X(rT, s) ds \right) dudcda \\
= & \frac{16\phi^{3/2}\sqrt{h_{n,T}}}{\sqrt{\varepsilon_{n,T}}} \left(\frac{\sigma'(x)}{\sigma(x)}\right)^2 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathbf{K}(u)\mathbf{K}(c) \left(\frac{1}{2} \int_{-\infty}^{\infty} \mathbf{1}_{\{|a-\phi c| \leq 1\}} \right) \\
& \times u (a - \phi c) \left(\int_{-\infty}^{\infty} \mathbf{1}_{(0 \leq b \leq u)} \mathbf{1}_{\left(0 \leq \frac{x + b h_{n,T} - x}{\varepsilon_{n,T}} \leq a\right)} L_X(rT, x + b h_{n,T}) db \right) dudcda
\end{aligned}
$$

$$= 16\phi^2 \left(\sigma'(x)\right)^2 \overline{L}_X(rT,x) \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathbf{K}(u)\mathbf{K}(c) \left(\frac{1}{2}\int_{-\infty}^{\infty} \mathbf{1}_{\{|a-\phi c|\le 1\}}\right) u\,(a-\phi c) \times$$
$$\times \left(\int_{-\infty}^{\infty} \mathbf{1}_{(0\le b\le u)}\mathbf{1}_{(0\le b\phi\le a)} db\right) dbdcda$$

Hence,

$$\begin{aligned}
&[\mathbf{A}^*_{n,T}, B^*]_1 \\
&= 16\phi^2 \left(\sigma'(x)\right)^2 \overline{L}_X(T,x) \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathbf{K}(u)\mathbf{K}(c) \left(\frac{1}{2}\int_{-\infty}^{\infty} \mathbf{1}_{\{|a-\phi c|\le 1\}}\right) \times \\
&\quad \times u\,(a-\phi c)\int_{-\infty}^{\infty} \mathbf{1}_{(0\le b\le u)}\mathbf{1}_{(0\le b\phi\le a)} dbdudcda
\end{aligned}$$

and

$$\eta_\phi = 4\phi^2 \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} u\,(a-\phi c)\,\mathbf{K}(u)\mathbf{K}(c)\mathbf{1}_{\{|a-\phi c|\le 1\}}\mathbf{1}_{(0\le b\le u)}\mathbf{1}_{(0\le b\phi\le a)} dbdudadc.$$

This completes the proof of the various elements of the stated result.

7.9. Proof of Theorem 5.11

As in Theorem 5.5 we start by considering

$$\frac{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^n \mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})\left(\tilde{\mu}_{n,T}(X_{i\Delta_{n,T}}) - \mu(X_{i\Delta_{n,T}})\right)}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^n \mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})} \tag{7.31}$$
$$+\frac{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^n \mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})\mu(X_{i\Delta_{n,T}})}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^n \mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})}. \tag{7.32}$$

Arguments like those in Theorem 5.5 enable us to prove that

$$\begin{aligned}
&\frac{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^n \mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})\mu(X_{i\Delta_{n,T}})}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^n \mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})} \\
&= \frac{\int_0^T \frac{1}{h_{n,T}}\mathbf{K}(\frac{X_s-x}{h_{n,T}})\mu(X_s)ds + o_{a.s.}\left(\frac{\overline{L}_X(T,x)}{h_{n,T}}(\Delta_{n,T})^{1/2-\epsilon}\right) + O_{a.s.}\left(\frac{\Delta_{n,T}}{h_{n,T}}\right)}{\int_0^T \frac{1}{h_{n,T}}\mathbf{K}(\frac{X_s-x}{h_{n,T}})ds + o_{a.s.}\left(\frac{\overline{L}_X(T,x)}{h_{n,T}}(\Delta_{n,T})^{1/2-\epsilon}\right) + O_{a.s.}\left(\frac{\Delta_{n,T}}{h_{n,T}}\right)} \\
&= \frac{\mu(x)\overline{L}_X(T,x) + o_{a.s.}(1)}{\overline{L}_X(T,x) + o_{a.s.}(1)} \overset{a.s.}{\to} \mu(x).
\end{aligned}$$

For (7.31), we just need to verify that

$$\widetilde{\mu}_{n,T}(X_{i\Delta_{n,T}}) = \mu(X_{i\Delta_{n,T}}) + o_{a.s.}(1). \tag{7.33}$$

To do so, we bound

$$\frac{1}{m_{n,T}(i\Delta_{n,T})} \sum_{j=0}^{m_{n,T}(i\Delta_{n,T})-1} \frac{[X_{t(i\Delta_{n,T})_j+\Delta_{n,T}} - X_{t(i\Delta_{n,T})_j}]}{\Delta_{n,T}} - \mu(X_{i\Delta_{n,T}})$$

using (7.18) and the Lipschitz property of μ as follows:

$$\begin{aligned}
&\frac{1}{m_{n,T}(i\Delta_{n,T})} \sum_{j=0}^{m_{n,T}(i\Delta_{n,T})-1} \frac{[X_{t(i\Delta_{n,T})_j+\Delta_{n,T}} - X_{t(i\Delta_{n,T})_j}]}{\Delta_{n,T}} - \mu(X_{i\Delta_{n,T}}) \\
= &\frac{1}{m_{n,T}(i\Delta_{n,T})\Delta_{n,T}} \sum_{j=0}^{m_{n,T}(i\Delta_{n,T})-1} \int_{t(i\Delta_{n,T})_j}^{t(i\Delta_{n,T})_j+\Delta_{n,T}} (\mu(X_s) - \mu(X_{i\Delta_{n,T}}))ds \\
&+ \frac{1}{m_{n,T}(i\Delta_{n,T})\Delta_{n,T}} \sum_{j=0}^{m_{n,T}(i\Delta_{n,T})-1} \int_{t(i\Delta_{n,T})_j}^{t(i\Delta_{n,T})_j+\Delta_{n,T}} \sigma(X_s)dB_s \\
= &\; O_{a.s.}(\kappa_{n,T}) + \frac{1}{m_{n,T}(i\Delta_{n,T})\Delta_{n,T}} \sum_{j=0}^{m_{n,T}(i\Delta_{n,T})-1} \int_{t(i\Delta_{n,T})_j}^{t(i\Delta_{n,T})_j+\Delta_{n,T}} \sigma(X_s)dB_s.
\end{aligned}$$

where $\kappa_{n,T}$ has its usual definition. From the proof of Theorem 5.5, we have

$$\frac{1}{m_{n,T}(i\Delta_{n,T})\Delta_{n,T}} \sum_{j=0}^{m_{n,T}(i\Delta_{n,T})-1} \int_{t(i\Delta_{n,T})_j}^{t(i\Delta_{n,T})_j+\Delta_{n,T}} \sigma(X_s)dB_s = O_{a.s.}\left(\frac{1}{\sqrt{\varepsilon_{n,T}\overline{L}_X(T, X_{i\Delta_{n,T}})}}\right).$$

But $\sqrt{\varepsilon_{n,T}\overline{L}_X(T, X_{i\Delta_{n,T}})} \overset{a.s.}{\to} \infty$ as $n, T \to \infty$ since we control $\varepsilon_{n,T}$ to ensure that this property holds. Hence, the bound vanishes in the limit, giving (7.33), and the stated result follows.

7.10. Proof of Theorem 5.12

Write the estimation error as

$$\begin{aligned}
&\frac{\frac{\Delta_{n,T}}{h_{n,T}} \sum_{i=1}^{n} \mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})\widetilde{\mu}_{n,T}(X_{i\Delta_{n,T}})}{\frac{\Delta_{n,T}}{h_{n,T}} \sum_{i=1}^{n} \mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})} - \mu(x) \\
= &\underbrace{\frac{\frac{\Delta_{n,T}}{h_{n,T}} \sum_{i=1}^{n} \mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})\widetilde{\mu}_{n,T}(X_{i\Delta_{n,T}})}{\frac{\Delta_{n,T}}{h_{n,T}} \sum_{i=1}^{n} \mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})} - \frac{\frac{\Delta_{n,T}}{h_{n,T}} \sum_{i=1}^{n} \mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})\mu(X_{i\Delta_{n,T}})}{\frac{\Delta_{n,T}}{h_{n,T}} \sum_{i=1}^{n} \mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})}}_{\text{term } V}
\end{aligned}$$

$$+\underbrace{\frac{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})\mu(X_{i\Delta_{n,T}})}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})}-\frac{\mu(x)\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})}}_{\text{term } B}$$

$$= \text{term } V + \text{term } B.$$

Consider term V first, viz.

$$\frac{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})\widetilde{\mu}_{n,T}(X_{i\Delta_{n,T}})}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})}-\frac{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})\mu(X_{i\Delta_{n,T}})}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})}$$

$$= \frac{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})\left(\widetilde{\mu}_{n,T}(X_{i\Delta_{n,T}})-\mu(X_{i\Delta_{n,T}})\right)}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})}.$$

The numerator can be written as

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

$$= \frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})\frac{\sum_{j=1}^{n-1}\mathbf{1}_{\{|X_{j\Delta_{n,T}}-X_{i\Delta_{n,T}}|\le\epsilon_{n,T}\}}\frac{n}{T}\left[\left(X_{(j+1)T/n}-X_{jT/n}\right)-\mu(X_{iT/n})\Delta_{n,T}\right]}{\sum_{j=1}^{n}\mathbf{1}_{\{|X_{j\Delta_{n,T}}-X_{i\Delta_{n,T}}|\le\epsilon_{n,T}\}}}.$$

Notice that

$$X_{(j+1)T/n}-X_{jT/n}=\int_{jT/n}^{(j+1)T/n}\mu(X_s)ds+\int_{jT/n}^{(j+1)T/n}\sigma(X_s)dB_s.$$

Hence,

$$\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})\times$$
$$\times\frac{\sum_{j=1}^{n-1}\mathbf{1}_{\{|X_{j\Delta_{n,T}}-X_{i\Delta_{n,T}}|\le\epsilon_{n,T}\}}\frac{n}{T}\left[\left(X_{(j+1)T/n}-X_{jT/n}\right)-\mu(X_{iT/n})\Delta_{n,T}\right]}{\sum_{j=1}^{n}\mathbf{1}_{\{|X_{j\Delta_{n,T}}-X_{i\Delta_{n,T}}|\le\epsilon_{n,T}\}}}$$

$$= \frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})\times$$
$$\times\frac{\sum_{j=1}^{n-1}\mathbf{1}_{\{|X_{j\Delta_{n,T}}-X_{i\Delta_{n,T}}|\le\epsilon_{n,T}\}}\frac{n}{T}\left[\int_{jT/n}^{(j+1)T/n}\left(\mu(X_s)-\mu(X_{iT/n})\right)ds+\int_{jT/n}^{(j+1)T/n}\sigma(X_s)dB_s\right]}{\sum_{j=1}^{n}\mathbf{1}_{\{|X_{j\Delta_{n,T}}-X_{i\Delta_{n,T}}|\le\epsilon_{n,T}\}}}$$

$$= \underbrace{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})\frac{\sum_{j=1}^{n-1}\mathbf{1}_{\{|X_{j\Delta_{n,T}}-X_{i\Delta_{n,T}}|\le\epsilon_{n,T}\}}\frac{n}{T}\left[\int_{jT/n}^{(j+1)T/n}\left(\mu(X_s)-\mu(X_{iT/n})\right)ds\right]}{\sum_{j=1}^{n}\mathbf{1}_{\{|X_{j\Delta_{n,T}}-X_{i\Delta_{n,T}}|\le\epsilon_{n,T}\}}}}_{(\mathbf{A}_{n,T})}$$

$$+\underbrace{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})\frac{\sum_{j=1}^{n-1}\mathbf{1}_{\{|X_{j\Delta_{n,T}}-X_{i\Delta_{n,T}}|\leq\varepsilon_{n,T}\}}\frac{n}{T}\left[\int_{jT/n}^{(j+1)T/n}\sigma(X_s)dB_s\right]}{\sum_{j=1}^{n}\mathbf{1}_{\{|X_{j\Delta_{n,T}}-X_{i\Delta_{n,T}}|\leq\varepsilon_{n,T}\}}}}_{(\mathbf{B}_{n,T}(1))}.$$

First examine $\sqrt{\varepsilon_{n,T}}\mathbf{B}_{n,T}(r)$, viz.

$$\sqrt{\varepsilon_{n,T}}\mathbf{B}_{n,T}(r)=\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{[nr]}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})\frac{\frac{1}{2\sqrt{\varepsilon_{n,T}}}\sum_{j=1}^{n-1}\mathbf{1}_{\{|X_{j\Delta_{n,T}}-X_{i\Delta_{n,T}}|\leq\varepsilon_{n,T}\}}\left[\int_{jT/n}^{(j+1)T/n}\sigma(X_s)dB_s\right]}{\frac{\Delta_{n,T}}{2\varepsilon_{n,T}}\sum_{j=1}^{n}\mathbf{1}_{\{|X_{j\Delta_{n,T}}-X_{i\Delta_{n,T}}|\leq\varepsilon_{n,T}\}}}.$$

The martingale $\mathbf{B}_{n,T}(r)$ has quadratic variation which can be analysed as in the proof of Theorem 5.6 as follows:

$$\begin{aligned}
&[\mathbf{B}_{n,T}]_r\\
=&\left(\frac{\Delta_{n,T}}{h_{n,T}}\right)^2\sum_{i=1}^{[nr]}\sum_{k=1}^{[nr]}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})\mathbf{K}(\frac{X_{k\Delta_{n,T}}-x}{h_{n,T}})\times\\
&\times\frac{\frac{1}{4}\left(\frac{1}{\sqrt{\varepsilon_{n,T}}}\right)^2\sum_{j=1}^{n-1}\mathbf{1}_{\{|X_{j\Delta_{n,T}}-X_{i\Delta_{n,T}}|\leq\varepsilon_{n,T}\}}\mathbf{1}_{\{|X_{j\Delta_{n,T}}-X_{k\Delta_{n,T}}|\leq\varepsilon_{n,T}\}}\left[\int_{jT/n}^{(j+1)T/n}\sigma^2(X_s)ds\right]}{\left(\frac{\Delta_{n,T}}{2\varepsilon_{n,T}}\sum_{j=1}^{n}\mathbf{1}_{\{|X_{j\Delta_{n,T}}-X_{i\Delta_{n,T}}|\leq\varepsilon_{n,T}\}}\right)\left(\frac{\Delta_{n,T}}{2\varepsilon_{n,T}}\sum_{j=1}^{n}\mathbf{1}_{\{|X_{j\Delta_{n,T}}-X_{k\Delta_{n,T}}|\leq\varepsilon_{n,T}\}}\right)}\\
=&\left(\frac{1}{h_{n,T}}\right)^2\int_0^{[Tr]}ds\int_0^{[Tr]}du\mathbf{K}(\frac{X_s-x}{h_{n,T}})\mathbf{K}(\frac{X_u-x}{h_{n,T}})\times\\
&\times\frac{\frac{1}{4\varepsilon_{n,T}}\int_0^T db\mathbf{1}_{\{|X_b-X_s|\leq\varepsilon_{n,T}\}}\mathbf{1}_{\{|X_b-X_u|\leq\varepsilon_{n,T}\}}\sigma^2(X_b+o_{a.s.}(1))}{\left(\frac{1}{2\varepsilon_{n,T}}\int_0^T\mathbf{1}_{\{|b-s|\leq\varepsilon_{n,T}\}}db\right)\left(\frac{1}{2\varepsilon_{n,T}}\int_0^T\mathbf{1}_{\{|b-u|\leq\varepsilon_{n,T}\}}db\right)}+o_{a.s}(1)\\
=&\left(\frac{1}{h_{n,T}}\right)^2\int_{-\infty}^{\infty}\int_{-\infty}^{+\infty}dsdu\mathbf{K}(\frac{s-x}{h_{n,T}})\mathbf{K}(\frac{u-x}{h_{n,T}})\times\\
&\times\frac{\frac{1}{4\varepsilon_{n,T}}\int_{-\infty}^{\infty}db\mathbf{1}_{\{|b-s|\leq\varepsilon_{n,T}\}}\mathbf{1}_{\{|b-u|\leq\varepsilon_{n,T}\}}\sigma^2(b)\overline{L}(T,b)\overline{L}(rT,s)\overline{L}(rT,u)}{\left(\frac{1}{2\varepsilon_{n,T}}\int_{-\infty}^{\infty}\mathbf{1}_{\{|b-s|\leq\varepsilon_{n,T}\}}\overline{L}(T,b)db\right)\left(\frac{1}{2\varepsilon_{n,T}}\int_{-\infty}^{\infty}\mathbf{1}_{\{|b-u|\leq\varepsilon_{n,T}\}}\overline{L}(T,b)db\right)}dsdudb\\
&+o_{a.s}(1).
\end{aligned}$$

Let

$$\frac{s-x}{h_{n,T}}=a\text{ and }\frac{u-x}{h_{n,T}}=e.$$

Then,

$$\begin{aligned}
&\int_{-\infty}^{\infty}da\int_{-\infty}^{+\infty}de\mathbf{K}(a)\mathbf{K}(e)\times\\
&\times\frac{\frac{1}{4\varepsilon_{n,T}}\int_{-\infty}^{\infty}db\mathbf{1}_{\{|b-x-ah_{n,T}|\leq\varepsilon_{n,T}\}}\mathbf{1}_{\{|b-x-eh_{n,T}|\leq\varepsilon_{n,T}\}}L_X(T,b)\overline{L}_X(rT,x+ah_{n,T})\overline{L}_X(rT,x+eh_{n,T})}{\left(\frac{1}{2\varepsilon_{n,T}}\int_{-\infty}^{\infty}\mathbf{1}_{\{|b-x-ah_{n,T}|\leq\varepsilon_{n,T}\}}\overline{L}_X(T,b)db\right)\left(\frac{1}{2\varepsilon_{n,T}}\int_{-\infty}^{\infty}\mathbf{1}_{\{|b-x-eh_{n,T}|\leq\varepsilon_{n,T}\}}\overline{L}_X(T,b)db\right)}
\end{aligned}$$

$$+o_{a.s}(1)$$

$$= \int_{-\infty}^{\infty} da \int_{-\infty}^{+\infty} de \mathbf{K}(a)\mathbf{K}(e) \times$$

$$\times \frac{\frac{1}{4\varepsilon_{n,T}} \int_{-\infty}^{\infty} db \mathbf{1}_{\{|\frac{b-x}{\varepsilon_{n,T}} - a\frac{h_{n,T}}{\varepsilon_{n,T}}| \leq 1\}} \mathbf{1}_{\{|\frac{b-x}{\varepsilon_{n,T}} - e\frac{h_{n,T}}{\varepsilon_{n,T}}| \leq 1\}} L_X(T,b)\overline{L}_X(rT, x + ah_{n,T})\overline{L}_X(rT, x + eh_{n,T})}{\left(\frac{1}{2\varepsilon_{n,T}} \int_{-\infty}^{\infty} \mathbf{1}_{\{|\frac{b-x}{\varepsilon_{n,T}} - a\frac{h_{n,T}}{\varepsilon_{n,T}}| \leq 1\}} \overline{L}_X(T,b)db\right)\left(\frac{1}{2\varepsilon_{n,T}} \int_{-\infty}^{\infty} \mathbf{1}_{\{|\frac{b-x}{\varepsilon_{n,T}} - e\frac{h_{n,T}}{\varepsilon_{n,T}}| \leq 1\}} \overline{L}_X(T,b)db\right)}$$

$$+o_{a.s}(1)$$

Setting

$$\frac{b-x}{\varepsilon_{n,T}} = z,$$

this last expression becomes

$$\int_{-\infty}^{\infty} da \int_{-\infty}^{+\infty} de \mathbf{K}(a)\mathbf{K}(e) \times$$

$$\times \frac{\frac{1}{4} \int_{-\infty}^{\infty} dz \mathbf{1}_{\{|z - a\frac{h_{n,T}}{\varepsilon_{n,T}}| \leq 1\}} \mathbf{1}_{\{|z - e\frac{h_{n,T}}{\varepsilon_{n,T}}| \leq 1\}} \sigma^2(x)\overline{L}_X(T, x + z\varepsilon_{n,T})\overline{L}_X(rT, x + ah_{n,T})\overline{L}_X(rT, x + eh_{n,T})}{\left(\frac{1}{2} \int_{-\infty}^{\infty} \mathbf{1}_{\{|z - a\frac{h_{n,T}}{\varepsilon_{n,T}}| \leq 1\}} \overline{L}_X(T, x + z\varepsilon_{n,T})dz\right)\left(\frac{1}{2} \int_{-\infty}^{\infty} \mathbf{1}_{\{|z - e\frac{h_{n,T}}{\varepsilon_{n,T}}| \leq 1\}} \overline{L}_X(T, x + z\varepsilon_{n,T})dz\right)}$$

$$+o_{a.s}(1).$$

Now, if $h_{n,T} = o(\varepsilon_{n,T})$, then

$$[\mathbf{B}_{n,T}]_r \xrightarrow{a.s.} \frac{1}{2}\sigma^2(x)\frac{\left(\overline{L}_X(rT,x)\right)^2}{\overline{L}_X(T,x)},$$

whereas if $h_{n,T} = O(\varepsilon_{n,T})$ with $h_{n,T}/\varepsilon_{n,T} \to \phi > 0$, then

$$[\mathbf{B}_{n,T}]_r \xrightarrow{a.s.} \frac{1}{2}\theta_\phi\sigma^2(x)\frac{\left(\overline{L}_X(rT,x)\right)^2}{\overline{L}_X(T,x)}.$$

where $\theta_\phi = \frac{1}{2}\int_{-\infty}^{\infty}\int_{(z-1)/\phi}^{(z+1)/\phi}\int_{(z-1)/\phi}^{(z+1)/\phi} \mathbf{K}(a)\mathbf{K}(e)dzdade$ again. Using standard arguments, as in Theorem 5.6, this implies that

$$\sqrt{\varepsilon_{n,T}}\left(\frac{\mathbf{B}_{n,T}(1)}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})}\right) \xrightarrow{d} MN\left(0, \frac{1}{2}\frac{\sigma^2(x)}{\overline{L}_X(T,x)}\right)$$

and

$$\sqrt{\varepsilon_{n,T}\overline{L}_X(T,x)}\left(\frac{\mathbf{B}_{n,T}(1)}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})}\right) \xrightarrow{d} N\left(0, \frac{1}{2}\sigma^2(x)\right)$$

provided $h_{n,T} = o(\varepsilon_{n,T})$. If $h_{n,T} = O(\varepsilon_{n,T})$ and $h_{n,T}/\varepsilon_{n,T} \to \phi > 0$, then

$$\sqrt{\varepsilon_{n,T}\overline{L}_X(T,x)}\left(\frac{\mathbf{B}_{n,T}(1)}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})}\right) \xrightarrow{d} N\left(0, \frac{1}{2}\theta_\phi\sigma^2(x)\right).$$

Next examine $\mathbf{A}_{n,T}$. We have

$$\begin{aligned}
&\mathbf{A}_{n,T} \\
= \;& \frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})\frac{\sum_{j=1}^{n-1}\mathbf{1}_{\{|X_{j\Delta_{n,T}}-X_{i\Delta_{n,T}}|\le\varepsilon_{n,T}\}}\frac{n}{T}\left[\int_{jT/n}^{(j+1)T/n}\left(\mu(X_s)-\mu(X_{iT/n})\right)ds\right]}{\sum_{j=1}^{n}\mathbf{1}_{\{|X_{j\Delta_{n,T}}-X_{i\Delta_{n,T}}|\le\varepsilon_{n,T}\}}} \\
= \;& \frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})\frac{\sum_{j=1}^{n-1}\mathbf{1}_{\{|X_{j\Delta_{n,T}}-X_{i\Delta_{n,T}}|\le\varepsilon_{n,T}\}}\frac{n}{T}\left[\int_{jT/n}^{(j+1)T/n}\left(\mu(X_s)-\mu(X_{jT/n})\right)ds\right]}{\sum_{j=1}^{n}\mathbf{1}_{\{|X_{j\Delta_{n,T}}-X_{i\Delta_{n,T}}|\le\varepsilon_{n,T}\}}} \\
& +\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})\frac{\frac{\Delta_{n,T}}{2\varepsilon_{n,T}}\sum_{j=1}^{n-1}\mathbf{1}_{\{|X_{j\Delta_{n,T}}-X_{i\Delta_{n,T}}|\le\varepsilon_{n,T}\}}\left(\mu(X_{jT/n})-\mu(X_{iT/n})\right)}{\frac{\Delta_{n,T}}{2\varepsilon_{n,T}}\sum_{j=1}^{n}\mathbf{1}_{\{|X_{j\Delta_{n,T}}-X_{i\Delta_{n,T}}|\le\varepsilon_{n,T}\}}} \\
= \;& O_{a.s}(1)\left(\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})\right)\kappa_{n,T} \\
& +\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})\frac{\frac{\Delta_{n,T}}{2\varepsilon_{n,T}}\sum_{j=1}^{n-1}\mathbf{1}_{\{|X_{j\Delta_{n,T}}-X_{i\Delta_{n,T}}|\le\varepsilon_{n,T}\}}\mu'(x_{ij}^*)\left(X_{jT/n}-X_{iT/n}\right)}{\frac{\Delta_{n,T}}{2\varepsilon_{n,T}}\sum_{j=1}^{n}\mathbf{1}_{\{|X_{j\Delta_{n,T}}-X_{i\Delta_{n,T}}|\le\varepsilon_{n,T}\}}}.
\end{aligned}$$

where $x_{ij}^* = f\left(X_{jT/n}, X_{iT/n}\right) \in [X_{jT/n}, X_{iT/n}]$. Then,

$$\begin{aligned}
&\mathbf{A}_{n,T} \\
= \;& \frac{1}{h_{n,T}}\int_0^T ds\mathbf{K}(\frac{X_s-x}{h_{n,T}})\frac{\frac{1}{2\varepsilon_{n,T}}\int_0^T du\mathbf{1}_{\{|X_u-X_s|\le\varepsilon_{n,T}\}}\mu'\left(f(X_u,X_s)\right)(X_u-X_s)}{\frac{1}{2\varepsilon_{n,T}}\int_0^T\mathbf{1}_{\{|X_u-X_s|\le\varepsilon_{n,T}\}}du}+o_{a.s}(1) \\
= \;& \frac{1}{h_{n,T}}\int_{-\infty}^{\infty} ds\mathbf{K}(\frac{s-x}{h_{n,T}})\frac{\frac{1}{2\varepsilon_{n,T}}\int_{-\infty}^{\infty} du\mathbf{1}_{\{|u-s|\le\varepsilon_{n,T}\}}\mu'\left(f(u,s)\right)(u-s)}{\frac{1}{2\varepsilon_{n,T}}\int_{-\infty}^{\infty}\mathbf{1}_{\{|u-s|\le\varepsilon_{n,T}\}}\overline{L}_X(T,u)du}\overline{L}_X(T,u)\overline{L}_X(T,s) \\
& +o_{a.s}(1).
\end{aligned}$$

Let

$$\frac{s-x}{h_{n,T}} = c \text{ and } \frac{u-x}{\varepsilon_{n,T}} = a.$$

Then

$$\frac{1}{\varepsilon_{n,T}}\mathbf{A}_{n,T}$$

$$= \frac{1}{\varepsilon_{n,T}}\int_{-\infty}^{\infty} dc\mathbf{K}(c)\frac{\frac{1}{2\varepsilon_{n,T}}\int_{-\infty}^{\infty} du\mathbf{1}_{\{|u-x-ch_{n,T}|\leq\varepsilon_{n,T}\}}\mu'\left(f(u,x+ch_{n,T})\right)(u-x-ch_{n,T})}{\frac{1}{2\varepsilon_{n,T}}\int_{-\infty}^{\infty}\mathbf{1}_{\{|u-x-ch_{n,T}|\leq\varepsilon_{n,T}\}}\overline{L}_X(T,u)du}\times$$

$$\times\overline{L}_X(T,u)\overline{L}_X(T,x+ch_{n,T})$$

$$= \int_{-\infty}^{\infty} dc\mathbf{K}(c)\frac{\frac{1}{2}\int_{-\infty}^{\infty} da\mathbf{1}_{\{|a-c\frac{h_{n,T}}{\varepsilon_{n,T}}|\leq 1\}}\mu'\left(f(x+a\varepsilon_{n,T},x+ch_{n,T})\right)\left(a-c\frac{h_{n,T}}{\varepsilon_{n,T}}\right)}{\frac{1}{2}\int_{-\infty}^{\infty}\mathbf{1}_{\{|a-c\frac{h_{n,T}}{\varepsilon_{n,T}}|\leq 1\}}\overline{L}_X(T,x+a\varepsilon_{n,T})da}\times$$

$$\times\overline{L}_X(T,x+a\varepsilon_{n,T})\overline{L}_X(T,x+ch_{n,T})+o_{a.s}(1)$$

$$= \int_{-\infty}^{\infty} dc\mathbf{K}(c)\frac{\frac{1}{2}\int_{-\infty}^{\infty} da\mathbf{1}_{\{|a-c\frac{h_{n,T}}{\varepsilon_{n,T}}|\leq 1\}}\mu'(x)\left(a-c\frac{h_{n,T}}{\varepsilon_{n,T}}\right)}{\overline{L}_X(T,x)+o_{a.s.}(1)}\left(\overline{L}_X(T,x+a\varepsilon_{n,T})-\overline{L}_X(T,x)\right)\overline{L}_X(T,x)$$

$$+o_{a.s}(1).$$

Hence, if $h_{n,T}=o(\varepsilon_{n,T})$

$$\frac{1}{\varepsilon_{n,T}^{3/2}}\mathbf{A}_{n,T}\xrightarrow{d} N\left(0,\varphi^*\left(\frac{\mu'(x)}{\sigma(x)}\right)^2\overline{L}_X(T,x)\right)$$

where $\varphi^*=-2\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}|a-b|ab\left(\frac{1}{2}\mathbf{1}_{\{|a|\leq 1\}}\right)\left(\frac{1}{2}\mathbf{1}_{\{|b|\leq 1\}}\right)dadb=0.2666$. Also,

$$\frac{\sqrt{\overline{L}_X(T,x)}}{\varepsilon_{n,T}^{3/2}}\left(\frac{\mathbf{A}_{n,T}}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})}\right)\xrightarrow{d} N\left(0,\varphi^*\left(\frac{\mu'(x)}{\sigma(x)}\right)^2\right).$$

Analogously, if $h_{n,T}=O(\varepsilon_{n,T})$ with $h_{n,T}/\varepsilon_{n,T}\to\phi>0$, then

$$\frac{\sqrt{\overline{L}_X(T,x)}}{\varepsilon_{n,T}^{3/2}}\left(\frac{\mathbf{A}_{n,T}}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})}\right)\xrightarrow{d} N\left(0,\vartheta_\phi\left(\frac{\mu'(x)}{\sigma(x)}\right)^2\right)$$

where

$$\vartheta_\phi=\frac{1}{8}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{\infty}^{\infty}\mathbf{K}(s)\mathbf{K}(c)(a-\phi s)\mathbf{1}_{\{|a-\phi s|\leq 1\}}(b-\phi s)\mathbf{1}_{\{|b-\phi s|\leq 1\}}(a-\phi c)\times$$
$$\times\mathbf{1}_{\{|a-\phi c|\leq 1\}}(b-\phi c)\mathbf{1}_{\{|b-\phi c|\leq 1\}}\mathbf{1}_{(0\leq v\leq a)}\mathbf{1}_{(0\leq v\leq b)}dadbdcdsdv$$

Now, we analyze term B. From previous results it is easy to prove that

$$\frac{\sqrt{\overline{L}_X(T,x)}}{h_{n,T}^{3/2}}(B)$$

$$=\frac{\sqrt{\overline{L}_X(T,x)}}{h_{n,T}^{3/2}}\left(\frac{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})\mu(X_{i\Delta_{n,T}})}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})}-\frac{\mu(x)\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})}\right)$$

$$=\frac{\sqrt{\overline{L}_X(T,x)}}{h_{n,T}^{3/2}}\left(\frac{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})\left(\mu(X_{i\Delta_{n,T}})-\mu(x)\right)}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})}\right)$$

$$\xrightarrow{d} N\left(0,\varphi\left(\frac{\mu'(x)}{\sigma(x)}\right)^2\right)$$

where $\varphi = -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} 2|a-b|ab\mathbf{K}(a)\mathbf{K}(b)dadb$. In consequence, defining the estimation error as

$$E = B + \frac{\mathbf{A}_{n,T}}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})} + \frac{\mathbf{B}_{n,T}(1)}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})}$$

we can write

$$\begin{aligned}
&\sqrt{\varepsilon_{n,T}\overline{L}_X(T,x)}\left(\frac{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})\widetilde{\mu}_{n,T}(X_{i\Delta_{n,T}})}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})} - \mu(x)\right) \\
&= \sqrt{\varepsilon_{n,T}\overline{L}_X(T,x)}(O_p\left(\frac{h_{n,T}^{3/2}}{\sqrt{\overline{L}_X(T,x)}}\right) + O_p\left(\frac{\varepsilon_{n,T}^{3/2}}{\sqrt{\overline{L}_X(T,x)}}\right) + \frac{\mathbf{B}_{n,T}(1)}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})} \\
&\quad + O_{a.s.}\left(\Delta_{n,T}\right)^{1/2-\delta}) \\
&= O_p\left(\varepsilon_{n,T}h_{n,T}^{3/2}\right) + O_p\left(\varepsilon_{n,T}^{2}\right) + \sqrt{\varepsilon_{n,T}\overline{L}_X(T,x)}\left(\frac{\mathbf{B}_{n,T}(1)}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})}\right) \\
&\quad + O_{a.s.}\left(\Delta_{n,T}\right)^{1/2-\delta}\sqrt{\varepsilon_{n,T}}\left(\overline{L}_X(T,x)\right)^{1/2} \\
&\xrightarrow{d} N\left(0, \frac{1}{2}\sigma^2(x)\right),
\end{aligned}$$

if $h_{n,T} = o(\varepsilon_{n,T})$. If $h_{n,T} = O(\varepsilon_{n,T})$ with $h_{n,T}/\varepsilon_{n,T} \to \phi > 0$, then

$$\begin{aligned}
&\sqrt{\varepsilon_{n,T}\overline{L}_X(T,x)}\left(\frac{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})\widetilde{\mu}_{n,T}(X_{i\Delta_{n,T}})}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})} - \mu(x)\right) \\
&\xrightarrow{d} N\left(0, \frac{1}{2}\theta_\phi\sigma^2(x)\right)
\end{aligned}$$

where $\theta_\phi = \frac{1}{2}\int_{-\infty}^{\infty}\int_{(z-1)/\phi}^{(z+1)/\phi}\int_{(z-1)/\phi}^{(z+1)/\phi}\mathbf{K}(a)\mathbf{K}(e)dzdade$. This proves the stated result.

8. Notation

$\rightarrow_{a.s.}$	almost sure convergence
$\rightarrow_p$	convergence in probability
$\Rightarrow, \rightarrow_d$	weak convergence
$:=$	definitional equality
$o_p(1)$	tends to zero in probability
$O_p(1)$	bounded in probability
$o_{a.s.}(1)$	tends to zero almost surely
$O_{a.s.}(1)$	bounded almost surely
$=_d$	distributional equivalence
$\sim_d$	asymptotically distributed as
$\mathrm{MN}(0, V)$	mixed normal distribution with variance V
$\mathbf{1}_A$	indicator function for the set A
$a \vee b$	$\max\{a, b\}$
$C_k, \quad k = 1, 2, \ldots$	constants

Part II
Short-Term Interest Rate Dynamics: a Spatial Approach

9. Introduction

A large body of recent literature has been devoted to the estimation of the short-term interest rate process. There are several reasons to take an interest in this issue. First, it is particularly relevant given the role of the short-term interest rate as a key economic variable linking real and monetary phenomena. Second, interest rate model specification (often in continuous-time) has implications for the pricing of fixed-income securities and derivatives. Finally, interest rate levels constitute a traditional benchmark to evaluate asset pricing, due to the fact that expected equilibrium returns are defined in terms of excess returns relative to the risk free rate.

In continuous-time finance, the dynamics of the spot interest rate process is usually modelled as a Markov stochastic differential equation. Stochastic differential equations are completely described by two functions, the drift and the diffusion function. Parametric approaches to the estimation of these two functions have yielded contradictory results. Aït-Sahalia (1996b), for example, suggests a semiparametric procedure to discriminate among alternative parametric specifications. He rejects every conventional one-factor model of the short rate but some recent evidence shows that his procedure has distorted size and low power in finite samples [Pritsker (1998)]. Fully nonparametric methods have been developed but they either rely on the existence of a time-invariant marginal density for the underlying process [Jiang and Knight (1997a), Jiang (1998)] or stationarity is assumed despite robustness to deviations from it [Stanton (1997)].

In this chapter, we implement a unified approach to the estimation of the drift and the diffusion function of the short-term interest rate process based on the estimation procedure proposed in Part I. As discussed earlier, we use functional methods. Minimal requirements are placed on the data generating mechanism allowing for both stationary and nonstationary systems. Cross-restrictions on the functional forms of the drift and diffusion function [as in Aït-Sahalia (1996a,b), Jiang and Knight (1997a), Jiang (1998)] are not imposed, nor is the existence of a time-invariant marginal data density either required or assumed [Stanton (1997)]. In consequence, the approach is robust against deviations from stationarity. The available data is taken to be a set of discrete sample observations. Econometric estimation proceeds by constructing refined sample analogues of unknown drift and diffusion functions.

The proposed methodology has several important features. First, as mentioned earlier, despite a flurry of theoretical contributions [c.f. Duffie (1992), for example], empirical results do not offer complete support for any specific parametrization. Given the importance of the short-term riskless rate in valuing and hedging a broad array of fixed-income contingent claims, fully nonparametric methods are particularly suitable to avoid potential misspecifications.

Second, even though the drift is theoretically harder to identify than the diffusion term [c.f. Aït-Sahalia (1996a) and Jiang and Knight (1997a) *inter alia*, and Part I], a unified and complete asymptotic theory for both estimated functions is crucial in fixed-income pricing. In effect, the drift of the underlying short rate process plays a role in assessing the value of fixed-income securities even under the no-arbitrage restrictions imposed by martingale pricing [c.f. Ingersoll (1987)].

Third, the evidence on the stationarity of the short rate process is quite ambiguous. Preliminary unit root tests either accept the null of nonstationarity or deliver results very close to the rejection threshold. This observation explains why high-frequency spot interest rate series are often modelled as nonstationary processes in macroeconomics [c.f. Aït-Sahalia (1996b)]. In continuous-time empirical finance, stationarity is generally assumed upfront to assist in developing a complete estimation theory. Some researchers have provided plausible ex-post justifications for this assumption based on estimated drift and diffusion functions. Aït-Sahalia (1996b) suggests that the spot rate can be locally nonstationary over the range of the process corresponding to a drift very close to zero. Nevertheless, a nonlinear mean-reverting drift at the edges of the range of the process can be sufficient to pull the series back into its middle region and determine global stationarity. Conley, Hansen, Luttmer and Scheinkman (1997) [CHLS, hereafter] suggest volatility-induced stationarity. Mean reversion at high rates can be small but increasing volatility is sufficient to import stationarity into the series.

Due to the mixed a-priori empirical evidence, estimation methods relying on stationarity can yield imprecise inference and suggest misleading conclusions. In consequence, we do not make the assumption of stationarity in this work.

We are interested in estimating the drift and the diffusion function at each point in the range of the sample interest rate process, so the density of the observations there plays a role in the operation of the asymptotics. This information is contained in the estimated *local time* of the spot rate process [c.f. Part I]. We review the definition of local time here, and some useful observations related to this crucial concept [classical references are Revuz and Yor (1994), Karatzas and Shreve (1988), Chung and Williams (1990)].

9.1 Definition *If X_t is a continuous semimartingale (SMG), then there exists a nondecreasing stochastic process (nondecreasing in t, that is) $L_X(t,a)$, called the local time of X at a. This process is defined, almost surely, as*

$$L_X(t,a) = \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \int_0^t 1_{[a,a+\varepsilon[}(X_s) d[X]_s. \tag{9.1}$$

$L_X(t,a)$ represents the amount of time that the process X spends in the vicinity of the point a. It is measured in units of the quadratic variation process ($[X]_t$). These are information units as they represent the amount of information that is being accumulated about the process. Consider, for simplicity, a Brownian motion B_t with variance σ^2. $L_X(t,a)$ reduces to

$$L_B(t,a) = \lim_{\varepsilon \to 0} \frac{1}{2\varepsilon} \int_0^t 1_{(|B_s - a| < \varepsilon)} \sigma^2 ds. \tag{9.2}$$

In this case we have integration with respect to the Lebesgue measure since the quadratic variation of Brownian motion is deterministic. If we now divide through by σ^2, we obtain

$$\overline{L}_B(t,a) = \frac{1}{\sigma^2} L_B(t,a) = \lim_{\varepsilon \to 0} \frac{1}{2\varepsilon} \int_0^t 1_{(|B_s - a| < \varepsilon)} ds. \tag{9.3}$$

$\overline{L}_B(t,a)$ can be called "chronological local time" [this notion was first introduced by Phillips and Park (1997)]. This formulation gives an interpretation of the local time in terms of amount of time, in real time units now, spent by the process in the spatial neighborhood of a point. Also, this definition shows the sense in which the local time, even though random in nature, is analogous to a probability density. In fact, it provides meaningful quantitative information about the locational features of the process, in just the same way as a probability density distribution can be used to characterize stationary time series.

Since the solution to a stochastic differential equation is a SMG, we can define the local time process of the short-term interest rate series in the usual fashion. Below, we interpret the local time process as a series of *spatial* densities, along lines pioneered in Phillips (1998).

Also, we show how to consistently estimate spatial densities using nonparametric density-like kernel estimators [c.f. Part I]. Our inference is based on a complete asymptotic theory for spatial densities of diffusion processes. The diffusion processes that we consider are potentially nonstationary solutions to possibly nonlinear stochastic differential equations.

The notion of spatial density assumes importance particularly when the underlying process is nonstationary as it furnishes the possibility of characterizing some of the features of the data, i.e. those related to the location of the process. In effect, in the presence of nonstationarity, conventional descriptive statistics fail to provide reliable information given the tendency of the data to drift away from a particular point. Spatial densities can then be regarded as a new descriptive tool for series that are nonstationary or whose stationarity can not be guaranteed [these observations were first made by Peter Phillips during the Irving Fisher Conference at Yale University, May 1998].

Based on estimated spatial densities, we discuss some of the features of the specific data set at hand. We study the annualized 7-day Eurodollar rate [June 1, 1973 - February 25, 1995]. This data was previously used in Aït-Sahalia (1996a,b).

Further, we define functionals of spatial densities, such as spatial hazard rates [c.f. Phillips (1998)]. Spatial hazard rates can be interpreted as spatial analogues to traditional hazard rates obtained from time-invariant marginal distributions. Again, a complete asymptotic theory for nonparametric estimates of spatial hazard rates assists our inference.

Finally, we carefully discuss the sense in which the information embodied in the spatial density of the interest rate process can be used to implement a flexible and rigorous approach to the nonparametric estimation of the two functions driving the interest rate dynamics in continuous-time, viz. the drift and the diffusion function.

As in many recent papers [Aït-Sahalia (1996a,b), Stanton (1997), Jiang (1998), for example], the main source of the rejection of traditional linear mean-reverting structures in the constant elasticity of variance class is the specification of the drift function. Our estimated drift is virtually zero up to about 15 percent. It mean-reverts in a nonlinear fashion only at the upper edge of the range of the sample process. Contrary to the existing literature, we emphasize the importance of the martingale behavior of the spot rate series over most of its range in disputing linear mean-reverting models. As for the marked nonlinearity of the drift at the upper edge of the sample process, its empirical relevance is clouded by the availability of few observations in this range. This idea can be phrased in a more rigorous fashion in our framework. We will show that in order to be able to draw precise inference on the drift of the process at a point, we require the estimated local time of the process at that point to be large. In other words, we require the time spent by the sample process in the spatial vicinity of that point to be large. Since the sample process barely visits interest rate levels at the upper edge of the empirical range, we cannot draw firm conclusions about the behavior of the drift at high interest rate levels, where nonlinearities arise. Even though rarely mentioned, the problem of the lack of sufficient observations at high rates affects most recent papers. We believe this issue should suggest a more cautious interpretation of the economic content of the estimated nonlinearities in the literature.

This chapter is organized as follows. Section 10 introduces the model and outlines existing parametric and nonparametric approaches to the problem. Section 11 reviews the estimation technique presented in the previous chapter, introduces new results and discusses the sense in which "spatial" arguments can be used to assist in developing a general approach to the functional analysis of the dynamics of the short-term interest rate process. In Section 12 we present the data and implement the method. Section 13 concludes. Technical details and proofs are provided in Section 14. Notation is laid out in section 15.

10. Conventional parametric specifications and estimation procedures

In one-factor term structure models, the spot interest rate is the only state variable on which the current yield curve depends. Its dynamic is usually modelled as a stochastic differential equation of the form

$$dr_t = \mu(r_t, t)dt + \sigma(r_t, t)dB_t \tag{10.1}$$

where $\mu(.,.) : \Re \times [0,T] \rightarrow \Re$, $\sigma(.,.) : \Re \times [0,T] \rightarrow \Re^+$ and $\{B_t, 0 \leq t \leq T\}$ is a standard Brownian motion. The functions $\mu(.,.)$ and $\sigma(.,.)$ are specified to guarantee the existence of a solution to (10.1) such that for all t the price $\Psi_{t,T} = \Im_t^{emm}[\exp(\int_t^T(-r_s)ds)]$ of the zero-coupon bond with maturity T is well defined [see Duffie (1992) for a discussion]. $\Im_t^{emm}$ represents the expectation with respect to the equivalent martingale measure. In particular, $\Psi_{t,T} = \Im_t^{emm}[\exp(\int_t^T(-r_s)ds)] = \Phi(r_t, t, T)$ for some function $\Phi : [0,T] \times [0,T] \times \Re \rightarrow \Re$. This implies that at each time t the term structure depends solely on the contemporaneous value of the short-term rate, on t and on the time to maturity T. Consider the following specification:

$$\mu(r_t,t) = \alpha_0(t)+\alpha_1(t)r_t+\alpha_2(t)r_t^2+\alpha_3(t)/r_t+\alpha_4(t)r_t\log r_t, \quad (10.2)$$

$$\sigma^2(r_t,t) = \beta_0(t)+\beta_1(t)r_t+\beta_2(t)r_t^{\beta_3(t)} \quad (10.3)$$

where $\alpha_0(.)$, $\alpha_1(.)$, $\alpha_2(.)$, $\alpha_3(.)$, $\alpha_4(.)$, $\beta_0(.)$, $\beta_1(.)$, $\beta_2(.)$ and $\beta_3(.)$ are continuous on $[0,T]$ into $\Re$. Formulae (10.2) and (10.3) encompass most models in the literature. Provided formulae (10.2) and (10.3) do not depend on time, the solution to (10.1) is a homogeneous Markov process. We can write the unrestricted homogeneous specification in the form

$$\begin{aligned} dr_t &= (\alpha_0+\alpha_1 r_t+\alpha_2 r_t^2+\alpha_3/r_t+\alpha_4 r_t\log r_t)dt \\ &\quad +(\sqrt{\beta_0+\beta_1 r_t+\beta_2 r_t^{\beta_3}})dB_t. \end{aligned} \quad (10.4)$$

Various restrictions on the parameters generate common single factor models.[1] For example,

$$\begin{aligned}
\alpha_2,\alpha_3,\alpha_4,\beta_0,\beta_1 &= 0 \text{ and } \beta_3=1\ [\textit{Cox- Ingersoll-Ross (1985)}],\\
\alpha_2,\alpha_3,\alpha_4,\beta_1,\beta_2 &= 0\ [\textit{Vasicek (1977)}],\\
\alpha_2,\alpha_3,\alpha_4,\beta_1 &= 0 \text{ and } \beta_3=1\ [\textit{Pearson and Sun (1994)}],\\
\alpha_2,\alpha_3,\alpha_4,\beta_0,\beta_1 &= 0 \text{ and } \beta_3=2\ [\textit{Brennan-Schwartz (1979)}],\\
\alpha_2,\alpha_3,\alpha_4,\beta_0,\beta_1 &= 0\ [\text{CEV diffusion-}\textit{Chan et al. (1992)}],\\
\alpha_1,\alpha_2,\alpha_3,\alpha_4,\beta_1,\beta_2 &= 0\ [\textit{Merton (1973)}],\\
\alpha_0,\alpha_1,\alpha_2,\alpha_3,\alpha_4,\beta_0,\beta_1 &= 0 \text{ and } \beta_3=2\ [\textit{Dothan (1978)}],\\
\alpha_0,\alpha_1,\alpha_2,\alpha_3,\alpha_4,\beta_0,\beta_1 &= 0 \text{ and } \beta_3=3\ [\textit{Constantinides-Ingersoll(1984)}],\\
\alpha_4 &= 0\ [\textit{Aït-Sahalia (1996b)}],\\
\alpha_0,\alpha_2,\alpha_3,\beta_0,\beta_1 &= 0 \text{ and } \beta_3=2\ [\textit{Homogeneous Black-Karasinski (1991)}],\\
\alpha_0,\alpha_2,\alpha_3,\alpha_4,\beta_0,\beta_1 &= 0\ [\textit{Cox and Ross (1976) / Cox (1985)}].
\end{aligned}$$

The Vasicek model, the Cox-Ingersoll-Ross (CIR, hereafter) model, the Merton model and the more general Pearson and Sun model belong to the so-called "affine-class" of term-structure models. They share the property that log bond prices, and hence bond yields, are affine in the underlying state variable, namely $\log\Phi(r_t,t,T)=\log\Phi(r_t,T-t)=A(T-t)r_t+$

[1] The following table summarizes standard approaches but does not, by any means, aim at being exhaustive given the magnitude of the theoretical literature on this topic.

$B(T-t)$[2] where A and B are continuously differentiable functions. They have a number of appealing features: their linearity allows a parsimonious representation of the term structure at each point in time, as a function of the system's state variable. Moreover, the model can be renormalized so that the yields themselves are the state variable. This observation is clear in one-factor models, but it is also true in set-ups with any number of unobservable factors. Longstaff and Schwartz (1992), for example, solve for the dynamics of the level of the short rate and the volatility of the short rate but the model can be also defined in terms of two bond yields of fixed maturities. In general, affine models permit closed-form solutions but impose strong restrictions on the term structure [e.g. see discussion in Campbell, Lo and MacKinlay (1997)].

The behavior of the spot interest rate implied by affine models is quite different. The Vasicek model specifies the short-term rate as a standard Ornstein-Uhlenbeck process. Provided $\alpha_1 < 0$ (i.e. mean-reversion occurs), it is strictly stationary in the steady state. Since linearity[3] holds, the process inherits the properties of the underlying Brownian motion, i.e. both the transition and the marginal density are normal. The CIR squared root model displays a noncentral chi-square transition distribution. If $\alpha_1 < 0$ and $2\alpha_0 > \beta_2$, the process is strictly stationary in the steady state with a gamma marginal [Feller (1951)]. The model in Merton (1973) is a standard Brownian motion with drift. The transition density function is normal and the process is nonstationary [see Part III for a thorough description of the statistical properties of these specifications].

The Merton model could be generalized to a specification with time-varying coefficients that is often called the Ho-Lee model (1986). The Ho-Lee model looks at movements in the yield curve consistent with the absence of arbitrage opportunities. The focus here is on the fitting, at a given time, of the underlying term structure, rather than on the description of the time series properties of the process driving it. This observation motivates time-dependent specifications in "arbitrage-free" models [see, also, Hull and White (1990)].

We will assume a time-invariant structure in what follows. Hence, we can rewrite (10.1) as,

$$dr_t = \mu(r_t)dt + \sigma(r_t)dB_t \quad (10.5)$$

[2] Duffie and Kan (1993) and Brown and Schaefer (1991) give conditions on the spot rate dynamics which deliver an affine structure in continuous-time. Technicalities aside, they show that the drift and diffusion term have to be affine, too.

[3] In the usual sense for stochastic differential equations.

where the functions $\mu(.) : \Re \rightarrow \Re$ and $\sigma(.) : \Re \rightarrow \Re^+$ satisfy the following expressions

$$\begin{aligned}
\mu(r_t) &= \lim_{h\to 0} E\{\frac{r_{t+h} - r_t}{h}|r_t = r\} = & (10.6)\\
&= \lim_{h\to 0} \frac{1}{h}\int_{[|r_{t+h}-r_t|<\varepsilon]} (r_{t+h} - r_t)dP(r_{t+h}|r_t = r),\\
\sigma^2(r_t) &= \lim_{h\to 0} E\{\frac{[r_{t+h} - r_t]^2}{h}|r_t = r\} = & (10.7)\\
&= \lim_{h\to 0} \frac{1}{h}\int_{[|r_{t+h}-r_t|<\varepsilon]} (r_{t+h} - r_t)^2 dP(r_{t+h}|r_t = r)
\end{aligned}$$

and, also

$$\lim_{h\to 0} h^{-1}P(|r_{t+h} - r_t| \geq \varepsilon|r_t = r) = 0.$$

We know that (10.6) and (10.7) represent the "instantaneous" conditional mean and the "instantaneous" conditional variance of the process when $r_t = r$. Specifically, (10.6) describes the conditional expected rate of change of the process for infinitesimal time changes, whereas (10.7) gives the conditional rate of change of volatility at r [see Part I, Section 2, Assumption 2.1, for conditions which guarantee the existence and uniqueness of a strong solution to (10.5)].

A further observation on the model is needed. We are working in a Markovian world. Some recent evidence [for example, Jeffrey (1997) and Aït-Sahalia (1998)] suggests that more degrees of freedom should be allowed for a better understanding of some financial time series, including interest rates. We believe that a more complete assessment of the empirical potential of path-dependent specifications is needed[4] before we dismiss models whose main virtues are simplicity and tractability.

We now turn to a concise review of the econometrics of stochastic differential equations.

[4] As mentioned earlier, it is hard to conclude that conventional short-term interest rate series are stationary on the basis of traditional unit root tests. The test of "Markovian nature" in Aït-Sahalia (1998) hinges theoretically on the stationarity of the series under analysis. It is based on the Chapman-Kolmogorov equation for Markov processes and consists of comparing a direct estimator of the 2Δ-interval conditional density to an indirect estimator obtained by iterating a direct Δ-interval estimator of the conditional density of the process. Transition densities are computed very naturally as ratios of joint and marginal distributions. Strict stationarity is necessary to evaluate the time-invariant marginal distribution of the process. The inspection of path dependent specifications in Jeffrey (1997) is thorough but, again, the asymptotic results supporting his GMM application rely on the stationarity of the underlying process.

10.1. A brief discussion of some relevant estimation methods in a continuous-time framework[5]

Nonparametric drift and diffusion functions have been proposed by Banon (1978), Geman (1979), Pham Dinh (1981), and Banon and Nguyen (1981) but they all assume continuous sampling observations. The first paper to deal with the nonparametric estimation of the diffusion term from a discrete record of observations is Florens-Zmirou (1993). She leaves the drift term unidentified and treats it as a nuisance parameter. The diffusion function is estimated by employing a sample analog estimator [compare to (10.7)] defined as follows,

$$\widehat{\sigma}^2_{(n)}(x) = \frac{1}{\Delta_n}\frac{\sum_{i=1}^{n-1} 1_{\{|X_{i/n}-x|<h_n\}}[X_{(i+1)\Delta_n} - X_{i\Delta_n}]^2}{\sum_{i=1}^{n} 1_{\{|X_{i/n}-x|<h_n\}}}. \tag{10.8}$$

We assume that we observe X_t at $\{t = t_1, t_2, .., t_n\}$ in the time interval $[0, T]$, with $T \geq T_0 > 0$, where T_0 is a positive constant. We also assume equispaced data. So, $\{X_t = X_{\Delta_n}, X_{2\Delta_n}, X_{3\Delta_n}, ..., X_{n\Delta_n}\}$ are n observations at $\{t_1 = \Delta_n, t_2 = 2\Delta_n, t_3 = 3\Delta_n, ..., t_n = n\Delta_n\}$, where $\Delta_n = T/n$. Asymptotic results are obtained as the sample frequency increases for a given ending time T. Under fairly regular conditions on the sample size (n) and the bandwidth (h_n), namely as $n \to \infty$, $h_n \to 0$, $nh_n \to \infty$ and $nh_3 \to 0$, Florens-Smirou proves the L^2-consistency of $\widehat{\sigma}^2_{(n)}(x)$ and convergence to a mixture of normal law, depending on the local time of the process [c.f. Theorem 5.9 in Part I].

The work by Jiang and Knight (1997a) relies heavily on the results in Florens-Zmirou (1993). Their diffusion function estimator is the same as the one suggested by Florens-Zmirou (1993), but is constructed from a general kernel function rather than from a discontinuous indicator function. Their drift estimator combines the estimator of the diffusion function along with the estimated nonparametric density of the underlying process under the assumption of stationarity. In effect, it is a well known result that, provided suitable regularity conditions are met, the marginal distribution of the process is fully characterized by the two functions of interest, i.e. the drift and the diffusion function [e.g. Karatzas and Shreve (1988), Karlin and Taylor (1981)]. More specifically, we can write

$$\mu(x) = \frac{1}{2}\left[\frac{\partial\sigma^2(x)}{\partial x} + \sigma^2(x)\frac{\frac{\partial p(x)}{\partial x}}{p(x)}\right]. \tag{10.9}$$

It is a simple task, by Slutsky's theorem, to define a consistent estimator for the "instantaneous" conditional mean given consistent estimates for the density, $p(x)$, the diffusion term,

[5] Again, due to the magnitude of the literature on the topic, the following discussion does not aim to be exhaustive.

$\sigma^2(x)$, and their first derivatives, $\frac{\partial \sigma^2(x)}{\partial x}$ and $\frac{\partial p(x)}{\partial x}$, respectively.

The same approach, though reversed, was previously utilized in the semiparametric estimation procedure proposed by Aït-Sahalia (1996a). He specifies a parametric, linear mean-reverting drift. Subsequently, given the theoretical cross-restriction on the diffusion based on the marginal density of a strongly stationary process and its drift function, he proves pointwise consistency and asymptotic normality for the semiparametric diffusion estimator constructed from the nonparametric estimate of the density and the parametric estimate of the drift. A similar semiparametric approach based on density-matching is contained in Aït-Sahalia (1996b). In a recent paper, Stanton (1997) suggests the use of nonparametric approximations to the true functions. This is the first work which attempts fully nonparametric identification of diffusions by use of discrete data without resorting to cross-restrictions as in (10.9).

Many authors have used parametric approaches to diffusion modeling, often employing maximum likelihood (ML) methods. Brown and Hewitt (1975), Lanska (1979) and Kutoyants (1984) are among those who assume continuous sampling observations. Dacunha-Castelle and Florens-Zmirou (1986) is the first paper concerned with the parametric estimation of nonlinear diffusions from a discrete record of data. Donhal (1987) proves the local asymptotic mixed normality property of the likelihood function of the diffusion term. Both papers use the expansion of the transition density of the underlying process for small changes in time. Lo (1988) discusses how to perform ML estimation of jump-diffusion processes. Pedersen (1995) proposes an approximate ML estimation procedure for multidimensional diffusion processes. In a recent paper, Aït-Sahalia (1998) illustrates the properties of consistency, normality and asymptotic efficiency for ML estimators obtained from maximizing a sequence of approximations to the true, but unknown, likelihood function of the discretely sampled process.

Generalized method of moments is employed in many papers often based on discretizations of the underlying process [Chan, Karolyi, Longstaff and Sanders (1992), CKLS hereafter, for example]. Hansen and Scheinkman (1995) rigorously derive moment conditions for continuous diffusions based on the infinitesimal generator and a stationarity assumption. Also promising are the simulation methods based on indirect inference [notably, Gourieroux, Monfort and Renault (1993) and Gallant and Tauchen (1996)]. This line of research is very close to the simulated moments procedure suggested in Duffie and Singleton (1993).

10.2. The issue

Modern asset pricing theory relies on continuous-time models, typically formulated in terms of stochastic differential equations, for the dynamics of the underlying state variables.[6] This mainstream approach motivates the design of appropriate estimation techniques in the same framework. Furthermore, the necessity of being as general as possible justifies the implementation of functional estimation methods capable of taking into account the discrete nature of the available data. As pointed out earlier, a wide array of contributions has been provided along these lines. The problem with conventional approaches is that they rely heavily on the stationarity of the process. The reason for this is that the assumption of stationarity allows the possibility of consistently estimating the marginal density of the series and, based on this, carrying out sophisticated estimation procedures [see, for example, Aït-Sahalia (1996 a,b) for semiparametric applications and Jiang and Knight (1997a) and Jiang (1998) for a fully nonparametric approach].[7]

Nevertheless, even though it is not an implausible theoretical assumption, the stationarity of short-term interest rate series can not be empirically guaranteed[8]. Therefore, more robust estimation methods are needed. A possible solution is to achieve identification of the functions of interest without resorting to cross-restrictions based on the marginal density of the process. In other words, it is important to estimate consistently both the drift and the diffusion function in situations where one of these is of primary concern and the other function is treated as a nuisance parameter.

Valuable, in this context, is the work by Stanton (1997) [see, also, Boudoukh, Richardson, Stanton and Whitelaw (1998) for a bivariate application]. His procedure is based on approximations to the true drift and diffusion obtained through the use of the infinitesimal generator [Revuz and Yor (1994) is a classical reference]. The econometric estimation hinges on the use of functional sample analogs. In the original paper, Stanton's methodology is

[6] For an overview, see Duffie (1992).

[7] Stationarity is also usually invoked in parametric studies. *Inter alia*, CKLS (1992) carry out a GMM procedure whose asymptotic theory is based on a (non-tested) assumption of stationarity and ergodicity for the underlying series. To generate moment conditions for continuous-time Markov processes, Hansen and Scheinkman (1995) utilize the Dynkin operator and impose the condition that the time derivative of the unconditional expectation of some 'well behaved' functions of the underlying process is equal to zero. Again, this comes from assuming the strict stationarity of the underlying process.

[8] In Aït-Sahalia (1996a,b), for example, the spot rate used is the seven-day Eurodollar deposit rate, bid-ask midpoint, from Bank of America. The data are daily from June 1, 1973 to February 25, 1995. The author rejects the null of nonstationarity at 90 percent using a standard Dickey-Fuller test (the value of the test statistic is -2.60 versus a critical value of -2.57). The robustness of this result can be disputed after a careful inspection of the same time series [see Subsection 12.1 and Tables 3 and 4].

presented in a heuristic fashion as no asymptotic theory is supplied to support his approach. One could argue that the existence of a complete limit theory is a valuable but, in some cases, not a necessary piece of information. Unfortunately, due to the difficulties posed by the availability of discrete data for the estimation of continuous systems [c.f. Aït-Sahalia (1996a,b), Florens-Smirou (1993), Jiang and Knight (1997a) and Part I], identification is truly an issue. Only a complete asymptotic theory can shed light on the capacity of an estimation methodology to capture the features of the functions of interest. A simple example will help clarify this point. We already described how to achieve identification of the diffusion function as in Florens-Zmirou (1993). As pointed out earlier, her sample analog estimator exploits the local dynamics of the process, i.e. her asymptotic results are obtained as the sample frequency increases for a given ending time T. The drift is harder to identify given its higher order of magnitude. Below, we specify a commonsense estimator for the drift function and use the same sampling method as in Florens-Smirou (1993), i.e. we let the data frequency increase for a given T. The following result proves that consistency can not be achieved.

10.1 Theorem *Given h_n such that $nh_n \to \infty$ and $nh_n^4 \to 0$, the sample analog estimator $\widehat{\mu}_{(n)}(x)$ [compare to (10.6)] defined as*

$$\widehat{\mu}_{(n)}(x) = \frac{1}{\Delta_n} \frac{\sum_{i=1}^{n-1} \mathbf{1}_{\{|X_{i/n}-x|<h_n\}}[X_{(i+1)\Delta_n} - X_{i\Delta_n}]}{\sum_{i=1}^{n} \mathbf{1}_{\{|X_{i/n}-x|<h_n\}}}$$

diverges at a rate given by the square root of the bandwidth as $n \to \infty$.

A more complex asymptotic theory is needed to identify the drift. A natural extension is to prolong the observation period, that is to let $T \to \infty$ as the interval between adjacent observations shrinks. Below, we discuss alternative estimators of the two functions of interest based on the necessity of reducing the local variability induced by discrete observations and enhancing the availability of information through the use of a longer time span[9]. We retain the sample analog structure and we do not impose cross-restrictions in order to avoid invoking strong requirements on the distribution of the underlying process. A complete description of the asymptotic theory is contained in Part I.

[9]Under the same conditions, that is as the frequency of observations increases over an enlarging time span, the estimates of the first order approximations to drift and diffusion function in Stanton (1997) are proven to be consistent and asymptotically normal [c.f. Bandi (1999) and Part III].

11. The Econometric Approach

We assume a Markov, possibly nonlinear, continuous data generating process as in (10.5) for the short-term interest rate process $\{r_t, t \geq 0\}$ [see Part I, Section 2, Assumption 2.1 for conditions imposed on (10.5)].

The process r_t is recorded at $\{t = t_1, t_2, .., t_n\}$ in the time interval $[0, T]$, with $T \geq T_0 > 0$, where T_0 is a positive constant. We assume equispaced data. Hence, $\{r_t = r_{\Delta_{n,T}}, r_{2\Delta_{n,T}}, r_{3\Delta_{n,T}}, ..., r_{n\Delta_{n,T}}\}$ are n observations at $\{t_1 = \Delta_{n,T}, t_2 = 2\Delta_{n,T}, t_3 = 3\Delta_{n,T}, ..., t_n = n\Delta_{n,T}\}$ where $\Delta_{n,T} = T/n$. Theoretically, we want the number of sampling points (n) to increase as the time span lengthens (T). Moreover, we want the frequency to increase with n. In Part I we explore the limit theory of the proposed estimators as $n \to \infty$, $T \to \infty$ and $\Delta_{n,T} = T/n \to 0$. We suggest the following estimators for (10.6) and (10.7).

$$\widehat{\sigma}^2_{(n,T)}(r) = \frac{\sum_{i=1}^n \mathbf{K}(\frac{r_{i\Delta_{n,T}} - r}{h_{n,T}})\widetilde{\sigma}^2_{n,T}(r_{i\Delta_{n,T}})}{\sum_{i=1}^n \mathbf{K}(\frac{r_{i\Delta_{n,T}} - r}{h_{n,T}})} \tag{11.1}$$

with

$$\widetilde{\sigma}^2_{n,T}(r_{i\Delta_{n,T}}) = \frac{1}{m_{n,T}(i\Delta_{n,T})\Delta_{n,T}} \sum_{j=0}^{m_{n,T}(i\Delta_{n,T})-1} [r_{t(i\Delta_{n,T})_j + \Delta_{n,T}} - r_{t(i\Delta_{n,T})_j}]^2 \tag{11.2}$$

and

$$\widehat{\mu}_{(n,T)}(r) = \frac{\sum_{i=1}^n \mathbf{K}(\frac{r_{i\Delta_{n,T}} - r}{h_{n,T}})\widetilde{\mu}_{(n,T)}(r_{i\Delta_{n,T}})}{\sum_{i=1}^n \mathbf{K}(\frac{r_{i\Delta_{n,T}} - r}{h_{n,T}})} \tag{11.3}$$

with

$$\widetilde{\mu}_{(n,T)}(r_{i\Delta_{n,T}}) = \frac{1}{m_{n,T}(i\Delta_{n,T})\Delta_{n,T}} \sum_{j=0}^{m_{n,T}(i\Delta_{n,T})-1} [r_{t(i\Delta_{n,T})_j + \Delta_{n,T}} - r_{t(i\Delta_{n,T})_j}]. \tag{11.4}$$

The symbols have the usual interpretation. The sequence $\{t(i\Delta_{n,T})_j\}$ is a sequence of random times defined as follows,

$$t(i\Delta_{n,T})_0 = \inf\{t \geq 0 : |r_t - r_{i\Delta_{n,T}}| \leq \epsilon_{n,T}\} \tag{11.5}$$

and

$$t(i\Delta_{n,T})_{j+1} = \inf\{t \geq t(i\Delta_{n,T})_j + \Delta_{n,T} : |r_t - r_{i\Delta_{n,T}}| \leq \varepsilon_{n,T}\}, \tag{11.6}$$

for all i. The number $m_{n,T}(i\Delta_{n,T}) \leq n$ counts the stopping times associated with the value $r_{i\Delta_{n,T}}$ and is defined as

$$m_{n,T}(i\Delta_{n,T}) = \sum_{j=1}^{n} \mathbf{1}_{[|r_{j\Delta_{n,T}} - r_{i\Delta_{n,T}}| \leq \varepsilon_{n,T}]}$$

where $\mathbf{1}_A$ denotes the indicator of A. The quantity $\varepsilon_{n,T}$ is a bandwidth-like parameter depending on the time span (T) and on the sample size (n). The function $\mathbf{K}(.)$ that appears in (11.1) and (11.3) is a kernel whose properties are described in Part I, Section 4, Assumption 4.1.

By looking at a long time span and high frequency observations (technically, we perform both *'infill'* and *'long span'* asymptotics) we aim at reconstructing as well as possible the theoretical path of the process.

The intuition underlying (11.1) and (11.3) is simple. As pointed out earlier, the idea is twofold. First, the use of stopping times in the algorithm is intended to replicate the instantaneous features of the theoretical functions. Notice that the averages $\tilde{\sigma}^2_{n,T}(r_{i\Delta_{n,T}})$ and $\tilde{\mu}_{n,T}(r_{i\Delta_{n,T}})$ in (11.1) and (11.3) are defined as empirical analogs to the values taken on by the true functions at the data points ($r_{i\Delta_{n,T}}$, for example, defines the i^{th} observation). The estimates $\tilde{\sigma}^2_{n,T}(r_{i\Delta_{n,T}})$ and $\tilde{\mu}_{n,T}(r_{i\Delta_{n,T}})$ are consistent for $\sigma^2(r_{i\Delta_{n,T}})$ and $\mu(r_{i\Delta_{n,T}})$ as $m_{n,T}(i\Delta_{n,T}) \rightarrow \infty$ $\forall i$. This is true, under suitable conditions on the bandwidths, when $T \rightarrow \infty$ provided the process is recurrent. We assume recurrence [by Assumption 2.1 in Part I], that is we require the t-continuous trajectory of the process to hit any point in its range an infinite number of times almost surely, i.e. $P_x\{r_t$ hits z at a sequence of times increasing to $\infty\} = 1$ $\forall x, z$.

Second, we apply the standard method of nonparametric smoothing to recover the two functions of interest from a scatter of estimates of the two functions at the data points. A few additional comments on the methodology are needed.

First, the assumption of recurrence is not restrictive and, indeed, makes economic sense because we expect interest rates to return to the values in their range over and over again.[10] In practice, we have at our disposal only finite datasets, therefore the risk of imprecise inference turns out to be larger in correspondence with data points that stand out as very

[10]Recurrence does not imply stationarity. Brownian motion is a typical example of a recurrent, nonstationary process.

different from the remaining observations. In Table 1 we display the time series of the annualized 7-day Eurodollar rate [June 1, 1973 - February 25, 1995]. As mentioned earlier, this is the data utilized in Aït-Sahalia (1996a,b) and in this work, too. Outliers can be detected in the period 1980-1982. We expect inference to be less reliable for interest rates in this time range. Later on, we will discuss this problem more carefully.

Furthermore, since we are not imposing cross-restrictions, our specification is robust to deviations from stationarity. We believe this is an important feature of our method since it allows us to accommodate interest rate processes that evolve over time in a general fashion, as we do not require the existence of a time-invariant marginal distribution, but tend to return to the values in their range.

Recall that we are increasing the observations frequency as $n \to \infty$. In other words, we are exploiting the local properties of the process over its dynamic range since, loosely speaking, we "almost" obtain a continuum of data points in the limit.

Having said this, we comment on the smoothing procedure. A rescaled version of the denominator in (11.1) and (11.3) converges to the stationary density of the process in conventional nonparametric smoothing. Here we do not use the information contained in the marginal density of the underlying process because the underlying process might not have a time-invariant marginal density. What we use is the information contained in the spatial density of the process. Spatial densities are well defined for stationary and nonstationary processes provided the semimartingale property is satisfied.

Theorem 5.1 in Part I, Section 5, gives the asymptotic expression of the (standardized) denominator in (11.1) and (11.3) in our specific case. We fix T for ease of analysis but the intuition does not change when we let T go to infinity provided $\frac{T}{n} = \Delta_{n,T} \to 0$ [c.f. Part I, Section 5, Corollary 5.3]. It is convenient to review the theorem here.

11.1 Theorem (Theorem 5.1, Part I, Section 5) *Given $n \to \infty$, T fixed $(= \overline{T})$, and $h_{n,\overline{T}} \to 0$ (as $n \to \infty$) in such a way that $\frac{1}{h_{n,\overline{T}}}(\Delta_{n,\overline{T}})^{\alpha} = O(1)$ for some $\alpha \in (0, \frac{1}{2})$, the estimator $\frac{\Delta_{n,\overline{T}}}{h_{n,\overline{T}}} \sum_{i=1}^{n} \mathbf{K}(\frac{r_i \Delta_{n,\overline{T}} - r}{h_{n,\overline{T}}})$ converges to $\overline{L}_r(\overline{T}, r) = \frac{1}{\sigma^2(r)} L_r(\overline{T}, r)$ a.s. where $L_r(.,.)$ is the local time of the process.*

In the formulation of (11.1) and (11.3) the local time estimator is used as a conventional kernel estimator for the probability density of a stationary process. Our discussion in the introduction should clarify the sense in which this amounts to spatial smoothing and guarantees protection against possible deviations from stationarity. Notice, in fact, that Theorem 11.1 holds regardless of the stationarity of the underlying process. This is due to the sampling technique, which is characterized by increasingly frequent data, and the nature of the data generating process, which is assumed to be a semimartingale.

These observations, in turn, open up the way for the definition of statistics, based on the local time of the underlying process, whose meaning and descriptive power do not depend on the existence of a time-invariant density function. In a recent paper, Peter C. B. Phillips (1998) discusses how to interpret the limit local times of standardized density-like kernel estimators in the presence of discrete-time integrated processes in terms of spatial densities. Further, he comments on the usefulness of spatial densities (and various functionals of them) in the descriptive analysis of time series which display random wandering characteristics. The same approach is used here.

The previous theorem gives us a way to estimate the spatial density of a process defined as in (10.5). The standardization $\frac{1}{\sigma^2(r)}$ permits us to interpret the spatial density at a point in terms of real time units spent by the process in the neighborhood of that point. It constitutes the stochastic differential equation analog to the $\frac{1}{\sigma^2}$ normalization in the Brownian motion context [c.f. formula (9.3)]. We now illustrate the rate of convergence to the true function and give the limit distribution. The latter is mixed normal (MN). The mixing variate is proportional to the spatial density itself.

11.2 Theorem *If $h_{n,\overline{T}} \to 0$ in such a way that $\frac{1}{h^{3/2}_{n,\overline{T}}}(\Delta_{n,\overline{T}})^{1/2-\epsilon} \to 0$ as $n \to \infty$ for an arbitrarily small ε, then*

$$\frac{1}{\sqrt{h_{n,\overline{T}}}}\left(\widehat{\overline{L}}_r(\overline{T},r) - \overline{L}_r(\overline{T},r)\right) \stackrel{d}{\Rightarrow} 4MN\left(0, \mathbf{k}\frac{1}{\sigma^2(r)}\overline{L}_r(\overline{T},r)\right)$$

where $\mathbf{k} = \int_0^\infty \int_0^\infty \min(s,q)\mathbf{K}(s)\mathbf{K}(q)dsdq$.

As pointed out by Phillips (1998) in the discrete-time context, this result enables us to construct asymptotic confidence intervals which closely resemble conventional intervals for probability densities obtained from kernel estimates. A 95% confidence interval for $\overline{L}_r(\overline{T},r)$ is given by

$$\widehat{\overline{L}}_r(\overline{T},r) \pm 1.96\left(16\mathbf{k}\frac{h_{n,\overline{T}}}{\sigma^2(r)}\widehat{\overline{L}}_r(\overline{T},r)\right)^{1/2}.$$

The spatial density estimator replaces the standard density function estimator. The scale factor 16**k** accounts for the time dependence in the observations as the Brownian covariance kernel appears in the definition of **k**.

Notice that the limit process $\overline{L}_r(\overline{T}, r)$ is a random variable.[11] Spatial densities have a time dimension, as opposed to probability densities. Their time dimension can be fruitfully explored.

Following the lead of Phillips (1998), we can define functionals of spatial densities such as local or spatial hazard rates. The spatial hazard $\overline{H}_r(\overline{T}, r)$ associated with a spatial density $\overline{L}_r(\overline{T}, r)$ can be represented as follows,

$$\overline{H}_r(\overline{T}, r) = \frac{\overline{L}_r(\overline{T}, r)}{\int_r^{\infty} \overline{L}_r(\overline{T}, y)dy}. \tag{11.7}$$

This definition allows us an easy interpretation in terms of spatial analogue to standard hazard functions obtained from time-invariant probability densities. Spatial hazard functions offer, once more, an easy generalization to the potentially nonstationary case by virtue of the use of the local information contained in the continuous process. Formula (11.7) has a standard meaning: it gives the conditional risk over the period $[0, \overline{T}]$ of an interest rate level of r, given that interest rates are at least as big as r. Again, the time dimension adds up to the traditional definition and permits us to investigate how the conditional risk evolves over time. The following theorem gives us a way to recover spatial hazards from sample analogues constructed from estimated spatial densities.

11.3 Theorem *Given $h_{n,\overline{T}} \to 0$ (as $n \to \infty$) in such a way that $\frac{1}{h^{3/2}_{n,\overline{T}}}(\Delta_{n,\overline{T}})^{1/2-\epsilon} \to 0$ for an arbitrarily small ε, the estimator $\widehat{\overline{H}}_r(\overline{T}, r)$ defined as $\frac{\widehat{\overline{L}}_r(\overline{T},r)}{\int_r^{\infty}\widehat{\overline{L}}_r(\overline{T},y)dy}$ converges to $\overline{H}_r(\overline{T}, r)$ a.s. Moreover,*

$$\frac{1}{\sqrt{h_{n,\overline{T}}}}\left(\widehat{\overline{H}}_r(\overline{T}, r) - \overline{H}_r(\overline{T}, r)\right) \overset{d}{\Rightarrow} 4MN\left(0, \frac{\mathbf{k}(\overline{H}_r(\overline{T}, r))^2}{\sigma^2(r)\overline{L}_r(\overline{T}, r)}\right)$$

where $\mathbf{k} = \int_0^{\infty}\int_0^{\infty} \min(s, q)\mathbf{K}(s)\mathbf{K}(q)dsdq$.

In what follows we will provide traditional descriptive statistics whose interpretation is straightforward should stationarity hold. Further, we will thoroughly examine the locational properties of the short-term interest rate series using the apparatus described above.

[11] In the stochastic processes jargon, we would say that it is a random variable (for fixed t) and a function of t (for a given $\varpi \in \Omega$).

Before moving to the implementation of our method, we turn to a concise description of the limit theory of the estimators in (11.1) and (11.3) [see Part I for details]. We start with the consistency of the diffusion function estimator.

11.4 Theorem (Theorem 5.5, Part I, Section 5) *Given $n \to \infty$, $T \to \infty$, $h_{n,T} \to 0$ (as $n, T \to \infty$) such that $\frac{\Delta_{n,T}}{h_{n,T}} \to 0$ and $\frac{\overline{L}_r(T,r)}{h_{n,T}}(\Delta_{n,T})^{\alpha} = O_{a.s.}(1)$ for some $\alpha \in (0, \frac{1}{2})$, and $\varepsilon_{n,T} \to 0$ (as $n, T \to \infty$) such that $\frac{\Delta_{n,T}}{\varepsilon_{n,T}} \to 0$ and $\frac{\overline{L}_r(T,r)}{\varepsilon_{n,T}}(\Delta_{n,T})^{\beta} = O_{a.s.}(1)$ for some $\beta \in (0, \frac{1}{2})$, the estimator*

$$\frac{\sum_{i=1}^{n} \mathbf{K}(\frac{r_{i\Delta_{n,T}} - r}{h_{n,T}})\tilde{\sigma}^2_{n,T}(r_{i\Delta_{n,T}})}{\sum_{i=1}^{n} \mathbf{K}(\frac{r_{i\Delta_{n,T}} - r}{h_{n,T}})} \stackrel{a.s.}{\to} \sigma^2(r)$$

where $\tilde{\sigma}^2_{n,T}(r_{i\Delta_{n,T}})$ is defined in (11.2).

The following theorem gives the asymptotic distribution and the rate of convergence.

11.5 Theorem (Theorem 5.6, Part I, Section 5) *Assume that $n \to \infty$, $T \to \infty$, $h_{n,T} \to 0$ (as $n, T \to \infty$) such that $\frac{\Delta_{n,T}}{h_{n,T}} \to 0$ and $\frac{\overline{L}_r(T,r)}{h_{n,T}}(\Delta_{n,T})^{\alpha} = O_{a.s.}(1)$ for some $\alpha \in (0, \frac{1}{2})$, and $\varepsilon_{n,T} \to 0$ (as $n, T \to \infty$) such that $\frac{\Delta_{n,T}}{\varepsilon_{n,T}} \to 0$ and $\frac{\overline{L}_r(T,r)}{\varepsilon_{n,T}}(\Delta_{n,T})^{\beta} = O_{a.s.}(1)$ for some $\beta \in (0, \frac{1}{2})$.*
If $\frac{\varepsilon^4_{n,T}}{\Delta_{n,T}} \to 0$, $\varepsilon_{n,T}\overline{L}_r(T,r) \stackrel{a.s.}{\to} 0$ and $h_{n,T} = o(\varepsilon_{n,T})$, then

$$\sqrt{\frac{\overline{L}_r(T,r)\varepsilon_{n,T}}{\Delta_{n,T}}}\left(\hat{\sigma}^2_{n,T}(r) - \sigma^2(r)\right) \stackrel{d}{\to} N\left(0, 2\sigma^4(r)\right). \tag{11.8}$$

If $\frac{\varepsilon^4_{n,T}}{\Delta_{n,T}} \to 0$,[12] *$\varepsilon_{n,T}\overline{L}_r(T,r) \stackrel{a.s.}{\to} 0$ and $h_{n,T} = O(\varepsilon_{n,T})$ with $h_{n,T}/\varepsilon_{n,T} \to \phi > 0$, then*

$$\sqrt{\frac{\overline{L}_r(T,r)\varepsilon_{n,T}}{\Delta_{n,T}}}\left(\hat{\sigma}^2_{n,T}(r) - \sigma^2(r)\right) \stackrel{d}{\to} N\left(0, 2\theta_{\phi}\sigma^4(r)\right). \tag{11.9}$$

where $\theta_{\phi} = \frac{1}{2}\int_{-\infty}^{\infty}\int_{(z-1)/\phi}^{(z+1)/\phi}\int_{(z-1)/\phi}^{(z+1)/\phi} \mathbf{K}(a)\mathbf{K}(e)dzdade$.

As pointed out earlier, we are letting T and n go to infinity. If we had at our disposal frequent data ($n \to \infty$) over a fixed time span ($T = \overline{T}$), we could still estimate the function consistently. T going to infinity is a technical device introduced to exploit the recurrence of the process in the estimation procedure. Recurrence is crucial in the estimation of the

[12] If $\frac{\varepsilon^4_{n,T}}{\Delta_{n,T}} \to \infty$, then (11.1) still converges to a mixed normal distribution. This is the distribution to which the "bias" term in the estimation error decomposition converges [see Part I].

drift function. In the case of diffusion estimation we do not need to require infinite passage times to identify the true function. This point is coherent with a statement in footnote n.1 of a 1979 paper by S. Geman, where we read

> "*$\sigma(\lambda)$ can, in principle, be determined from any interval containing a single crossing of λ*".

For a fixed T $(=\overline{T})$, we can rewrite (11.8) as follows:

$$\sqrt{n\varepsilon_{n,\overline{T}}}\{\hat{\sigma}^2_{(n,\overline{T})}(r) - \sigma^2(r)\} \xrightarrow{d} MN\left(0, 2\frac{\sigma^4(r)}{(\overline{L}_r(\overline{T},r)/\overline{T})}\right).$$

The conditions on the bandwidths $h_{n,\overline{T}}$ and $\varepsilon_{n,\overline{T}}$ reduce to

$$\varepsilon_{n,\overline{T}} \propto n^{-k_1} \text{ with } k_1 \in \left(\frac{1}{4}, \frac{1}{2}\right)$$

and

$$h_{n,\overline{T}} \propto n^{-k_2} \text{ with } k_2 \in \left(0, \frac{1}{2}\right)$$

with

$$h_{n,\overline{T}} = o(\varepsilon_{n,\overline{T}}).$$

If $h_{n,\overline{T}} = O(\varepsilon_{n,\overline{T}})$ with $h_{n,\overline{T}}/\varepsilon_{n,\overline{T}} \to \phi > 0$, what changes in the limiting distribution is simply the constant of proportionality in the asymptotic variance term. In consequence, we just need to multiply the scalar 2 by a scalar whose value depends on the value that the constant ϕ takes on, that is $\theta_\phi = \frac{1}{2}\int_{-\infty}^{\infty}\int_{(z-1)/\phi}^{(z+1)/\phi}\int_{(z-1)/\phi}^{(z+1)/\phi}\mathbf{K}(a)\mathbf{K}(e)dzdade$. In both cases the rate of convergence is $\sqrt{n\varepsilon_{n,\overline{T}}}$, which is the standard rate in functional regression analysis, and the asymptotic variance is inversely related to $\overline{L}_r(\overline{T},r)$ or, equivalently, to the number of time units spent by the process in the spatial neighborhood of a point. Intuitively, even though repeated visits are not necessary for consistent estimation of the diffusion function, the larger the number of visits to a point in the range of the process, the more precise will be the estimation of the diffusion at that point.

We now consider the drift estimator. Again, we state consistency for the true function and give the asymptotic distribution.

11.6 Theorem (Theorem 5.11, Part I, Section 5) *Given $n \to \infty$, $T \to \infty$, $h_{n,T} \to 0$ (as $n, T \to \infty$) such that $\frac{\Delta_{n,T}}{h_{n,T}} \to 0$ and $\frac{\overline{L}_r(T,r)}{h_{n,T}}(\Delta_{n,T})^{\alpha} = O_{a.s.}(1)$ for some $\alpha \in (0, \frac{1}{2})$, and $\varepsilon_{n,T} \to 0$ (as $n, T \to \infty$) such that $\frac{\Delta_{n,T}}{\varepsilon_{n,T}} \to 0$ and $\frac{\overline{L}_r(T,r)}{\varepsilon_{n,T}}(\Delta_{n,T})^{\beta} = O_{a.s.}(1)$ for some $\beta \in (0, \frac{1}{2})$, and provided $\varepsilon_{n,T}\overline{L}_r(T,r) \overset{a.s.}{\to} \infty$, the estimator*

$$\frac{\sum_{i=1}^{n} \mathbf{K}(\frac{r_{i\Delta_{n,T}}-r}{h_{n,T}})\widetilde{\mu}_{n,T}(r_{i\Delta_{n,T}})}{\sum_{i=1}^{n} \mathbf{K}(\frac{r_{i\Delta_{n,T}}-r}{h_{n,T}})} \overset{a.s.}{\to} \mu(r)$$

where $\widetilde{\mu}_{n,T}(r_{i\Delta_{n,T}})$ is defined in (11.4).

11.7 Theorem (Theorem 5.12, Part I, Section 5) *Assume that $n \to \infty$, $T \to \infty$, $h_{n,T} \to 0$ (as $n, T \to \infty$) such that $\frac{\Delta_{n,T}}{h_{n,T}} \to 0$ and $\frac{\overline{L}_r(T,r)}{h_{n,T}}(\Delta_{n,T})^{\alpha} = O_{a.s.}(1)$ for some $\alpha \in (0, \frac{1}{2})$, and $\varepsilon_{n,T} \to 0$ (as $n, T \to \infty$) such that $\frac{\Delta_{n,T}}{\varepsilon_{n,T}} \to 0$ and $\frac{\overline{L}_r(T,r)}{\varepsilon_{n,T}}(\Delta_{n,T})^{\beta} = O_{a.s.}(1)$ for some $\beta \in (0, \frac{1}{2})$, and $\varepsilon_{n,T}\overline{L}_r(T,r) \overset{a.s.}{\to} \infty$.*

If $h_{n,T} = o(\varepsilon_{n,T})$, then

$$\sqrt{\overline{L}_r(T,r)\varepsilon_{n,T}}\left(\widehat{\mu}_{n,T}(r) - \mu(r)\right) \overset{d}{\to} MN\left(0, \frac{1}{2}\sigma^2(r)\right). \tag{11.10}$$

If $h_{n,T} = O(\varepsilon_{n,T})$ with $h_{n,T}/\varepsilon_{n,T} \to \phi > 0$, then

$$\sqrt{\overline{L}_r(T,r)\varepsilon_{n,T}}\left(\widehat{\mu}_{n,T}(r) - \mu(r)\right) \overset{d}{\to} MN\left(0, \frac{1}{2}\theta_{\phi}\sigma^2(r)\right) \tag{11.11}$$

where $\theta_{\phi} = \frac{1}{2}\int_{-\infty}^{\infty}\int_{(z-1)/\phi}^{(z+1)/\phi}\int_{(z-1)/\phi}^{(z+1)/\phi} \mathbf{K}(a)\mathbf{K}(e)dzdade$.

The drift term cannot be identified nonparametrically over a fixed time interval, no matter how frequently the data is sampled [c.f. Theorem 10.1] unless cross-restrictions are imposed. Here we are lengthening the sample span ($T \to \infty$) as the frequency of observations increases ($n \to \infty$). We do so to gain information on the theoretical function through repeated visits to a point. In other words, when estimating the drift, we need to use the information on the trajectory of the process over its entire dynamic range. Since local arguments as in the case of diffusion estimation can not be utilized, three are the main consequences [c.f. Part I for details]:

[1] Contrary to diffusion estimation, the stochastic properties of the underlying process play a vital role in the drift estimation.

[2] The rate of convergence of the diffusion estimator ($\sqrt{\frac{\overline{L}_r(T,r)\varepsilon_{n,T}}{\Delta_{n,T}}}$) is faster than the rate of convergence of the drift estimator ($\sqrt{\overline{L}_r(T,r)\varepsilon_{n,T}}$).

[3] The admissible spatial bandwidth $\varepsilon_{n,T}$ must converge to zero at a slower pace than the corresponding bandwidth in the diffusion case. Furthermore, apart from specific cases such as Brownian motion [c.f. Part I, Section 5, Remark 5.16], it is not possible to express in closed-form the main condition that the window width $\varepsilon_{n,T}$ needs to satisfy, that is $\overline{L}_r(T,r)\varepsilon_{n,T} \overset{a.s.}{\to} \infty$.

Should T be fixed, then the local time factor, $\overline{L}_r(T,r)$, would be $O_p(1)$ rather than tend to infinity almost surely and the estimator would diverge at a speed equal to $\frac{1}{\sqrt{\varepsilon_{n,T}}}$ [c.f. Theorem 10.1 in this chapter].

The idea behind the consistency of the drift estimator is simple. Even in finite samples, it is possible to identify the drift function in a situation where the diffusion term is treated as a nuisance parameter, provided the sampled data is "sufficiently" recurrent. In our case, we require the data set to be characterized by repeated observations in the neighborhood of each interest rate level. We can phrase this condition differently and say that we require the data set to be affected by few outliers. This is a standard requirement when we are interested in reliable statistical inference. Then, the issue is how to assess the recurrence properties of the specific data set at hand. A possible rule of thumb in our framework is to estimate the spatial density of the sample process at each point and verify its order of magnitude. A large spatial density at an interest rate level implies repeated passages in the vicinity of that level and, in consequence, satisfactory inference. In what follows we will study the local time process to describe the locational features of the short-term interest rate series and detect areas where inference might be imprecise due to the lack of sufficient data points.

A final observation on the limit theories is needed. Recall from the previous chapter that the normalizations in theorems (11.5) and (11.7) are random because of the presence of the local time factor $\overline{L}_r(T,r)^{\frac{1}{2}}$. As T diverges to infinity, the chronological local time $\overline{L}_r(T,r)$ of the process r_t diverges to infinity as well since r_t is recurrent. Hence, the rate of convergence depends on the asymptotic divergence characteristics of the chronological local time $\overline{L}_r(T,r)$. As a consequence, the rate of convergence is, in general, path dependent [Part I, Section 5, Remark 5.10].[13]

We now turn to the implementation of the method.

[13] This is not the case if r_t is Brownian motion. In this case the convergence rates are $\sqrt{\varepsilon_{n,T}T^{\frac{1}{2}}/\Delta_{n,T}}$ and $\sqrt{\varepsilon_{n,T}T^{\frac{1}{2}}}$ and are not path dependent [Part I, Section 5, Remark 5.10]

12. Implementation

12.1. The data

The proposed methodology requires a long time series of high frequency observations. Nevertheless, due to the risk of spurious microstructure contaminations, the proper use of such series is an issue which goes beyond the scope of this work. We compromise by using a less ideal but still suitable series, namely the seven-day Eurodollar deposit spot rate, bid-ask midpoint.[14] This data was previously used in Aït-Sahalia (1996 a,b). Our description of the data is based on Aït-Sahalia's work and we refer the reader to his papers for a complete discussion. The data points are daily observations from June 1, 1973 to February 25, 1995.[15] The total number of observations is 5505. The rates quoted were originally bond-equivalent yields. They were transformed to continuously-compounded yields to maturity (annualized rate). Weekends and holidays do not receive special treatment but weekend effects [e.g. French and Roll (1997)] do not seem to be a major concern for money-market instruments. Monday is taken as the first day after Friday. A time series plot of the data is contained in Figure 1. Figure 2 provides a time series plot of the first differences. Table 1 contains a summary of the characteristics of the data whereas Table 2 gives some standard descriptive statistics.

We stressed before the necessity of devising estimation procedures robust to deviations from strong distributional assumptions. In effect, it is hard to claim that the data in question is, without any doubt, stationary. We perform conventional Augmented Dickey-Fuller (ADF) tests both with a constant term and a trend, and with only a constant in the deterministic part of the fitted regression. The order of the time polynomial is set equal to one whereas the number of lagged first differences is set equal to five.[16] The test systematically accepts the null at all conventional levels [Tables 3 and 4]. In the literature, even slight rejections (and this is not the case here) are interpreted as evidence in favor of stationarity due to the low power of the test [for example, Aït-Sahalia (1996 a,b) and Jiang and Knight (1997a)]. We verify the previous outcome by implementing a different

[14] As Aït-Sahalia (1996a) points out, choosing a seven-day rate is "*...a compromise between: (i) literally taking an "instantaneous" rate and (ii) avoiding some of the spurious microstructure effects associated with overnight rates...*".

[15] The Monte Carlo evidence in Jiang and Knight (1997b) and in Part III suggests that daily frequencies are good approximations to frequent sampling for estimators relying on increasingly frequent observations.

[16] Different plausible values do not affect the results. For the sake of comparison, we worked with up to 30 lagged first differences. This is the number of lags used in Aït-Sahalia (1996a). In the constant/trend case the test accepts at all conventional levels. In the constant case, the first rejection occurs with 30 lags, at the 10 percent level. This outcome is consistent with the result reported in Aït-Sahalia (1996a).

testing procedure based on Phillips' (1987) Z_α and Z_t statistics. Z tests generally have more power than the ADF test.[17] Again, we consider both the constant/time trend case and the constant case. The number of autocovariance terms to compute the spectrum at frequency zero is set equal to five.[18] The automated "optimal" bandwidth case is also evaluated as it usually delivers quite different results. In all cases the test accepts at the 1 percent critical level. Results are mixed at the 5 percent level and generally in favor of stationarity at the 10 percent level, with the only exception being the Z_t test when both a constant and a trend are included [Tables 3 and 4].

Since the series displays time intervals of fairly regular behavior and relatively low volatility (1973-1980 and 1982-1995), we apply the same tests to subperiods in order to assess the influence of higher volatility periods on the test results. A pattern seems to emerge: higher volatility periods inject stationarity in the data. Even though the null is very rarely rejected, the test values appear to be closer to the corresponding critical values in the presence of more volatile data. For instance, over the period 1973-1980 the null of nonstationarity is never rejected and the statistic values are very safely located in the acceptance region. When we add the more volatile data between 1980 and 1982, the overall picture remains quite unequivocally nonstationary, but our statistics appear to deliver values closer to the rejection threshold. The same applies to the data in the period 1982-1995 versus the longer period 1980-1995.

For the time being, it seems safe to say that standard testing procedures do not offer unambiguous support to the stationarity assumption and justify the use of investigation methods robust to deviations from it.

12.2. A look at the spatial characteristics of the data

Some authors have recently modelled the spot interest rate process as a randomly shifting process with no fixed mean [c.f. Das (1994) and Naik and Lee (1993), *inter alia*]. Aït-Sahalia (1996b) proposes a simpler modelling alternative to time-inhomogeneity which is based on the nonlinearity of the drift and diffusion functions. The idea is that sufficiently general specifications in the stationary time-homogeneous class can determine multimodal densities resembling regime shifts. Here, we suggest a direct way to study data that display irregular behavior. We look at the time spent by the sample process in each point of

[17] As far as size is concerned, the ADF test is generally less subject to distortions, especially in the presence of $MA(1)$ errors with parameter close to one [Phillips and Perron (1988)].

[18] Results are not qualitatively altered by different specifications.

its range and examine how this evolves over time. Our underlying process maintains the simplicity of time-homogeneous specifications but, at the same time, is general enough to allow for nonstationarities.

We consider three different time horizons: 1973-1980, 1973-1982 and 1973-1995 [c.f. Section 14 for details on the implementation]. Our interest in the change that occurred between 1980 and 1982 is motivated by the corresponding "atypical" behavior of the series. As briefly mentioned earlier, this period is characterized by high volatility [see Figure 2] and high interest rates [see Figure 1]. These features make inference more difficult. Below, we clarify the nature of these difficulties.

We start with the period 1973-1980. The spatial density of the process appears to be bimodal [Figure 3 (a)]. The modes show up at around 6 percent and 11 percent. When we add the data from 1980 to 1982 [Figure 3 (b)], we recognize persistence in the previous features and the emerging of two additional modes associated with higher interest rate levels around 15.5 percent and 18.5 percent. As we move to considering the whole data set, we expect to find evidence of a prolonged passage of the series below the 4 percent line but still in its vicinity. This is confirmed by our estimated spatial density displaying a minor, additional mode at the corresponding value [Figure 3 (c)].

Given the features of the estimation procedure, in a finite sample we expect to be able to identify very well the true functions of interest at points that are visited often. After a quick look at the graph of the estimated local time in the full period 1973-1995 [Figure 3 (c)] we anticipate that problems will arise for interest rates in the 20-24 percent range, as the time spent by the sample process in this range is fairly small.

In Figure 4 we report our results for the spatial hazard rate process. In the period 1973-1980 we recognize a non-monotonic increase in the interest rate risk [Figure 4 (a)]. Two peaks can be detected, around 6 percent and 11 percent. Notice that they correspond to the modes of the *sojourn* density in the same period. When we include the observations from 1980 to 1982 [Figure 4 (b)], the already identified peaks survive and two new ones emerge roughly at 16 percent and 20 percent. Notice, also, that the confidence bands are very broad at interest rates above the 20 percent threshold, implying unreliable inference. It is worth recalling that the same information is contained in the estimated spatial density. In effect, the empirical process appears not to spend much time in the 20-24 percent range. This justifies the uncertainty embodied in the wide confidence bands.[19] The same considerations

[19]A similar feature will characterize the estimates of the two functions of interest, whose asymptotic

apply to the full period 1973-1995 [Figures 4 (c)].

A simple accounting exercise further clarifies the information contained in the estimated spatial density. Between 16 percent and 18 percent we have at our disposal 80 observations. The number increases to 114 between 18 percent and 20 percent but it is only equal to 28 between 20 percent and 22 percent. The number of observations in the 22-24 percent range is 3. Most of the data is concentrated in the 6-8 percent and 8-10 percent ranges: 1197 and 1516, respectively. In what follows we will inspect the drift and diffusion functions in the range up to 22 percent. In fact, the dimension of the empirical local time of the process for interest rate values above the 22 percent threshold would make inference very unreliable.

12.3. Non parametric results

Moving to the estimation [c.f. Section 14, Subsection 14.1] of the curves of interest, Figure 5 (a) and Figure 5 (b) plot the nonparametric estimators, with 95 percent asymptotic confidence bands. The drift term appears to be nonlinear as opposed to most conventional parametric specifications [c.f. Section 10 and the next subsection]. Nevertheless, it mean-reverts strongly only when it approaches the upper bound of its range. Up to values close to 15 percent, the short-term rate virtually behaves as a martingale, since the drift is statistically significant but economically negligible. In Aït-Sahalia (1996b) the drift reverts at both ends of the theoretical domain due to the adopted parametric specification but is very close to zero between 3 percent and 24 percent, that is over the sample domain. Aït-Sahalia (1996a) assumes a linear mean-reverting drift from the start. Stanton (1997) and Jiang (1998) work with a different series[20] but their drift dynamics resemble the general features of our findings. Similar to the results in Aït-Sahalia (1996b) is the drift estimate in Jiang and Knight (1997).[21] A common feature of this literature is to imply the unpredictability of the short rate over most of its range since the process evolves over time as a martingale. Mean-reversion comes into play only at the extremities of the sample range.

The nonlinearity of the drift could account for atypical dynamics. In effect, during the 1980 to 1982 period a change in parameters seems to occur [c.f. Figure 1].[22] The estimated

confidence bands will be broad around the upper bound of the range of the sample process.

[20] They both use daily values of the secondary market yields on 3-month U.S Treasury Bills. The time horizon is January 1965 - July 1995 in Stanton (1997) and January 1962 - January 1996 in Jiang (1998). The series is converted from discounts to annualized compound rates.

[21] They use daily data of the Canadian 3-month treasury bill rate from January 2, 1982 to January 31, 1995.

[22] Aït-Sahalia (1996b) reports that "*...the mean α of the process with drift $\mu(r,\theta) = \beta(\alpha - r)$ estimated over 1980 to 1982 is significantly higher than the mean estimated on the rest of the sample...*"

drift appears to capture this shift as it displays nonlinearities for high interest rate values, largely corresponding to the same period. In general, as Aït-Sahalia (1996b) points out, misspecified linear models for the drift can hide nonlinearities as changes in parameters. We will come back to this point later on.

We now turn to the estimates of the diffusion function. The diffusion term exhibits quite conventional dynamics (CEV diffusions, for example) up to interest rate values around 16 percent. Volatility tapers off between 18 percent and 20 percent and then rises again. A comparable result is contained in Jiang (1998). Up to 16 percent, his nonparametric diffusion mimics the behavior of an estimated parametric CEV diffusion. Above 16 percent his estimates suggest less volatility than implied by a monotonically increasing parametric function in the CEV class. The diffusion function in Jiang and Knight (1997) displays a "smile" with lower volatilities associated with interest rates at the lower end of the range and in the middle. In Aït-Sahalia (1996b) the "instantaneous" conditional volatility of the process has U-shaped dynamics. It is equal to about 1.7 percent at 0. It decreases to about 0 at 11 percent and then rises steadily. Its domain of variation over the sample range is between 1.25 percent (at 4 percent) and about 4.3 percent (at 24 percent). The estimated diffusion in Aït-Sahalia (1996a) is increasing over the range of the sample process but this increase is non-monotonic. The function has an absolute peak at about 17 percent. The diffusion function in Stanton (1997) is monotonically increasing.

In the next subsection we will see that a CEV diffusion with an exponent equal to 3, i.e. $\sigma^2(r) = const.r^3$, fits our data very well up to about 16 percent. When paired with a drift function which is almost zero over the same range, such a diffusion function determines possibly nonstationary dynamics. In effect, the model

$$dr_t = const.r_t^{3/2} dB_t \tag{12.1}$$

which appears in Cox (1975) and CIR (1980), implies recurrent [Assumption 2.1 in Section 2, Part I, is satisfied] and nonstationary [c.f. Table 1 in Aït-Sahalia (1996a)] behavior for the spot interest rate series over $(0, \infty)$.

Here we are not taking a stance on stationarity. We simply notice that using a methodology that is robust to deviations from the existence of a time-invariant marginal density, we obtain shapes for the two functions of interest that give support to the necessity of being cautious about the stationarity of the series in question. This point is coherent with the results of our preliminary unit-root tests.

We now discuss the significance of the estimated nonlinearities at high rates.

12.4. A parametric comparison

A quick look at the overall shape of the two functions suggests that some conventional parametric models might not be severely misspecified. For instance, between 0 and about 16 percent the dynamics of the interest rate diffusion function can be well replicated by a conventional CEV model for the instantaneous variance. This observation is important. By not specifying a particular parametric structure, functional methods avoid misspecifications, but do so at the expense of a greater estimation error than their parametric counterparts.

Here, we undertake a simple exercise to assess whether credible traditional parametric models lead to reliable inference in the presence of well behaved functions like the ones we have just estimated. The estimation method we use is a conventional GMM [see CKLS (1992) for a well-known application]. Notice that we do not attempt to estimate consistently parametric continuous models by use of discretely sampled data. We only intend to compare the outcome of a naive[23] but very commonplace technique to our previous findings. We assume, for simplicity, a continuous-time model as in CKLS (1992), namely (10.4) with $\alpha_2, \alpha_3, \alpha_4, \beta_0, \beta_1 = 0$, that is

$$\begin{aligned} dr_t &= (\alpha_0 + \alpha_1 r_t)dt + (\sqrt{\beta_2 r_t^{\beta_3}})dB_t \\ &= (\alpha_0 + \alpha_1 r_t)dt + (\gamma_0 r_t^{\gamma_1})dB_t \end{aligned} \tag{12.2}$$

where r is the spot interest rate and B is a standard Brownian motion. As usual, alternative conventional models of the short-term riskless rate of interest can be nested in the specification in (12.2) with appropriate parameter restrictions [c.f. Section 10]. For instance, $\alpha_1 = 0$ and $\gamma_1 = 0$ deliver the Merton model (1973). Provided $\alpha_1 < 0$ and $\gamma_1 = 0$, formula (12.2) gives the Ornstein-Uhlenbeck process in Vasicek (1977), whereas $\gamma_1 = 1/2$ characterizes the process introduced by CIR (1985). The constant elasticity of variance (CEV) specification

[23] Discretizations are approximations. The relationship between parameters in the continuous-time format and in the discrete time analog is not straightforward [see, for instance, Drost and Werker (1996) and Nelson (1990)]. Recall, also, the temporal aggregation problem in Grossman, Melino and Shiller (1987), Breeden, Gibbons and Litzenberger (1989) and Longstaff (1989b, 1990a). Despite all this, there is a tendency to think that the error introduced by discretizing is of second order importance if changes are measured over short periods of time [CKLS (1992) and Campbell (1986)]. This point provides a justification for using the procedure with daily data [note that CKLS (1992) use monthly observations].

proposed by Cox (1985) and Cox and Ross (1976) requires $\alpha_0 = 0$. In Brennan-Schwartz (1982), γ_1 is equal to 1. We consider the discrete-time econometric specification,

$$\begin{aligned} r_{t+1} - r_t &= \alpha_0 + \alpha_1 r_t + \varepsilon_{t+1} \\ E[\varepsilon_{t+1}] &= 0,\ E[\varepsilon_{t+1}^2] = \gamma_0^2 r_t^{2\gamma_1} \end{aligned} \tag{12.3}$$

We follow CKLS (1992) in defining the relevant moment conditions.[24] The results of this exercise are reported in Table 5.

We start with the drift function. All the parameter estimates are significant. In CKLS (1992) the drift parameter estimates are statistically insignificant, implying that a linear mean-reverting structure fits our data set better than the data set examined in CKLS (1992). A graphical comparison between our functional estimates and their parametric counterparts is contained in Figures 6 (a) and 6 (b). Obviously, nonlinearities cannot be captured by a linear parametric structure. Nevertheless, nonlinear dynamics do not play a substantial role up to the upper extremity of the range of the sample process. Note that up to about 15 percent our nonparametric drift is measured precisely in a tight neighborhood of zero. Still, the parametric specification displays mild mean-reversion. Where nonlinearities arise, the unrestricted parametric model seems to mimick sufficiently well the behavior of the functional estimates with the exception of interest rate levels above 20 percent. Moreover, the parametric curve lies within our asymptotic 95 percent bands. This is an important point. A vanishing nonparametric drift up to about 15 percent implies that the interest rate process behaves as a martingale over a region of its range. Further, the tight nonparametric confidence bands in the same region and the shape of the asymptotic bands of the parametric drift function[25] suggest that the difference between our nonparametric specification and the

[24] Define a vector θ with elements α_0, α_1, γ_0^2 and γ_1. Given $\varepsilon_{t+1} = r_{t+1} - r_t - \alpha_0 - \alpha_1 r_t$, let the vector $f_t(\theta)$ be

$$f_t(\theta) = \begin{bmatrix} \varepsilon_{t+1} \\ \varepsilon_{t+1} r_t \\ \varepsilon_{t+1}^2 - \gamma_0^2 r_t^{2\gamma_1} \\ (\varepsilon_{t+1}^2 - \gamma_0^2 r_t^{2\gamma_1}) r_t \end{bmatrix}$$

If the restrictions implied by the discrete time model hold, then $E[\ f_t(\theta)] = 0$. This observation provides us with a set of four moment conditions. In what follows we utilize the optimal weighting matrix $\widehat{W}_T(\theta) = \frac{1}{T}(\sum_{t=1}^{T} f_t(\theta) f_t(\theta)')$. The asymptotic covariance matrix can be consistently estimated by

$$\frac{1}{T}([\frac{1}{T}\sum_{t=1}^{T}\frac{\partial f_t(\theta)}{\partial \theta}]'[\widehat{W}_T(\theta)][\frac{1}{T}\sum_{t=1}^{T}\frac{\partial f_t(\theta)}{\partial \theta}])^{-1}.$$

[25] We use the $\Delta-$method to compute parametric confidence bands.

parametric model is statistically significant [Figure 7]. Less clear-cut is the behavior of the drift function at higher interest rates, that is from 15 to 20 percent, and around the upper bound of the range of the sample process, that is above 20 percent. Nonlinearities come into play in a region where the available data is fairly thin. Here the linear parametric model can hardly be rejected on purely statistical grounds. Nevertheless, as demonstrated by the large nonparametric confidence bands and by the relatively large parametric bands, the overall uncertainty in this region suggest caution in interpreting our results.

The parametric model is more satisfactory for the estimation of the diffusion function. The specification is nonlinear. The CEV structure only fails to capture dynamics such as those detected for interest rate levels around the upper bound of the sample process. Up to 16 percent, the nonparametric and the parametric curves overlap almost perfectly. Not surprisingly, the parametric exponent in the variance term is estimated very precisely. This is a conventional result. In CKLS (1992) γ_1 is about 1.5 and almost two standard errors above one. Further, all models which make $\gamma_1 \leq 1$ are rejected. Our estimated γ_1 takes on a similar value. Even though a more complex specification is needed to fit the diffusion curve around the upper bound of the interest rate range, the estimated parametric diffusion lies almost everywhere within our estimated asymptotic bands.

Some observations on the behavior of the functions at high rates are in order at this point. If we believe the dynamics of our interest rate series can be described by a stochastic differential equation then, as pointed out by many authors including Aït-Sahalia (1996b) and CHLS (1997), traditional parametric structures do not completely capture the overall behavior of the series in the relevant domain. One could also argue that the series cannot be described by an homogeneous process in the period from 1980 to 1982, corresponding to high interest rates, since regime shifts are likely to occur (the rest of the series is fairly well behaved). We believe this issue should not be a major concern. We are not inclined to support the shift-in-regime point of view because nonlinearities such as those detected for high interest rate values (in the period 1980-1982) can produce dynamics resembling time-inhomogeneous changes. Thus, our findings appear to support one of the conclusions in Aït-Sahalia (1996b), namely

> *"..models, like linear drift and CEV diffusion, will mask nonlinearities as changes in parameters..."*[26].

[26]CKLS (1992) conclude that the shift in the Federal Reserve monetary policy in October 1979 did not

We now look at the plausibility of a nonlinear mean-reverting drift from a statistical perspective. Some complications arise. First, even assuming stationarity, the drift function is not constrained to display a specific shape at high rates provided the elasticity of the CEV instantaneous variance is sufficient to balance the drift dynamics and determine reversion to the center of the stationary distribution of the process. This point is made in CHLS (1997). They show that what matters for mean-reversion is a "pull" measure defined as the ratio between the drift and two times the diffusion function. Stationarity can be volatility-induced. In consequence, should the series be stationary, the uncertainty related to the lack of sufficient observations at high rates would add up to the absence of a strong theoretical motivation for drift-induced mean-reversion. This would make conclusions on the dynamics of the drift at high rates quite arbitrary. The potential nonstationarity of the series complicates matters even further.[27]

Second, at the edges of the sample range the drift is more easily biased in small samples. As for the upper bound, we know with near certainty that the maximum value achieved by the interest rate process in our sample is almost surely lower than the theoretical maximum value. Hence, the drift term is almost surely too low at the upper edge of the distribution of the data. The contrary is true for interest rate values at the lower bound. Still, the size of the bias depends on the volatility of the sample process. The volatility is very low at the lower edge of the sample but, as we discussed before, quite high at the upper edge. Therefore, a downward bias is more likely to occur at high interest rates,[28] thus strengthening our concerns related to the thin data available.[29]

Finally, as pointed out earlier, the estimated parametric curves lie within our asymptotic confidence bands for high interest rate values. Hence, the simple unrestricted parametric model proposed here cannot be statistically rejected for values in the vicinity of the upper bound of the sample process.

---

result in a structural break in the interest rate process. They interpret this result as suggesting that their volatility structure (our unrestricted CEV specification) is rich enough to capture the apparent change in the interest rate behavior in the post-1979 period.

[27] These observations are somewhat stronger than one of the conclusions put forward in concurrent work by Jones (1998): in a Bayesian framework, he shows that the use of uninformative Jeffreys priors does not result in statistical evidence for a nonlinear drift unless stationarity is imposed.

[28] I thank Chris Sims for pointing this out to me.

[29] In independent and parallel work, Chapman and Pearson (1998) reach a similar conclusion using a weighted least-squares estimation procedure applied to the data set in this paper and in Stanton (1997). They also show that the estimation methods proposed by Stanton (1997) and Aït-Sahalia (1996b) suggest nonlinearities of the type reported in the corresponding papers even when applied to sample paths simulated from a (linear mean-reverting) square root process. They conclude that the nonlinearity of the short term interest rate drift is not a "stylized fact".

To summarize, certainly small sample biases do not affect our functional estimation procedure for the drift in the range up to 16 percent. The spot rate behaves as a martingale up to about 16 percent. At higher values it mean-reverts nonlinearly but a standard parametric linear mean-reverting model for the drift cannot be rejected in this region. Our results appear statistically unreliable, even though economically sensible, for interest rate values between 20 percent and 22 percent. This observation can be applied to most papers in the literature since nonlinearities usually play a role in scarcely populated regions of the spot interest rate domain. Notice, though, that if we were interested in pricing, then potentially imprecise inference for interest rate levels that are hardly ever visited should not be a major concern.

13. Conclusion

A new descriptive and estimation approach for possibly nonstationary diffusion processes is implemented. As for the descriptive side of the suggested methodology, we extend results recently illustrated by Phillips and Park (1998) and Phillips (1998) in the unit root econometrics literature, to tackle the investigation of the spatial characteristics of time series modelled as solutions to potentially nonlinear and nonstationary stochastic differential equations. We construct and study spatial densities and spatial hazard rates. Further, we discuss how to use the information contained in the spatial dynamics of the underlying process to help develop a flexible approach to the functional estimation of stochastic differential equations under minimal assumptions on the distribution of the underlying process and using only a discrete sample of observations [c.f. Part I for a rigorous treatment]. We believe it is of primary importance to be able to achieve identification of both the drift and the diffusion function in situations where one of these is of primary concern and the other function is treated as a nuisance parameter. In effect, estimation without resorting to cross-restrictions permits us to obtain reliable inference when restrictions are hard to impose, namely when the solution to the stochastic differential equation is nonstationary. As a matter of fact, the evidence regarding the stationarity of some crucial financial time series, such as interest rates and exchange rates, is quite ambiguous.

We apply the new methodology to the analysis of the dynamics of the short-term interest rate process in continuous-time. We use a well-known data set in empirical finance, namely the Bank of America 7-day Eurodollar spot rate (midpoint bid ask) [Aït-Sahalia (1996a,b)].

In this work we provide statistical evidence of martingale behavior for our interest rate series over the restricted range up to about 16 percent. This suggests that the short-term process displays less predictability than implied by a linear mean-reverting structure for the drift. As for the nonlinearity of the drift at values between 16 percent and 22 percent, we cannot interpret it as a purely economically driven phenomenon, due to the possibility of a small sample bias in this range. This is particularly true in the range between 20 percent and 22 percent. In effect, "the time spent" by the empirical process at values about its upper edge is very small, thus making conclusions based on statistical inference potentially arbitrary. In consequence, we do not attempt to put forward conclusions on the dynamics of the drift about the upper extreme of its range. We only suggest that, after we take into consideration the potential for statistical artifacts, the non-linearity of the drift can account for the atypical dynamics often interpreted in terms of structural breaks or regime shifts. A parametric CEV structure mimics very well the behavior of the diffusion function over most of its domain.

In a recent paper, Pritsker (1998) points out that methods based on the estimation of the marginal density of the interest rate process [for example, Siddique (1994), Aït-Sahalia (1996a,b) and Stanton (1997)] fail to account for the effects of time-dependence in finite samples. The method suggested in this work relies on a more general notion of density and the temporal dependence in the trajectory of the short-rate process plays a role.

Note that our functional drift and diffusion functions can be used to test alternative parametric models of the short-term interest rate process based on a testing methodology that matches parametric specifications to their nonparametric counterparts. Due to the larger identifying information and the generality of "spatial" methods, this procedure is likely to have better size properties and more power than testing methods based on density-matching [c.f. Pritsker (1998)].

14. Proofs and Technical Details

14.1. The choice of the kernel and window widths

The smoothing parameter $h_{n,\overline{T}=1}$ and the spatial smoothing parameter $\varepsilon_{n,\overline{T}=1}$ are set as

$$\begin{aligned} h_{n,\overline{T}=1} &= c_h \widehat{\sigma}_r n^{-k_h} \\ \varepsilon_{n,\overline{T}=1} &= c_\varepsilon \widehat{\sigma}_r n^{-k_\varepsilon} \end{aligned}$$

where $\widehat{\sigma}_r$ is the estimated unconditional standard deviation of the series over the period of interest, n is the number of observations, k_h, k_ε are positive exponents depending on the limit theory of the specific estimator and c_h, c_ε are constants of proportionality. For practical purposes, T is set equal to 1.

14.1.1. Estimation of the sojourn times and spatial hazard rates

We use a second order Epanechnikov kernel as it simplifies calculations for the hazard functions, i.e.

$$\mathbf{K}(x) = 3/4(1 - x^2)\mathbf{1}_{\{-1,1\}}$$

We set $c_h^{ltime} = 3.5$ and $k_h^{ltime} = \frac{1}{4}$. The value for the constant is chosen to guarantee informativeness. In the periods 1973-1980, 1973-1982 and 1973-1995, the numerical values of the smoothing parameters are 1.455 percent, 2.05 percent and 1.459 percent, respectively.

14.1.2. Estimation of the drift and diffusion function

We employ a Gaussian kernel, but the use of an Epanechnikov or an exponential kernel would not change qualitatively the results. The drift window widths are set as follows:

$$\begin{aligned} c_h^{drift} &= 3.5 \\ k_h^{drift} &= \frac{1}{4} \\ c_\varepsilon^{drift} &= 4 \\ k_\varepsilon^{drift} &= \frac{1}{4}. \end{aligned}$$

As for the diffusion window widths, we set

$$\begin{aligned} c_h^{diff} &= 3.5 \\ k_h^{diff} &= \frac{1}{4} \\ c_\varepsilon^{diff} &= 3.6 \\ k_\varepsilon^{diff} &= \frac{1}{4}. \end{aligned}$$

The pairs c_h^{drift}, k_h^{drift} and c_h^{diff}, k_h^{diff} are chosen equal to 3.5 and $\frac{1}{4}$ for coherence with c_h^{ltime} and k_h^{ltime}. The values $c_\varepsilon^{diff} = 3.6$ and $c_\varepsilon^{drift} = 4$ are set larger than $c_h^{diff} = 3.5 =$

c_h^{ltime}. In consequence, the limiting distributions that we use in the paper are (11.8) and (11.10). The constant $c_\varepsilon^{diff} = 3.6$ is set smaller than $c_\varepsilon^{drift} = 4$ for consistence with the asymptotic theories [c.f. discussion in Section 11]. Different (but plausible) values for c_ε^{drift} that are chosen to be larger than c_ε^{diff} do not affect the results qualitatively.

The numerical value of $h^{drift}_{n,\overline{T}=1}$ and $h^{diff}_{n,\overline{T}=1}$ is 1.459 percent. The numerical values of the spatial bandwidths $\varepsilon^{drift}_{n,\overline{T}=1}$ and $\varepsilon^{diff}_{n,\overline{T}=1}$ are 1.6 percent and 1.5 percent, respectively.

14.2. Proof of Theorem 10.1

Assume the same model and the same properties as in Florens-Zmirou (1993). We follow Florens-Zmirou (1993) in the proof. Consider the quantity

$$\widehat{\mu}_{(n)}(x) = n\frac{\sum_{i=1}^{n-1} \mathbf{1}_{\{|X_{i/n}-x|<h_n\}}[X_{(i+1)\Delta_n} - X_{i\Delta_n}]}{\sum_{i=1}^{n} \mathbf{1}_{\{|X_{i/n}-x|<h_n\}}}.$$

We want to assess the asymptotic properties of

$$\sqrt{\frac{\sum_{i=1}^{[nt]} \mathbf{1}_{\{|X_{i/n}-x|<h_n\}}}{n}}\left(\frac{\sum_{i=1}^{[nt]-1} \mathbf{1}_{\{|X_{i/n}-x|<h_n\}} n[(X_{(i+1)\Delta_n} - X_{i\Delta_n}) - \mu(x)/n]}{\sum_{i=1}^{[nt]} \mathbf{1}_{\{|X_{i/n}-x|<h_n\}}}\right)$$

$$= \frac{1}{\sqrt{n}}\frac{\sum_{i=1}^{[nt]-1} \mathbf{1}_{\{|X_{i/n}-x|<h_n\}} n[(X_{(i+1)\Delta_n} - X_{i\Delta_n}) - \mu(x)/n]}{\sqrt{\sum_{i=1}^{[nt]} \mathbf{1}_{\{|X_{i/n}-x|<h_n\}}}}$$

$$= \frac{1}{\sqrt{n}}\frac{n^{\frac{1}{2}}}{\sqrt{2h_n}}\frac{\sum_{i=1}^{[nt]-1} \mathbf{1}_{\{|X_{i/n}-x|<h_n\}}[(X_{(i+1)\Delta_n} - X_{i\Delta_n}) - \mu(x)/n]}{\sqrt{\widehat{L}_X(x)}}$$

where $\mu(x)$ is a bounded function. Define

$$M_{(n)}(t) = \frac{1}{\sqrt{n}}\frac{n^{\frac{1}{2}}}{\sqrt{2h_n}}\sum_{i=1}^{[nt]-1} \mathbf{1}_{\{|X_{i/n}-x|<h_n\}}[(X_{(i+1)\Delta_n} - X_{i\Delta_n}) - \mu(x)/n]$$

and

$$m_{i+1} = \frac{1}{\sqrt{n}}\frac{n^{\frac{1}{2}}}{\sqrt{2h_n}}\mathbf{1}_{\{|X_{i/n}-x|<h_n\}}[(X_{(i+1)\Delta_n} - X_{i\Delta_n}) - \mu(x)/n].$$

Hence,

$$M_{(n)}(t) = \sum_{i=1}^{[nt]-1} m_{i+1}.$$

We denote by $\Im_{i/n}$ the conditional expectation with respect to $\Im^X_{i/n} = \sigma(X_s; s \leq i/n)$. We know that $\Im^X_{i/n} \subseteq \Im^B_{i/n}$ where $\Im^B_{i/n}$ is the filtration generated by B_s with $s \leq i/n$. If under conditions on h_n compatible with $h_n \to 0$, $nh_n \to \infty$ and $nh_n^4 \to 0$ the following four expressions hold, namely

$$\text{(A)} \sum_{i=1}^{[nt]-1} \Im_{i/n}(m_{i+1}) \xrightarrow{p} 0$$

$$\text{(B)} \sum_{i=1}^{[nt]-1} \Im_{i/n}(m_{i+1}^2) \xrightarrow{p} \Psi(x) L_X(t,x)$$

$$\text{(C)} \sum_{i=1}^{[nt]-1} \Im_{i/n}|m_{i+1}|^3 \xrightarrow{p} 0$$

$$\text{(D)} \sum_{i=1}^{[nt]-1} \Im_{i/n}(m_{i+1} b_{i+1}) \xrightarrow{p} 0$$

where $\Psi(x)$ is a generic bounded function of x, then the sequence of processes $(M_{(n)}(t), B_{(n)}(t))$, with $M_{(n)}(t)$ defined as before and $B_{(n)}(t) = \sum_{i=1}^{[nt]-1} b_{i+1}$ with $b_{i+1} = B(\frac{i+1}{n}) - B(\frac{i}{n})$, converges in distribution to the process $(U(t)_{\Psi(x)L(t,x)}, B(t))$ where $U(t)$ and $B(t)$ are independent Brownian motions. We start by verifying condition (A).

$$\begin{aligned}
\Im_{i/n}(m_{i+1}) &\leq |\Im_{i/n}(m_{i+1})| \\
&= \frac{n^{\frac{1}{2}}}{\sqrt{2nh_n}} |\mathbf{1}_{\{|X_{i/n}-x|<h_n\}} \Im_{i/n}[\int_{i/n}^{(i+1)/n} \mu(X_s)ds + \int_{i/n}^{(i+1)/n} \sigma(X_s)dB_s - \mu(x)/n]| \\
&= \frac{n^{\frac{1}{2}}}{\sqrt{2nh_n}} |\mathbf{1}_{\{|X_{i/n}-x|<h_n\}} \Im_{i/n}[\int_{i/n}^{(i+1)/n} (\mu(X_s) - \mu(x))ds| \\
&= \frac{n^{\frac{1}{2}}}{\sqrt{2nh_n}} |\mathbf{1}_{\{|X_{i/n}-x|<h_n\}} \Im_{i/n}[\int_{i/n}^{(i+1)/n} (\mu(X_s) - \mu(X_{i/n}))ds + \\
&\quad + \int_{i/n}^{(i+1)/n} (\mu(X_{i/n}) - \mu(x))ds]| \\
&\leq \frac{n^{\frac{1}{2}}}{\sqrt{2nh_n}} |\mathbf{1}_{\{|X_{i/n}-x|<h_n\}} \Im_{i/n}[\int_{i/n}^{(i+1)/n} (\mu'(X_{i/n})|X_s - X_{i/n}| + o(|X_s - X_{i/n}|))ds \\
&\quad + \int_{i/n}^{(i+1)/n} (const.|X_{i/n} - x| + o(|X_{i/n} - x|))ds]| \\
&\leq const.\frac{n^{\frac{1}{2}}}{\sqrt{2nh_n}} |\mathbf{1}_{\{|X_{i/n}-x|<h_n\}} \Im_{i/n}[\sup_{s\leq(i+1)/n} |X_s - X_{i/n}|]\frac{1}{n}| + \\
&\quad \frac{n^{\frac{1}{2}}}{\sqrt{2nh_n}} |\mathbf{1}_{\{|X_{i/n}-x|<h_n\}} \Im_{i/n}[o(\sup_{s\leq(i+1)/n} |X_s - X_{i/n}|)\frac{1}{n}]| \\
&\quad + const.\frac{n^{\frac{1}{2}}}{\sqrt{2nh_n}} |\mathbf{1}_{\{|X_{i/n}-x|<h_n\}} \frac{h_n}{n}| + const.\frac{n^{\frac{1}{2}}}{\sqrt{2nh_n}} |\mathbf{1}_{\{|X_{i/n}-x|<h_n\}} o(\frac{h_n}{n})|
\end{aligned}$$

We apply Burkholder-Davis-Gundy (BDG) inequality to the random quantity $\Im_{i/n}[\sup_{s\leq(i+1)/n}|X_s - X_{i/n}|]$ to obtain

$$
\begin{aligned}
\Im_{i/n}(m_{i+1}) &\leq const.\frac{n^{\frac{1}{2}}}{\sqrt{2nh_n}}|\mathbf{1}_{\{|X_{i/n}-x|<h_n\}}\frac{1}{n}\Im_{i/n}[\int_{i/n}^{(i+1)/n}\sigma^2(X_s)ds]^{\frac{1}{2}}| \\
&+const.\frac{n^{\frac{1}{2}}}{\sqrt{2nh_n}}|\mathbf{1}_{\{|X_{i/n}-x|<h_n\}}\Im_{i/n}[o(\sup_{s\leq(i+1)/n}|X_s - X_{i/n}|)\frac{1}{n}]| \\
&+const.\frac{n^{\frac{1}{2}}}{\sqrt{2nh_n}}|\mathbf{1}_{\{|X_{i/n}-x|<h_n\}}\frac{h_n}{n}| + const.\frac{n^{\frac{1}{2}}}{\sqrt{2nh_n}}|\mathbf{1}_{\{|X_{i/n}-x|<h_n\}}o(\frac{h_n}{n})|
\end{aligned}
$$

Summing up over i's and neglecting the smaller order of magnitude, we get

$$
\begin{aligned}
\sum_{i=1}^{[nt]-1}\Im_{i/n}(m_{i+1}) &\leq const.\left(\frac{\sum_{i=1}^{[nt]-1}\mathbf{1}_{\{|X_{i/n}-x|<h_n\}}}{2nh_n}\right)(nh_n)^{1/2}\frac{1}{n} + \\
&+const.\left(\frac{\sum_{i=1}^{[nt]-1}\mathbf{1}_{\{|X_{i/n}-x|<h_n\}}}{2nh_n}\right)({h_n}^{3/2}) \\
&\leq const.O_p(1)\left(\frac{{h_n}^{1/2}}{n^{1/2}} + {h_n}^{3/2}\right)
\end{aligned}
$$

We have just used the result $\frac{\sum_{i=1}^{[nt]-1}\mathbf{1}_{\{|X_{i/n}-x|<h_n\}}}{2nh_n} = O_p(1)$. This is proved in Florens-Zmirou (1993).

Proposition (Florens-Zmirou (1993)) *If $nh_n^4 \to 0$ and $nh_n \to \infty$ as $n \to \infty$, then $\frac{\sum_{i=1}^{[nt]-1}\mathbf{1}_{\{|X_{i/n}-x|<h_n\}}}{2nh_n}$ converges (in the L^2 sense) to the local time $L_X(t,x)$.*

We now verify condition (B). We add and subtract $\frac{\sigma^2(x)}{n}$ to obtain

$$
\begin{aligned}
\Im_{i/n}(m_{i+1}^2) &= \frac{1}{2h_n}\mathbf{1}_{\{|X_{i/n}-x|<h_n\}}\Im_{i/n}[(X_{(i+1)\Delta_n} - X_{i\Delta_n})^2 - \frac{\sigma^2(x)}{n} + \\
&-2\frac{\mu(x)}{n}(X_{(i+1)\Delta_n} - X_{i\Delta_n}) + \frac{\sigma^2(x)}{n} + \frac{\mu^2(x)}{n^2}] \\
&\leq \frac{1}{2h_n}\mathbf{1}_{\{|X_{i/n}-x|<h_n\}}\Im_{i/n}[\int_{i/n}^{(i+1)/n} 2(X_s - X_{i/n})\mu(X_s)ds \\
&+\int_{i/n}^{(i+1)/n} 2(X_s - X_{i/n})\sigma(X_s)dB_s + \int_{i/n}^{(i+1)/n}(\sigma^2(X_s) - \sigma^2(x))ds] \\
&+|\frac{\mu(x)}{n}|\frac{1}{2h_n}\Im_{i/n}\mathbf{1}_{\{|X_{i/n}-x|<h_n\}}(X_{(i+1)\Delta_n} - X_{i\Delta_n}) + \\
&+\frac{1}{2h_n}\mathbf{1}_{\{|X_{i/n}-x|<h_n\}}\frac{\mu^2(x)}{n^2} + \frac{1}{2h_n}\mathbf{1}_{\{|X_{i/n}-x|<h_n\}}\frac{\sigma^2(x)}{n}
\end{aligned}
$$

We examine the first term. By applying Cauchy-Schwartz (CS) inequality and BDG inequality and neglecting terms with smaller order of magnitude, we write

$$
\begin{aligned}
& \frac{1}{2h_n}\mathbf{1}_{\{|X_{i/n}-x|<h_n\}}\Im_{i/n}\Big[\int_{i/n}^{(i+1)/n} 2(X_s - X_{i/n})\mu(X_s)ds + \int_{i/n}^{(i+1)/n} 2(X_s - X_{i/n})\sigma(X_s)dB_s \\
& + \int_{i/n}^{(i+1)/n} (\sigma^2(X_s) - \sigma^2(x))ds\Big] \\
= & \frac{1}{2h_n}\mathbf{1}_{\{|X_{i/n}-x|<h_n\}}\Im_{i/n}\Big[\int_{i/n}^{(i+1)/n} 2(X_s - X_{i/n})\mu(X_s)ds + \int_{i/n}^{(i+1)/n} (\sigma^2(X_s) - \sigma^2(x))ds\Big] \\
\leq & \frac{1}{2h_n}\mathbf{1}_{\{|X_{i/n}-x|<h_n\}}\Big[(2\Im_{i/n}[\sup_{s\leq(i+1)/n} (X_s - X_{i/n})]^2)^{1/2}(\Im_{i/n}[\int_{i/n}^{(i+1)/n} |\mu(X_s)|ds]^2)^{1/2} \\
& + \int_{i/n}^{(i+1)/n} (\sigma^2(X_s) - \sigma^2(X_{i/n}))ds + \int_{i/n}^{(i+1)/n} (\sigma^2(X_{i/n}) - \sigma^2(x))ds\Big] \\
\leq & \; const.\frac{1}{2h_n}\mathbf{1}_{\{|X_{i/n}-x|<h_n\}}\frac{1}{n}(\Im_{i/n}[\int_{i/n}^{(i+1)/n} \sigma^2(X_s)ds])^{1/2} \\
& + const.\frac{1}{2h_n}\mathbf{1}_{\{|X_{i/n}-x|<h_n\}}\frac{h_n}{n} \\
\leq & \; const.\frac{1}{\sqrt{n}}\frac{\mathbf{1}_{\{|X_{i/n}-x|<h_n\}}}{2nh_n} + const.h_n\frac{\mathbf{1}_{\{|X_{i/n}-x|<h_n\}}}{2nh_n}
\end{aligned}
$$

Now we sum up over i's and the previous expression can be bounded by $const.(\frac{1}{\sqrt{n}} + h_n)O_p(1)$. The second term is

$$
\begin{aligned}
& |\frac{\mu(x)}{n}|\frac{1}{2h_n}\mathbf{1}_{\{|X_{i/n}-x|<h_n\}}\Im_{i/n}(X_{(i+1)\Delta_n} - X_{i\Delta_n}) \\
= & |\frac{\mu(x)}{n}|\frac{1}{2h_n}\mathbf{1}_{\{|X_{i/n}-x|<h_n\}}\Im_{i/n}\Big[\int_{i/n}^{(i+1)/n} \mu(X_s)ds \\
& + \int_{i/n}^{(i+1)/n} \sigma(X_s)dB_s\Big] \\
= & |\frac{\mu(x)}{n}|\frac{1}{2h_n}\mathbf{1}_{\{|X_{i/n}-x|<h_n\}}\Im_{i/n}\Big[\int_{i/n}^{(i+1)/n} \mu(X_s)ds\Big] \\
\leq & \; const.\frac{1}{n}\frac{1}{2nh_n}\mathbf{1}_{\{|X_{i/n}-x|<h_n\}}.
\end{aligned}
$$

Summing up over i's, we obtain

$$
\begin{aligned}
& |\frac{\mu(x)}{n}|\frac{1}{2h_n}\sum_{i=1}^{[nt]-1}\mathbf{1}_{\{|X_{i/n}-x|<h_n\}}\Im_{i/n}(X_{(i+1)\Delta_n} - X_{i\Delta_n}) \\
\leq & \; const.\frac{1}{n}O_p(1).
\end{aligned}
$$

The third term is

$$\frac{1}{2h_n}1_{\{|X_{i/n}-x|<h_n\}}\frac{\mu^2(x)}{n^2} \le const.\frac{1}{n}\frac{1}{2nh_n}1_{\{|X_{i/n}-x|<h_n\}}$$

and then,

$$\frac{1}{2h_n}\sum_{i=1}^{[nt]-1} 1_{\{|X_{i/n}-x|<h_n\}}\frac{\mu^2(x)}{n^2} \le const.\frac{1}{n}O_p(1).$$

The fourth term is

$$\frac{1}{2h_n}1_{\{|X_{i/n}-x|<h_n\}}\frac{\sigma^2(x)}{n} = \sigma^2(x)\left(\frac{1}{2nh_n}1_{\{|X_{i/n}-x|<h_n\}}\right).$$

Hence,

$$\sigma^2(x)\sum_{i=1}^{[nt]-1}\frac{1}{2nh_n}1_{\{|X_{i/n}-x|<h_n\}} \xrightarrow{p} \sigma^2(x)L_X(t,x).$$

Thus,

$$\sum_{i=1}^{[nt]-1}\Im_{i/n}(m_{i+1}^2) \xrightarrow{p} \sigma^2(x)\overline{L}(t,x).$$

We now verify condition (C). By using previous results, it is easy to prove that

$$\Im_{i/n}|m_{i+1}|^3 \le const.\frac{1}{2(nh_n)^{3/2}}1_{\{|X_{i/n}-x|<h_n\}}.$$

Hence,

$$\begin{aligned}\sum_{i=1}^{[nt]}\Im_{i/n}|m_{i+1}|^3 &\le const.\frac{1}{(nh_n)^{1/2}}\frac{\sum_{i=1}^{[nt]}1_{\{|X_{i/n}-x|<h_n\}}}{2nh_n}\\ &= const.\frac{1}{(nh_n)^{1/2}}O_p(1).\end{aligned}$$

But $\frac{1}{(nh_n)^{1/2}} \to 0$ as $n\to\infty$ since $nh_n\to\infty$. Along the same lines, we prove that (D) holds. Write

$$\begin{aligned}\Im_{i/n}(m_{i+1}b_{i+1}) &\le |\Im_{i/n}(m_{i+1}b_{i+1})| =\\ &= \frac{1}{\sqrt{2h_n}}1_{\{|X_{i/n}-x|<h_n\}}\times\\ &\quad\times|\Im_{i/n}[((X_{(i+1)\Delta_n}-X_{i\Delta_n})-\mu(x)/n)(B_{(i+1)\Delta_n}-B_{i\Delta_n})]|\\ &\le \frac{1}{\sqrt{2h_n}}1_{\{|X_{i/n}-x|<h_n\}}|\Im_{i/n}[(X_{(i+1)\Delta_n}-X_{i\Delta_n})(B_{(i+1)\Delta_n}-B_{i\Delta_n})]|+\\ &\quad+\frac{1}{\sqrt{2h_n}}1_{\{|X_{i/n}-x|<h_n\}}\frac{|\mu(x)|}{n}|\Im_{i/n}(B_{(i+1)\Delta_n}-B_{i\Delta_n})|.\end{aligned}$$

Since $\Im_t^X \subseteq \Im_t^B$, the Brownian motion B_t is a martingale with respect to the filtration $\Im_t^X$ generated by $\{X_s; s \leq t\}$, hence $\Im_{i/n}(B_{(i+1)\Delta_n} - B_{i\Delta_n}) = 0$. A simple application of CS inequality gives

$$\begin{aligned}\Im_{i/n}(m_{i+1}b_{i+1}) &\leq \frac{1}{\sqrt{2h_n}}\mathbf{1}_{\{|X_{i/n}-x|<h_n\}} \times \\ &\quad \times|(\Im_{i/n}(X_{(i+1)\Delta_n} - X_{i\Delta_n})^2)^{1/2}(\Im_{i/n}(B_{(i+1)\Delta_n} - B_{i\Delta_n})^2)^{1/2}| \\ &\leq const.\frac{1}{\sqrt{2nh_n}}\mathbf{1}_{\{|X_{i/n}-x|<h_n\}}(\Im_{i/n}(X_{(i+1)\Delta_n} - X_{i\Delta_n})^2)^{1/2}.\end{aligned}$$

From previous results we know that the order of magnitude of $\Im_{i/n}(X_{(i+1)\Delta_n} - X_{i\Delta_n})^2$ is $\frac{1}{n}$, then

$$\begin{aligned}\Im_{i/n}(m_{i+1}b_{i+1}) &\leq const.\frac{1}{\sqrt{2nh_n}}\mathbf{1}_{\{|X_{i/n}-x|<h_n\}}\frac{1}{\sqrt{n}} \\ &\leq const.\frac{h_n^{1/2}}{2nh_n}\mathbf{1}_{\{|X_{i/n}-x|<h_n\}}.\end{aligned}$$

Hence,

$$\begin{aligned}\sum_{i=1}^{[nt]}\Im_{i/n}(m_{i+1}b_{i+1}) &\leq const.h_n^{1/2}\sum_{i=1}^{[nt]}\frac{\mathbf{1}_{\{|X_{i/n}-x|<h_n\}}}{2nh_n} \\ &\leq const.h_n^{1/2}O_p(1)\end{aligned}$$

and this last inequality verifies condition (D).

To conclude, if h_n is such that as $n \to \infty$, $h_n \to 0$, $nh_n \to \infty$ and $nh_n^4 \to 0$, then conditions (A) through (C) guarantee that $M_{(n)}(t)$ converges to a Brownian motion U_t with quadratic variation $[U]_t = \sigma^2(x)L_X(t,x)$. Further, condition (D) implies that $U_{\sigma^2(x)\bar{L}(t,x)}$ and B_t are independent Brownian motions. Then,

$$\sqrt{\frac{\sum_{i=1}^{[nt]}\mathbf{1}_{\{|X_{i/n}-x|<h_n\}}}{n}}\left(\frac{\mu_{(n)}(x)-\mu(x)}{\sigma(x)}\right) \xrightarrow{d} N(0,1).$$

But this last expression gives

$$\sqrt{h_n}\left(\widehat{\mu}_{(n)}(x)\right) \xrightarrow{d} \sigma(x)(L_X(t,x))^{-\frac{1}{2}}N(0,1)$$

and, so

$$\widehat{\mu}_{(n)}(x) = O_p\left(\frac{1}{\sqrt{h_n}}\right).$$

This proves the stated result.

14.3. Proof of Theorem 11.1

See Part I, Section 5, Theorem 5.1.

14.4. Proof of Theorem 11.2

We write the estimation error as follows.

$$\frac{1}{\sqrt{h_{n,\overline{T}}}}\left(\widehat{\overline{L}}_r(\overline{T},r)-\overline{L}_r(\overline{T},r)\right)$$

$$= \frac{1}{\sqrt{h_{n,\overline{T}}}}\left(\frac{\overline{T}}{nh_{n,\overline{T}}}\sum_{i=1}^{n}\mathbf{K}\Big(\frac{r_{i\Delta_{n,\overline{T}}}-r}{h_{n,\overline{T}}}\Big)-\overline{L}_r(\overline{T},r)\right)$$

$$= \frac{1}{\sqrt{h_{n,\overline{T}}}}\left(\frac{1}{h_{n,\overline{T}}}\int_0^{\overline{T}}\mathbf{K}\Big(\frac{r_s-r}{h_{n,\overline{T}}}\Big)ds-\overline{L}_r(\overline{T},r)+o_{a.s.}\left(\frac{1}{h_{n,\overline{T}}}(\Delta_{n,\overline{T}})^{\frac{1}{2}-\epsilon}\right)\right)$$

$$= \frac{1}{\sqrt{h_{n,\overline{T}}}}\left(\frac{1}{h_{n,\overline{T}}}\int_{-\infty}^{\infty}\mathbf{K}\Big(\frac{a-r}{h_{n,\overline{T}}}\Big)\left(\frac{1}{\sigma^2(a)}\right)L_r(\overline{T},a)da-\overline{L}_r(\overline{T},r)\right)$$

$$+o_{a.s.}\left(\frac{1}{h^{3/2}_{n,\overline{T}}}(\Delta_{n,\overline{T}})^{\frac{1}{2}-\epsilon}\right)$$

$$= \frac{1}{\sqrt{h_{n,\overline{T}}}}\Big[\int_{-\infty}^{\infty}\mathbf{K}(q)\left(\frac{1}{\sigma^2(h_{n,\overline{T}}q+r)}\right)L_r(\overline{T},h_{n,\overline{T}}q+r)dq-\overline{L}_r(\overline{T},r)\Big]$$

$$+o_{a.s.}\left(\frac{1}{h^{3/2}_{n,\overline{T}}}(\Delta_{n,\overline{T}})^{\frac{1}{2}-\epsilon}\right)$$

We omit the stochastic order term since it is negligible in the limit under the assumptions made on the bandwidth parameter $h_{n,\overline{T}}$. Then,

$$\frac{1}{\sqrt{h_{n,\overline{T}}}}\left(\int_{-\infty}^{\infty}\mathbf{K}(q)\left(\frac{1}{\sigma^2(h_{n,\overline{T}}q+r)}\right)L_r(\overline{T},h_{n,\overline{T}}q+r)dq-\overline{L}_r(\overline{T},r)\right)$$

$$= \frac{1}{\sqrt{h_{n,\overline{T}}}}\left(\int_{-\infty}^{\infty}\mathbf{K}(q)\left(\frac{1}{\sigma^2(h_{n,\overline{T}}q+r)}\right)\left(L_r(\overline{T},h_{n,\overline{T}}q+r)-L_r(\overline{T},r)\right)dq\right)$$

$$+\frac{1}{\sqrt{h_{n,\overline{T}}}}\left(\int_{-\infty}^{\infty}\mathbf{K}(q)\left(\frac{\sigma^2(r)-\sigma^2(h_{n,\overline{T}}q+r)}{\sigma^2(h_{n,\overline{T}}q+r)\sigma^2(r)}\right)L_r(\overline{T},r)dq\right)$$

$$= \Big[\int_{-\infty}^{\infty}\mathbf{K}(q)\left(\frac{1}{\sigma^2(h_{n,\overline{T}}q+r)}\right)2\frac{1}{2\sqrt{h_{n,\overline{T}}}}\left(L_r(\overline{T},h_{n,\overline{T}}q+r)-L_r(\overline{T},r)\right)dq\Big]$$

$$+\frac{1}{\sqrt{h_{n,\overline{T}}}}[\int_{-\infty}^{\infty}\mathbf{K}(q)\left(\frac{\sigma^2(r)-\sigma^2(h_{n,\overline{T}}q+r)}{\sigma^2(h_{n,\overline{T}}q+r)\sigma^2(r)}\right)L_r(\overline{T},r)dq].$$

The second term is negligible as $n\to\infty$ (and $h_{n,\overline{T}}\to 0$). In fact,

$$\begin{aligned}&\frac{1}{\sqrt{h_{n,\overline{T}}}}[\int_{-\infty}^{\infty}\mathbf{K}(q)\left(\frac{\sigma^2(r)-\sigma^2(h_{n,\overline{T}}q+r)}{\sigma^2(h_{n,\overline{T}}q+r)\sigma^2(r)}\right)L_r(\overline{T},r)dq]\\ \leq\ &\frac{1}{\sqrt{h_{n,\overline{T}}}}[\int_{-\infty}^{\infty}\mathbf{K}(q)\left(\frac{const.h_{n,\overline{T}}q}{\sigma^2(h_{n,\overline{T}}q+r)\sigma^2(r)}\right)L_r(\overline{T},r)dq]\overset{a.s.}{\to}0\text{ when }h_{n,\overline{T}}\to 0.\end{aligned}$$

We now examine the first term

$$[\int_{-\infty}^{\infty}\mathbf{K}(q)\left(\frac{1}{\sigma^2(h_{n,\overline{T}}q+r)}\right)2\frac{1}{2\sqrt{h_{n,\overline{T}}}}\left(L_r(\overline{T},h_{n,\overline{T}}q+r)-L_r(\overline{T},r)\right)dq].$$

We use Lemma 3.5 in Part I, Section 3.

Lemma [Limit theory for the local time of a diffusion] *Let X satisfy the properties in Assumption 2.1, Section 2, Part I. Let r and a be fixed real numbers and treat $\{L_X(t,r+\frac{a}{\lambda})-L_X(t,r)\}$ as a double indexed stochastic process in (t,a). Then, as $\lambda\to\infty$*

$$\frac{1}{2}\sqrt{\lambda}\{L_X(t,r+\frac{a}{\lambda})-L_X(t,r)\}\overset{d}{\to}\mathfrak{B}(L_X(t,r),a)$$

where $\mathfrak{B}(t,a)$ is a standard Brownian sheet.

Then,

$$\begin{aligned}&[\int_{-\infty}^{\infty}\mathbf{K}(q)\left(\frac{1}{\sigma^2(r)+o_{a.s.}(1)}\right)2\frac{1}{2\sqrt{h_{n,\overline{T}}}}\left(L_r(\overline{T},h_{n,\overline{T}}q+r)-L_r(\overline{T},r)\right)dq]\\ &\overset{d}{\Rightarrow}2\int_{-\infty}^{\infty}\mathbf{K}(q)\left(\frac{1}{\sigma^2(r)}\right)\mathfrak{B}(L_r(\overline{T},r),q)dq\\ &\overset{d}{=}2\int_{-\infty}^{\infty}\mathbf{K}(q)\mathfrak{B}\left(\frac{1}{\sigma^2(r)}\overline{L}_r(\overline{T},r),q\right)dq\\ &\overset{d}{=}2\left(\frac{1}{\sigma(r)}\right)\left(\overline{L}_r(\overline{T},r)\right)^{1/2}\int_{-\infty}^{\infty}\mathbf{K}(q)\mathfrak{B}(1,q)dq\\ &\overset{d}{=}4\left(\frac{1}{\sigma(r)}\right)\left(\overline{L}_r(\overline{T},r)\right)^{1/2}\int_{0}^{\infty}\mathbf{K}(q)B(q)dq\\ &\overset{d}{=}4\left(\frac{1}{\sigma(r)}\right)\left(\overline{L}_r(\overline{T},r)\right)^{1/2}N\left(0,\int_0^{\infty}\int_0^{\infty}\min(s,q)\mathbf{K}(q)\mathbf{K}(s)dqds\right)\end{aligned}$$

This proves the stated result.

14.5. Proof of Theorem 11.3

We know that

$$\widehat{\overline{L}}_r(\overline{T},r) \stackrel{a.s.}{\to} \overline{L}_r(\overline{T},r)$$

provided $\frac{1}{h_{n,\overline{T}}}(\Delta_{n,\overline{T}})^{\alpha} = O(1)$ for some $\alpha \in (0,\frac{1}{2})$. Hence, a simple application of Slutsky's theorem gives

$$\widehat{\overline{H}}_r(\overline{T},r) = \frac{\widehat{\overline{L}}_r(\overline{T},r)}{\int_r^{\infty}\widehat{\overline{L}}_r(\overline{T},s)ds} \stackrel{a.s.}{\to} \frac{\overline{L}_r(\overline{T},r)}{\int_r^{\infty}\overline{L}_r(\overline{T},s)ds} = \overline{H}_r(\overline{T},r).$$

We now study the asymptotic distribution. First, we examine the term $\int_r^{\infty}\widehat{\overline{L}}_r(t,s)ds$.

$$\begin{aligned}
&\int_r^{\infty}\widehat{\overline{L}}_r(\overline{T},s)ds \\
&= \frac{\overline{T}}{nh_{n,\overline{T}}}\sum_{i=1}^{n}\int_r^{\infty}\mathbf{K}\Big(\frac{s-r_{i\Delta_{n,\overline{T}}}}{h_{n,\overline{T}}}\Big)ds \\
&= \frac{\overline{T}}{n}\sum_{i=1}^{n}\mathbb{K}\Big(\frac{r-r_{i\Delta_{n,\overline{T}}}}{h_{n,\overline{T}}}\Big) \\
&= \int_0^{\overline{T}}\mathbb{K}\Big(\frac{r-r_s}{h_{n,\overline{T}}}\Big)ds + o_{a.s.}\left(\frac{1}{h_{n,\overline{T}}}(\Delta_{n,\overline{T}})^{\frac{1}{2}-\epsilon}\right) \\
&= \int_{-\infty}^{\infty}\mathbb{K}\Big(\frac{r-a}{h_{n,\overline{T}}}\Big)\frac{1}{\sigma^2(a)}L(\overline{T},a)da + o_{a.s.}\left(\frac{1}{h_{n,\overline{T}}}(\Delta_{n,\overline{T}})^{\frac{1}{2}-\epsilon}\right) \\
&= \int_r^{\infty}\overline{L}(\overline{T},a)da + o_{a.s.}\left(\frac{1}{h_{n,\overline{T}}}(\Delta_{n,\overline{T}})^{\frac{1}{2}-\epsilon}\right)
\end{aligned}$$

The last equality derives from noticing that

$$\mathbb{K}\Big(\frac{r-a}{h_{n,\overline{T}}}\Big) = \int_{\frac{r-a}{h_{n,\overline{T}}}}^{\infty}\mathbf{K}(s)ds = \left\{\begin{array}{c} O(h_{n,\overline{T}}^{q+1}) \text{ for } r>a \\ 1-O(h_{n,\overline{T}}^{q+1}) \text{ for } r\le a\end{array}\right\}$$

where the order of magnitude depends on the features of the kernel function [c.f. Part I, Section 4, Assumption 4.1]. Hence,

$$\begin{aligned}
&\frac{1}{\sqrt{h_{n,\overline{T}}}}\left[\widehat{\overline{H}}_r(\overline{T},r) - \overline{H}_r(\overline{T},r)\right] \\
&= \frac{1}{\sqrt{h_{n,\overline{T}}}}\left[\frac{\widehat{\overline{L}}_r(\overline{T},r)}{\int_r^{\infty}\widehat{\overline{L}}_r(\overline{T},s)ds} - \frac{\overline{L}_r(\overline{T},r)}{\int_r^{\infty}\overline{L}_r(\overline{T},s)ds}\right] \\
&= \frac{1}{\sqrt{h_{n,\overline{T}}}}\left[\frac{\widehat{\overline{L}}_r(\overline{T},r)-\overline{L}_r(\overline{T},r)}{\int_r^{\infty}\overline{L}_r(\overline{T},s)ds + o_{a.s.}(1)}\right] \text{ since } \frac{1}{h_{n,\overline{T}}}(\Delta_{n,\overline{T}})^{\frac{1}{2}-\epsilon} \to 0 \text{ as } n\to\infty
\end{aligned}$$

Then, by a simple application of Theorem 11.2 it follows that

$$
\begin{aligned}
&= \frac{1}{\sqrt{h_{n,\overline{T}}}}\left(\frac{\widehat{\overline{L}}_r(\overline{T},r)-\overline{L}_r(\overline{T},r)}{\int_r^\infty \overline{L}_r(\overline{T},s)ds}\right) \\
&\overset{d}{\Rightarrow} 4MN\left(0,\left(\frac{1}{\sigma(r)}\right)^2 \frac{\mathbf{k}\overline{L}_r(\overline{T},r)}{\left(\int_r^\infty \overline{L}_r(\overline{T},s)ds\right)^2}\right) \\
&\overset{d}{=} 4MN\left(0,\left(\frac{1}{\sigma(r)}\right)^2 \frac{\mathbf{k}(\overline{H}_r(\overline{T},r))^2}{\overline{L}_r(\overline{T},r)}\right)
\end{aligned}
$$

where $\mathbf{k} = \int_0^\infty \int_0^\infty \min(s,q)\mathbf{K}(s)\mathbf{K}(q)dsdq$. This proves the stated result.

14.6. Proof of Theorem 11.4

See Part I, Section 5, Theorem 5.5.

14.7. Proof of Theorem 11.5

See Part I, Section 5, Theorem 5.6

14.8. Proof of Theorem 11.6

See Part I, Section 5, Theorem 5.11

14.9. Proof of Theorem 11.7

See Part I, Section 5, Theorem 5.12

15. Notation

$\rightarrow_{a.s.}$	almost sure convergence
$\rightarrow_p$	convergence in probability
$\Rightarrow, \rightarrow_d$	weak convergence
$:=$	definitional equality
$o_p(1)$	tends to zero in probability
$O_p(1)$	bounded in probability
$o_{a.s.}(1)$	tends to zero almost surely
$O_{a.s.}(1)$	bounded almost surely
$=_d$	distributional equivalence
$\sim_d$	asymptotically distributed as
$MN(0,V)$	mixed normal distribution with variance V
$\mathbf{1}_A$	indicator function for the set A

Table 1. Summary of the features of the data set

Source	**Bank of America 7-Day Eurodollar (midpoint bid-ask)**
Frequency	Daily
Sample Period	1 June 1973-25 February 1995
Sample Size	5505 observations
Type	Continuously compounded yield-to-maturity (annualized rate)

Table 2. Traditional descriptive statistics for the data set in Table 1

	Spot Interest Rate	**First Differences**
Mean	0.0836	-0.0000035
Standard Deviation	0.0359	0.004070
Daily $\rho(1)$	0.9936	-0.2710
$\rho(2)$	0.9908	-0.0347
$\rho(3)$	0.9883	-0.0377
$\rho(4)$	0.9863	0.0297
$\rho(5)$	0.9839	-0.1789
$\rho(10)$	0.9779	-0.0173

Monthly autocorrelations are reported in Aït-Sahalia (1996a,b)

Table 3. Summary of the results of two nonstationarity tests for the series in Table 1. We implement the augmented Dickey-Fuller test (ADF) and the Z tests in Phillips (1987). We consider a constant and a trend in the fitted regression.

Constant and trend in the fitted regression					
	Auto parameter	Test statistic	1% value	5% value	10% value
ADF test	0.9963	-2.3447	-3.9978	-3.4318	-3.1617
Z(a) test	0.9923	-19.4572	-28.9388	-21.2162	-17.9117
Z(t) test	0.9923	-3.1383	-3.9978	-3.4318	-3.1617
Automatic window width					
Z(a) test	0.9923	-21.0867	-28.9388	-21.2162	-17.9117
Z(t) test	0.9923	-3.2655	-3.9978	-3.4318	-3.1617

Note: In the ADF test the number of lagged first differences in the fitted regression is equal to 5. In the Z(a) and Z(t) tests the number of autocovariance terms to compute the spectrum at frequency zero is equal to 5.

Table 4. Summary of the results of two nonstationarity tests for the series in Table 1. We implement the augmented Dickey-Fuller test (ADF) and the Z tests in Phillips (1987). We consider a constant in the fitted regression.

Constant in the fitted regression					
	Auto parameter	Test statistic	1% value	5% value	10% value
ADF test	0.997	-2.0918	-3.4583	-2.8710	-2.5936
Z(a) test	0.9935	-16.2600	-19.8270	-13.7251	-11.0755
Z(t) test	0.9935	-2.8608	-3.4583	-2.8710	-2.5936
Automatic window width					
Z(a) test	0.9935	-17.6300	-19.8270	-13.7251	-11.0755
Z(t) test	0.9935	-2.9781	-3.4583	-2.8710	-2.5936

Note: In the ADF test the number of lagged first differences in the fitted regression is equal to 5. In the Z(a) and Z(t) tests the number of autocovariance terms to compute the spectrum at frequency zero is equal to 5.

Table 5. We estimate a linear mean-reverting model for the drift in the CEV class as in Chan, Karolyi, Longstaff and Sanders (1992). We impose various restrictions on the parameters (bold figures) to obtain conventional parametric models for the short-term interest rate process.

	$\alpha(0)$	$\alpha(1)$	$\gamma(0)^2$	$\gamma(1)$	χ^2	d.f.
Model						
Unrestricted CEV	0.1320	-1.5918	4.1295	1.49		
	(0.0494)	(0.6981)	(2.0264)	(0.1095)		
CIR	0.0776	-0.8549	0.0323	**0.5**	38.37	1
	(0.0488)	(0.6901)	(0.0023)		[5.83e−0.10]	
Vasicek	0.0621	-0.6614	0.0021	**0**	59	1
	(0.0488)	(0.6903)	(0.0001)		[1.5e−0.14]	
Restricted CEV	**0**	0.2375	3.3289	1.44	6.98	1
		(0.1340)	(1.7105)	(0.113)	[0.0082]	
Brennan and Schwartz	0.1035	-1.1919	0.4136	**1**	14.27	1
	(0.0489)	(0.6911)	(0.0279)		[0.000158]	

Notes: The parameters are estimated by GMM [we follow CKLS (1992) in the implementation, see text]. Standard deviations are in parentheses. The results of Hansen's (1982) test of overidentifying restrictions are reported. P-values are in squared brackets.

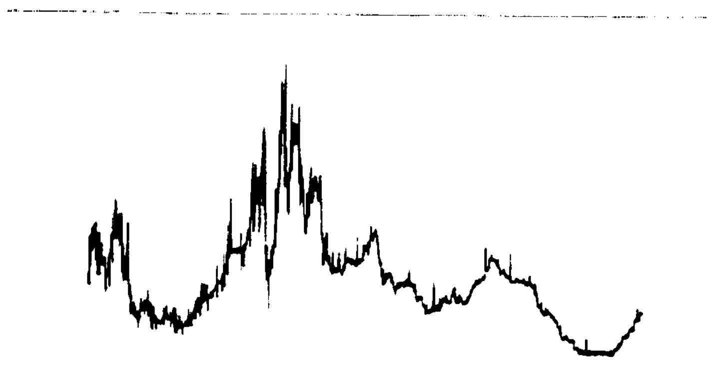

Figure 1: Graph of the series.

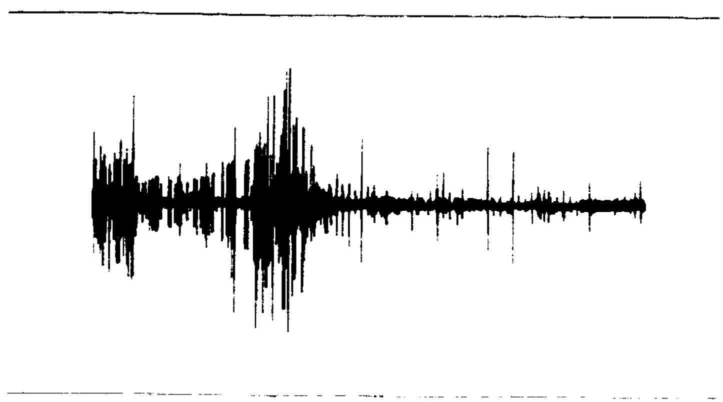

Figure 2: Graph of the first differences of the series.

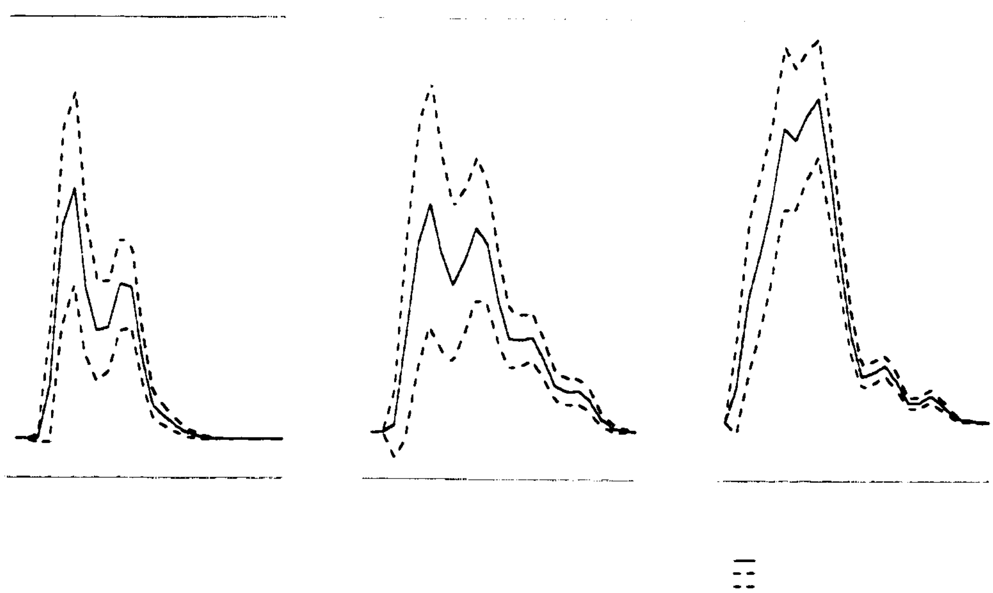

Figure 3: Functional estimates of the local time process.

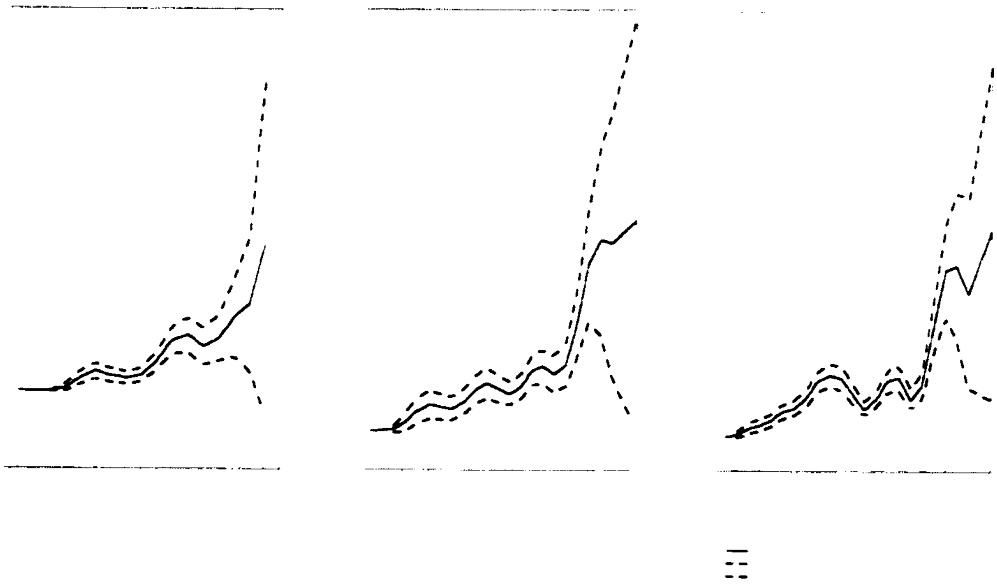

Figure 4: Functional estimates of the hazard process.

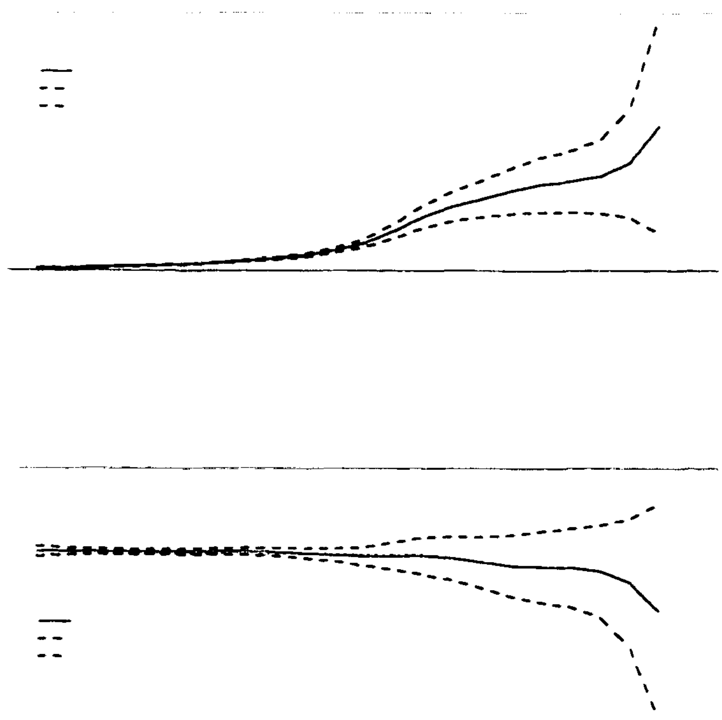

Figure 5: Functional estimates of drift and diffusion function.

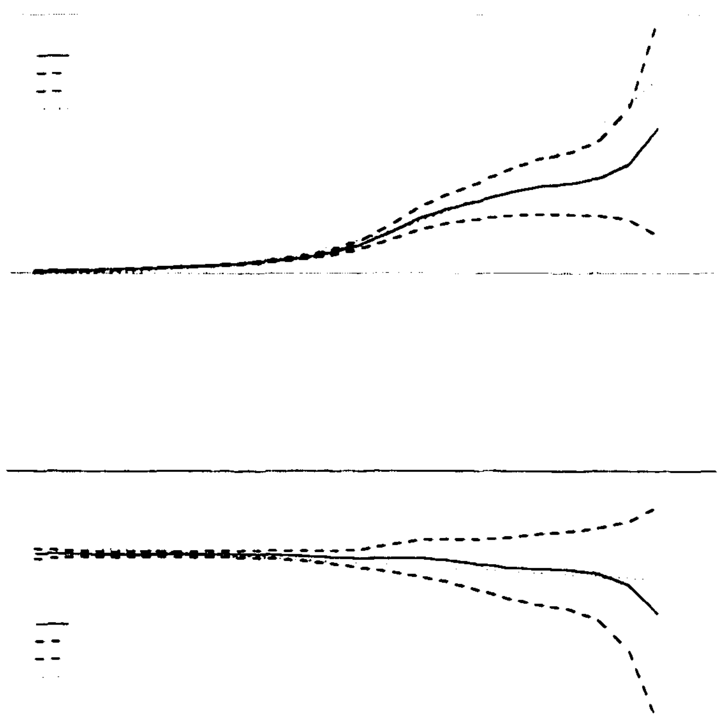

Figure 6: Comparison between the functional estimates and a parametric linear mean-reverting model for the drift in the CEV class as in Chan, Karolyi, Longstaff and Sanders (1992).

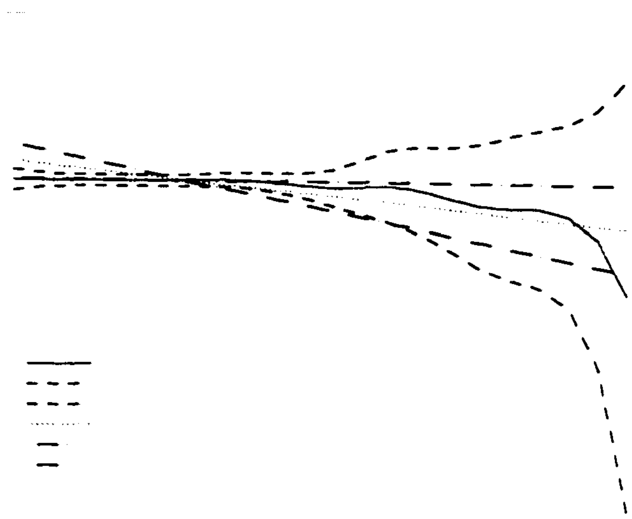

Figure 7: Comparison between the functional drift and a linear mean-reverting drift as in Chan, Karolyi, Longstaff and Sanders (1992).

Part III
Fully Nonparametric Estimators for Diffusion Models: a Small Sample Analysis [with Thong H. Nguyen]

16. Introduction

Stochastic differential equations play an important role in modeling economic time series. They are largely used in continuous-time finance, for example. Unfortunately, even in finance and for many processes of interest, relatively little is known about how to correctly parametrize the functions that describe the solution to the stochastic differential equation of interest, that is the infinitesimal first and second moments [$\mu(.)$ and $\sigma(.)$ in (17.1) below]. This issue has led many researchers to design estimation methods that do not rely on the necessity of imposing a parametric structure up front.

Fully functional estimation procedures have been recently proposed by Jiang and Knight (1997) [JK, henceforth], Stanton (1997) and Bandi and Phillips [BP, hereafter] (c.f. procedure in Part I).[30] While their limiting properties have been discussed at length [c.f. Florens-Zmirou (1993), JK (1997), Bandi (1999) and Part I in this thesis], no work has been done on the analysis of their performance in finite sample.[31] We believe this issue is particularly important due to the widespread use of continuous-time modeling in economics, and especially in finance. In consequence, the goal of the present chapter is to investigate the finite sample properties of the above-mentioned estimators in the presence of several simulated underlying processes.

We address six main questions.

[1] How well alternative methods capture the main features of the functions of interest, that is drift and diffusion [i.e. analysis of the small sample bias]?

[2] How volatile are the finite sample estimates?

[3] How well the asymptotic theories approximate the finite sample distributions?

[4] How important is the choice of the kernel?

[30] In this chapter, the definitions "the BP estimators" and "the estimators in Part I" are used interchangeably.

[31] In a recent paper, Jiang and Knight (1997) compare their functional approach to existing *parametric* methods in the literature.

[5] How important is the choice of the bandwidth(s)?

[6] How crucial are some of the statistical properties of the underlying processes, such as stationarity and temporal persistence?

The main results of our analysis can be briefly outlined as follows.

We confirm that the infinitesimal volatility of a process is easier to identify than the infinitesimal first moment in finite sample. Furthermore, we stress that identification of the drift is intimately related to the appropriate choice of the bandwidth parameter(s). We know that for estimators that are robust to deviations from stationarity [c.f Stanton (1997) and Part I], the admissible bandwidth for diffusion estimation can be theoretically smaller than the admissible bandwidth for drift estimation [c.f. Bandi (1999) and Part I for discussions]. The reason is that local information suffices for consistent estimation of the diffusion but is not sufficient for consistent estimation of the drift [c.f. Aït-Sahalia (1996a) and Part I, for instance]. This reality is confirmed by our simulations. Consistently with Bandi (1999), we point out that an appropriately chosen, larger (than for diffusion estimation) window width for the drift generally would not determine oversmoothing. Rather, it would help capture the salient features of the infinitesimal first moment of the underlying process. In effect, for most of the processes examined in this chapter [i.e. non explosive and persistent processes], we expect the drift to be fairly smooth and quite flat. This, in turn, implies that the extent of the estimated nonlinearities in the finance literature on the estimation of the short-term interest rate process [c.f. Stanton (1997), for instance] can be partly induced by erroneous choices of the smoothing parameter(s) causing undersmoothing [c.f. Bandi (1999) for a discussion of this point]. In effect, nonexplosion and persistence are typical features of US interest rate data. Undersmoothing can then be invoked as a explanation for the estimated nonlinearities that complements alternative theories in the literature [c.f. Jones (1997), Chapman and Pearson (1998) and Part II in this work].

Theoretically, the drift estimation of processes which revert towards the middle of their stationary distribution at a fast pace [low persistence processes] requires optimal bandwidths whose magnitude is close to the optimal magnitude in the case of diffusion estimation. Below, we discuss the technical reasons for this result. A simple intuition is sufficient here. Processes whose speed of return to the long run mean is fast, display a fairly steep, negatively sloped drift function. Excessively large bandwidths can oversmooth the drift in this case and introduce too much bias. This observation implies that a window width for the drift that is roughly as large as (or slightly larger than) the one used for diffusion estimation can be justifiable in the presence of low persistence. For highly persistent processes a substantially larger smoothing parameter is needed due to the flat nature of the infinitesimal first moment across points in the range of the process.

Being the drift estimator proposed by JK (1997) intimately related to the estimation of the marginal density of the process and the diffusion function, the theoretical requirements on the underlying process that are needed for this method to be well defined (i.e. stationarity) are tighter than in Stanton (1997) and BP. In the presence of stationarity, this estimator tends to underperform slightly the alternative methods analyzed here for comparable choices of the window widths. On the other hand, it does not lead to excessively misleading inference when stationarity is not satisfied.

The potential need for a larger smoothing parameter for the drift when straight sample analogues to conditional expectations are used as in Stanton (1997) and BP lies at the heart of the potential trade-off between optimal bandwidth for the drift and optimal bandwidth for the diffusion coefficient. An excessively large window width might oversmooth the diffusion but be suitable for the drift. Unfortunately, the asymptotic condition that the admissible smoothing parameter for the drift needs to satisfy depends on the stochastic properties of the underlying continuous-time process, through its *chronological local time*. This is a random quantity that provides an assessment of the time that the process spends in the vicinity of every spatial point [c.f. Phillips and Park (1998), Part I and Part II, for example]. In particular, the rate of convergence to zero of the window width depends on the rate of divergence to infinity of the local time of the process. The later can not be assessed in closed form apart from few specific processes, such as Brownian motion [c.f. Part I]. This implies a fundamental difficulty in choosing the correct smoothing parameter for the drift. The same difficulty does not occur when estimating the diffusion function. Volatility can be identified locally. Hence, the stochastic properties of the process do not play a vital role.

Interestingly, though, the implementation of estimation procedures based on double-smoothing as in Part I seems to improve the above mentioned trade-off, thus making the need for finding the correct window width for the drift less compelling. The idea is very simple. Sample analogues to conditional expectations based on *convoluted* kernel functions can achieve in finite sample the level of smoothing for the drift that weighted averages based on *simple* kernels would guarantee with relatively more appropriate choices of the bandwidth. Of course, double-smoothing might impose a cost in terms of oversmoothed second moments, but our finite sample analysis suggests that the benefit outweighs the cost considerably.

Finally, the asymptotic theories generally reproduce quite accurately the characteristics of the finite sample distributions. Even though a bias can be present, small sample distributions are sufficiently close to normals with variances that are numerically similar to the variances that the limit theories predict, mainly at spatial points that are often visited by the sample process. As expected, the larger is the number of observations at a point, the smaller is the bias of the estimated drift and diffusion function at that point and the closer is the finite sample distribution to its asymptotic counterpart. This result generally holds across estimators and underlying processes.

This chapter is organized as follows. Section 17 discusses two alternative assumptions on the underlying continuous-time process. The asymptotic features of the estimators that we examine in this work depend on the validity of either one of these two assumptions. In Section 18 we describe the estimators in JK (1997) and Stanton (1997) and provide an outline of their limiting properties. Section 19 illustrates five simulated process. We simulate processes that have been, or could be, employed as descriptions of the short-term interest rate process in continuous-time finance. This will allow us to comment on the practical implications of our findings and relate our results to previous results on the empirical estimation of diffusions. In Section 20 we discuss the simulation exercises. Section 21 concludes. Proofs are in Section 22. Notation is in Section 23.

17. The model

As in Part I, we consider the process $\{X_t;\, t \geq 0\}$ that is the solution to the homogeneous stochastic differential equation (SDE)

$$dX_t = \mu(X_t)dt + \sigma(X_t)dB_t, \tag{17.1}$$

with initial condition $X_0 = \overline{X}$. B_t is a standard Brownian motion defined on the filtered probability space $(\Omega, \Im^B, (\Im_t^B)_{t\geq 0}, P)$. The initial condition $\overline{X} \in L^2$ and is taken to be independent of $\{B_t : t \geq 0\}$. We define the left-continuous filtration

$$\overline{\Im}_t := \sigma(\overline{X}) \vee \Im_t^B = \sigma(\overline{X}, B_s; 0 \leq s \leq t) \qquad 0 \leq t < \infty$$

and the collection of null sets

$$\aleph := \{N \subseteq \Omega; \exists G \in \overline{\Im}_\infty \text{ with } N \subseteq G \text{ and } P(G) = 0\}.$$

We create the *augmented* filtration

$$\tilde{\Im}_t^X := \sigma(\overline{\Im}_t \cup \aleph) \qquad 0 \le t < \infty.$$

We impose Assumption 17.1 and Assumption 17.2, below, in the study of (17.1). Both Assumption 17.1 and Assumption 17.2 ensure the existence and pathwise uniqueness of a nonexplosive solution to (17.1) that is adapted to the *augmented* filtration $\{\tilde{\Im}_t^X\}$. Assumption 17.2 guarantees stationarity by ensuring the existence of a time-invariant stationary distribution for the process $\{X_t; t \ge 0\}$ and by making the process start out in the stationary distribution.

Assumption 17.1

(A) *$\mu(\cdot)$ and $\sigma(\cdot)$ are time-homogeneous, $\mathfrak{B}$-measurable functions on $\mathfrak{D} = (l, u)$ with $-\infty \le l < u \le \infty$ where $\mathfrak{B}$ is the σ-field generated by Borel sets on $\mathfrak{D}$. Both functions are at least once continuously differentiable. Hence, they satisfy local Lipschitz and growth conditions. Thus, for every compact subset $J = [1/H, H]$ with $H > 0$ of the range of the process, there exist constants C_1 and C_2 such that, for all x and y in J,*

$$|\mu(x) - \mu(y)| + |\sigma(x) - \sigma(y)| \le C_1|x - y|,$$

and

$$|\mu(x)| + |\sigma(x)| \le C_2\{1 + |x|\}.$$

(B) *$\sigma^2(\cdot) > 0$ on $\mathfrak{D}$.*

(C) [Feller's (1952) necessary and sufficient condition for nonexplosion]. *We define $V(\alpha)$ as*

$$\int_0^\alpha S'(y)\{\int_0^y \left[\frac{2}{S'(x)\sigma^2(x)}\right] dx\}dy$$

where $S'(x)$ is the first derivative of the natural scale measure,

$$S(\alpha) = \int_0^\alpha \exp\{\int_0^y \left[-\frac{2\mu(x)}{\sigma^2(x)}\right] dx\}dy.$$

We require $V(\alpha)$ to diverge at the boundaries of $\mathfrak{D}$, i.e.

$$\lim_{\alpha \to l^+} V(\alpha) = \lim_{\alpha \to u^-} V(\alpha) = \infty.$$

As discussed in Part I, under Assumption 17.1 the stochastic differential equation has a strong solution X_t that is unique, recurrent and continuous in $t \in [0, T]$. X_t satisfies

$$X_t = \overline{X} + \int_0^t \mu(X_s)ds + \int_0^t \sigma(X_s)dB_s$$

a.s., with $\int_0^T E[X_t^2]dt < \infty$.

Assumption 17.2

(A') *is defined as* (A) *above.*

(B') *is defined as* (B) *above.*

(C') *Let $M(\alpha)$, the speed measure, defined as*

$$\int_0^\alpha \frac{1}{\sigma^2(y)} \exp\{\int_0^y \left[\frac{2\mu(x)}{\sigma^2(x)}\right] dx\}dy$$

converge at both boundaries of $\mathfrak{D}$. Further, let $S(\alpha)$, the natural scale measure defined in Assumption 17.1 (C), diverge at both boundaries of $\mathfrak{D}$.

(D') *X_0 is distributed as π_0, the time-invariant density of the process.*

17.3 Remark (C') imposes a structure on the solution to the stochastic differential equation which is substantially different from the structure determined by (C). The condition on the speed measure guarantees the existence of a time-invariant marginal density for X_t. Under Assumption 17.1 the process might not have a stationary distribution function. Further, contrary to the conditions on $V(\alpha)$ that are necessary and sufficient for nonexplosion, the conditions on the natural scale measure are only sufficient for nonexplosion. In fact, the following implications can be easily derived [c.f. Karatzas and Shreve (1991, Problem 5.5.27, page 348)]:

$$S(l^+) = -\infty \Rightarrow V(l^+) = \infty$$

and

$$S(u^-) = +\infty \Rightarrow V(u^-) = \infty.$$

Under Assumption 17.2, the stochastic differential equation has a strong solution X_t that is unique, stationary, recurrent and continuous in $t \in [0, T]$ [c.f. Aït-Sahalia (1996a,b) for a discussion]. X_t satisfies

$$X_t = \overline{X} + \int_0^t \mu(X_s)ds + \int_0^t \sigma(X_s)dB_s$$

a.s., with $\int_0^T E[X_t^2]dt < \infty$.

17.1. The objects of interest

We are interested in the functions $\mu(.)$ and $\sigma^2(.)$ in (17.1) above. They represent the first and the second moment of the infinitesimal conditional distribution of X_t. More specifically,

$$E\{X_{t+h} - X_t | X_t\} = \mu(X_t)h + o(h) \tag{17.2}$$

and

$$E\{(X_{t+h} - X_t)^2 | X_t\} = \sigma^2(X_t)h + o(h) \tag{17.3}$$

where h is an arbitrarily small time step and $o(1)$ is a standard order symbol such that $o(h)$ denotes a function converging to zero at a faster rate than h.

The functions $\mu(.)$ and $\sigma^2(.)$ drive the dynamics of the solution to the SDE (17.1) since the transition density is written as a function of both $\mu(.)$ and $\sigma^2(.)$. In fact, the transition density $\pi(X_t = x | X_0 = x_0)$ is the unique solution to both the *Kolmogorov backward equation,*

$$\frac{\partial \pi(X_t = x|X_0 = x_0)}{\partial t} = \mu(x_0)\frac{\partial \pi(X_t = x|X_0 = x_0)}{\partial x_0} + \frac{1}{2}\sigma^2(x_0)\frac{\partial^2 \pi(X_t = x|X_0 = x_0)}{\partial^2 x_0} \tag{17.4}$$

and the *Kolmogorov forward* (or *Fokker-Plank*) *equation,*

$$\frac{\partial \pi(X_t = x|X_0 = x_0)}{\partial t} = -\frac{\partial(\mu(x)\pi(X_t = x|X_0 = x_0))}{\partial x} + \frac{1}{2}\frac{\partial^2(\sigma^2(x)\pi(X_t = x|X_0 = x_0))}{\partial^2 x} \tag{17.5}$$

with initial condition $\pi(X_0 = x|X_0 = x_0) = \delta(x - x_0)$ where δ is the Dirac delta function.

17.4 Remark A heuristic derivation of the Kolmogorov backward equation is instructive and can be easily laid out. We follow Chang (1999). Let the function $\Phi(x_0, t)$ be $E^{x_0}[f(X_t)]$. We can write

$$\Phi(x_0, t+h) = E^{x_0} f(X_{t+h}) = E^{x_0} E^{x_0}[f(X_{t+h})|X_h] = E^{x_0}\Phi(X_h, t).$$

Hence,

$$\begin{aligned}
\frac{\partial \Phi(x_0,t)}{\partial t} &= \lim_{h\to 0}\frac{1}{h}[\Phi(x_0, t+h) - \Phi(x_0,t)] \\
&= \lim_{h\to 0}\frac{1}{h}[E^{x_0}\Phi(X_h,t) - \Phi(x_0,t)] \\
&= \lim_{h\to 0}\frac{1}{h}E^{x_0}[\frac{\partial \Phi(x_0,t)}{\partial x_0}(X_h - x_0) + \frac{1}{2}\frac{\partial^2\Phi(x_0,t)}{\partial^2 x_0}(X_h - x_0)^2 + o(h)] \\
&= \mu(x_0)\frac{\partial \Phi(x_0,t)}{\partial x_0} + \frac{1}{2}\sigma^2(x_0)\frac{\partial^2\Phi(x_0,t)}{\partial^2 x_0}. \qquad (17.6)
\end{aligned}$$

If $x_0 = X_t$, then

$$\begin{aligned}
\mathcal{A}[f(X_t)] &= \frac{\partial E^{x_t}[f(X_t)]}{\partial t} \\
&= \frac{\partial f(X_t)}{\partial t} \\
&= \lim_{h\to 0}\frac{E[f(X_{t+h})|X_t] - f(X_t)}{h} \\
&= \mu(X_t)\frac{\partial f(X_t)}{\partial X_t} + \frac{1}{2}\sigma^2(X_t)\frac{\partial^2 f(X_t)}{\partial^2 X_t}. \qquad (17.7)
\end{aligned}$$

where $\mathcal{A}$ is the infinitesimal generator associated with the strong Markov process $\{X_t; t \geq 0\}$. We will refer to this important concept later on. Suppose now that the function $f(X_t)$ is equal to $\mathbf{1}\{X_t \leq y\}$ for a fixed y. Then,

$$\Phi(x_0, t) = E^{x_0} f(X_t) = P^{x_0}\{X_t \leq y\} = F(t, x_0, y)$$

where $F(t, x_0, y)$ is the conditional cumulative distribution function of X_t at y. $F(t, x_0, y)$ can be plugged into (17.6). We can subsequently differentiate both sides with respect to y and obtain the required result, that is equation (17.4).

Under Assumption 17.2 the solution to the SDE (17.1) is a stationary process. Drift and diffusion function completely describe the time-invariant stationary distribution of the process which is given by

$$\pi_0(x) = \frac{C_3}{\sigma^2(x)} \exp\{\int_{x^*}^{x} \frac{2\mu(u)}{\sigma^2(u)} du\}. \tag{17.8}$$

The choice of the lower bound of integration x^* in the interior of $\mathfrak{D}$ is irrelevant. It only affects the choice of the constant C_3 which is determined to guarantee that the density integrates to one.

Formula (17.8) can be derived from the forward Kolmogorov equation (17.5) [c.f. Karlin and Taylor (1981, 15.6.22)]. It implies that

$$\mu(X_t) = \frac{1}{2\pi(X_t)} \frac{\partial[\sigma^2(X_t)\pi(X_t)]}{\partial X_t} \tag{17.9}$$

and

$$\sigma^2(X_t) = \frac{2}{\pi(X_t)} \int_0^{X_t} \mu(u)\pi(u)du. \tag{17.10}$$

It is worth emphasizing that the assumption of stationarity (Assumption 17.2 above) determines the possibility of expressing either one of the two objects of interest as a function of the time-invariant distribution of the process and the other function. This is of substantial help if we are interested in estimating continuous systems by use of discretely sampled data [c.f. discussion in Part II, Section 10]. We will return to this issue later on.

More general expressions for $\mu(.)$ and $\sigma^2(.)$ which are valid for stationary and nonstationary processes can be easily derived from (17.2) and (17.3), namely

$$\mu(X_t) = \lim_{h \to 0} \frac{1}{h} E\{X_{t+h} - X_t | X_t\} \tag{17.11}$$

and

$$\sigma^2(X_t) = \lim_{h \to 0} \frac{1}{h} E\{(X_{t+h} - X_t)^2 | X_t\}. \tag{17.12}$$

18. Nonparametric estimators

This section is devoted to the discussion of the fully nonparametric estimators for drift and diffusion function in JK (1997) and Stanton (1997). As for the BP estimators, we refer the reader to the discussions in Part I and Part II. As a caveat, we will not dwell on asymptotic results since they are readily available elsewhere [c.f. Florens-Zmirou (1993), Jiang and Knight (1997) and Bandi (1999)]. We will simply provide a brief illustration of them and refer to the relevant theorems in the original papers.[32] Further, we will comment on the features of the estimators that are relevant for a complete understanding of their finite sample properties.

The symbols below have the usual interpretation in standard nonparametric analysis, that is $\mathbf{K}(.)$ is a smooth kernel function and h is a bandwidth controlling the amount of smoothing [c.f. Härdle (1990)]. We impose Assumption 4.1 in Part I, Section 4, on the kernel function.

18.1. Jiang and Knight (1997)

We assume that we observe X_t at $\{t = t_1, t_2, .., t_n\}$ in the fixed time interval $[0, \overline{T}]$, with $\overline{T} \geq T_0 > 0$, where T_0 is a positive constant. We also assume equispaced data. So, $\{X_t = X_{\Delta_n}, X_{2\Delta_n}, X_{3\Delta_n}, ..., X_{n\Delta_n}\}$ are n observations at $\{t_1 = \Delta_n, t_2 = 2\Delta_n, t_3 = 3\Delta_n, ..., t_n = n\Delta_n\}$, where $\Delta_n = \overline{T}/n$. The diffusion function estimator is defined as

$$\widehat{\sigma}^2_{(n)}(x) = \frac{1}{\Delta_n} \frac{\sum_{i=1}^{n-1} \mathbf{K}(\frac{X_{i\Delta_n} - x}{h_n})[X_{(i+1)\Delta_n} - X_{i\Delta_n}]^2}{\sum_{i=1}^{n} \mathbf{K}(\frac{X_{i\Delta_n} - x}{h_n})}. \tag{18.1}$$

From (17.9), the functional drift estimator is

$$\widehat{\mu}_{(n)}(x) = \frac{1}{2}\left[\frac{\partial \widehat{\sigma}^2_{(n)}(x)}{\partial x} + \widehat{\sigma}^2_{(n)}(x)\frac{\frac{\partial \widehat{\pi}_{(n)}(x)}{\partial x}}{\widehat{\pi}_{(n)}(x)}\right], \tag{18.2}$$

where $\widehat{\pi}_{(n)}(x) = \frac{1}{nh_n}\sum_{i=1}^{n} \mathbf{K}(\frac{X_{i\Delta_n} - x}{h_n})$ is an estimate of the stationary distribution of the process.

18.1 Remark The intuition behind (18.1) and (18.2) is simple. The diffusion function has a lower order of magnitude than the drift function for infinitesimal time changes. Hence, the local dynamics of the underlying process reflect more of the characteristics of the diffusion than those of the drift. In consequence, the diffusion function can be estimated consistently over a fixed time span provided the observation frequency is high, even in situations where the drift is treated as a nuisance parameter [as in (17.12) and its sample analog (18.1)]. This idea is contained in a paper by Florens-Zmirou (1993) where (18.1) is originally suggested with $\mathbf{K}(.)$ being replaced by a discontinuous indicator function.

[32]For completeness and clarity, though, Section 22 reports the proofs of the theorems contained in the unpublished paper by Bandi (1999).

The drift term can not be estimated over a fixed time span unless cross-restrictions relying on the knowledge of the time-invariant density of the process are imposed [c.f. Aït-Sahalia (1996a), JK (1997), Part I and Part II, for instance]. Assumption 17.2 needs to hold to be able to construct the drift estimator based on (17.9). In fact, under Assumption 17.2, formula (18.2) defines a consistent estimator of the true theoretical function (17.9) [which, in turn, is equal to (17.11)].

Limit theory for the diffusion function estimator [c.f. Bandi (1999, Theorem 2)] *Provided $n \to \infty$ [the number of observations increases over a fixed time span], $h_n \to 0$ and $n^{1-\varepsilon}h_n^2 \to \infty$ with ε arbitrarily small, the estimator is consistent almost surely. If $nh_n^4 \to 0$,[33] then we have convergence to a mixture of normal law (MN), depending on the chronological local time of the underlying process [see Part I and Part II], i.e.*

$$\sqrt{nh_n}\left(\widehat{\sigma}^2_{(n)}(x) - \sigma^2(x)\right) \xrightarrow{d} MN\left(0, 4\left(\int_{-\infty}^{\infty} \mathbf{K}^2(s)ds\right)\frac{\sigma^4(x)}{\overline{L}_X(\overline{T},x)/\overline{T}}\right),$$

where $\overline{L}_X(\overline{T},x) = \frac{1}{\sigma^2(x)}\lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\int_0^{\overline{T}} \mathbf{1}_{[x,x+\varepsilon[}(X_s)d[X]_s$.

The conditions on the bandwidth can be easily rewritten. If $h_n \propto n^{-k}$, then $k \in (\frac{1}{4}, \frac{1}{2})$ for consistency and weak convergence.

Consistency of the drift is proved in JK (1997, Corollary to Theorem 2, page 627) by virtue of Slutsky's theorem and provided the estimators of the marginal density function and its first derivative are consistent for the theoretical functions, but this is a standard result in nonparametric statistics [see Silverman (1986) for a general discussion and Aït-Sahalia (1996a,b) for empirical applications in finance]. In effect, if $nh_n \to \infty$ and $nh_n^5 \to 0$, then the distribution function estimator is consistent for the true function and asymptotically normal. The conditions that ensure consistency and asymptotic normality of the estimator of the first derivative of the marginal density function are $nh_n^3 \to \infty$ and $nh_n^5 \to 0$. The above assumptions can be rewritten in the form $h_n \propto n^{-k}$ with $k \in (\frac{1}{5}, 1)$ for the marginal density estimator and $k \in (\frac{1}{5}, \frac{1}{3})$ for the estimator of the first derivative of the distribution function [c.f. Jiang (1998)]. Weak convergence of the drift estimator to a Gaussian distribution follows from applying the delta method to (18.2)

[33] If $nh_n^4 \to \infty$, then weak convergence still holds, but the limiting distribution of the diffusion function estimator is driven by the bias term in the estimation error decomposition [c.f. Bandi (1999, Theorem 2)].

18.2. Stanton (1997)

X_t is recorded at $\{t = t_1, t_2, .., t_n\}$ in the time interval $[0, T]$, with $T \geq T_0 > 0$, where T_0 is a positive constant. We assume equispaced data. Hence,

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

are n observations at

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

where $\Delta_{n,T} = T/n$. The diffusion function estimator is

$$\widehat{\sigma}^2_{(n,T)}(x) = \frac{1}{\Delta_{n,T}} \frac{\sum_{i=1}^{n-1} \mathbf{K}(\frac{X_{i\Delta_{n,T}} - x}{h_{n,T}})[X_{(i+1)\Delta_{n,T}} - X_{i\Delta_{n,T}}]^2}{\sum_{i=1}^{n} \mathbf{K}(\frac{X_{i\Delta_{n,T}} - x}{h_{n,T}})} \tag{18.3}$$

The drift function estimator is

$$\widehat{\mu}_{(n,T)}(x) = \frac{1}{\Delta_{n,T}} \frac{\sum_{i=1}^{n-1} \mathbf{K}(\frac{X_{i\Delta_{n,T}} - x}{h_{n,T}})[X_{(i+1)\Delta_{n,T}} - X_{i\Delta_{n,T}}]}{\sum_{i=1}^{n} \mathbf{K}(\frac{X_{i\Delta_{n,T}} - x}{h_{n,T}})}. \tag{18.4}$$

18.2 Remark Stanton's approach is presented in the original paper [Stanton (1997)] as a methodology based on approximations to the true functions. The approximations can be made finer and finer using the infinitesimal generator of the process (17.7). As already commonplace in the finance literature, we will only analyze the performance of estimates of the first order approximations, that is (18.3) and (18.4) above. Theoretically, it is entirely appropriate to restrict ourselves to the first order approximations since they are proven to be consistent for the true functions and have nice asymptotic properties as $T \to \infty$, $n \to \infty$ and $\frac{T}{n} \to 0$ [c.f. Bandi (1999)].

18.3 Remark As in Part I, Bandi (1999) derives the limit theory assuming that the time span becomes larger (i.e. *long span* asymptotics) as the distance between adjacent observations decreases (i.e. *infill* asymptotics). The implementation of *infill* asymptotics is necessary to estimate consistently continuous processes using fully functional methods. Enlarging the time span is crucial to be able to identify the drift while avoiding the need to impose stationarity on the underlying process [c.f. Remark 18.1 above, and Part II for a discussion]. All that is needed to identify the drift is to assume that the process is recurrent, as it is given Assumption 17.1. Note that (18.4) is defined as a straight sample analogue to (17.11). Since the theoretical drift is simply a conditional expectation, we wish to be able to visit the conditioning level [or an arbitrarily small neighborhood of it] a large number of times over time [in the limit an infinite number of times] and recurrence ensures that this holds asymptotically almost surely. Clearly, the recurrence of the process can not be exploited over a fixed time span ($\overline{T}$). Hence, recurrence is not a necessary condition for diffusion estimation since the diffusion function can be identified over a fixed time span provided the number of observations increases [technically, the infinitesimal volatility is part of the quadratic variation of the process and the quadratic variation is defined over a fixed time span].

18.4 Remark Note that the BP estimators in Part I can be regarded as the product of a general approach to the functional estimation of diffusions that encompasses specifications based on simple smoothing of which the Stanton's approach is an example. The relative merits of double-smoothing in finite sample will be one of the objects of our investigation.

The details of the asymptotic results are contained in Bandi (1999). We provide a concise illustration here and report proofs in Section 22.

Limit theory for the diffusion function estimator [Bandi (1999, Theorem 2)] *If $n \to \infty$, $T \to \infty$, $h_{n,T} \to 0$ (as $n, T \to \infty$) such that $\frac{\Delta_{n,T}}{h_{n,T}} \to 0$ and $\frac{\overline{L}_X(T,x)}{h_{n,T}}(\Delta_{n,T})^{\alpha} = O_{a.s.}(1)$ for some $\alpha \in (0, \frac{1}{2})$, than the estimator (18.3) is consistent almost surely for the true function. Further, if $\frac{h^4_{n,T}}{\Delta_{n,T}} \to 0$[34] and $h_{n,T}\overline{L}_X(T,x) \overset{a.s.}{\to} 0$, then*

$$\sqrt{\frac{\overline{L}_X(T,x)h_{n,T}}{\Delta_{n,T}}}\left(\widehat{\sigma}^2_{(n,T)}(x) - \sigma^2(x)\right) \overset{d}{\to} MN\left(0, 4\left(\int_{-\infty}^{\infty} \mathbf{K}^2(s)ds\right)\sigma^4(x)\right).$$

In the fixed T case, the conditions on the bandwidth reduce to $h_n \propto n^{-k}$ with $k \in (\frac{1}{4}, \frac{1}{2})$.

Limit theory for the drift function estimator [Bandi (1999, Theorem 1)] *If $n \to \infty$, $T \to \infty$, $h_{n,T} \to 0$ (as $n, T \to \infty$) such that $\frac{\Delta_{n,T}}{h_{n,T}} \to 0$ and $\frac{\overline{L}_X(T,x)}{h_{n,T}}(\Delta_{n,T})^{\alpha} = O_{a.s.}(1)$ for*

[34]The usual caveat applies in the case $\frac{h^4_{n,T}}{\Delta_{n,T}} \to \infty$.

some $\alpha \in (0, \frac{1}{2})$, than the estimator (18.4) is consistent almost surely for the true function provided $h_{n,T}\overline{L}_X(T,x) \overset{a.s.}{\to} \infty$. Also,

$$\sqrt{\overline{L}_X(T,x)h_{n,T}}\left(\widehat{\mu}_{(n,T)}(x) - \mu(x)\right) \overset{d}{\to} MN\left(0, \left(\int_{-\infty}^{\infty} \mathbf{K}^2(s)ds\right)\sigma^2(x)\right).$$

18.5 Remark The comments that we made earlier on the BP estimators [c.f. Part I and II] readily extend to (18.3) and (18.4) with $\varepsilon_{n,T}$ being replaced by $h_{n,T}$. More explicitly:

[1] The rate of convergence of the diffusion function estimator ($\sqrt{\frac{\overline{L}_X(T,x)h_{n,T}}{\Delta_{n,T}}}$) is faster than the rate of convergence of the drift estimator ($\sqrt{\overline{L}_X(T,x)h_{n,T}}$).

[2] In the case of diffusion estimation, the admissible bandwidth $h_{n,T}$ can converge to zero at a faster pace that in the case of drift estimation. In fact, when estimating the drift, the condition $\overline{L}_X(T,x)h_{n,T} \overset{a.s.}{\to} \infty$ imposes a tight requirement on the sequence $h_{n,T}$. Furthermore, since the diffusion function can be identified in every neighborhood, it is straightforward to find closed form solutions for the admissible bandwidths [c.f. discussion above with $T = \overline{T}$]. Being the rate of divergence of the local time factor related to the characteristics of the underlying process, the same possibility is ruled out in the drift case. Then, small sample analysis can be particularly useful to find informal guidelines to choose the relevant window widths [c.f. Subsection 20.1 below].

[3] The features of the limiting distributions clarify the sense in which enlarging the time span is necessary only to estimate the drift. If T were fixed, than $\overline{L}_X(\overline{T},x) = O_p(1)$ and (18.4) would diverge at a speed equal to $\sqrt{\frac{1}{h_{n,T}}}$ [c.f. Theorem 10.1 in Part II]. Later we will comment on the ability of the asymptotic theory to capture the salient features of the small sample distributions.

18.6 Remark Note that if we use the same kernel functions and the same bandwidths, the diffusion function estimators in Stanton (1997) and JK (1997) deliver the same outcome in finite sample. In consequence, we will discuss only one estimated curve in what follows.

18.3. A final observation on the functional estimates

Since the diffusion function can be identified over a fixed time span, Assumptions 17.1 and 17.2 are excessively stringent. The diffusion function estimators considered in this chapter and in Part I maintain their limiting properties, even if the Feller's condition for nonexplosion does not hold. In other words, the instantaneous volatility of a *transient* process can be estimated consistently using the apparatus described above. As far as the drift is concerned, Assumption 17.1 (i.e. recurrence) is necessary for consistency of the estimators in BP and Stanton (1997), whereas Assumption 17.2. (i.e. stationarity) is necessary for the estimator in JK (1997) to be well defined.

19. The simulated processes

In this section we discuss the choice of the simulated processes. We simulate processes that have been, or might be, used as descriptions of the short-term interest rate process in continuous-time finance. Our aim is to verify the finite sample properties of the BP estimators and of the estimators outlined above in the presence of underlying series that display various statistical properties. In fact, we consider processes that satisfy either Assumption 17.1 or Assumption 17.2 above. Their infinitesimal first and second moments are parametrized as either linear or nonlinear functions. We specify parameter values that are coherent with recent results in the finance literature on the estimation of the short-term interest rate process [see Aït-Sahalia (1996a,b) and Part II, *inter alia*, for discussions]. The initial value is set equal to 0.067. The distance between observations Δ_t is set equal to 1/250, that is we simulate daily observations.

19.1. Brownian motion [linear, nonstationary and recurrent process]

We start with the simplest nonstationary continuous-time process satisfying Assumption 17.1. We will be interested in evaluating how well methods robust to deviations from stationarity, that is Stanton's (1997) and BP's, capture the features of the theoretical functions. In this case,

$$\mu(.) = 0 \text{ and } \sigma^2(.) = \sigma^2 = \text{constant}$$

Then,

$$dX_t = \sigma dB_t \tag{19.1}$$

where $\{B_t; t \geq 0\}$ is a standard Brownian motion. The process can be simulated easily because the transitional density is known, i.e.

$$\pi(X_{t+1} = x_{t+1}|X_t = x_t) = \frac{1}{\sqrt{2\pi\sigma^2\Delta_t}} \exp\{-\frac{(x_{t+1} - x_t)^2}{2\sigma^2\Delta_t}\}.$$

We set σ^2 equal to 0.002.

19.2. Vasicek (1977) [linear, stationary and recurrent process]

The drift is specified as a linear mean-reverting function, i.e.

$$dX_t = \kappa(\theta - X_t)dt + \sigma dB_t \tag{19.2}$$

with θ corresponding to the long-run mean of the process and κ corresponding to the speed of reversion to the long run mean. The solution to (19.2) satisfies Assumption 17.2 (A')-(C') provided $\kappa > 0$. If $\kappa > 0$, than the stationary marginal distribution is normal with unconditional mean θ and variance $\frac{\sigma^2}{2\kappa}$. The parameter κ determines the persistence of the process by controlling the rate at which X_t reverts toward the unconditional mean. Lowering the value of κ increases persistence because it slows the rate of mean reversion, which increases the correlation between observations. Usually, higher persistence negatively affects inference [c.f. Pritsker (1998) and Chapman and Pearson (1998)]. We will verify this result.

As in the Brownian motion case, the process is easy to simulate since the transitional density is known. In fact,

$$\pi(X_{t+1} = x_{t+1}|X_t = x_t) = \frac{1}{\sqrt{2\pi s^2(\Delta_t)}} \exp\{-\frac{(x_{t+1} - \theta - (x_t - \theta)e^{-\kappa\Delta_t})^2}{2s^2(\Delta_t)}\}$$

where $s^2(\Delta_t) = \frac{\sigma^2}{2\kappa}[1 - e^{-2\kappa\Delta_t}]$.

We consider two different experiments. The second set of parameter values corresponds to an increase in persistence [$\kappa \downarrow$] that does not affect the second moment of the stationary distribution of the process [$\sigma^2 \downarrow$ to keep $\frac{\sigma^2}{2\kappa}$ constant as $\kappa \downarrow$]. We set

$$(i)\ \kappa = 0.85837,\ \theta = 0.089102,\ \sigma^2 = 0.0021854$$

and

$$(ii)\ \kappa = 0.214592,\ \theta = 0.089102,\ \sigma^2 = 0.000546.$$

The values are taken from Pritsker (1998) and are consistent with the estimated first and second moments of the short-term interest rate data used in Ait-Sahalia (1996a,b).

19.3. Cox-Ingersoll-Ross (1985) [affine, stationary and recurrent process]

The drift is linear mean-reverting as in Vasicek (1977) and the diffusion function depends linearly on the underlying process, i.e.

$$dX_t = \kappa(\theta - X_t)dt + \sigma\sqrt{X_t}dB_t. \tag{19.3}$$

The parameters have the usual interpretation [c.f. Vasicek (1977)]. The process satisfies Assumption 17.2 (A')-(C') if $\kappa > 0$ and $2\kappa\theta > \sigma^2$. Under this assumption, the stationary marginal distribution is gamma with unconditional mean θ and variance $\frac{\theta\sigma^2}{2\kappa}$, i.e.

$$\pi(x) = \frac{\varpi}{\Gamma(\nu)}x^{\nu-1}\exp(-\varpi x)$$

where $\varpi = \frac{2\kappa}{\sigma^2}$, $\nu = \frac{2\kappa\theta}{\sigma^2}$ and $\Gamma(.)$ is the gamma function. The transitional density is noncentral chi-square,

$$\pi(X_{t+1} = x_{t+1}|X_t = x_t) = c\exp(-u-v)\left(\frac{v}{u}\right)^{q/2} I_q(2\sqrt{uv})$$

where

$$\begin{aligned} c &= \frac{2\kappa}{\sigma^2[1-\exp(-\kappa(\Delta_t))]}, \\ u &= cx_t\exp(-\kappa(\Delta_t)), \\ v &= cx_{t+1}, \\ q &= \nu - 1 = \frac{2\kappa\theta - \sigma^2}{\sigma^2}. \end{aligned}$$

and I_q is the modified Bessel function of the first kind of order q. The degrees of freedom are $2q+2$ and the noncentrality parameter is $2u$. As usual, the knowledge of the transitional density permits to simulate the process in a straightforward fashion.

We consider two different experiments. As in the Vasicek case, the second set of parameter values corresponds to an increase in persistence [$\kappa \downarrow$] that does not affect the second moment of the stationary distribution of the process [$\sigma^2 \downarrow$ to keep $\frac{\theta\sigma^2}{2\kappa}$ constant as $\kappa \downarrow$]. Specifically,

$$(i)\ \kappa = 0.85837,\ \theta = 0.085711,\ \sigma = 0.15660$$

and

(*ii*) $\kappa = 0.214592,\ \theta = 0.085711,\ \sigma = 0.07830.$

The values are taken from Chapman and Pearson (1998) and are (approximately) consistent with the estimated first and second moments of the data set used in Aït-Sahalia (1996a,b). Notice that in this model [and in Vasicek (1977)], the autocorrelation can be expressed as

$$Corr[X_{t+\Delta_t}, X_t] = \exp(-\kappa\Delta_t).$$

As Chapman and Pearson (1998, Subsection 3.1, page 11) point out, a κ equal to 0.85837 implies a first-order monthly autocorrelation coefficient which is equal to that of the Eurodollar data in Aït-Sahalia (1996a,b), i.e. ≈ 0.938. The second value, $\kappa = 0.214592$ that is, implies a monthly autocorrelation of 0.982, which is consistent with the upper end of the estimated values for US interest rate data.

19.4. Aït-Sahalia (1996) [nonlinear, stationary and recurrent process]

Both the drift and the diffusion function are nonlinear, i.e.

$$dX_t = \left(\alpha_0 + \alpha_1 X_t + \alpha_2 X_t^2 + \alpha_3/X_t\right) dt + \left(\sqrt{\beta_0 + \beta_1 X_t + \beta_2 X_t^{\beta_3}}\right) dB_t \tag{19.4}$$

The process satisfies Assumption 17.2 provided the parameter values meet specific requirements [see Aït-Sahalia (1996b)]. The parameter values specified below make Assumption 17.2 hold, that is they imply stationarity. Notice that neither the marginal stationary distribution nor the transitional density are known. In consequence, we simulate the model using a discretization scheme. We use a scheme with order-two error belonging to the class of *Milshstein approximations* [c.f. Milshstein (1974) and Duffie (1996)]. *Milshstein approximations* are schemes of higher order than Euler discretization schemes.

The parameter values are set according to the estimated values in Aït-Sahalia (1996b, Table 4, page 412), namely

$$\begin{aligned}\alpha_0 &= -5.652 \times 10^{-3}\\ \alpha_1 &= 9.648 \times 10^{-2}\\ \alpha_2 &= -5.349 \times 10^{-1}\end{aligned}$$

$$\begin{aligned}\alpha_3 &= 1.041 \times 10^{-4} \\ \beta_0 &= 1.099 \times 10^{-4} \\ \beta_1 &= -2.007 \times 10^{-3} \\ \beta_2 &= 1.329 \times 10^{-2} \\ \beta_3 &= 2.051.\end{aligned}$$

19.5. Experimental process [Nonlinear, nonstationary and recurrent process]

Both the drift and the diffusion function are nonlinear, i.e.

$$dX_t = (\alpha_0 + \alpha_1 X_t + \alpha_2 / X_t)\,dt + \left(\sqrt{\beta_0 + \beta_1 X_t + \beta_2 X_t^2}\right) dB_t \qquad (19.5)$$

The process is nonstationary and non explosive (that is it satisfies Assumption 17.1) for the parameter choice specified below. We set the parameter values as follows:

$$\begin{aligned}\alpha_0 &= -4.643 \times 10^{-3} \\ \alpha_1 &= 4.333 \times 10^{-2} \\ \alpha_2 &= 1.304 \times 10^{-4} \\ \beta_0 &= 1.108 \times 10^{-4} \\ \beta_1 &= 1.883 \times 10^{-3} \\ \beta_2 &= 9.681 \times 10^{-3}.\end{aligned}$$

As before, we use an order-two *Milshstein approximation* to simulate the model.

20. The simulation results

In this section we report the outcome of our simulations by addressing the six questions that we posed in the introduction to this chapter. A subsection is devoted to each question.

It is worth emphasizing from the beginning that we simulate 5000 daily observations ($\Delta_t = 1/250$) from the five processes described above. We set the number of observations equal to 5000 to replicate almost 20 years of data. This is consistent with the magnitude of the data sets recently used in the investigation of the short-term interest rate process. We use 1000 repetitions. In our preliminary analysis the estimates converge quickly even after a few hundred repetitions, and 1000 repetitions is certainly deemed sufficient.

20.1. Nonparametric estimates: small sample bias.

Two variables are to be chosen when implementing functional methods: the kernel function(s) and the window width(s).

We begin our experiments by setting the same kernel function across different methods. The function that we use is a first order Gaussian kernel.

It is well known that the choice of the bandwidth(s) is crucial in nonparametric statistics. Still, how to choose it properly is an open issue. While asymptotic results give us limit conditions (i.e. orders of convergence) that the bandwidths, specified as sequences, have to satisfy, little is known about how to set them in practice. In effect, what we choose is a number rather than a sequence. Then, scaling plays a vital role. We now clarify this concept.

In nonparametric regression analysis various methods have been proposed to select the "optimal" bandwidth. Some of these methods have been used in the analysis of diffusion models [Jiang (1998), *inter alia*]. For example, consider the standard regression model

$$y_i = m(x_i) + \varepsilon_i$$

where y_i is the *i-th* observation of the vector of dependent variables, x_i is the *i-th* observation of the vector of exogenous variables, $m(x_i)$ is a function of x_i and ε_i is a regression error satisfying certain properties. In regression analysis the so-called *least-squares cross validation* bandwidth is chosen as the solution to the following criterion

$$\min_h \frac{1}{n} \sum_{i=1}^{n} [y_i - \widehat{m}_h(x_i)]^2 \varpi(x_i) \tag{20.1}$$

where $\varpi(x_i)$ is a weighting function introduced to reduce the impact of boundary biases by giving less weight to observations that are at the extremes of the distribution of the sampled process. The estimates $\widehat{m}_h(x_i)$ are leave-one kernel estimates of $m(x_i)$ obtained by considering every data point apart from the *i-th* observation in the estimation of the function at the *i-th* observation. Under some assumptions, it can be proven that the LSCV bandwidth is optimal with respect to performance criteria such as the average squared error, the integrated squared error and the conditional mean squared error [c.f. Härdle and Marron (1985) and Stone (1984)]. Typically, we set the bandwidth equal to n, the number of observations, raised to some negative power, i.e. $h_n = cn^{-k}$. The proportionality constant c is chosen to satisfy the criterion

$$\min_c \frac{1}{n}\sum_{i=1}^{n}[y_i - \widehat{m}_c(x_i)]^2\varpi(x_i) \tag{20.2}$$

which is just a version of (20.1). This procedure is consistent with the underlying asymptotic theory, provided k is chosen accordingly, and allows us to determine the appropriate scaling factor. Effectively, we choose a number. If (20.1) rather than (20.2) is implemented, then k and c can not be identified separately. Generally, more complicated criteria are used. Consider

$$\min_c \frac{1}{n}\sum_{i=1}^{n}[y_i - \widehat{m}_c(x_i)]^2\varpi(x_i)\Xi(\frac{1}{n}, c) \tag{20.3}$$

where $\Xi(\frac{1}{n}, c) = 1+\frac{2}{n}\frac{1}{h}\mathbf{K}(0)$ and $h = cn^{-k}$ for some k consistent with the asymptotic theory of the estimator $\widehat{m}_c(x_i)$ being investigated. The function $\Xi(\frac{1}{n}, c)$ is known as Shibata penalty function. It is introduced to penalize excessively small bandwidths. Should this function not be in the formula, then the criterion would determine overfitting and introduce too much variation in the estimates[35].

Unfortunately, procedures with a firm theoretical justification in the regression context, such as (20.2) or (20.3) above, are not fully applicable in the case of diffusion models. Even though we specify the drift and diffusion function estimators as sample analogues to conditional expectations, the nature of the underlying process and of the true functions makes the problem different from pure regression analysis. Still, criteria such as (20.3) are widely used in the functional analysis of diffusions. Typically, y_i is set equal to $x_{i+1} - x_i$ in the case of drift estimation and equal to $(x_{i+1} - x_i)^2$ in the case of diffusion function estimation. Accordingly, $\widehat{m}_c(x_i)$ is either $\widehat{\mu}_c(x_i)dt$ or $\widehat{\sigma^2}_c(x_i)dt$. While this way of proceeding has some appeal, it seems to be inaccurate and does not have a convincing theoretical justification. Diffusion estimation resembles a typical regression analysis problem more closely than drift estimation. This observation is embodied in the limit theories for the diffusion estimators in Section 18 and Part I and derives from the rate of convergence which is equal to the square root of the number of observations times the bandwidth (over a fixed time span). Note, in fact, that this is the same rate that would emerge from a standard regression context. Not surprisingly, then, when we implement the selection criterion (20.3) in the diffusion case we experience nice convergence properties. A full-blown minimization of (20.3) in the drift case tends to deliver a broad array of different values for the constant

[35] Criterion (20.3) has been recently employed by Chapman and Pearson (1998) in the analysis of diffusion processes.

according to the specific trajectory being examined. Such values are generally quite large. We regard the poor convergence properties as a sign that the criterion is misspecified (a different standardization is likely to be needed). Further, we interpret the large values that are generally delivered as a signal that the optimal window width for the drift is generally larger than conventionally believed [c.f. comments in Section 18]. We will come back to these observations. It is worth saying here that, in the light of our previous discussion, in this chapter we use criterion (20.3) only for the diffusion and simply as an informal check. The rigorous design of data-driven selection criteria for diffusions is certainly a topic of interest for future research but goes beyond the scope of the present work.

Finally, we describe how the bandwidths are selected. Due to the difficulties in choosing the correct window width for the drift we decided to implement a conservative selection procedure. We now clarify what we mean by "conservative".

We start with the diffusion function estimator in Stanton (1997) and JK (1997) [S-JK diffusion, hereafter]. We set the bandwidth equal to $h^{diff(S-JK)}_{n,\overline{T}=1} = c\widehat{\sigma}n^{-k}$ where $\widehat{\sigma}$ is the standard deviation of the observations and n is the number of observations (which is equal to 5000). The time span T is set equal to 1. We choose k and c according to the limit theory and criterion (20.3). The exponent k is chosen equal to $\frac{1}{4}$. Theoretically, the bandwidth $\varepsilon^{diff(BP)}_{n,\overline{T}=1}$ in BP plays the same role as $h^{diff(S-JK)}_{n,\overline{T}=1}$ provided

$$h^{diff(BP)}_{n,\overline{T}=1} = o(\varepsilon^{diff(BP)}_{n,\overline{T}=1})$$

In consequence, we set $\varepsilon^{diff(BP)}_{n,\overline{T}=1} = 2h^{diff(BP)}_{n,\overline{T}=1} = h^{diff(S-JK)}_{n,\overline{T}=1}$.

As far as the drift functions are concerned, we start our experiments using the same bandwidths as for the diffusion function estimators, that is we choose window widths that are potentially suboptimal [c.f. Section 18]. This is a useful way to proceed because it gives us a feel for the extent of the suboptimality and for the corrections to be implemented.

We begin commenting on the diffusion estimators [c.f Figures 8-14, first column]. We are not surprised to verify that both the S-JK estimator [Figures 8-14, first column and first row] and the BP estimator [Figures 8-14, first column and third row] capture the underlying functions quite well. They also deliver very similar outcomes across processes. The magnitude of the differences is numerically minimal but differences do occur mostly at the upper boundaries of the empirical range of the processes, that is where observations are thinner. The relative merit of the two procedures in finite sample depends on the smoothness of the true functions. Specifically, in the presence of flat diffusions [i.e. Brownian motion and Vasicek in Figures 8, 9 and 10] the BP estimator displays a smaller bias at the upper boundary. The opposite result occurs in situations where the diffusion function is not a constant. In small sample, this is due to the double-smoothing procedure that characterizes the BP estimator.

We now discuss the drift estimators [c.f. Figures 8-14, second column]. The differences across methods and processes are clearly more marked than in the diffusion function case. Nonetheless, two main stylized facts can be detected. First, the suboptimal choice of the window widths translates into undersmoothing. The selected bandwidths are generally too small to capture underlying functions that are quite flat. This is not surprising. In the case of the estimators in Stanton (1997) [Figures 8-14, second column and first row] and Part I [Figures 8-14, second column and third row] the asymptotic theories allow the leading window widths for the diffusion (i.e. $h_{n,T}^{diff(S)}$ in the former case and $\varepsilon_{n,T}^{diff(BP)}$ in the later) to have admissible rates of convergence that can be faster then the corresponding rates for the drift bandwidths (respectively, $h_{n,T}^{drift(S)}$ and $\varepsilon_{n,T}^{drift(BP)}$). The reason is that local information is not sufficient for identification of the infinitesimal first moment of a diffusion, unless stationarity is invoked as in the case of the JK estimator for the drift [Figures 8-14, second column and second row]. Being the JK estimator based on the restrictions imposed by the distribution function of the process and its diffusion function on the theoretical drift, pointwise identification based on local information is possible. Further, it is perfectly appropriate to set $h_{n,\overline{T}=1}^{drift(JK)}$ proportional to n^{-k} with $k = \frac{1}{4}$ as in the diffusion case. The constant of proportionality can be calibrated to achieve better fit across models.

Second, with the sole exception of the Vasicek process in correspondence with low levels of persistence [c.f. Figure 9] and for very high levels of the process itself, the drift estimator proposed by BP does systematically better that the alternative methods in reproducing the underlying drift function. This result is perfectly understandable and represents the "flip side of the coin" of our findings in the diffusion case. Since the drifts are generally quite flat, double-smoothing induces better fit. In other words, using BP we seem to gain in terms of drift estimation while we slightly lose in terms of diffusion function estimation. Note, though, that the gain clearly outweighs the loss. Hence, the trade-off between optimal smoothing for the drift and optimal smoothing for the diffusion appears to be less severe when a convolution of kernels is employed.

It is worth emphasizing again that we are using potentially suboptimal values for the bandwidths. Therefore, we are not claiming that double-smoothing is a prerequisite for bias reduction. We are simply saying that a naive choice of the window widths seems to penalize statistical fit to a lesser extent in the presence of convoluted kernel weights. Due to the difficulties posed by the appropriate choice of the smoothing parameter for the drift, this is a valuable information for empirical work. On the other hand, less naive choices of the smoothing parameters determine minimal biases even when simple kernels are employed [see below].

So far, two lessons can be drawn from our analysis.

[1] Choices of the bandwidth that are optimal for diffusion estimation generally determine undersmoothing for the drift.

[2] The trade-off between optimal bandwidths is less severe when we use convoluted kernels. The price paid for oversmoothing the diffusion function is minimal compared to the gain from bias reduction for the drift. Double-smoothing is in some sense more "forgiving": suboptimal choices of the smoothing parameters have a smaller effect on inference.

Of course, the analysis is contingent on the specific processes being used. On the other hand, we study a wide array of specifications with different statistical properties. Many of the proposed specifications have been, or could be, used to model the short-term interest rate process in continuous-time. Interestingly, some of our findings allow us to reassess the informational content of one of the outcomes of the empirical literature on the spot interest rate process, that is the estimated nonlinearities of the drift at the upper boundary of the range of the sample process. Note that undersmoothing often implies nonlinear behavior at the boundaries even in the presence of linear drifts. In a recent paper, Bandi (1999) points out that nonlinearities might be partly due to erroneous choices of the smoothing parameter. When using functional methods, imprecise choices of the window width would, in fact, exacerbate the finite sample bias that naturally arises at the boundaries of the empirical process due to the truncation of its finite sample distribution [c.f. Part II and Chapman and Pearson (1998)]. Our simulations give support to the importance of the bandwidth and somehow reduce the role played by the truncation in explaining nonlinear dynamics.

We now increase the leading bandwidths for the drift, that is $h_{n,\overline{T}=1}^{drift(S)}$, $h_{n,\overline{T}=1}^{drift(JK)}$ and $\varepsilon_{n,\overline{T}=1}^{drift(BP)}$. We implement two different experiments. We take the original "optimal values" of the window widths in the diffusion case and multiply them by 1.5 [see Figures 8-14, third column] and 2 [see Figures 8-14, fourth column], respectively.[36] The results confirm our intuition. Increasing the bandwidth generally determines better fit. Exceptions occur in the case of the Vasicek and CIR processes for low levels of persistence [c.f. Figures 9 and 11]. In this cases, in fact, the original window widths appear to guarantee sufficiently small biases over the entire range of the sampled process. The intuition goes as follows. Recall that in the case of Stanton (1997) and BP, the asymptotic condition on the smoothing parameters for the drift depends on the rate of divergence of local time to infinity. The slower is this rate, the slower should be the rate of convergence of the leading bandwidth sequence to zero. This should translate into larger numerical values. Theoretically, higher persistence determines a slower rate of diverge to infinity of the time spent by the process in neighborhoods, i.e. local time. In small sample, higher persistence determines fewer visits to levels that are far from the initialization of the process and hence less precise estimates at the boundaries of the empirical range of the process.

Then, the recipe is simple: "relatively small" bandwidths should be used in the presence of low persistence, whereas "relatively larger" bandwidths should be employed when persistence is high [we will later clarify what we mean by "relatively small" and "relatively large"]. This is perfectly understandable if we take into consideration the fact that the drift of very persistent processes is quite flat. This remark has clear empirical implications. It is a well-known fact that U.S. short-term interest rate processes are very persistent [c.f. Pritsker (1998)]. The second value of the mean-reversion parameter κ that we use in our simulations of the Vasicek [Figure 10] and CIR [Figure 12] processes is equal to 0.2145 and implies a monthly autocorrelation of 0.982 which is equal to the upper end of the estimated values for US interest rate data [c.f. Chapman and Pearson (1998)]. Being the U.S. interest rates slowly reverting, identification of their infinitesimal first moments requires larger bandwidths. As mentioned earlier, inaccurate smoothing might, in fact, determine spurious nonlinearities [c.f. Bandi (1999)].

These observations suggest a rough rule-of-thumb to choose the "leading" bandwidths in the drift case:

[1] Start from a sensible choice of the diffusion bandwidth based on credible criteria such

[36] By virtue of the clear pattern that is determined by our (increasing) choices of the leading window widths, the outcome of different choices can be easily deduced. As expected, the larger is the bandwidth, the flatter is the estimated drift.

as least-squares cross validation.

[2] Use the bandwidth choice in the diffusion case as a lower bound to select the drift bandwidth.

[3] The numerical distance from the lower bound needs to be directly related to the persistence of the process. Higher persistence requires larger values. Persistence can be assessed by using a simple descriptive statistic such as the first-order autocorrelation.

20.2. Nonparametric estimates: small sample volatility

Apparently, the finite sample dispersion of the functional estimates is inversely related to the amount of smoothing being implemented. On the other hand, increased smoothing generally implies better fit (reduced bias) in the case of drift estimation. In other words, for many of the processes considered here [namely for highly persistent processes], the trade-off between bias and volatility, which is a typical feature of regression analysis, does not play an important role in the estimation of the infinitesimal first moment. Thus, consistently with the limit theory that dictates potentially larger bandwidths for drift estimation than for diffusion estimation, our simulations show that more smoothing (whether determined by larger window widths or convoluted kernels) generally has the appealing feature of determining more accurate fit and reduced volatility in finite sample.

The usual care must be exercised when dealing with processes whose reversion to the mean is very fast. The drift of a stationary process can be identified locally [see JK (1997)]. Hence, we expect the bias-volatility trade-off to be more severe than for nonstationary processes. Generally speaking, the extent to which the trade-off matters in estimating the drift depends on the level of persistence. The drift of a process with high persistence can be accurately estimated at a low cost in terms of dispersion by appropriately choosing a relatively large smoothing parameter.

20.3. Nonparametric estimates: asymptotic distributions versus finite sample distributions

We are mainly interested in studying the finite sample properties of the estimators proposed by Stanton (1997) and BP. We confine ourselves to these two approaches since they are theoretically more flexible than the approach suggested by JK (1997). In effect, they do not rely on stationarity. Furthermore, even when the process is stationary our results show that no benefit seems to arise from using the JK drift estimator (recall that the JK diffusion estimates coincide with Stanton's) [see Subsection 20.1].

In Figures 15-17 we report the finite sample and asymptotic distributions of the Stanton and BP estimators. The first two rows report the drift estimates in Stanton (1997) and BP (1998) respectively, whereas the last two rows report the corresponding diffusion estimates. The underlying processes are the CIR process (for two different levels of persistence) [Figures 15-16] and the Aït-Sahalia process [Figure 17]. The results are qualitatively similar when using different processes.[37] We compare the pointwise distributions at numerous levels in the empirical range of the underlying process. The bandwidths are "optimally" chosen. For the CIR process with low persistence we use the benchmark bandwidths [Figure 11, columns 1 and 2]. For the Aït-Sahalia process and the CIR process with high persistence we use the benchmark case for the diffusion estimates and the largest leading bandwidths [from experiment 2 in Subection 20.1] for the drift estimates [Figures 12-13, columns 1 and 4].

The results confirm the validity of functional methods that do not rely on stationarity for inference in small sample [c.f. Pritsker (1998) for a discussion of the limitations of stationarity-based nonparametric and semiparametric procedures]. The finite sample distributions are sufficiently close to normals. The similarity to normals is very satisfactory at levels in the middle of the empirical range of the process. Skewness plays a role at levels that are close to the ends of the distribution. Consistently with Figures 8-14, biases arise mostly at the extremities of the empirical range, that is where observations are thinner. The asymptotic variances replicate quite well the finite sample variances. As in the bias case, this is particularly true for values that are central to the empirical distribution.

The similarity between finite sample and asymptotic distributions is generally more accurate in the drift case. In the diffusion case, we experience a relatively more marked tendency of the limit theory to underestimate the true finite sample dispersions. A simple pattern can be detected. The closer is the analyzed level to the lower end of the empirical range, the further away from a symmetric distribution is the finite sample distribution and the smaller is the limiting variance compared to the true finite sample variability. These results generally hold across methods, processes and degrees of persistence. Of course, at values whose corresponding local time is large (generally values in the middle of the empirical range), the limiting distributions guarantee a relatively more accurate description

[37]Similar graphs for additional processes can be provided on request.

of their finite sample counterparts.

20.4. The choice of the kernel

Different choices of the kernel function $\mathbf{K}(.)$ that is used to weigh observations [or weighted averages of observations as in the case of BP] do not affect significantly our inference in terms of bias. The small sample variability is slightly affected and the direction of the changes is generally coherent with the theoretical properties of the kernels being used. Also, the relative performances of the estimators analyzed here does not depend on the choice of the kernel.

As mentioned earlier, though, if we interpret the procedure suggested by BP as based on *convoluted* kernel weights, it appears that double-smoothing might be advantageous to estimate the drift without imposing a severe cost on the estimation of the diffusion function [c.f. Subsection 20.1]. A possible explanation for this result is the following. There are no theoretical reasons why Stanton (1997) and BP should deliver different results in the limit. To phrase it differently, there are no theoretical reasons why double-smoothing and single-smoothing should deliver different results asymptotically. The problem is that, as far as the point estimates are concerned, "the limit" is "closer" for diffusion estimation than for drift estimation. Volatility is easier to identify than the conditional mean. In consequence, in the case of diffusion estimation, methods that guarantee consistent estimation deliver very similar results even in finite sample [for comparable choices of the smoothing parameters]. This is clearly not the case when estimating the drift. Unless close-to-optimally-chosen window widths are selected, different estimators for the drift generally produce quite different outcomes. Then, the use of convoluted kernels might be beneficial for the reasons outlined in Subsection 20.1.

Note that the estimators in BP are originally defined based on Gaussian kernels convoluted with discontinuous indicators. The use of a smooth function replacing the indicator function is perfectly legitimate and, coherently with the asymptotic theory [c.f. Part I], does not change the results qualitatively.

20.5. The choice of the bandwidth

As discussed earlier, the choice of the smoothing parameter is crucial and particularly difficult when dealing with the drift estimator. The asymptotic theories and the simulations presented here are consistent in paving the way for optimally choosing generally larger [than for diffusion estimation] bandwidths for the drifts. However, the stochastic conditions that the drift bandwidths have to satisfy, i.e. $\varepsilon_{n,T}^{BP}\overline{L}_X(T,x) \overset{a.s.}{\to} \infty$ and $h_{n,T}^{S}\overline{L}_X(T,x) \overset{a.s.}{\to} \infty$, render the selection problem quite hard to solve explicitly due to the lack of closed-form solutions for the rate of divergence of the local time factor $[\overline{L}_X(T,x)]$ to infinity. Since the drift of a general process cannot be identified locally, the stochastic properties of the underlying process play a vital role through the above-mentioned conditions. In consequence, we believe among the most important contributions to the functional estimation of scalar diffusions is now the design of sensible automatic selection criteria for the smoothing parameters.

Note that persistence affects the rate of divergence of the local time factor to infinity [c.f. next Subsection]. Hence, coherently with Conley, Hansen and Liu (1997) we emphasize that the optimal bandwidth should depend on the dynamic properties of the underlying process [see Pritsker (1998) for a comment in the same spirit]. Our previous discussion should also clarify that this point is particularly valid as far as drift estimation is concerned. Being the infinitesimal second moment of a diffusion locally identifiable, it is certainly less valid in the case of diffusion estimation.

20.6. The statistical properties of the underlying process

We already discussed the importance of temporal dependence and its qualitative effect on the optimal window widths for the drift through the local time-based conditions in the previous sections. The general rule goes as follows. Since stationary processes are locally identifiable [see JK (1998)], they require smaller bandwidths for the drift than nonstationary processes, i.e. the rate of divergence to ∞ of the local time factor is faster for stationary than for nonstationary processes. Further, within the class of stationary processes, higher temporal dependence triggers larger bandwidths.

As mentioned earlier, for comparable choices of the smoothing parameters, the drift estimator in JK (1998) does not outperform the more flexible estimators in Stanton (1997) and BP in the presence of stationarity. Further, the JK estimator generally entails larger biases at the boundaries. This is due to the poor convergence properties of the derivative estimates at the upper and lower ends of the empirical range of the process. Interestingly, when the underlying process is nonstationary [i.e. Brownian motion and experimental process in Figure 8 and Figure 14], the JK estimator delivers outcomes that are quite comparable to those of the Stanton estimator [with the exception of the usual larger biases at the boundaries].

21. Conclusion

When little is known about how to parametrize stochastic differential equations, accurately implemented functional methods can represent valuable descriptive tools.

Theoretical justifications for employing functional methods rely on limit arguments based on increasingly frequent observations. Coherently with JK (1998), we show that daily data represent a good approximation to increasingly frequent observations for estimators that hinge on high frequency observations [c.f. Part II]. In finance, for example, daily data are readily available. In consequence, the difference between limit requirements and finite sample data sets does not represent a major empirical problem. Hence, we regard the fully nonparametric identification of stochastic differential equations as meaningful statistical inference.

Local identification of the diffusion function using data that are sampled at daily frequencies is easy to implement. Also, being the diffusion locally identifiable, the stochastic properties of the underlying diffusion process do not play a vital role in inference.

As opposed to diffusion estimation, identification of the drift for general classes of processes [as in Stanton (1997) and BP] is intimately related to the their stochastic behavior over time and requires observation of the trajectories over a long time horizon. Two are the consequences. First, larger smoothing parameters for drift estimation than for diffusion estimation are generally needed. Second, given the importance of stochastic asymptotic conditions based on the local time factor in choosing the optimal smoothing parameter for the drift, the actual choice of the bandwidth is particularly cumbersome. As a general rule, the more persistent the process is, the larger should be the window width. In effect, persistence negatively affect the speed of divergence of the time spent by a recurrent process at a point (i.e. the chronological local time of a process at a point) to infinity. Being the relationship between the admissible rate of convergence of the bandwidth sequence and the rate of divergence of the local time factor inverse, a smaller local time requires a larger smoothing parameter. Intuitively, more persistent processes have relatively flatter drifts and, hence, demand increased smoothing. As discussed earlier, due to the significant persistence of U.S. interest rate series, the nonlinear behavior of the drift in the short-term interest rate literature might be partly due to inaccurate choices of the smoothing parameter determining undersmoothing [c.f. Bandi (1999)].

Interestingly, even though the drift of a *stationary* process can be identified locally using the information contained in the time-invariant distribution function of the process [as in JK (1997)], no clear benefit in terms of inference seems to arise with respect to the outcomes of methods that are theoretically robust to deviations from stationarity [Stanton (1997) and BP].

To summarize, we believe the main issue in the functional estimation of diffusions is how to correctly choose the window widths. A potential trade-off characterizes drift and diffusion estimation. Should the same bandwidth be employed, the optimal drift bandwidth would generally oversmooth the diffusion, whereas the optimal diffusion bandwidth would generally undersmooth the drift. Still, as discussed earlier, automated criteria "borrowed" from regression analysis provide some valuable indications in selecting a sensible smoothing parameter for the diffusion function. The same consideration does not apply to the drift due to the difference between drift estimation and regression analysis, thus making the design of meaningful data-driven selection criteria more valuable in this case. This paper provides some guidelines and a simple rule of thumb to choose the optimal window width for the drift. We also point out that *double-smoothing* [as in Part I] based on a convolution of kernels can, in finite sample, improve the trade-off between sensible magnitudes of the smoothing parameters for drift and diffusion function estimation. Despite this, further research on bandwidth selection criteria for diffusions is certainly needed.

The comparison between the finite sample and asymptotic distributions is very encouraging. As discussed in Part II and Bandi (1999), contrary to functional estimators relying on the assumption of stationarity [c.f. Aït-Sahalia (1996) and JK (1997), *inter alia*], the estimators in Part I and Stanton (1997) have asymptotic distributions that depend on the dynamic characteristics of the underlying process (such as persistence) through the local time factor. This is a particularly nice feature since it is apparent that the finite sample distributions depend on the stochastic features of the process. This observation provides a possible answer to one of the observations in Pritsker (1998) where it is noted that kernel estimators based on stationarity have the potential for misleading inference in studying the short-term interest rate process due to the fact that "...the asymptotic distributions do not depend on persistence although the finite sample distributions do". As shown in this chapter, estimators that are robust to deviations from stationarity have asymptotic distributions that sufficiently well approximate their finite sample counterparts. Hence, when using estimation methods such as those in Stanton (1997) and BP, the need for computing standard errors and confidence bands using bootstrapping procedures is certainly less compelling than suggested by Pritsker (1998).

22. Proofs

In this section we report the proofs of Theorem 1 and 2 in Bandi (1999).

22.1. Proof of Theorem 1 [Bandi (1999)]

We use, in parts, results contained in Part I. Consider the estimator

$$\widehat{\mu}_{n,T}(x) = \frac{1}{\Delta_{n,T}} \frac{\sum_{i=1}^{n-1} \mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})[X_{(i+1)\Delta_{n,T}} - X_{i\Delta_{n,T}}]}{\sum_{i=1}^{n} \mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})}.$$

We start by proving that

$$\widehat{\mu}_{n,T}(x) - \mu(x) \stackrel{a.s.}{\to} 0.$$

Since

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

then $\widehat{\mu}_{n,T}(x)$ can be written as follows

$$\begin{aligned}\widehat{\mu}_{n,T}(x) &= \underbrace{\frac{1}{\Delta_{n,T}} \frac{\sum_{i=1}^{n-1} \mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})[\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}} \mu(X_s)ds]}{\sum_{i=1}^{n} \mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})}}_{(\alpha)} \\ &+ \underbrace{\frac{1}{\Delta_{n,T}} \frac{\sum_{i=1}^{n-1} \mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})[\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}} \sigma(X_s)dB_s]}{\sum_{i=1}^{n} \mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})}}_{(\beta)}.\end{aligned}$$

First, we analyze α. By the Lipschitz property of $\mu(.)$ we can write

$$\begin{aligned}\alpha &= \frac{1}{\Delta_{n,T}} \frac{\sum_{i=1}^{n-1} \mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})[\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}} (\mu(X_s) - \mu(X_{i\Delta_{n,T}}))\, ds]}{\sum_{i=1}^{n} \mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})} \\ &+ \frac{1}{\Delta_{n,T}} \frac{\sum_{i=1}^{n-1} \mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})[\mu(X_{i\Delta_{n,T}})\Delta_{n,T}]}{\sum_{i=1}^{n} \mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})} \\ &\leq const.\kappa_{n,T} + \frac{1}{\Delta_{n,T}} \frac{\sum_{i=1}^{n-1} \mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})[\mu(X_{i\Delta_{n,T}})\Delta_{n,T}]}{\sum_{i=1}^{n} \mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})}\end{aligned}$$

where $\kappa_{n,T} = \max_{i \leq n} \sup_{i\Delta_{n,T} \leq s \leq (i+1)\Delta_{n,T}} |X_s - X_{i\Delta_{n,T}}|$ as in Part I. We know that $\kappa_{n,T} = o_{a.s.}(\Delta_{n,T}^{1/2-\delta})$ with δ arbitrarily small, then the bound becomes

$$const.o_{a.s.}(\Delta_{n,T}^{1/2-\delta}) + \frac{1}{\Delta_{n,T}} \frac{\sum_{i=1}^{n-1} \mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})[\mu(X_{i\Delta_{n,T}})\Delta_{n,T}]}{\sum_{i=1}^{n} \mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})}.$$

By a straightforward application of the results in Part I [Section 5, Theorem 5.11], the second term converges to $\mu(x)$ provided $\frac{L_X(T,x)}{h_{n,T}}(\Delta_{n,T})^{1/2-\varepsilon} \xrightarrow{a.s.} 0$ and $\frac{\Delta_{n,T}}{h_{n,T}} \xrightarrow{a.s.} 0$. In order to prove $a.s.$ consistency it remains to prove that $\beta \xrightarrow{a.s.} 0$. Notice that the martingale $y_{(i+1)\Delta_{n,T}} = \int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}} \mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})\sigma(X_s)dB_s$ is measurable with respect to $\Im_{(i+1)\Delta_{n,T}}$, where $\Im_{(i+1)\Delta_{n,T}} = \{A \in \Im : A\{(i+1)\Delta_{n,T} \leq t^*\} \in \Im_{t^+} \forall t \geq 0\}$. Further,

$$\mathbf{E}(y_{(i+1)\Delta_{n,T}}) = 0$$

and, by the Ito isometry

$$\theta_{(i+1)\Delta_{n,T}} = var(y_{(i+1)\Delta_{n,T}}) = \mathbf{E}\left(\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}} \mathbf{K}^2(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})\sigma^2(X_s)ds\right) < \infty.$$

Also, $(y_{(i+1)\Delta_{n,T}}, \Im_{(i+1)\Delta_{n,T}})$ is a martingale difference sequence with zero mean and variance $\theta_{(i+1)\Delta_{n,T}}$. We now invoke a strong law of large numbers for martingale differences [e.g. Hall and Heyde (1980, Theorem 2.19, page 36)] to obtain

$$\frac{\sum_{i=1}^{n-1} \mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})[\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}} \sigma(X_s)dB_s]}{\sum_{i=1}^{n-1} \mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})} \xrightarrow{a.s} 0.$$

The rate of convergence is derived below. In particular, we prove that

$$\frac{\frac{1}{\Delta_{n,T}}\frac{\Delta_{n,T}}{h_{n,T}} \sum_{i=1}^{n-1} \mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}}) \int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}} \sigma(X_s)dB_s}{\frac{\Delta_{n,T}}{h_{n,T}} \sum_{i=1}^{n} \mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})} = O_p\left(\frac{1}{\overline{L}_X(T,x)h_{n,T}}\right).$$

This implies that β is negligible in the limit provided $\overline{L}_X(T,x)h_{n,T} \xrightarrow{a.s.} \infty$. Now we study the asymptotic distribution. We are interested in the limiting properties of

$$\begin{aligned} & \widehat{\mu}_{n,T}(x) - \mu(x) \\ = & \frac{1}{\Delta_{n,T}} \frac{\frac{\Delta_{n,T}}{h_{n,T}} \sum_{i=1}^{n-1} \mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})[(X_{(i+1)\Delta_{n,T}} - X_{i\Delta_{n,T}}) - \mu(x)\Delta_{n,T}]}{\frac{\Delta_{n,T}}{h_{n,T}} \sum_{i=1}^{n} \mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})}. \end{aligned}$$

First, we examine the numerator

$$\frac{1}{h_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})[(X_{(i+1)\Delta_{n,T}}-X_{i\Delta_{n,T}})-\mu(x)\Delta_{n,T}].$$

This can be written as follows,

$$\begin{aligned}
&\frac{1}{h_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})\left((X_{(i+1)\Delta_{n,T}}-X_{i\Delta_{n,T}})-\mu(x)\Delta_{n,T}\right)\\
=\;&\frac{1}{h_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})\left(\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}[\mu(X_s)-\mu(x)]ds+\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}\sigma(X_s)dB_s\right)\\
=\;&\underbrace{\frac{1}{h_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}[\mu(X_s)-\mu(x)]ds}_{(\mathbf{A}_{n,T})}\\
&+\underbrace{\frac{1}{h_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}\sigma(X_s)dB_s}_{(\mathbf{C}_{n,T}(1))}\,.
\end{aligned}$$

Consider $\mathbf{U}_{n,T}(r)=\sqrt{h_{n,T}}\mathbf{C}_{n,T}(r)$ which is equal to

$$\mathbf{U}_{n,T}(r)=\frac{1}{\sqrt{h_{n,T}}}\sum_{i=1}^{[nr]-1}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}\sigma(X_s)dB_s.$$

$\mathbf{U}_{n,T}$ is a continuous martingale whose quadratic variation process $[\mathbf{U}_{n,T}]_r$ is

$$\begin{aligned}
[\mathbf{U}_{n,T}]_r &= \frac{1}{h_{n,T}}\sum_{i=1}^{[nr]-1}\mathbf{K}^2(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}\sigma^2(X_s)ds\\
&= \frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{[nr]-1}\mathbf{K}^2(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})\sigma^2(X_{i\Delta_{n,T}}+o_{a.s.}(1))\\
&= \frac{1}{h_{n,T}}\int_0^{rT}\mathbf{K}^2(\frac{X_s-x}{h_{n,T}})\sigma^2(X_s)ds+O_{a.s}\left(\frac{\overline{L}_X(rT,x)}{h_{n,T}}(\Delta_{n,T})^{1/2-\epsilon}\right)\\
&= \frac{1}{h_{n,T}}\int_{-\infty}^{\infty}\mathbf{K}^2(\frac{a-x}{h_{n,T}})\sigma^2(a)\overline{L}_X(rT,a)da+o_{a.s}(1)\\
&= \int_{-\infty}^{\infty}\mathbf{K}^2(c)\sigma^2(x+h_{n,T}c)\overline{L}_X(rT,x+h_{n,T}c)dc+o_{a.s}(1)\\
&\xrightarrow{d}\left(\int_{-\infty}^{\infty}\mathbf{K}^2(c)dc\right)\sigma^2(x)\overline{L}_X(rT,x)+o_{a.s}(1).
\end{aligned}$$

Also, the covariation process $[\mathbf{U}_{n,T},B]_r\to 0$ $a.s$. Then, let

$$\rho_{n,T}(r)=\inf\{s:[\mathbf{U}_{n,T}]_s>r\}$$

be a sequence of time changes. Define $\mathbf{V}_{n,T}(r) = \mathbf{U}_{n,T}(\rho_{n,T}(r))$. The process $\mathbf{V}_{n,T}(r)$ is the *Dambis, Dubins-Schwartz* Brownian motion of the martingale $\mathbf{U}_{n,T}$ [see, for example, Revuz and Yor (1991, Theorem 1.6, page 170)]. The conditions on the variation and covariation process are sufficient to ensure that

$$(\mathbf{V}_{n,T}, B) \xrightarrow{d} (\mathbf{V}, B)$$

where $(\mathbf{V}, B)$ is a vector of independent Brownian motions. Hence,

$$\begin{aligned}\mathbf{U}_{n,T}(r) &= \sqrt{h_{n,T}}\left(\frac{1}{h_{n,T}}\sum_{i=1}^{[nr]-1}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}\sigma(X_s)dB_s\right)\\ &\xrightarrow{d} \mathbf{V}\left(\left(\int_{-\infty}^{\infty}\mathbf{K}^2(c)dc\right)\sigma^2(x)\overline{L}_X(rT,x)\right).\end{aligned}$$

This, in turn, implies that

$$\mathbf{U}_{n,T}(1) \xrightarrow{d} \mathbf{V}\left(\left(\int_{-\infty}^{\infty}\mathbf{K}^2(c)dc\right)\sigma^2(x)\overline{L}_X(T,x)\right).$$

Further,

$$\frac{\mathbf{U}_{n,T}(1)}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})} \xrightarrow{d} \mathbf{V}\left(\left(\int_{-\infty}^{\infty}\mathbf{K}^2(c)dc\right)\frac{\sigma^2(x)}{\overline{L}_X(T,x)}\right).$$

Hence,

$$\begin{aligned}&\sqrt{\overline{L}_X(T,x)h_{n,T}}\left(\frac{\frac{1}{\Delta_{n,T}}\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}\sigma(X_s)dB_s}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})}\right)\\ &\xrightarrow{d} \mathbf{V}\left(\left(\int_{-\infty}^{\infty}\mathbf{K}^2(c)dc\right)\sigma^2(x)\right).\end{aligned}$$

We now study $\mathbf{A}_{n,T}$.

$$\begin{aligned}&\frac{1}{h_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}[\mu(X_s)-\mu(x)]ds\\ =\;&\underbrace{\frac{1}{h_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}\left(\mu(X_s)-\mu(X_{i\Delta_{n,T}})\right)ds}_{\mathbf{A}^1_{n,T}}\\ &+\underbrace{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})\left(\mu(X_{i\Delta_{n,T}})-\mu(x)\right)}_{\mathbf{A}^2_{n,T}}.\end{aligned}$$

By the Lipschitz property of $\mu(.)$, the first term can be bounded as follows,

$$\begin{aligned}
\mathbf{A}^1_{n,T} &= \frac{1}{h_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}\left(\mu(X_s)-\mu(X_{i\Delta_{n,T}})\right)ds \\
&\leq const.\kappa_{n,T}\left(\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})\right)
\end{aligned}$$

where $\kappa_{n,T}$ is, as before, equal to $\max_{i\leq n}\sup_{i\Delta_{n,T}\leq s\leq(i+1)\Delta_{n,T}}|X_s-X_{i\Delta_{n,T}}|$. The bound becomes

$$\mathbf{A}^1_{n,T}\leq\left(o_{a.s.}\left(\Delta_{n,T}^{1/2-\delta}\right)\right)\left(\overline{L}_X(T,x)+o_{a.s}\left(\frac{\overline{L}_X(T,x)}{h_{n,T}}(\Delta_{n,T})^{1/2-\epsilon}\right)\right).$$

Further, by the *mean-value theorem* and the occupation time formula we can write,

$$\begin{aligned}
\mathbf{A}^2_{n,T} &= \frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})\mu'(x_i^*)\left(X_{i\Delta_{n,T}}-x\right) \\
&= \frac{1}{h_{n,T}}\int_0^T\mathbf{K}(\frac{X_s-x}{h_{n,T}})\mu'\left(f(X_s,x)\right)(X_s-x)+o_{a.s.}\left((\Delta_{n,T})^{1/2-\epsilon}\overline{L}_X(T,x)\right) \\
&= \frac{1}{h_{n,T}}\int_{-\infty}^{\infty}\mathbf{K}(\frac{a-x}{h_{n,T}})\mu'\left(f(a,x)\right)(a-x)\overline{L}_X(T,a)da+o_{a.s.}\left((\Delta_{n,T})^{1/2-\epsilon}\overline{L}_X(T,x)\right)
\end{aligned}$$

where $x_i^*=f(X_{i\Delta_{n,T}},x)\in[X_{i\Delta_{n,T}},x]\ \forall i$. If we multiply by $\frac{1}{h_{n,T}}$, then the first term becomes

$$\begin{aligned}
&\frac{1}{h_{n,T}}\left(\frac{1}{h_{n,T}}\int_{-\infty}^{\infty}\mathbf{K}(\frac{a-x}{h_{n,T}})\mu'\left(f(a,x)\right)(a-x)\overline{L}_X(T,a)da\right) \\
&= \frac{1}{h_{n,T}}\int_{-\infty}^{\infty}\mathbf{K}(\frac{a-x}{h_{n,T}})\mu'\left(f(a,x)\right)\left(\frac{a-x}{h_{n,T}}\right)\overline{L}_X(T,a)da \\
&= \int_{-\infty}^{\infty}c\mathbf{K}(c)\mu'\left(f(x,x+h_{n,T}c)\right)\overline{L}_X(T,x+h_{n,T}c)dc+o_{a.s.}(1) \\
&= \int_{-\infty}^{\infty}c\mathbf{K}(c)\left(\frac{\mu'(x)}{\sigma^2(x)}\right)L_X(T,x+h_{n,T}c)dc+o_{a.s.}(1) \\
&= \int_{-\infty}^{\infty}c\mathbf{K}(c)\left(\frac{\mu'(x)}{\sigma^2(x)}\right)\left(L_X(T,x+h_{n,T}c)-L_X(T,x)\right)dc\ +o_{a.s}(1).
\end{aligned}$$

By Lemma 3.5 in Part I and neglecting the smaller order of magnitude,

$$\int_{-\infty}^{\infty}c\mathbf{K}(c)2\left(\frac{\mu'(x)}{\sigma^2(x)}\right)\frac{1}{2\sqrt{h_{n,T}}}\left(L_X(T,x+h_{n,T}c)-L_X(T,x)\right)dc$$

$$
\begin{aligned}
&= 2\left(\frac{\mu'(x)}{\sigma^2(x)}\right)\int_{-\infty}^{\infty} c\mathbf{K}(c)\frac{1}{2\sqrt{h_{n,T}}}\left(L_X(T,x+h_{n,T}c)-L_X(T,x)\right)dc \\
&\xrightarrow{d} 2\left(\frac{\mu'(x)}{\sigma^2(x)}\right)\int_{-\infty}^{\infty} c\mathbf{K}(c)\mathfrak{B}(L_X(T,x),c)dc \\
&\overset{d}{=} 2\left(\frac{\mu'(x)}{\sigma^2(x)}\right)\sqrt{L_X(T,x)}\int_{-\infty}^{\infty} c\mathbf{K}(c)\mathfrak{B}(1,c)dc.
\end{aligned}
$$

Now, define $G(u) = \int_{-\infty}^{u} c\mathbf{K}(c)dc$. We can integrate $\int_{-\infty}^{\infty} c\mathbf{K}(c)\mathfrak{B}(1,c)dc$ by parts and obtain,

$$
\begin{aligned}
&\int_{-\infty}^{\infty} c\mathbf{K}(c)\mathfrak{B}(1,c)dc \\
&= G(c)\mathfrak{B}(1,c)|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} G(c)d\mathfrak{B}(1,c)dc \\
&= -\int_{-\infty}^{\infty} G(c)d\mathfrak{B}(1,c)dc \\
&\overset{d}{=} \mathbf{B}(\int_{-\infty}^{\infty} G(c)^2dc) \\
&\overset{d}{=} \mathbf{B}(\varphi/4)
\end{aligned}
$$

where $\varphi = -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} 2|a-b|ab\mathbf{K}(a)\mathbf{K}(b)dadb$. In consequence,

$$
\begin{aligned}
&\frac{1}{(h_{n,T})^{3/2}}\mathbf{A}_{n,T}^2 = \frac{1}{(h_{n,T})^{3/2}}\left(\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})\left(\mu(X_{i\Delta_{n,T}})-\mu(x)\right)\right) \\
&\xrightarrow{d} \mathbf{B}\left(\varphi\left(\frac{\mu'(x)}{\sigma(x)}\right)^2\overline{L}_X(T,x)\right)
\end{aligned}
$$

where $\mathbf{B}$ is a standard Brownian motion independent of $\overline{L}_X(T,x)$ and φ is a constant of proportionality equal to $2\langle f,f\rangle$ where $\langle f,f\rangle$ is the "energy" of the function $f(s) = s\mathbf{K}(s)$, i.e. $\langle f,f\rangle = -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}|a-b|ab\mathbf{K}(a)\mathbf{K}(b)dadb$. In turn,

$$
\begin{aligned}
&\frac{\sqrt{\overline{L}_X(T,x)}}{(h_{n,T})^{3/2}}\left(\frac{\mathbf{A}_{n,T}^2}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})}\right) \xrightarrow{d} N\left(\varphi\left(\frac{\mu'(x)}{\sigma(x)}\right)^2\right) \\
&\overset{d}{=} N\left(0,\varphi\left(\frac{\mu'(x)}{\sigma(x)}\right)^2\right).
\end{aligned}
$$

Hence,

$$\sqrt{h_{n,T}\overline{L}_X(T,x)}\left(\frac{\mathbf{A}_{n,T}}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})}+\frac{\mathbf{C}_{n,T}(1)}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})}\right)$$

$$= \sqrt{h_{n,T}\overline{L}_X(T,x)}(\frac{\mathbf{A}^1_{n,T}}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})}+\frac{\mathbf{A}^2_{n,T}}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})}$$

$$+\frac{\mathbf{C}_{n,T}(1)}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})})$$

$$= \sqrt{h_{n,T}\overline{L}_X(T,x)}\left(O_{a.s.}\left(\Delta_{n,T}^{1/2-\delta}\right)+O_p\left(\frac{(h_{n,T})^{3/2}}{\sqrt{\overline{L}_X(T,x)}}\right)+\frac{\mathbf{C}_{n,T}(1)}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})}\right)$$

$$= O_p\left(h_{n,T}{}^2\right)+2o_{a.s.}(1)+\sqrt{h_{n,T}\overline{L}_X(T,x)}\left(\frac{\mathbf{C}_{n,T}(1)}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})}\right)$$

$$+O_{a.s.}\left(\Delta_{n,T}^{1/2-\delta}\left(\overline{L}_X(T,x)\right)^{1/2}\sqrt{h_{n,T}}\right)\xrightarrow{d} N\left(\left(\int_{-\infty}^{\infty}\mathbf{K}^2(c)dc\right)\sigma^2(x)\right).$$

This proves the stated result.

22.2. Proof of Theorem 2 [Bandi (1999)]

The proof is similar to the proof of Theorem 1. Consider the estimator

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

We start by proving that

$$\widehat{\sigma}^2_{n,T}(x)-\sigma^2(x)\xrightarrow{a.s.}0.$$

Since

$$\left(X_{(i+1)\Delta_{n,T}}-X_{i\Delta_{n,T}}\right)^2$$

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

then $\widehat{\sigma}^2_{n,T}(x)$ can be written as follows

$$\widehat{\sigma}^2_{n,T}(x) = \underbrace{\frac{1}{\Delta_{n,T}}\frac{\sum_{i=1}^{n-1}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})[\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}2\left(X_s-X_{i\Delta_{n,T}}\right)\mu(X_s)ds]}{\sum_{i=1}^{n}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})}}_{(a)}$$

$$
+\underbrace{\frac{1}{\Delta_{n,T}}\frac{\sum_{i=1}^{n-1}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})[\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}2\left(X_s-X_{i\Delta_{n,T}}\right)\sigma(X_s)dB_s]}{\sum_{i=1}^{n}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})}}_{(\beta)}
$$

$$
+\underbrace{\frac{1}{\Delta_{n,T}}\frac{\sum_{i=1}^{n-1}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})[\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}\sigma^2(X_s)ds]}{\sum_{i=1}^{n}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})}}_{(\gamma)}.
$$

First, we analyze α.

$$
\begin{aligned}
\alpha &= \frac{1}{\Delta_{n,T}}\frac{\sum_{i=1}^{n-1}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})[\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}2\left(X_s-X_{i\Delta_{n,T}}\right)\mu(X_s)ds]}{\sum_{i=1}^{n}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})}\\
&\leq \frac{2\kappa_{n,T}}{\Delta_{n,T}}\frac{\sum_{i=1}^{n-1}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})[\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}\mu(X_s)ds]}{\sum_{i=1}^{n}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})}
\end{aligned}
$$

where $\kappa_{n,T}=\max_{i\leq n}\sup_{i\Delta_{n,T}\leq s\leq(i+1)\Delta_{n,T}}|X_s-X_{i\Delta_{n,T}}|$ as earlier. Hence, by the Lipschitz property of $\mu(.)$ we can write

$$
\begin{aligned}
\alpha &\leq \frac{2\kappa_{n,T}}{\Delta_{n,T}}\frac{\sum_{i=1}^{n-1}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})[\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}\left(\mu(X_s)-\mu(X_{i\Delta_{n,T}})\right)ds]}{\sum_{i=1}^{n}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})}\\
&\quad+\frac{2\kappa_{n,T}}{\Delta_{n,T}}\frac{\sum_{i=1}^{n-1}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})[\mu(X_{i\Delta_{n,T}})\Delta_{n,T}]}{\sum_{i=1}^{n}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})}\\
&\leq const.\,(\kappa_{n,T})^2+\frac{2\kappa_{n,T}}{\Delta_{n,T}}\frac{\sum_{i=1}^{n-1}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})[\mu(X_{i\Delta_{n,T}})\Delta_{n,T}]}{\sum_{i=1}^{n}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})}.
\end{aligned}
$$

Since $\kappa_{n,T}=o_{a.s.}(\Delta_{n,T}^{1/2-\delta})$, then the bound becomes

$$
o_{a.s.}(\Delta_{n,T}^{1-2\delta})+\frac{2\kappa_{n,T}}{\Delta_{n,T}}\frac{\sum_{i=1}^{n-1}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})[\mu(X_{i\Delta_{n,T}})\Delta_{n,T}]}{\sum_{i=1}^{n}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})}.
$$

By a straightforward application of the results in Part I [Theorem 5.11], the second term converges to $2\kappa_{n,T}\mu(x)$ provided $\frac{\overline{L}_X(T,x)}{h_{n,T}}(\Delta_{n,T})^{1/2-\epsilon}\overset{a.s.}{\to}0$ and $\frac{\Delta_{n,T}}{h_{n,T}}\to 0$. In order to prove $a.s.$ convergence it remains to prove that $\beta\overset{a.s.}{\to}0$ and $\gamma\overset{a.s.}{\to}\sigma^2(x)$. We analyse β first. As before, the martingale

$$y_{(i+1)\Delta_{n,T}} = \int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}} \mathbf{K}(\frac{X_{i\Delta_{n,T}} - x}{h_{n,T}}) 2\left(X_s - X_{i\Delta_{n,T}}\right)\sigma(X_s)dB_s$$

is measurable with respect to $\Im_{(i+1)\Delta_{n,T}}$, where $\Im_{(i+1)\Delta_{n,T}} = \{A \in \Im : A\{(i+1)\Delta_{n,T} \leq t^*\} \in \Im_{t+} \forall t \geq 0\}$. Further,

$$\mathbf{E}(y_{(i+1)\Delta_{n,T}}) = 0$$

and, by the Ito isometry

$$\theta_{(i+1)\Delta_{n,T}} = var(y_{(i+1)\Delta_{n,T}}) = \mathbf{E}\left(\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}} 4\mathbf{K}^2(\frac{X_{i\Delta_{n,T}} - x}{h_{n,T}})\left(X_s - X_{i\Delta_{n,T}}\right)^2 \sigma^2(X_s)ds\right) < \infty.$$

Also, $(y_{(i+1)\Delta_{n,T}}, \Im_{(i+1)\Delta_{n,T}})$ is a martingale difference sequence with zero mean and variance $\theta_{(i+1)\Delta_{n,T}}$. We now invoke a strong law of large numbers for martingale differences [e.g. Hall and Heyde (1980, Theorem 2.19, page 36)] to obtain

$$\frac{\sum_{i=1}^{n-1} \mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})[\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}} \sigma(X_s)dB_s]}{\sum_{i=1}^{n-1} \mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})} \stackrel{a.s}{\to} 0.$$

Below we derive the rate of convergence. In particular, we prove that

$$\frac{\frac{1}{\Delta_{n,T}}\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n-1} \mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}} 2\left(X_s - X_{i\Delta_{n,T}}\right)\sigma(X_s)dB_s}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n} \mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})} = O_p\left(\sqrt{\frac{\Delta_{n,T}}{\overline{L}_X(T,x)h_{n,T}}}\right).$$

Now consider

$$\gamma = \frac{1}{\Delta_{n,T}}\frac{\sum_{i=1}^{n-1} \mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})[\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}} \sigma^2(X_s)ds]}{\sum_{i=1}^{n} \mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})}.$$

By the Lipschitz property of $\sigma^2(.)$,

$$\begin{aligned}
\gamma &= \frac{1}{\Delta_{n,T}}\frac{\sum_{i=1}^{n-1} \mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})[\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}} \left(\sigma^2(X_s) - \sigma^2(X_{i\Delta_{n,T}})\right)ds]}{\sum_{i=1}^{n} \mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})} \\
&\quad + \frac{1}{\Delta_{n,T}}\frac{\sum_{i=1}^{n-1} \mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})[\sigma^2(X_{i\Delta_{n,T}})\Delta_{n,T}]}{\sum_{i=1}^{n} \mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})} \\
&\leq const.\kappa_{n,T} + \frac{1}{\Delta_{n,T}}\frac{\sum_{i=1}^{n-1} \mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})[\sigma^2(X_{i\Delta_{n,T}})\Delta_{n,T}]}{\sum_{i=1}^{n} \mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})}.
\end{aligned}$$

The bound becomes

$$o_{a.s.}(\Delta_{n,T}^{1/2-\delta}) + \frac{1}{\Delta_{n,T}} \frac{\sum_{i=1}^{n-1} \mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})[\sigma^2(X_{i\Delta_{n,T}})\Delta_{n,T}]}{\sum_{i=1}^{n} \mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})}.$$

Applying the results in Part I [Theorem 5.5], it is easy to prove that the second term converges to $\sigma^2(x)$ provided $\frac{\overline{L}_X(T,x)}{h_{n,T}}(\Delta_{n,T})^{1/2-\epsilon} \to 0$ and $\frac{\Delta_{n,T}}{h_{n,T}} \to 0$. This proves *a.s.* convergence. Now we study the asymptotic distribution. We are interested in the limiting properties of

$$\begin{aligned} &\widehat{\sigma}^2_{n,T}(x) - \sigma^2(x) \\ = &\frac{1}{\Delta_{n,T}} \frac{\frac{\Delta_{n,T}}{h_{n,T}} \sum_{i=1}^{n-1} \mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})[\left(X_{(i+1)\Delta_{n,T}} - X_{i\Delta_{n,T}}\right)^2 - \sigma^2(x)\Delta_{n,T}]}{\frac{\Delta_{n,T}}{h_{n,T}} \sum_{i=1}^{n} \mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})}. \end{aligned}$$

First, we examine the numerator

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

This can be written as,

$$\begin{aligned} &\frac{1}{h_{n,T}} \sum_{i=1}^{n-1} \mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}}) \left(\left(X_{(i+1)\Delta_{n,T}} - X_{i\Delta_{n,T}}\right)^2 - \sigma^2(x)\Delta_{n,T}\right) \\ = &\underbrace{\frac{1}{h_{n,T}} \sum_{i=1}^{n-1} \mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}}) \int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}} 2\left(X_s - X_{i\Delta_{n,T}}\right)\mu(X_s)ds}_{(\mathbf{B}_{n,T})} \\ &+ \underbrace{\frac{1}{h_{n,T}} \sum_{i=1}^{n-1} \mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}}) \int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}} \left(\sigma^2(X_s) - \sigma^2(x)\right)ds}_{(\mathbf{A}_{n,T})} \\ &+ \underbrace{\frac{1}{h_{n,T}} \sum_{i=1}^{n-1} \mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}}) \int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}} 2\left(X_s - X_{i\Delta_{n,T}}\right)\sigma(X_s)dB_s}_{(\mathbf{C}_{n,T}(1))}. \end{aligned}$$

Consider $\mathbf{U}_{n,T}(r) = \frac{\sqrt{h_{n,T}}}{\sqrt{\Delta_{n,T}}}\mathbf{C}_{n,T}(r)$ which is equal to

$$\mathbf{U}_{n,T}(r) = \frac{1}{\sqrt{h_{n,T}}} \sum_{i=1}^{[nr]-1} \mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}}) \frac{1}{\sqrt{\Delta_{n,T}}} \int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}} 2\left(X_s - X_{i\Delta_{n,T}}\right)\sigma(X_s)dB_s.$$

$\mathbf{U}_{n,T}$ is a continuous martingale whose quadratic variation process $[\mathbf{U}_{n,T}]_r$ is

$$
\begin{aligned}
[\mathbf{U}_{n,T}]_r &= \frac{1}{h_{n,T}} \sum_{i=1}^{[nr]-1} \mathbf{K}^2(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}}) \frac{1}{\Delta_{n,T}} \int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}} 4\left(X_s - X_{i\Delta_{n,T}}\right)^2 \sigma^2(X_s)ds \\
&= 4\frac{\Delta_{n,T}}{h_{n,T}} \sum_{i=1}^{[nr]-1} \mathbf{K}^2(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}}) \sigma^4(X_{i\Delta_{n,T}} + o_{a.s.}(1)) \\
&= 4\frac{1}{h_{n,T}} \int_0^{rT} \mathbf{K}^2(\frac{X_s - x}{h_{n,T}}) \sigma^4(X_s)ds + o_{a.s}\left(\frac{\overline{L}_X(rT,x)}{h_{n,T}}(\Delta_{n,T})^{1/2-\epsilon}\right) \\
&= 4\frac{1}{h_{n,T}} \int_{-\infty}^{\infty} \mathbf{K}^2(\frac{a-x}{h_{n,T}}) \sigma^4(a) \overline{L}_X(rT,a)da + o_{a.s}(1) \\
&= 4\int_{-\infty}^{\infty} \mathbf{K}^2(c)\sigma^4(x + h_{n,T}c)\overline{L}_X(rT, x + h_{n,T}c)dc + o_{a.s}(1) \\
&\xrightarrow{d} \left(4\int_{-\infty}^{\infty} \mathbf{K}^2(c)dc\right)\sigma^4(x)\overline{L}_X(rT,x) + o_{a.s}(1).
\end{aligned}
$$

Also, the covariation process $[\mathbf{U}_{n,T}, B]_r \to 0$ $a.s$. Then, let

$$\rho_{n,T}(r) = \inf\{s : [\mathbf{U}_{n,T}]_s > r\}$$

be a sequence of time changes. Define $\mathbf{V}_{n,T}(r) = \mathbf{U}_{n,T}(\rho_{n,T}(r))$. The process $\mathbf{V}_{n,T}(r)$ is the *Dambis, Dubins-Schwartz* Brownian motion of the martingale $\mathbf{U}_{n,T}$ [see, for example, Revuz and Yor (1991, theorem 1.6, p.170)]. The conditions on the variation and covariation process are sufficient to ensure that

$$(\mathbf{V}_{n,T}, B) \xrightarrow{d} (\mathbf{V}, B)$$

where $(\mathbf{V}, B)$ is a vector of independent Brownian motions. Hence,

$$
\begin{aligned}
\mathbf{U}_{n,T}(r) &= \frac{\sqrt{h_{n,T}}}{\sqrt{\Delta_{n,T}}}\left(\frac{1}{h_{n,T}} \sum_{i=1}^{[nr]-1} \mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}}) \int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}} 2\left(X_s - X_{i\Delta_{n,T}}\right)\sigma(X_s)dB_s\right) \\
&\xrightarrow{d} \mathbf{V}\left(\left(4\int_{-\infty}^{\infty}\mathbf{K}^2(c)dc\right)\sigma^4(x)\overline{L}_X(rT,x)\right).
\end{aligned}
$$

This, in turn, implies that

$$\mathbf{U}_{n,T}(1) \xrightarrow{d} \mathbf{V}\left(\left(4\int_{-\infty}^{\infty}\mathbf{K}^2(c)dc\right)\sigma^4(x)\overline{L}_X(T,x)\right).$$

Further,

$$\frac{\mathbf{U}_{n,T}(1)}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})} \xrightarrow{d} \mathbf{V}\left(\left(4\int_{-\infty}^{\infty}\mathbf{K}^2(c)dc\right)\frac{\sigma^4(x)}{\overline{L}_X(T,x)}\right).$$

Hence,

$$\frac{\sqrt{\overline{L}_X(T,x)h_{n,T}}}{\sqrt{\Delta_{n,T}}}\frac{\left(\frac{1}{\Delta_{n,T}}\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{[nr]-1}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}2\left(X_s-X_{i\Delta_{n,T}}\right)\sigma(X_s)dB_s\right)}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})}$$

$$\xrightarrow{d}\mathbf{V}\left(\left(4\int_{-\infty}^{\infty}\mathbf{K}^2(c)dc\right)\sigma^4(x)\right).$$

We now study $\mathbf{A}_{n,T}$.

$$\begin{aligned}
&\frac{1}{h_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}\left(\sigma^2(X_s)-\sigma^2(X_{i\Delta_{n,T}})\right)ds\\
=&\underbrace{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}\left(\sigma^2(X_s)-\sigma^2(X_{i\Delta_{n,T}})\right)ds}_{\mathbf{A}^1_{n,T}}\\
&+\underbrace{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})\left(\sigma^2(X_{i\Delta_{n,T}})-\sigma^2(x)\right)}_{\mathbf{A}^2_{n,T}}.
\end{aligned}$$

The first term can be bounded as follows

$$\begin{aligned}
\mathbf{A}^1_{n,T} &= \frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}\left(\sigma^2(X_s)-\sigma^2(X_{i\Delta_{n,T}})\right)ds\\
&\leq \left(\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})\right)\kappa_{n,T}.
\end{aligned}$$

Hence, the bound becomes

$$\mathbf{A}^1_{n,T}\leq\left(o_{a.s.}\left(\Delta_{n,T}^{1/2-\delta}\right)\right)\left(\overline{L}_X(T,x)+o_{a.s}\left(\frac{\overline{L}_X(T,x)}{h_{n,T}}(\Delta_{n,T})^{1/2-\epsilon}\right)\right).$$

Further, by the *mean-value theorem* and the occupation time formula we can write,

$$\begin{aligned}
\mathbf{A}^2_{n,T} &= \frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})\sigma'\left(x_i^*\right)\sigma\left(x_i^*\right)\left(X_{i\Delta_{n,T}}-x\right)\\
&= \frac{1}{h_{n,T}}\int_0^T\mathbf{K}(\frac{X_s-x}{h_{n,T}})\sigma'\left(f(X_s,x)\right)\sigma\left(f(X_s,x)\right)\left(X_s-x\right)+O_{a.s.}\left((\Delta_{n,T})^{1/2-\epsilon}\overline{L}_X(T,x)\right)\\
&= \underbrace{\frac{1}{h_{n,T}}\int_{-\infty}^{\infty}\mathbf{K}(\frac{a-x}{h_{n,T}})\sigma'\left(f(a,x)\right)\sigma\left(f(a,x)\right)(a-x)\overline{L}_X(T,a)da}_{\mathbf{A}^{2(1)}_{n,T}}+O_{a.s.}\left((\Delta_{n,T})^{1/2-\epsilon}\overline{L}_X(T,x)\right)
\end{aligned}$$

where $x_i^* = f(X_{i\Delta_{n,T}}, x) \in [X_{i\Delta_{n,T}}, x]$. If we multiply by $\frac{1}{h_{n,T}}$, then the first term becomes

$$
\begin{aligned}
& \frac{1}{h_{n,T}}\left(\frac{1}{h_{n,T}}\int_{-\infty}^{\infty}\mathbf{K}(\frac{a-x}{h_{n,T}})\sigma'(f(a,x))\,\sigma(f(a,x))\,(a-x)\overline{L}_X(T,a)da\right) \\
= & \frac{1}{h_{n,T}}\int_{-\infty}^{\infty}\mathbf{K}(\frac{a-x}{h_{n,T}})\sigma'(f(a,x))\,\sigma(f(a,x))\left(\frac{a-x}{h_{n,T}}\right)\overline{L}_X(T,a)da \\
= & \int_{-\infty}^{\infty} c\mathbf{K}(c)\left(\sigma'(f(a,x))\,\sigma(f(a,x))\right)\overline{L}_X(T,x+h_{n,T}c)dc + o_{a.s.}(1) \\
= & \int_{-\infty}^{\infty} c\mathbf{K}(c)\left(\frac{\sigma'(x)}{\sigma(x)}\right)L_X(T.x+h_{n,T}c)dc + o_{a.s.}(1) \\
= & \int_{-\infty}^{\infty} c\mathbf{K}(c)\left(\frac{\sigma'(x)}{\sigma(x)}\right)(L_X(T,x+h_{n,T}c)-L_X(T,x))\,dc \; + o_{a.s}(1).
\end{aligned}
$$

By Lemma 3.5 in Part I and neglecting the smaller order of magnitude,

$$
\begin{aligned}
& \int_{-\infty}^{\infty} c\mathbf{K}(c)2\left(\frac{\sigma'(x)}{\sigma(x)}\right)\frac{1}{2\sqrt{h_{n,T}}}(L_X(T,x+h_{n,T}c)-L_X(T,x))\,dc \\
= & 2\left(\frac{\sigma'(x)}{\sigma(x)}\right)\int_{-\infty}^{\infty} c\mathbf{K}(c)\frac{1}{2\sqrt{h_{n,T}}}(L_X(T,x+h_{n,T}c)-L_X(T,x))\,dc \\
& \xrightarrow{d} 2\left(\frac{\sigma'(x)}{\sigma(x)}\right)\int_{-\infty}^{\infty} c\mathbf{K}(c)\mathfrak{B}(L_X(T,x),c)dc \\
& \stackrel{d}{=} 2\left(\frac{\sigma'(x)}{\sigma(x)}\right)\sqrt{L_X(T,x)}\int_{-\infty}^{\infty} c\mathbf{K}(c)\mathfrak{B}(1,c)dc.
\end{aligned}
$$

Now, define $G(u) = \int_{-\infty}^{u} c\mathbf{K}(c)dc$. We can integrate $\int_{-\infty}^{\infty} c\mathbf{K}(c)\mathfrak{B}(1,c)dc$ by parts and obtain,

$$
\begin{aligned}
& \int_{-\infty}^{\infty} c\mathbf{K}(c)\mathfrak{B}(1,c)dc \\
= & G(c)\mathfrak{B}(1,c)|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} G(c)d\mathfrak{B}(1,c)dc \\
= & -\int_{-\infty}^{\infty} G(c)d\mathfrak{B}(1,c)dc \\
& \stackrel{d}{=} \mathbf{B}(\int_{-\infty}^{\infty} G(c)^2dc) \\
& \stackrel{d}{=} \mathbf{B}(\varphi/4)
\end{aligned}
$$

where $\varphi = -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} 2|a-b|ab\mathbf{K}(a)\mathbf{K}(b)dadb$. In consequence,

$$\frac{1}{(h_{n,T})^{3/2}}\mathbf{A}^2_{n,T} = \frac{1}{(h_{n,T})^{3/2}}\left(\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})\left(\mu(X_{i\Delta_{n,T}})-\mu(x)\right)\right)$$
$$\xrightarrow{d} \mathbf{B}\left(4\varphi\left(\sigma^{'}(x)\right)^2\overline{L}_X(T,x)\right)$$

where $\mathbf{B}$ is a standard Brownian motion independent of $\overline{L}_X(T,x)$ and φ is a constant of proportionality equal to $2\langle f,f\rangle$ where $\langle f,f\rangle$ is the "energy" of the function $f(s)=s\mathbf{K}(s)$, i.e. $\langle f,f\rangle = -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}|a-b|ab\mathbf{K}(a)\mathbf{K}(b)dadb$. In turn,

$$\frac{\sqrt{\overline{L}_X(T,x)}}{(h_{n,T})^{3/2}}\left(\frac{\mathbf{A}^2_{n,T}}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})}\right) \xrightarrow{d} \mathbf{B}\left(\varphi\left(\sigma^{'}(x)\right)^2\right)$$
$$\stackrel{d}{=} N\left(0,4\varphi\left(\sigma^{'}(x)\right)^2\right).$$

Also,

$$\frac{\mathbf{B}_{n,T}}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})}$$
$$= \frac{\frac{1}{h_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}2\left(X_s-X_{i\Delta_{n,T}}\right)\mu(X_s)ds}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})} \leq o_{a.s}\left(\frac{\sqrt{\Delta_{n,T}}}{\sqrt{h_{n,T}\overline{L}_X(T,x)}}\right).$$

Hence,

$$\frac{\sqrt{h_{n,T}\overline{L}_X(T,x)}}{\Delta_{n,T}^{1/2}}\left(\frac{h^{3/2}}{\sqrt{\overline{L}_X(T,x)}}+o_{a.s.}\left(\Delta_{n,T}^{1/2-\delta}\right)+\frac{\mathbf{C}_{n,T}(1)}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})}\right.$$
$$\left.+o_{a.s.}\left(\frac{\sqrt{\Delta_{n,T}}}{\sqrt{h_{n,T}\overline{L}_X(T,x)}}\right)\right)\xrightarrow{d}\left(\left(4\int_{-\infty}^{\infty}\mathbf{K}^2(c)dc\right)\sigma^4(x)\right)$$

provided $h_{n,T}\overline{L}_X(T,x)\xrightarrow{a.s.}0$ and $\frac{h^4_{n,T}}{\Delta_{n,T}}\to 0$. If $\frac{h^4_{n,T}}{\Delta_{n,T}}\to\infty$ and $\frac{\sqrt{\overline{L}_X(T,x)\Delta_{n,T}}}{h^{3/2}}\xrightarrow{a.s.}0$, then

$$\frac{\sqrt{\overline{L}_X(T,x)}}{h^{3/2}}\left(o_{a.s.}\left(\Delta_{n,T}^{1/2-\delta}\right)+O_{a.s.}\left(\frac{\sqrt{\Delta_{n,T}}}{\sqrt{h_{n,T}\overline{L}_X(T,x)}}\right)+\frac{\mathbf{A}^{2(1)}_{n,T}}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}})}\right)$$
$$\xrightarrow{d} N\left(0,4\varphi\left(\sigma^{'}(x)\right)^2\right)$$

where $\varphi = -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}2|a-b|ab\mathbf{K}(a)\mathbf{K}(b)dadb$. This proves the stated result.

23. Notation

$\to_{a.s.}$	almost sure convergence
$\to_{p}$	convergence in probability
$\Rightarrow, \to_d$	weak convergence
$:=$	definitional equality
$o_p(1)$	tends to zero in probability
$O_p(1)$	bounded in probability
$o_{a.s.}(1)$	tends to zero almost surely
$O_{a.s.}(1)$	bounded almost surely
$=_d$	distributional equivalence
$\sim_d$	asymptotically distributed as
$\mathrm{MN}(0, V)$	mixed normal distribution with variance V
$\mathbf{1}_A$	indicator function for the set A
$a \vee b$	$\max\{a, b\}$
C_i with $i = 1, 2, ...$	constants

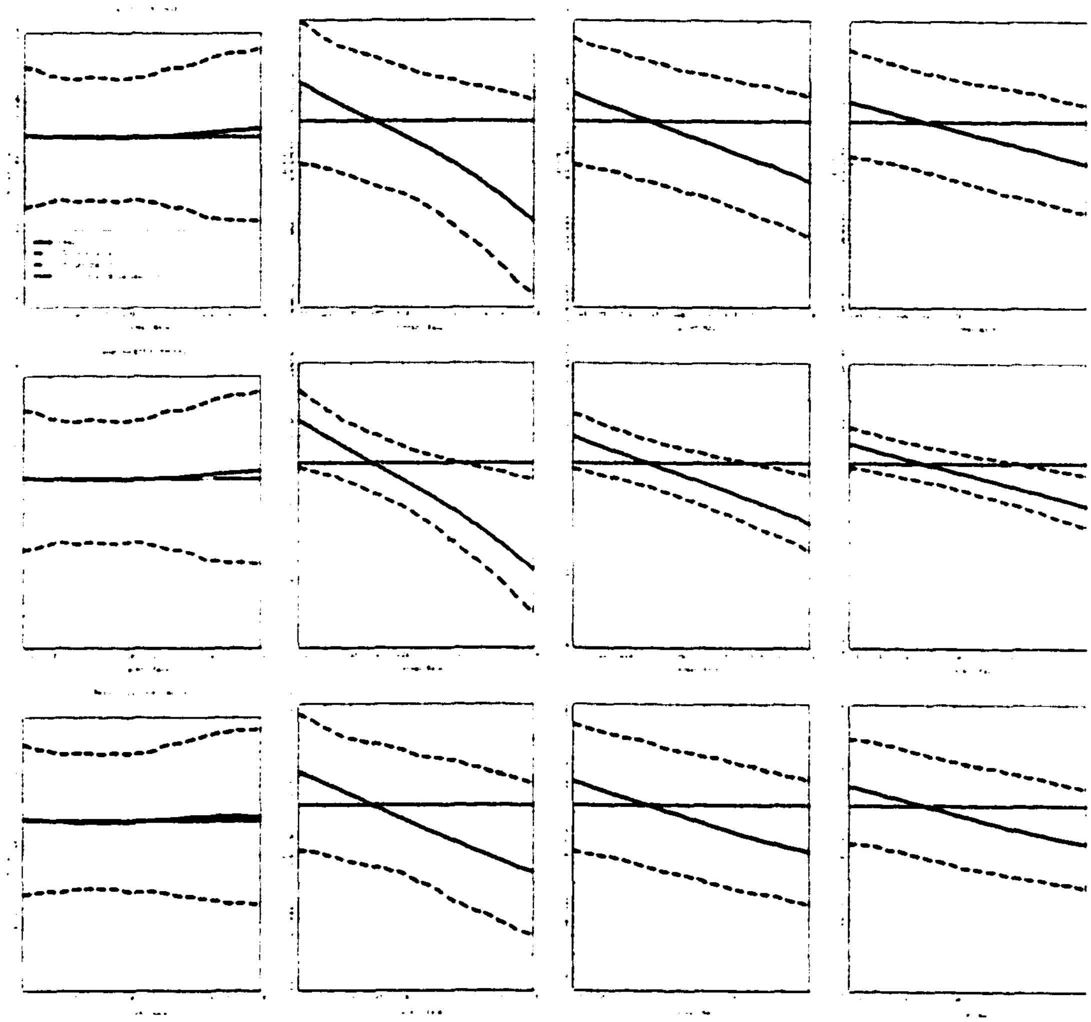

Figure 8: The underlying process is Brownian motion. We simulate it by antithetic-variate method for 5000 daily observations, with 1000 repetitions. Diffusion and drift estimates obtained using the methods in Stanton (1997), JK (1997) and BP (1998) are in row 1, row 2 and row 3, respectively. The solid lines are the estimated functions averaged across the 1000 repetitions. The dotted lines are the true functions. The dashed lines are 75 and 25 percentiles. Diffusion estimates and related curves are in the first column. Notice that the estimated diffusions in Stanton (1997) and JK (1997) coincide [see text]. Drift estimates and related curves for increasing numerical values of the leading bandwidths are in column 2, 3 and 4. For a discussion of the bandwidth choices see text.

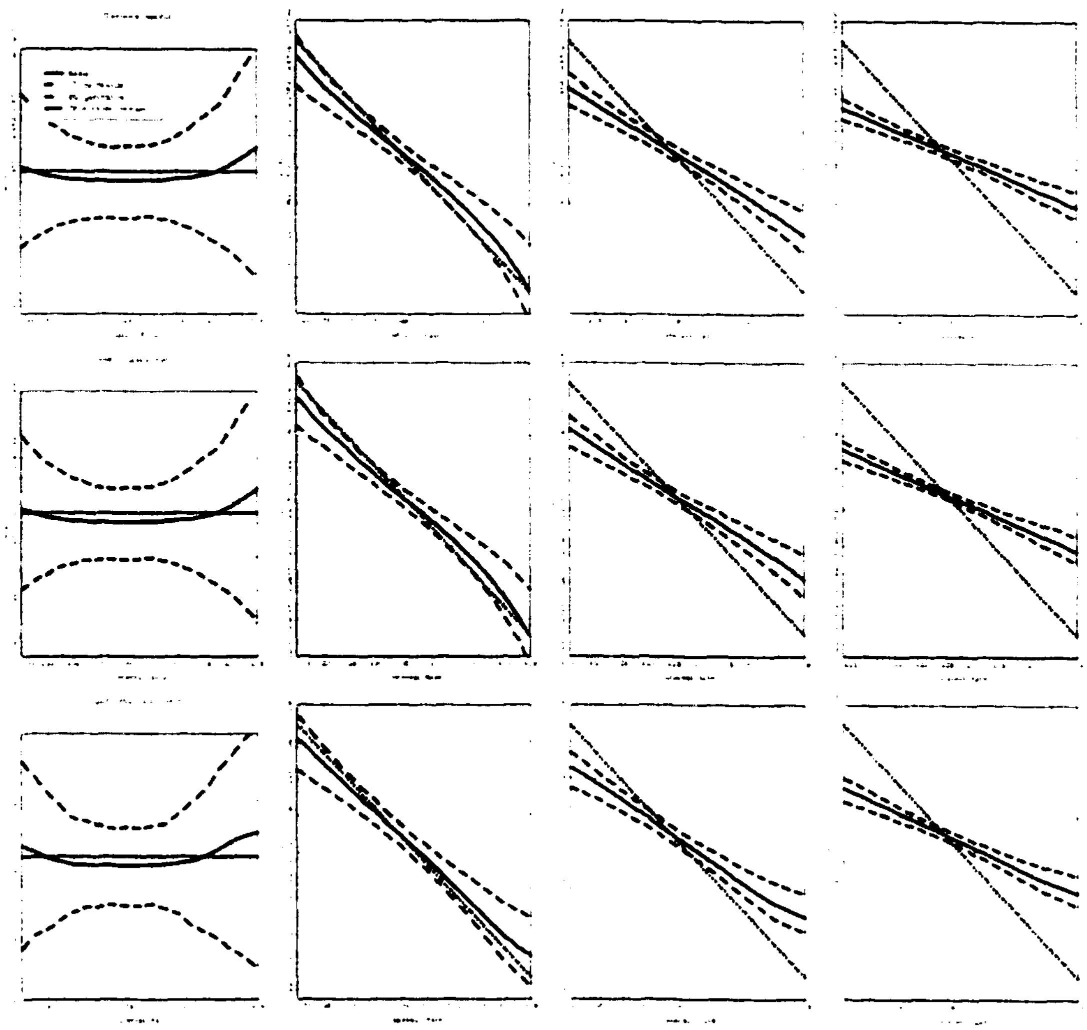

Figure 9: The underlying process is the Vasicek process 1. We simulate it by antithetic-variate method for 5000 daily observations, with 1000 repetitions. Diffusion and drift estimates obtained using the methods in Stanton (1997), JK (1997) and BP (1998) are in row 1, row 2 and row 3, respectively. The solid lines are the estimated functions averaged across the 1000 repetitions. The dotted lines are the true functions. The dashed lines are 75 and 25 percentiles. Diffusion estimates and related curves are in the first column. Notice that the estimated diffusions in Stanton (1997) and JK (1997) coincide [see text]. Drift estimates and related curves for increasing numerical values of the leading bandwidths are in column 2, 3 and 4. For a discussion of the bandwidth choices see text.

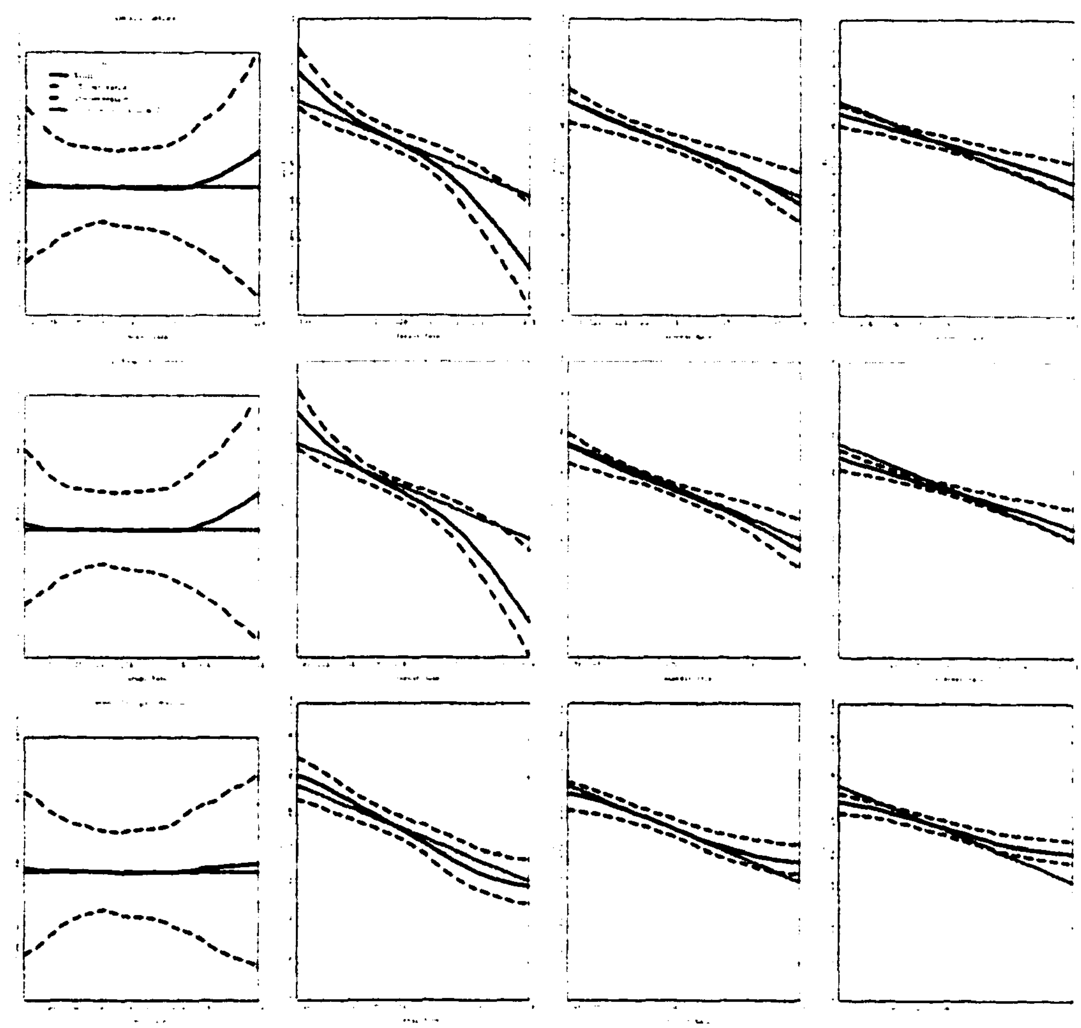

Figure 10: The underlying process is the Vasicek process 2. We simulate it by antithetic-variate method for 5000 daily observations, with 1000 repetitions. Diffusion and drift estimates obtained using the methods in Stanton (1997), JK (1997) and BP (1998) are in row 1, row 2 and row 3, respectively. The solid lines are the estimated functions averaged across the 1000 repetitions. The dotted lines are the true functions. The dashed lines are 75 and 25 percentiles. Diffusion estimates and related curves are in the first column. Notice that the estimated diffusions in Stanton (1997) and JK (1997) coincide [see text]. Drift estimates and related curves for increasing numerical values of the leading bandwidths are in column 2, 3 and 4. For a discussion of the bandwidth choices see text.

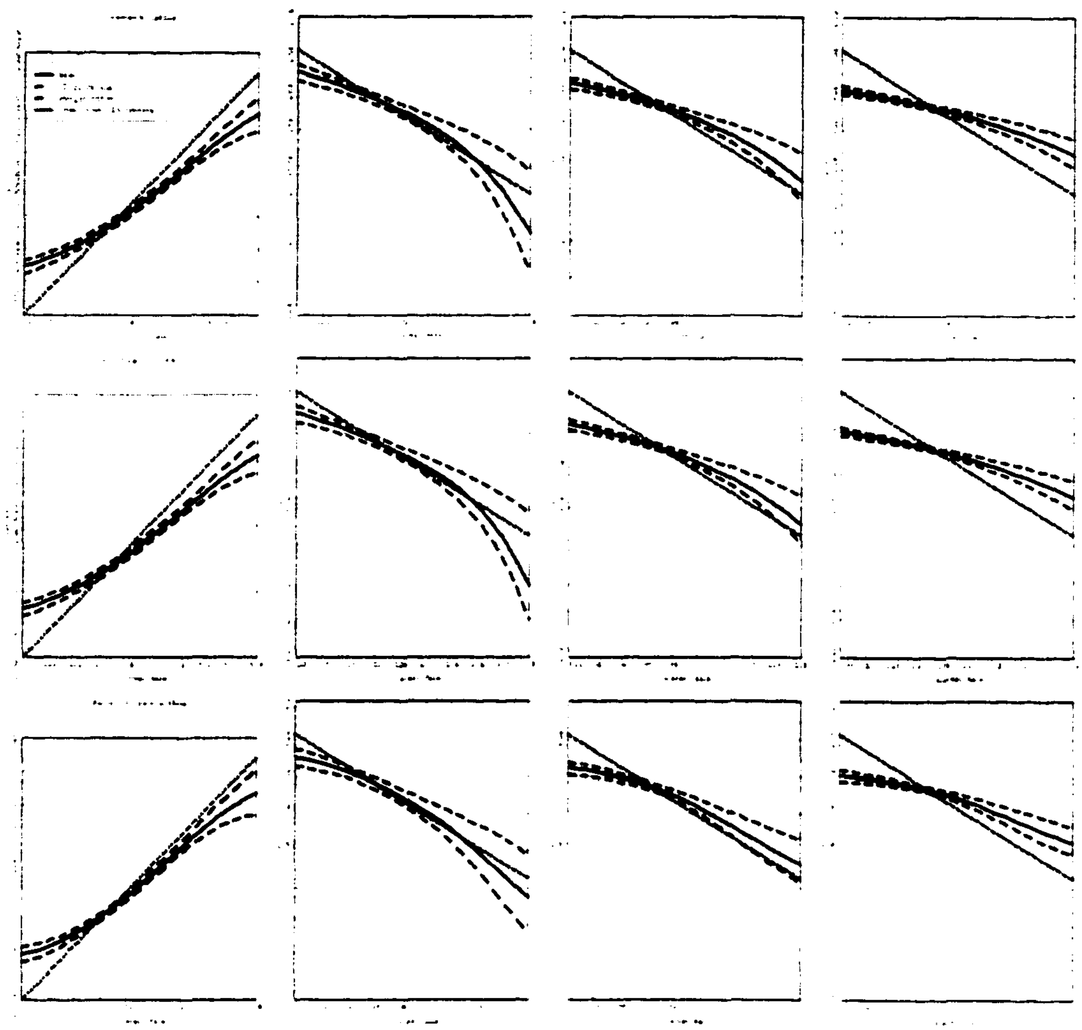

Figure 11: The underlying process is the CIR process 1. We simulate it by antithetic-variate method for 5000 daily observations, with 1000 repetitions. Diffusion and drift estimates obtained using the methods in Stanton (1997), JK (1997) and BP (1998) are in row 1, row 2 and row 3, respectively. The solid lines are the estimated functions averaged across the 1000 repetitions. The dotted lines are the true functions. The dashed lines are 75 and 25 percentiles. Diffusion estimates and related curves are in the first column. Notice that the estimated diffusions in Stanton (1997) and JK (1997) coincide [see text]. Drift estimates and related curves for increasing numerical values of the leading bandwidths are in column 2, 3 and 4. For a discussion of the bandwidth choices see text.

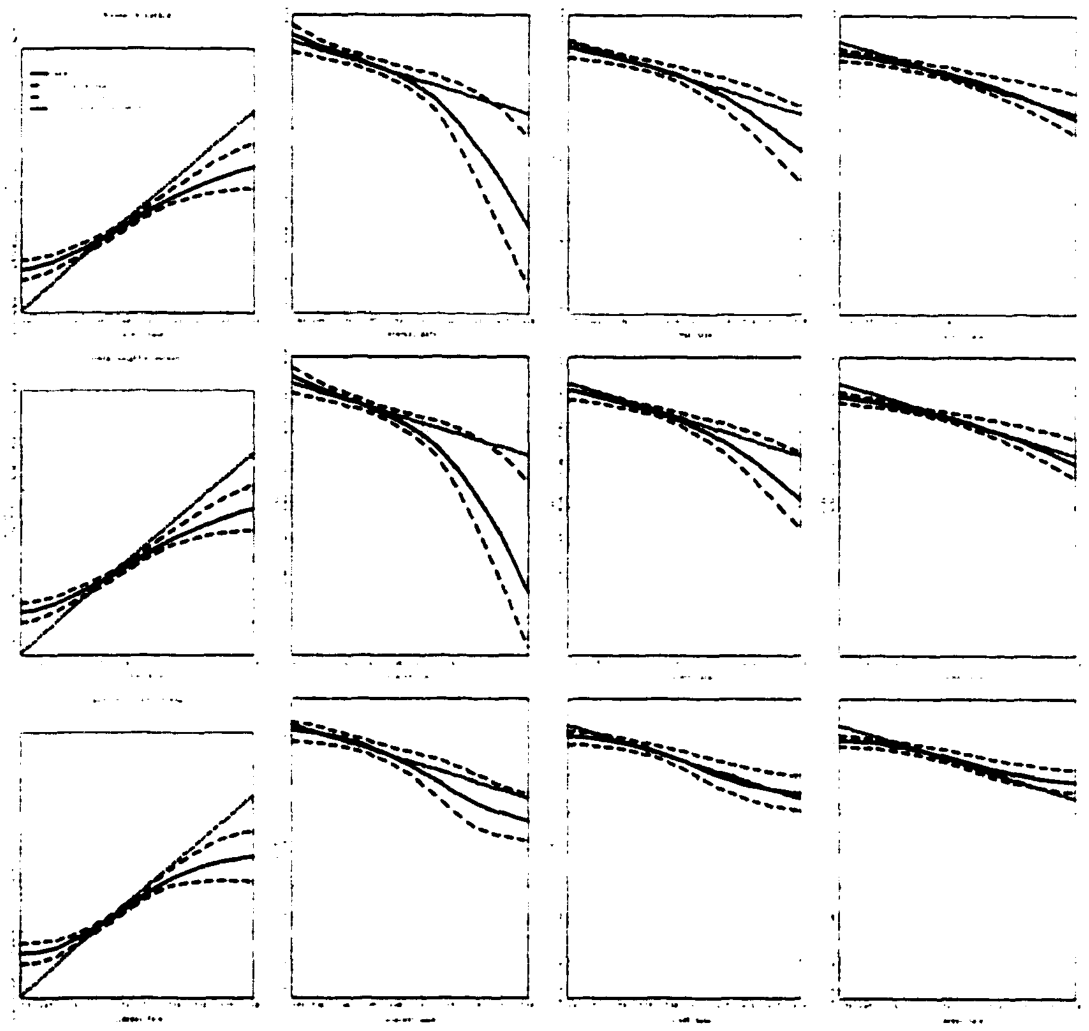

Figure 12: The underlying process is the CIR process 2. We simulate it by antithetic-variate method for 5000 daily observations, with 1000 repetitions. Diffusion and drift estimates obtained using the methods in Stanton (1997), JK (1997) and BP (1998) are in row 1, row 2 and row 3, respectively. The solid lines are the estimated functions averaged across the 1000 repetitions. The dotted lines are the true functions. The dashed lines are 75 and 25 percentiles. Diffusion estimates and related curves are in the first column. Notice that the estimated diffusions in Stanton (1997) and JK (1997) coincide [see text]. Drift estimates and related curves for increasing numerical values of the leading bandwidths are in column 2, 3 and 4. For a discussion of the bandwidth choices see text.

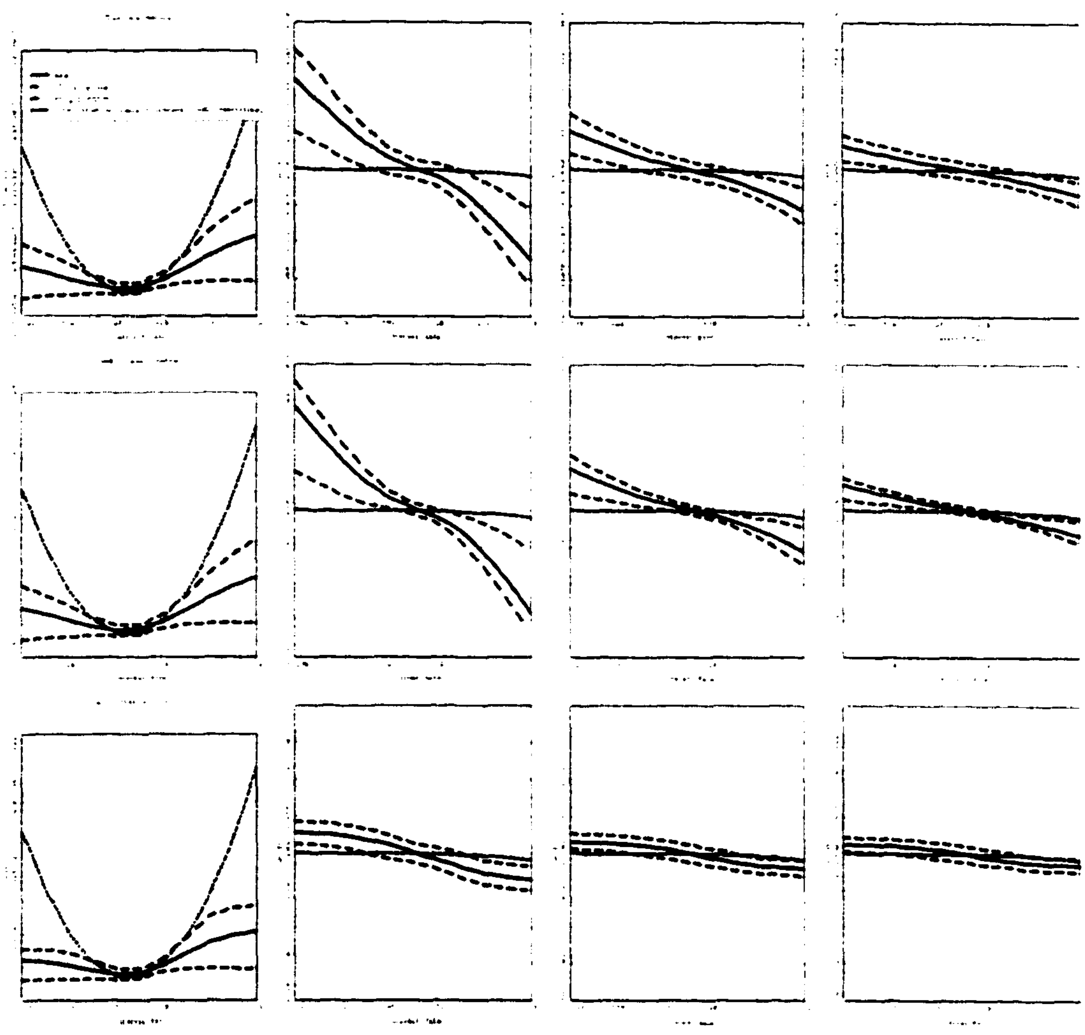

Figure 13: The underlying process is the Aït-Sahalia process. We simulate it by antithetic-variate method for 5000 daily observations, with 1000 repetitions. Diffusion and drift estimates obtained using the methods in Stanton (1997), JK (1997) and BP (1998) are in row 1, row 2 and row 3, respectively. The solid lines are the estimated functions averaged across the 1000 repetitions. The dotted lines are the true functions. The dashed lines are 75 and 25 percentiles. Diffusion estimates and related curves are in the first column. Notice that the estimated diffusions in Stanton (1997) and JK (1997) coincide [see text]. Drift estimates and related curves for increasing numerical values of the leading bandwidths are in column 2, 3 and 4. For a discussion of the bandwidth choices see text.

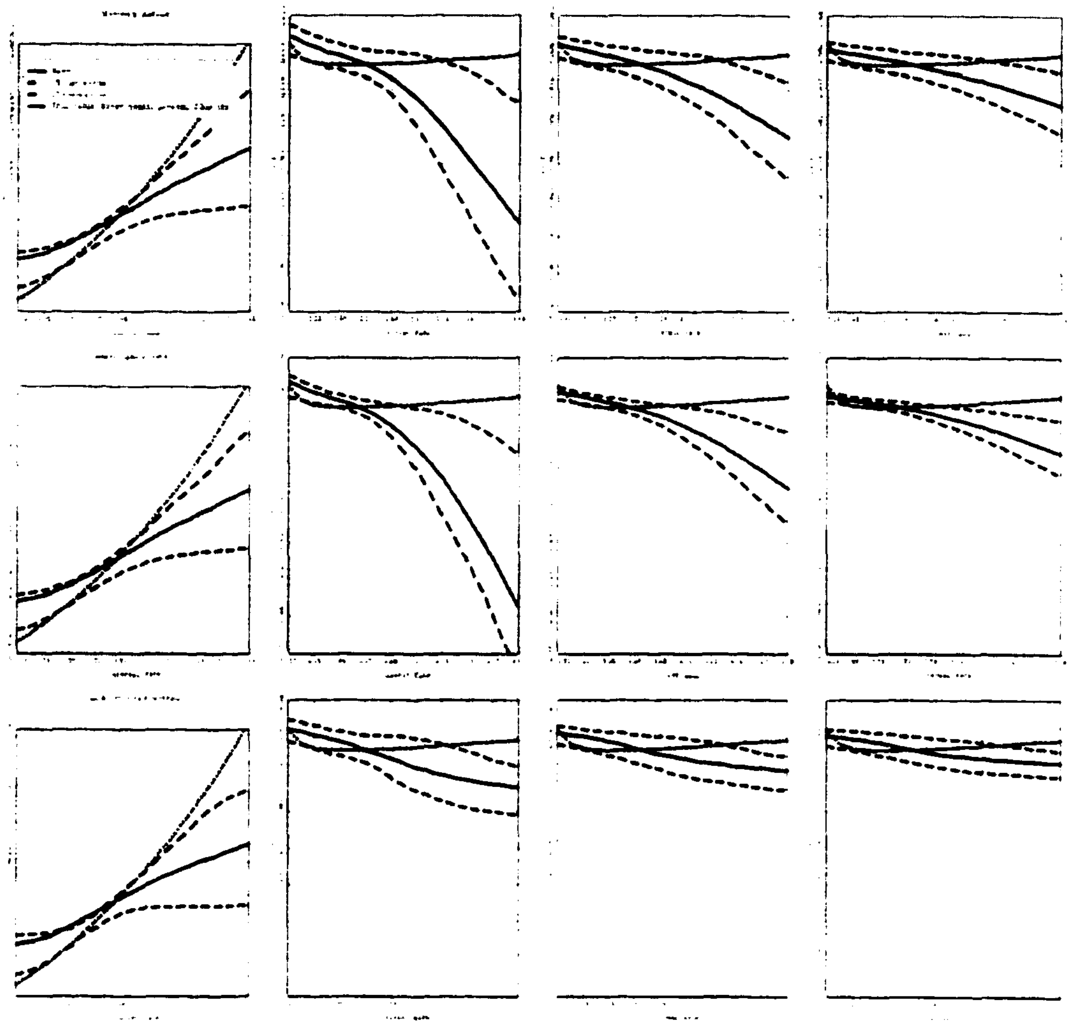

Figure 14: The underlying process is the "experimental" process. We simulate it by antithetic-variate method for 5000 daily observations, with 1000 repetitions. Diffusion and drift estimates obtained using the methods in Stanton (1997), JK (1997) and BP (1998) are in row 1, row 2 and row 3, respectively. The solid lines are the estimated functions averaged across the 1000 repetitions. The dotted lines are the true functions. The dashed lines are 75 and 25 percentiles. Diffusion estimates and related curves are in the first column. Notice that the estimated diffusions in Stanton (1997) and JK (1997) coincide [see text]. Drift estimates and related curves for increasing numerical values of the leading bandwidths are in column 2, 3 and 4. For a discussion of the bandwidth choices see text.

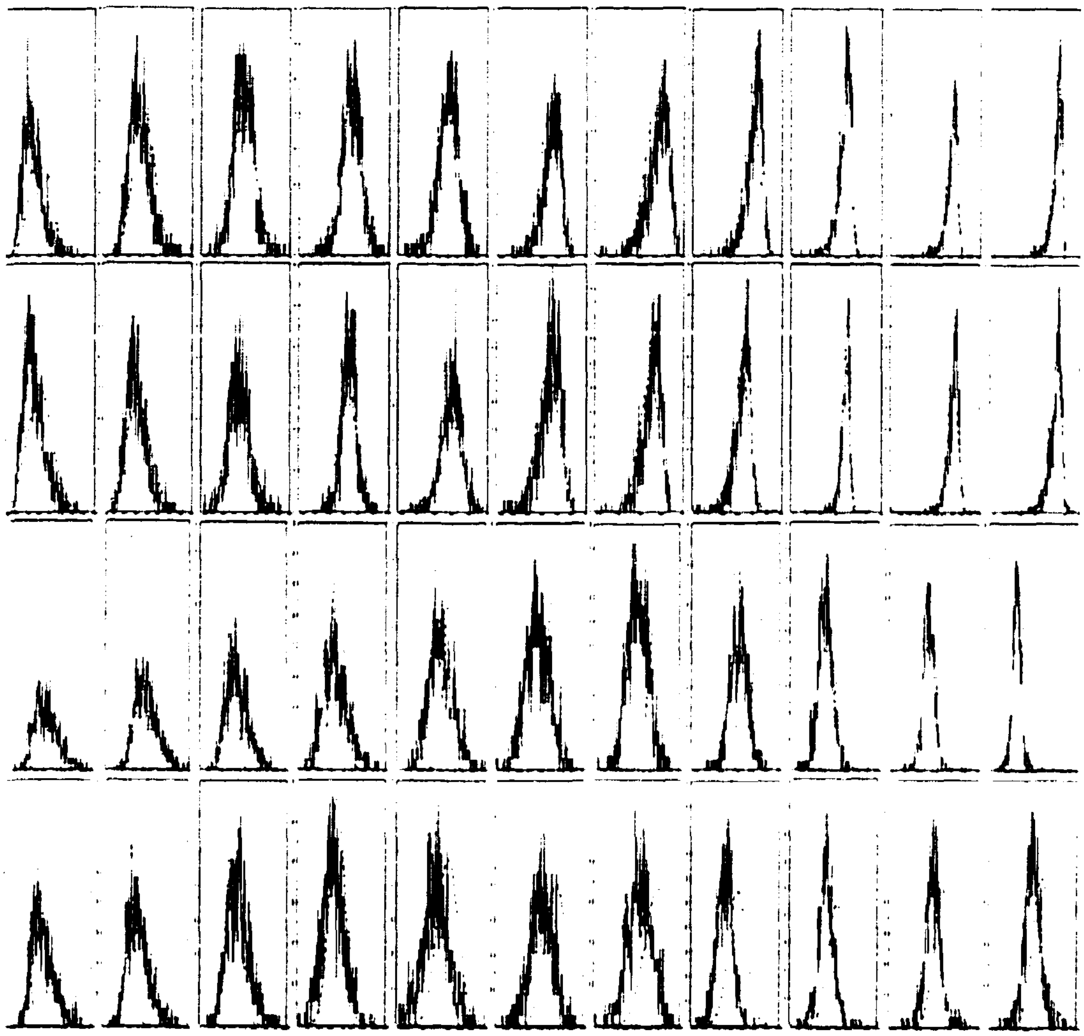

Figure 15: Finite sample and asymptotic distributions for the Stanton (1997) and BP (1998) estimators. The underlying process is the CIR process 1. Row 1 and row 2 contain the pointwise limiting densities [dashed lines] of the drift estimates and their finite sample counterparts in the case of Stanton (1997) and BP (1998), respectively. In row 3 and row 4 are plots of the pointwise limiting and finite sample distributions of the diffusion estimates in the Stanton's and BP's case, respectively. We plot the distributions for values of the underlying process that range from 5% up to 15%.

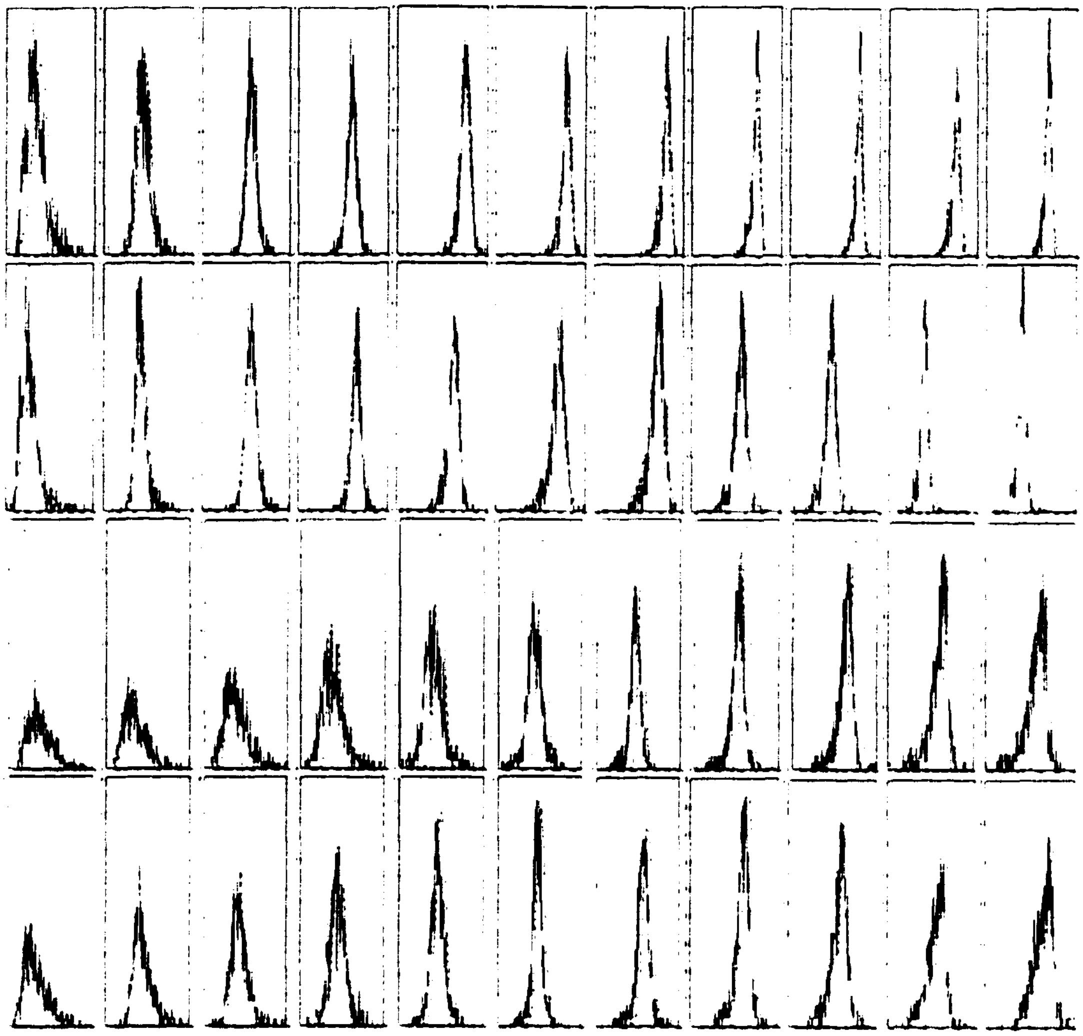

Figure 16: Finite sample and asymptotic distributions for the Stanton (1997) and BP (1998) estimators. The underlying process is the CIR process 2. Row 1 and row 2 contain the pointwise limiting densities [dashed lines] of the drift estimates and their finite sample counterparts in the case of Stanton (1997) and BP (1998), respectively. In row 3 and row 4 are plots of the pointwise limiting and finite sample distributions of the diffusion estimates in the Stanton's and BP's case, respectively. We plot the distributions for values of the underlying process that range from 5% up to 15%.

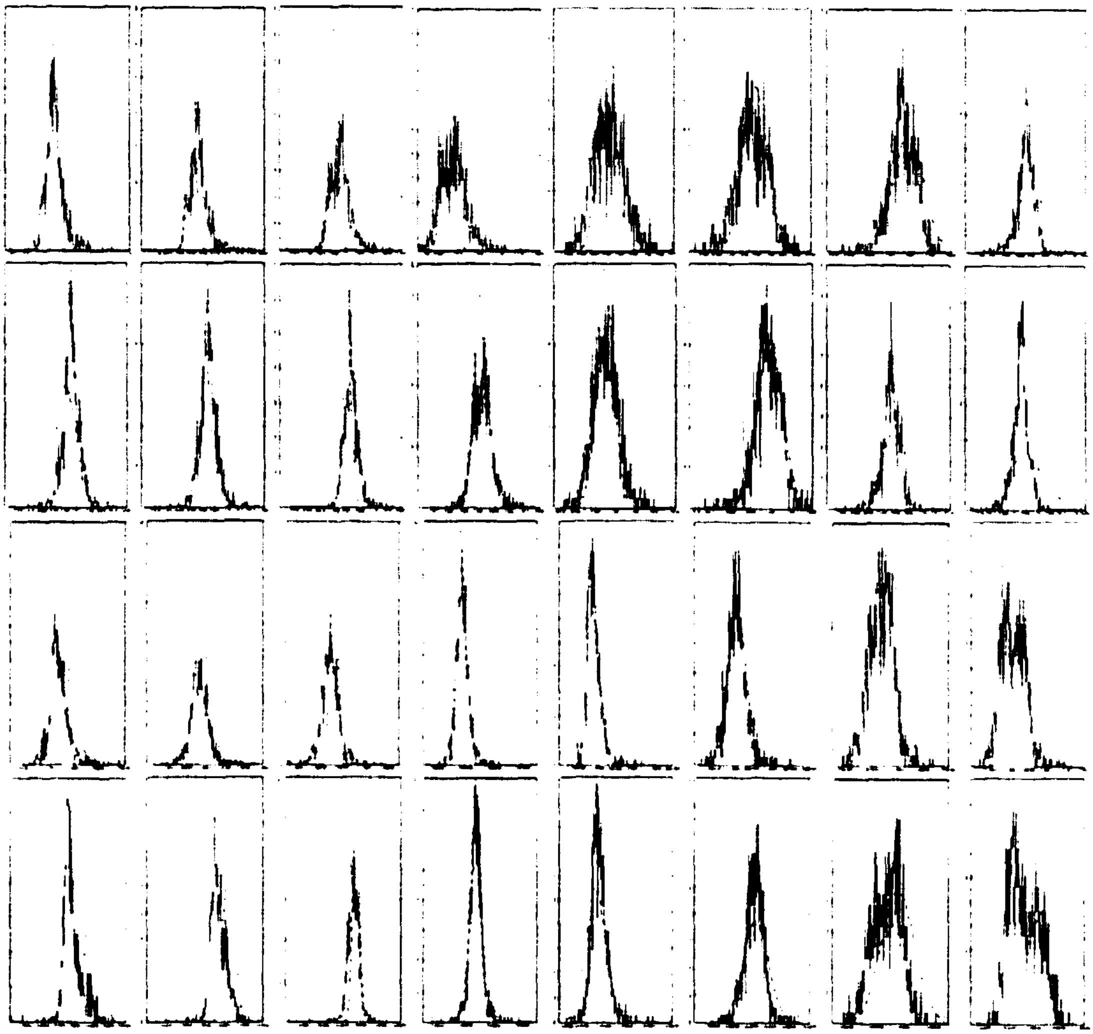

Figure 17: Finite sample and asymptotic distributions for the Stanton (1997) and BP (1998) estimators. The underlying process is the Aït-Sahalia process. Row 1 and row 2 contain the pointwise limiting densities [dashed lines] of the drift estimates and their finite sample counterparts in the case of Stanton (1997) and BP (1998), respectively. In row 3 and row 4 are plots of the pointwise limiting and finite sample distributions of the diffusion estimates in the Stanton's and BP's case, respectively. We plot the distributions for values of the underlying process that range from 5% up to 12%.

References

[1] Aït-Sahalia, Y., 1996a. Nonparametric pricing of interest rate derivative securities. Econometrica 64, 527-560.

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

[3] Aït-Sahalia Y., 1998a. Do interest rates really follow continuous-time markov diffusions? Unpublished working paper, University of Chicago.

[4] Aït-Sahalia, Y., 1998b. Maximum likelihood estimation of discretely sampled diffusions: A Closed-Form Approach. Unpublished working paper, University of Chicago.

[5] Andersen, G.T., Lund, J., 1997. Estimating continuous-time stochastic volatility models of the short-term interest rate. Journal of Econometrics 77, 343-377.

[6] Bandi, F., 1998. Short-term interest rate dynamics: a spatial approach. Unpublished working paper, Yale University.

[7] Bandi, F., 1999. Stanton's (1997) approximations to the first and second moment of a diffusion: statistical properties and implications for pricing. Unpublished working paper, Yale University.

[8] Bandi, F., Nguyen T., 1999. Fully nonparametric estimators for diffusions: a small sample analysis. Unpublished working paper, Yale University.

[9] Bandi, F., Phillips P.C.B., 1998. Econometric estimation of diffusion models. Unpublished working paper, Yale University.

[10] Bannon, G., 1978. Nonparametric identification for diffusion processes. SIAM Journal of Optimal Control and Optimization 16, 380-395.

[11] Bannon, G., Nguyen, H.T., 1981. Recursive estimation in diffusion models. SIAM Journal of Optimal Control and Optimization 19, 676-685.

[12] Baxter, M., Rennie, A., 1996. Financial Calculus. An Introduction to Derivative Pricing. Cambridge University Press, Cambridge.

[13] Bergstrom, A.R., 1998. A history of continuous-time econometric models. Econometric theory 4, 365-383.

[14] Black, F., Karasinsky, P., 1991. Bond and option pricing when short rates are lognormal. Financial Analysts Journal, 52-59.

[15] Boudoukh, J., Richardson, M., Stanton R., Whitelaw, R.F., 1998. The stochastic behavior of interest rates: implications from a multifactor, nonlinear continuous-time model. Unpublished working paper.

[16] Breeden, D. T., Gibbons, M., Litzenberger, R., 1989. Empirical tests of the consumption-oriented CAPM. Journal of Finance 44, 231-262.

[17] Brennan, R.J., Schwartz, E.S., 1979. A continuous-time approach to the pricing of bonds. Journal of Banking and Finance 3, 133-155.

[18] Brown, B.M., Hewitt, J.I., 1975. Asymptotic likelihood theory for diffusion processes. Journal of Applied Probability 12, 228-238.

[19] Brown, R., Schaefer, S., 1991. Interest rate volatility and the term structure. Unpublished working paper, London School of Economics.

[20] Campbell, J., 1986. A defense of traditional hypotheses about the term structure of interest rates. Journal of Finance 41, 373-400.

[21] Campbell, J., Lo A.W., Mackinlay A.C., 1997. The Econometrics of Financial Markets. Princeton University Press.

[22] Chan, K.C., Karolyi G.A., Longstaff F.A., Sanders A.B., 1992. An empirical comparison of alternative models of the short-term interest rate. Journal of Finance 47, 1209-1227.

[23] Chang, J., 1998. Lecture Notes on Stochastic Calculus. Yale University.

[24] Chapman, D.A., Pearson, N.D., 1998. Is the short rate drift actually nonlinear? Unpublished working paper.

[25] Chung, K.L., Williams, R.J., 1990. Introduction to Stochastic Integration. Birkhäuser, Boston.

[26] Conley, T.G., Hansen, L.P., Liu, W.F. Bootstrapping the long run. Macroeconomic Dynamics 1, 279-311.

[27] Conley, T.G., Hansen, L.P., Luttmer, E.G.J., Scheinkman, J.A. 1997. Short-term interest rates as subordinated diffusions. Review of Financial Studies 3, 525-577.

[28] Cox, J.C., 1975. Notes on Option Pricing I: Constant Elasticity of Variance Diffusion. Lecture Notes, Stanford University.

[29] Cox, J.C., 1985. A theory of the term structure of interest rates. Econometrica 53, 385-407.

[30] Cox, J.C., Ingersoll J.E., Ross, S.A., 1980. An analysis of variable rate loan contracts. Journal of Finance 35, 389-403.

[31] Cox, J.C., Ingersoll J.E., Ross, S.A., 1985a. An intertemporal general equilibrium model of asset pricing. Econometrica 53, 363-384.

[32] Cox, J.C., Ingersoll J.E., Ross, S.A., 1985b. A theory of the term structure of interest rates. Econometrica 53, 385-407.

[33] Cox, J.C., Ross, S.A., 1976. The valuation of options for alternative stochastic processes. Journal of Financial Economics 6, 145-166.

[34] Dacunha-Castelle, D., Florens-Zmirou, D., 1986. Estimation of the coefficients of a diffusion from discrete observations. Stochastics 19, 263-284.

[35] Das, S.R., 1994. Mean rate shifts and alternative models of the interest rate: theory and evidence. Unpublished working paper, New York University.

[36] Donhal, G., 1981. On estimating the diffusion coefficient. Journal of Applied Probability 24, 105-114.

[37] Dothan, L.U., 1978. On the term structure of interest rates. Journal of Financial Economics 6, 59-69.

[38] Drost, F.C., Werker, B.J.M., 1996. Closing the GARCH gap: continuous-time GARCH Modeling. Journal of Econometrics 74, 31-57.

[39] Duffie, D., 1992. Dynamic Asset Pricing Theory. Princeton University Press.

[40] Duffie, D., Kan, 1992. A yield-factor model of interest rates. Unpublished working paper. Graduate School of Business, Stanford University.

[41] Duffie, D., Singleton, K.J., 1993. Simulated moments estimation of Markov models for asset prices. Econometrica 61, 929-952.

[42] Durrett, R., 1996. Stochastic Calculus. A Practical Introduction. CRC press, New York.

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

[44] French, K.R., Roll, R., 1986. Stock return variances: the arrival of information and the reaction of traders. Journal of Financial Economics 17. 5-26.

[45] Gallant, A.R., Tauchen, G., 1996. Which moments to match. Econometric Theory 12, 657-681.

[46] Geman, S.A., 1979. On a common sense estimator for the drift of a diffusion. Unpublished working paper, Brown University.

[47] Gourieroux, C., Monfort, A., Renault, E., 1993. Indirect inference. Journal of Applied Econometrics 8, S85-S118.

[48] Grossman, S.J., Melino, A. Shiller, R., 1987. Estimating the continuous-time consumption based asset pricing model. Journal of Business and Economic Statistics 5, 315-327.

[49] Hansen, L., Scheinkman, J., 1995. Back to the future: generating moment implications for continuous-time Markov processes. Econometrica 63, 767-804.

[50] Härdle, W., 1990. Applied Nonparametric Regression. Cambridge University Press, Cambridge.

[51] Härdle, W., Marron, J.S., 1985. Optimal bandwidth selection in nonparametric regression function estimation. Annals of Statistics 13, 1465-1481.

[52] Ho, T.S.Y., Lee, S., 1986. Term structure movements and pricing interest rate contingent claims. Journal of Finance, 41, 1011-1029.

[53] Hull, J.C., White, A., 1990. Pricing interest-rate derivative securities. Review of Financial Studies 3, 573-592.

[54] Ingersoll, J., 1987. Theory of Financial Decision Making. Rowman and Litterfield Publishers, Inc.

[55] Jacod, J., 1997. Nonparametric kernel estimation of the diffusion coefficient of a diffusion. Prépublication No. 405 du Laboratoire de Probabilités de l'Université Paris VI.

[56] Jeffrey, A., 1997. An empirical examination of a path dependent term structure model. Unpublished working paper, Yale University.

[57] Jiang, G.J., Knight, J., 1996. Parametric versus nonparametric estimation of diffusion processes - A Monte Carlo comparison. Unpublished working paper, University of Western Ontario.

[58] 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.

[59] Jiang, G.J., 1998. Nonparametric modeling of U.S interest rate term structure dynamics and implications on the prices of derivative securities. Forthcoming Journal of Financial and Quantitative Analysis.

[60] Jones, C., 1998. Nonlinear mean reversion in the short-term interest rate. Unpublished working paper, the Wharton School, University of Pennsylvania.

[61] Karatzas, I., Shreve, S.E., 1988. Brownian Motion and Stochastic Calculus. Springer-Verlag, New York.

[62] Karlin, S., Taylor, 1981. A Second Course in Stochastic Processes. Academic Press, New York.

[63] Kutoyants, Y., 1984. Parameter Estimation for Stochastic Processes. Heldermann.

[64] Lanska, V., 1979. Minimum contrast estimation in diffusion processes. Journal of Applied Probability 16, 65-75.

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

[66] Longstaff F., 1989b. Temporal aggregation and the continuous-time CAPM. Journal of Finance 44, 871-887.

[67] Longstaff, F., 1990a. Time varying term premia and traditional hypothesys about the term structure. Journal of Finance 45, 1307-1314.

[68] Longstaff, F., Schwartz, E., 1992. Interest rate volatility and the term structure: A two-factor general equilibrium model. The Journal of Finance, 1259-1282.

[69] Merton, R., 1973. Theory of rational option pricing. Bell Journal of Economics and Management Science 4, 141-183.

[70] Naik, V., Lee, M.H., 1993. The yield curve and bond option prices with discrete shifts in economic regimes. Unpublished working paper, University of British Columbia.

[71] Nelson, D.B., 1990. ARCH models as diffusions approximations. Journal of Econometrics 45, 7-38.

[72] Park, J., Phillips P.C.B., 1998a. Asymptotics for nonlinear transformations of integrated time series, unpublished paper. Yale University.

[73] Park, J., Phillips P.C.B., 1998b. Nonlinear regressions with integrated time series. Unpublished working paper, Yale University.

[74] Pedersen, A.R., 1995. Consistency and asymptotic normality of an approximate maximum likelihood estimator for discretely observed diffusion processes. Bernoulli 1, 257-279.

[75] Pearson, N., Sun, T.S., 1994. An empirical examination of the Cox, Ingersoll, Ross Model of the term structure of interest rates using the method of maximum likelihood. Journal of Finance 54, 929-959.

[76] Pham Dinh, T.D., 1981. Nonparametric estimator of the drift coefficient in the diffusion equation. Math. Operations Statist. Ser Statist., 12, 61-73.

[77] Phillips, P.C.B., 1987. Time series regression with a unit root. Econometrica 55, 277-301.

[78] Phillips, P.C.B., 1998. Econometric analysis of Fisher's equation. Unpublished paper prepared for the Fisher Conference, Yale University, 1998.

[79] Phillips, P.C.B., Park, J., 1997. Nonstationary density estimation and kernel autoregression. Unpublished paper, Yale University.

[80] Phillips, P.C.B., Perron, P., 1988. Testing for a unit root in time series regression. Biometrika 75, 335-346.

[81] Phillips, P.C.B., Ploberger, W., 1996. An Asymptotic Theory of Bayesian Inference for Time Series. Econometrica 64, 381-413.

[82] Pritsker, M., 1998. Nonparametric density estimation and tests of continuous-time interest rate models. The Review of Financial Studies 3.

[83] Protter, P., 1990. Stochastic Integration and Differential Equations. Springer-Verlag. New York.

[84] Revuz, D., Yor, M., 1994. Continuous Martingales and Brownian Motion. Second Edition, Springer-Verlag, New York.

[85] Siddique, A.R., 1994. Nonparametric estimation of mean and variance and pricing of securities. Unpublished working paper, Georgetown University.

[86] Silverman, B.W., 1986. Density Estimation for Statistics and Data Analysis. Chapman and Hall, London.

[87] Stanton, R., 1998. A nonparametric model of term strucure dynamics and the market price of interest rate risk. Journal of Finance 5, 1973-2002.

[88] Vasicek. O., 1977. An equilibrium characterization of the term structure. Journal of Financial Economics 5, 177-188.

[89] Yor, M., 1983. Le drap brownien comme limite en loi de temps locaux linéaires. Sém. Prob. XVII. Lecture Notes in Mathematics, vol. 986, Springer, Berlin Heidelberg, 1983, 89-105.