Nonstationary Continuous-Time Processes*

Federico M. Bandi†
Booth School of Business,
University of Chicago

Peter C. B. Phillips‡
Yale University, University of Auckland
University of York & Singapore Management University

First draft: September 2001
This draft: July 2008

Abstract

We survey local nonparametric methods applied to the estimation of scalar and multidimensional continuous-time Markov processes (with continuous and discontinuous sample paths). In doing so, particular emphasis is placed on recently-proposed identification methods and asymptotic approaches which do not necessitate strict stationarity but hinge only on recurrence, possibly of the nonstationary kind.

JEL Classifications: C14, C22, C32

*Paper written for the *Handbook of Financial Econometrics* edited by Yacine Aït-Sahalia and Lars Peter Hansen. We thank Darrell Duffie, Jianqing Fan, Benoit Perron, Mark Watson, the Editor Lars Peter Hansen, and two anonynous referees for useful suggestions. We are grateful to Seoyeon (Suzanne) Lee for research assistance on a previous draft. In attempting to survey what is an extremely vast field relating to inference for continuous-time Markov processes, we adopt the perspective of focusing on (local) nonparametric methods and mild assumptions regarding specification and conditions needed for identification. We sought to keep this survey updated from its original completed form in 2001 and apologize for any relevant omissions given the intervening time gap. Bandi acknowledges financial support from the IBM Corporation Faculty Research Fund at the University of Chicago. Phillips thanks the NSF for support under grant nos. SES 00-92509 and SS 06-47086.

†federico.bandi@chicagobooth.edu
‡peter.phillips@yale.edu

Table of Contents

1 Introduction

A large body of recent asset pricing theory is written in continuous-time, for which Merton (1990) and Duffie (1996) are classic references. Notwithstanding the evident benefit of continuous-time tools for modelling purposes and recent advances in the econometric treatment of continuous-time models,[1] the use of stochastic processes with continuous (in time) sample paths still poses important challenges when it comes to the econometric estimation and empirical implementation of modern asset pricing models. Campbell *et al.* (1997, CLM henceforth), Gourieroux and Jasiak (2001), and Tsay (2005) are recent textbooks on the general topic of financial econometrics and outline some of the relevant issues.

Perhaps the most basic econometric problem arises because, while the relevant series are often specified as processes that evolve continuously in time, observations of the process occur only at discrete points in time. The discrete nature of the data has forced researchers to design estimation methodologies which are capable of circumventing the so-called "aliasing problem" and that can uniquely identify the fine grain structure of the underlying process from a sample of observations located along the continuous sample path rather than from a continuous record of the process over that path. (Readers are referred to the chapters by Aït-Sahalia *et al.,* 2009, Gallant and Tauchen, 2009, Jacod, 2009, and Johannes and Polson, 2009, in the present volume for a treatment of these issues). Such methodologies generally, but not exclusively (c.f., Phillips, 1973; Hansen and Sargent, 1983; McCrorie, 2009), rely on stationarity. The reason is clear. Should the underlying process be endowed with a stationary probability density, then the information extracted from the discrete data can fruitfully be employed to identify the probability measure and thereby, hopefully, characterize the continuous dynamics of the system. In this way, stationarity can be a powerful aid to identification and estimation.

Despite the advantages of assuming the existence of a time-invariant probability distribution, it appears that for many empirical applications in continuous-time asset pricing it would be more appropriate to allow for martingale and other forms of nonstationary behavior, while not ruling out stationarity either. In such cases, an additional layer of complication in estimation comes from the necessity of achieving identification without resorting to the restrictions that are provided by the existence of a stationary probability density for the process of interest.

This chapter discusses techniques which have been recently introduced to identify potentially nonstationary, time-homogeneous, continuous-time Markov processes. The focus will be on classes of processes which are widely used in continuous-time asset pricing, namely scalar and multivariate diffusion processes as well as jump-diffusion processes. Such

[1] In his survey on continuous-time methods in finance appeared in the Papers and Proceedings of the Sixtieth Annual Meeting of the American Finance Association, Sundaresan (2000) writes "Perhaps the most significant development in the continuous-time field during the last decade has been the innovations in econometric theory and in the estimation techniques for models in continuous-time."

processes, irrespective of their stationarity properties, have infinitesimal conditional moment definitions. Their infinitesimal moments are known to fully characterize the temporal evolution of the corresponding system and, in consequence, readily lend themselves to estimation for the purpose of the identification of the system's dynamics. Consider a standard scalar diffusion (i.e., the solution to (14) below), but a similar argument holds for more involved continuous-time Markov processes of the type reviewed in this chapter. Its transition density (which is, in general, not known in closed form) is fully determined by the two functions that are commonly known as the drift, $\mu(.)$, and the diffusion, $\sigma^2(.)$. The drift represents the conditional expected rate of change of the process for infinitesimal time changes, i.e.,

$$\mu(a) = \lim_{t \to 0} \frac{1}{t} \mathbf{E}[X_t - X_0 | X_0 = a] = \lim_{t \to 0} \frac{1}{t} \mathbf{E}^a [X_t - a], \tag{1}$$

while the diffusion gives the conditional rate of change of volatility for infinitesimal variations in time, i.e.,

$$\sigma^2(a) = \lim_{t \to 0} \frac{1}{t} \mathbf{E}[(X_t - X_0)^2 | X_0 = a] = \lim_{t \to 0} \frac{1}{t} \mathbf{E}^a [(X_t - a)^2]. \tag{2}$$

Formulae (1) and (2) are suggestive in that one could hope to identify the functions of interests, which are defined as conditional expectations over infinitesimal time distances, using sample analogues to conditional expectations as in standard nonparametric inference for conditional moments in discrete time. For example, it is natural to estimate the drift at a by differencing the data and then averaging the first differences $X_{t+\Delta} - X_t$ corresponding to observations X_t in the spatial neighborhood of the generic level a. Provided the level a is visited an infinite number of times over time so that an infinite number of differences can be averaged asymptotically, we would expect the procedure to be consistent in the limit (i.e., the sample average converges, in probability at least, to the conditional moment). Interestingly, the underlying process (and, under some conditions, the sampled process) visits the level a an infinite number of times provided recurrence is satisfied. Recurrence implies return of the sample path of the underlying process to any spatial set of non-zero Lebesgue measure with probability one and is known to be a milder assumption than stationarity (Section 2 provides a definition and additional discussions).

Some recent papers have pursued the econometric implications of these observations and designed spatial estimation methods for various classes of continuous-time, time-homogeneous, recurrent Markov processes. The methods are easy to implement and have some natural appeal because they are based on commonly used nonparametric (and semiparametric) estimation procedures for conditional moments in more conventional stationary, discrete-time, frameworks. However, they have the additional attraction that their statistical properties apply even though stationarity of the underlying continuous-time model is never assumed. These methods have been introduced in research by Bandi and Phillips (2003, BP hereafter), Bandi and Nguyen (2003, BN hereafter), and Bandi and Moloche (2004, BM hereafter) which develops Nadaraya-Watson (N-W) kernel estimation procedures for recurrent scalar diffusions, scalar jump-diffusions and multivariate diffusions, respectively, and have subsequently been extended in a variety of directions.

Specifically, following work by Brugiére (1991, 1993), Florens-Zmirou (1993), and Jacod (1997) in nonparametric volatility estimation for diffusions, this literature lays theoretical foundations for using well-understood and conventional nonparametric and semiparametric methods in the estimation of all the infinitesimal moment functionals driving the evolution of continuous-time Markov processes (with or without discontinuities in the sample path). The literature explores conditions (like recurrence) under which consistency and weak convergence results can be obtained in the (potential) absence of a stationary distribution for the process, and it provides results that can be evaluated in a manner that is closely related to conventional interpretations of nonparametric estimates for stationary discrete-time series. While the findings that emerge from this literature contain the stationary case as a subcase, their more general form reflects the fact that a stationary density of the underlying process may not exist, which leads to important issues of interpretation. We discuss these issues and indicate avenues for future research both in the estimation of potentially nonstationary continuous-time processes (which are, as said, the core subject of our review) and in the estimation of potentially nonstationary (recurrent) discrete-time series.

The chapter is organized as follows. Section 2 provides some intuition for the methodology, introduces the notion of recurrence, and discusses the asymptotic features of our adopted sampling scheme. In particular, consistency is shown to hinge on the joint implementation of "infill" and "long span" asymptotics. The latter is crucial in exploiting the recurrence properties of the process under investigation. The former is vital in replicating the infinitesimal features of the functions of interest. Both conditions are necessary for the identification of continuous-time Markov processes under minimal assumptions on their dynamic properties and parametric form. Accordingly, the discussion in this chapter focuses on estimation procedures which impose mild assumptions on the stochastic nature of the underlying process but require the presence, at least in the limit, of "frequent" observations[2] to achieve consistent estimation. In this regard, our review can be viewed as complementary to those by Aït-Sahalia *et al.* (2009), Phillips and Yu (2008), and Johannes and Polson (2009). The former two papers discuss functional and parametric estimation methods for diffusions that do not require infill asymptotics but rely on stationarity and

[2]The appropriateness of this asymptotic approximation is an empirical issue which depends on the application. However, it is known to be a realistic approximation in fields, such as finance, where data sets are often characterized by a large number of observations sampled at relatively high frequencies. Importantly, the highest frequency we consider here is generally the daily frequency (see Subsections 3.6 and 4.3 for exceptions). While higher (intra-daily) frequencies would pose additional theoretical and empirical complications induced by the presence of market microstructure noise contaminations (see, e.g., Bandi and Russell, 2007, for discussions), it is well-known that daily data are good approximations to very frequent observations for estimators relying on very frequent observations (see, e.g., the simulation study of Jiang and Knight, 1999).

mixing for identification and estimation. The latter reviews Bayesian simulation procedures that are sufficiently flexible to deal with nonstationarities but impose a tight parametric structure on the process of interest. This review is also complementary to recent survey articles on selected nonparametric methods in continuous-time asset pricing (Cai and Hong, 2003, and Fan, 2005).

Sections 3, 4, and 5 specialize the analysis to the estimation of recurrent scalar diffusions, recurrent jump-diffusions and recurrent multivariate diffusions, respectively. Some emphasis is also placed on stochastic volatility modelling (with and without discontinuities in the volatility's sample path) which, in light of the latent nature of volatility, poses additional challenges and is the subject of much research currently under way.

This chapter is largely self-contained and its discussion is kept at a fairly intuitive level. Nonetheless, some basic notions of stochastic process theory and functional estimation in discrete-time econometrics will help the reader. Karatzas and Shreve (1991), Protter (1995) and Revuz and Yor (1998) are standard references for the former and, while not providing all the background material for the present chapter, are strongly recommended references. Thorough discussions of functional methods for discrete-time series are contained in Härdle (1990), Fan and Yao (2003), Pagan and Ullah (1999), and Li and Racine (2006). A concise and highly accessible introduction to nonparametric techniques is Härdle and Linton (1994). Chapter 12 of the book by CLM (1997) also provides accessible discussions of kernel regression methods similar to those employed here.

2 Intuition and Conditions[3]

As noted in the Introduction, the existence of conditional moments for interesting classes of continuous-time Markov models provides a mechanism for inference based on the construction of sample analogues (i.e., weighted averages) to infinitesimal conditional expectations. To fix ideas, consider a simple example in discrete time. Suppose the observations $X_1, X_2, ..., X_n$ are generated by a time-homogeneous Markov process X. One might be interested in estimating the conditional moment functional

$$M(a) = \mathbf{E}^a[f(X_1, a)], \tag{3}$$

where $X_0 = a$ is a generic initial condition and f is some integrable function. A crude (but intuitively appealing) sample analogue estimator for $M(a)$ is

$$\widehat{M}_{(n)}(a) = \frac{\sum_{i=1}^n \mathbf{1}_{X_i=a} f(X_{i+1}, X_i)}{\sum_{i=1}^n \mathbf{1}_{X_i=a}}, \tag{4}$$

[3]Parts of this section are based on the discussion of the paper "On the functional estimation of jump-diffusion processes" (BN, 2003) given by Durrell Duffie at the 2001 Winter Meetings of the Econometric Society (New Orleans, January 9, 2001).

where $\mathbf{1}_A$ is the indicator function of the set A. Formula (4) implies identification of the conditional expectation (at $X_0 = a$) of the function $f(X_1, X_0)$ (as in the definition of $M(a)$) through a (weighted) sample average of functions of the observations ($f(X_{i+1}, X_i)$) taken at values X_i which are equal to a. Simple intuitive arguments based on the law of large numbers suggest that the level a ought to be visited an infinite number of times to achieve consistency. In consequence, it appears that the condition $\#\{i : X_i = a\} = \sum_{i=1}^{n} \mathbf{1}_{X_i=a} \to \infty$ as $n \to \infty$[4] is, in general, necessary to obtain asymptotic convergence of $\widehat{M}_{(n)}(a)$ to $M(a)$.

We now turn to a similar example in the context of a continuous-time Markov process X not necessarily endowed with a continuous sample path (one such case will be covered in Section 4 below). Suppose we are interested in estimating the infinitesimal conditional moment

$$M(a) = \lim_{t\to 0} \frac{1}{t}\mathbf{E}[f(X_t, X_0)|X_0 = a] = \lim_{t\to 0} \frac{1}{t}\mathbf{E}^a[f(X_t, a)]. \tag{5}$$

Note that if X is a scalar diffusion and $f(y, a)$ is equal to either $(y-a)$ or $(y-a)^2$, then $M(a)$ coincides with either the drift in Eq. (1) or the diffusion function in Eq. (2), respectively. Then, coherently with (4) above and the previous discussion, one could estimate (5) by using

$$\widehat{M}^{(1)}_{(n,\Delta,\varepsilon)}(a) = \frac{\sum_{i=1}^{n} \mathbf{1}_{X_i\in(a-\varepsilon,a+\varepsilon)} f(X_{i+\Delta}, X_i)/\Delta}{\sum_{i=1}^{n} \mathbf{1}_{X_i\in(a-\varepsilon,a+\varepsilon)}}, \tag{6}$$

where Δ is the time distance between discretely-observed observations and ε is a bandwidth parameter according to which an interval around a on the sample path of the process is determined. Asymptotically, we send Δ to 0 in order to replicate the limit operation in the definition of $M(a)$ (i.e. $\lim_{t\to 0}$). Furthermore, (i) we let the bandwidth ε vanish so as to obtain averages of functions $f(X_{i+\Delta}, X_i)$ such that X_i is in a close neighborhood of a and (ii) we let n grow to infinity in order to guarantee that the number of observations X_i in the actual vicinity of a (i.e., $\#\{i : X_i \in (a-\varepsilon, a+\varepsilon)\} = \sum_{i=1}^{n} \mathbf{1}_{X_i\in(a-\varepsilon,a+\varepsilon)}$) diverges to infinity for identification. Again, we expect $\widehat{M}^{(1)}_{(n,\Delta,\varepsilon)}(a)$ to converge to $M(a)$ as $n \to \infty$, $\Delta \to 0$, and $\varepsilon \to 0$.[5]

Clearly, the function $\sum_{i=1}^{n} \mathbf{1}_{X\in(a-\varepsilon,a+\varepsilon)}$ counts the number of observations inside the window $(a-\varepsilon, a+\varepsilon)$ and weighs them equally. It seems plausible, however, that observations that are closer to a contain more useful information than more distant ones. In consequence, it might be worth replacing the so-called indicator kernel, i.e. $\mathbf{1}_{X\in(a-\varepsilon,a+\varepsilon)}$, with a function that is centered at a and converges monotonically to 0 as $|X| \to \infty$. Such a function would give a higher weight to observations that are closer to a, thereby increasing efficiency. This is typically achieved by using smooth kernels $\mathbf{K}(.)$ satisfying $\int \mathbf{K}(u)du = 1$ (see, e.g., Härdle

[4] We are purposely non-specific about the mode of divergence at this point. We will be clear about it in what follows.

[5] Naturally, any notion of consistency (and convergence in distribution) requires appropriate (limiting) relations between n, Δ, and ε. In what follows, we will make these conditions explicit for all estimators.

and Linton, 1994, for example, and the assumptions below). The ubiquitous second-order Gaussian kernel is an example. Hence, we may write

$$\widehat{M}^{(2)}_{(n,\Delta,\varepsilon)}(a) = \frac{\sum_{i=1}^{n} \mathbf{K}\left(\frac{X_i - a}{\varepsilon}\right) f(X_{i+\Delta}, X_i)/\Delta}{\sum_{i=1}^{n} \mathbf{K}\left(\frac{X_i - a}{\varepsilon}\right)}, \tag{7}$$

which is simply a version of (6). As earlier, we expect $\widehat{M}^{(2)}_{(n,\Delta,\varepsilon)}(a)$ to be consistent for $M(a)$ as $n \to \infty$, $\Delta \to 0$, and $\varepsilon \to 0$ at appropriate rates.

We now summarize the features of the asymptotic requirements which appear to be necessary for consistency. In general, we will need to assume that the distance between observations Δ vanishes in the limit (i.e., infill asymptotics) while the time span (T, say) diverges to infinity (i.e., long span asymptotics) along with the number of observations n. As briefly mentioned in the Introduction and illustrated above in the context of simple examples, the former assumption (i.e., $\Delta \to 0$) is important to replicate the infinitesimal features of the theoretical quantities. The latter (i.e., $T, n \to \infty$) is crucial in order to guarantee that the number of visits that the sampled process makes in the neighborhood of a generic point a diverges to infinity in the limit (i.e., $\sum_{i=1}^{n} \mathbf{1}_{X_i \in (a-\varepsilon, a+\varepsilon)} \to \infty$ or $\sum_{i=1}^{n} \mathbf{K}\left(\frac{X_i - a}{\varepsilon}\right) \to \infty$), provided the path of the underlying process does so. Of course, the additional assumption $\varepsilon \to 0$ permits proper (asymptotic) conditioning at a.

Coherently with our discussion, the following sampling scheme has been adopted by the recent literature on the functional estimation of continuous-time Markov processes and will be used throughout this chapter. We will assume that we observe the process of interest X_t at points $\{t = t_1, t_2, .., t_n\}$ in the time interval $[0, T]$, with $T \geq T_0$, where T_0 and T are positive constants. Also, the data will be taken to be equispaced. Thus,

$$\{X_{\Delta_{n,T}}, X_{2\Delta_{n,T}}, X_{3\Delta_{n,T}}, ..., X_{n\Delta_{n,T}}\} \tag{8}$$

will be n observations at

$$\{t_1 = \Delta_{n,T}, t_2 = 2\Delta_{n,T}, t_3 = 3\Delta_{n,T}, ..., t_n = n\Delta_{n,T}\}, \tag{9}$$

where $\Delta_{n,T} = T/n$. In the limit, we will let $n \to \infty$, $T \to \infty$, and $\Delta_{n,T} = T/n \to 0$. In a few instances, T will be fixed at $\overline{T}$. In the sequel, we will be explicit about the limiting behaviour of the time span T.

Based on our discussion, it appears that the only requirements that we have to impose on the dynamic properties of the processes of interest for identification are those that guarantee divergence of the number of visits in the spatial vicinity of points in the range of the process. This is a typical feature of recurrent processes. Specifically, the sample path of a recurrent process returns to sets of non-zero Lebesgue measure an infinite number of times over time with probability one. We now rigorously state the definitions of recurrence used in this review (the interested reader is referred to the classical treatment in Meyn and Tweedie, 1993, for additional discussions).

Definition (Null and Positive Harris Recurrence) *Let A be a measurable set of the range $\mathfrak{D}$ of the process of interest. Define the first hitting time of A as* $\tau_A = \inf\{t \geq 0 : X_t \in A\}$.

The process X_t is called null Harris recurrent if there is a $\sigma-$finite measure $m^*(dx)$ such that $m^*(A) > 0$ implies $P^a[\tau_A < \infty] = 1$ for every $a \in \mathfrak{D}/\overline{A}$, where $\overline{A}$ is the closure of the set A.

It is called positive Harris recurrent (ergodic) if there is a $\sigma-$finite measure $m^*(dx)$ such that $m^*(A) > 0$ implies $\mathbf{E}^a[\tau_A] < \infty$ for every $a \in \mathfrak{D}/\overline{A}$.

Define the occupation time measure of the set A of positive Lebesgue measure as

$$\eta_A^T = \int_0^T \mathbf{1}_{\{X_s \in A\}} ds. \tag{10}$$

The quantity η_A^T gives the amount of time spent by the process at A between 0 and T. Under both notions of recurrence, we obtain $\mathbf{P}^a\left[\lim_{T\to\infty} \eta_A^T = \infty\right] = 1$ for $\forall a \in \mathfrak{D}/\overline{A}$. Specifically, starting from a level a not belonging to the generic set A, the process X_t visits A an infinite number of times as $T \to \infty$, almost surely. This property is of course crucial for (point-wise) identification.

Null and positive recurrence are milder assumptions than stationarity. Stationary processes are recurrent but recurrent processes do not have to be stationary. In particular, recurrent processes do not have to be endowed with a stationary probability measure. Null recurrent processes, in fact, do not possess a time-invariant probability measure. Nonetheless, null Harris recurrence implies the existence of a unique invariant measure $m(dx)$ ($= m^*(dx)$ in the Definition). Assume $X^{(x)}$ is the process' unique strong solution with initial condition $X_0^{(x)} = x \in \mathfrak{D}$. The invariant measure is such that

$$m(A) = \int_{\mathfrak{D}} P\left(X_t^{(x)} \in A\right) m(dx) \quad \forall A \subset \mathfrak{B}(\mathfrak{D}), \tag{11}$$

for every $0 \leq t < \infty$ (see, e.g., Azéma *et al.*, 1967, and Karatzas and Shreve, 1991, Exercise 6.18, page 362).[6] If the invariant measure is finite on $\mathfrak{D}$ ($m(\mathfrak{D}) < \infty$, that is), then the process is positive recurrent (ergodic) and has a time-invariant stationary probability measure (distribution) to which it converges, at least in the limit. Such measure is given by $f(dx) = \frac{m(dx)}{m(\mathfrak{D})}$. A positive-recurrent process that is started in its stationary distribution remains in the stationary distribution and, as a consequence, is strictly stationary. Examples will be provided in the sequel. For now it suffices to say that Brownian motion in one (c.f. Example 1 in Section 3) and two dimensions are classical examples of null recurrent processes. Classical Vasicek (Ornstein-Uhlenbeck) and CIR processes (c.f. Vasicek, 1977,

[6] Alternatively, one could write

$$m = P_t m \qquad \forall t \geq 0,$$

where $(P_t)_{t\geq 0}$ is the semigroup of the process X_t (c.f. Ethier and Kurtz, 1986, and Aït-Sahalia *et al.*, 2009, in this volume).

and Cox *et al.*, 1985) are strictly stationary or positive recurrent depending on whether they are initiated at their stationary measures or not (c.f. Example 5 in Section 3).

To conclude, the estimators reviewed in this chapter are nonparametric or semiparametric in nature and either follow the general form of (7) above or are constructed based on it. Null recurrence is all that we require to guarantee consistency of the estimates for the infinitesimal moments of interest. Importantly, positive recurrence and stationarity, which are clearly more stringent assumptions than null recurrence, will be shown to only yield an increase in the rates of convergence of the estimators to the corresponding moments.

We now turn to a detailed analysis of the specific processes mentioned in the introduction, namely scalar diffusion processes (SDPs), scalar jump-diffusion processes (SJDPs), and multivariate diffusion processes (MDPs). A separate section will be devoted to each of these. In what follows we will not review the definitions of recurrence that were laid out previously but simply list conditions under which the processes display either ergodic or null recurrent behavior.

The following, rather standard, assumptions will be imposed on the kernel function $\mathbf{K}(.)$ throughout the present chapter.

The kernel $\mathbf{K}(.)$ *is a continuously differentiable, symmetric, and nonnegative function on the real line so that*

$$\int \mathbf{K}(s)ds = 1, \quad \mathbf{K}_1 = \int s^2\mathbf{K}(s)ds < \infty \tag{12}$$

and

$$\mathbf{K}_2 = \int \mathbf{K}^2(s)ds < \infty, \ \int |\mathbf{K}'(s)|ds < \infty. \tag{13}$$

3 Scalar Diffusion Processes (SDPs)

In this section we model a generic time series as the solution X_t to the stochastic differential equation

$$dX_t = \mu(X_t)dt + \sigma(X_t)dB_t, \tag{14}$$

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 $X_0 = \overline{X}$ belongs to L^2 and is taken to be independent of $\{B_t : t \geq 0\}$. Define the left-continuous filtration

$$\overline{\Im}_t := \sigma(\overline{X}) \vee \Im_t^B = \sigma(\overline{X}, B_s; 0 \leq s \leq t) \qquad 0 \leq t < \infty, \tag{15}$$

and the collection of null sets

$$\aleph := \{N \subseteq \Omega; \exists G \in \overline{\Im}_{\infty} \text{ with } N \subseteq G \text{ and } P(G) = 0\}. \tag{16}$$

Create the augmented filtration

$$\widetilde{\Im}_t^X := \sigma(\overline{\Im}_t \cup \aleph) \qquad 0 \leq t < \infty, \tag{17}$$

and impose Assumption 1 through 3 (1 through 3b) below, to assure the existence and pathwise uniqueness of a null recurrent (positive recurrent) and $\{\widetilde{\Im}_t^X\}$-adapted solution to (14).

(1) *$\mu(.)$ and $\sigma(.)$ 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 satisfy local Lipschitz and growth conditions. Thus, for every compact subset J of the range of the process, there exist constants C_1 and C_2 such that, for all x and y in J,*

$$|\mu(x) - \mu(y)| + |\sigma(x) - \sigma(y)| \leq C_1|x - y|, \tag{18}$$

and

$$|\mu(x)| + |\sigma(x)| \leq C_2\{1 + |x|\}. \tag{19}$$

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

(3) *(Null recurrence) Define the second-order elliptic operator*[7]

$$\mathfrak{L}\varphi(.) = \varphi'(.)\mu(.) + \frac{1}{2}\varphi''(.)\sigma^2(.). \tag{20}$$

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

$$\mathfrak{L}\varphi(.) \leq 0 \qquad \text{on} \qquad \mathfrak{R}\backslash\{0\} \tag{21}$$

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

[7] The operator L is generally called the infinitesimal generator of the SDP X_t. We refer the interested reader to Aït-Sahalia *et al.* (2009), in the present volume, and Hansen and Scheinkman (1995) for a discussion of estimation methods for strictly stationary diffusions based on the properties of the infinitesimal generator.

(3b) *(Positive recurrence) There is a function* $\varphi(.) : \mathfrak{R} \backslash \{0\} \to \mathfrak{R}$ *of class* C^2 *in the domain of the operator that satisfies*

$$\mathfrak{L}\varphi(.) \leq -1 \qquad \text{on} \qquad \mathfrak{R} \backslash \{0\}, \tag{22}$$

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

Under Assumptions 1, 2 and 3 (3b) the stochastic differential equation (14) yields a strong solution X_t which is unique, null recurrent (positive recurrent) and continuous in $t \in [0, T]$. In particular, the process X_t satisfies

$$X_t = \overline{X} + \int_0^t \mu(X_s) ds + \int_0^t \sigma(X_s) dB_s \tag{23}$$

almost surely, with $\int_0^t \mathbf{E}\left(X_s^2\right) ds < \infty$, and is a semimartingale. The dynamics of X_t are determined by the functions $\mu(.)$ and $\sigma(.)$. These functions are the object of econometric interest.

Assumptions 3 and 3b are vital in determining recurrent behavior for X_t. As pointed out earlier, null recurrence is a sufficient condition for the existence of σ-finite invariant measure $m(dx)$. Such a measure is unique up to multiplication by a constant and, in the case of SDPs, is known to be equal (up to a proportionality factor) to the so-called *speed measure*, i.e.,

$$m(dx) = \frac{2dx}{S'(x)\sigma^2(x)} \qquad \forall x \in \mathfrak{D} \subseteq \mathfrak{R}, \tag{24}$$

where $S'(x)$ is the first derivative of the *scale function*, namely

$$S(x) = \int_c^x \exp\left\{\int_c^y \left[-\frac{2\mu(s)}{\sigma^2(s)}\right] ds\right\} dy, \tag{25}$$

where $c \in \mathfrak{D}$. Under Assumption 3b, the SDP is positive recurrent (i.e., $m(\mathfrak{D}) < \infty$) and admits a time-invariant probability measure. In particular, the normalized speed measure, i.e. $m(dx)/m(\mathfrak{D})$, is the limiting stationary probability measure of X_t implying

$$\lim_{t\to\infty} P^x(X_t < z) = \frac{m((l, z))}{m(\mathfrak{D})} \qquad \forall x, z \in \mathfrak{D} \subseteq \mathfrak{R}, \tag{26}$$

(see, e.g., Pollack and Siegmund, 1985, and Karatzas and Shreve, 1991, Exercise 5.40, page 353). More explicitly, we can write the stationary probability density of the process as

$$\begin{aligned} f(x) &= \frac{m(x)}{m(\mathfrak{D})} = \frac{1}{m(\mathfrak{D})} \frac{\exp\left\{\int_c^x \left[\frac{2\mu(s)}{\sigma^2(s)}\right] ds\right\}}{\sigma^2(x)} \\ &= \left(\int_{\mathfrak{D}} \frac{\exp\left\{\int_c^x \left[\frac{2\mu(s)}{\sigma^2(s)}\right] ds\right\}}{\sigma^2(x)} dx\right)^{-1} \frac{\exp\left\{\int_c^x \left[\frac{2\mu(s)}{\sigma^2(s)}\right] ds\right\}}{\sigma^2(x)}. \end{aligned} \tag{27}$$

We now provide some examples.

Example 1 (Natural scale diffusions) *For general scalar diffusions, if the scale function $S(x)$ is such that $\lim_{x \to l+} S(x) = -\infty$ and $\lim_{x \to u-} S(x) = \infty$, then the process is recurrent, that is it satisfies $\mathfrak{L}\varphi(.) \leq 0$, where φ is defined in (3) and (3b) above (c.f. Khasminskii, 1980). Apparently, the solution to $dX_t = \sigma(X_t)dB_t$, with $\sigma(.)$ continuous and strictly positive, is Harris recurrent over $\Re$ with scale function $S(x) = x - c$ and invariant measure $m(dx) = \frac{2dx}{\sigma^2(x)}$. Chen et al. (1999), for example, discuss the mixing properties of the natural scale diffusion with $\sigma^2(x) = (1 + x^2)^\gamma$ for $\frac{1}{2} < \gamma < 1$ (see, also, Aït-Sahalia et al., 2009, in this volume). If $0 \leq \gamma \leq \frac{1}{2}$, the process is null recurrent on $\Re$. For values of γ strictly larger than $\frac{1}{2}$ the process is positive recurrent. This is a case of "volatility-induced" reversion to the mean (Conley et al., 1998). Trivially, standard Brownian motion (i.e., $\gamma = 0$) is null recurrent.*

Example 2 (Brownian motion with drift) *Assume X_t is the solution to $dX_t = \mu dt + \sigma dB_t$ with $\sigma > 0$. The scale function and the speed measure are $S(x) = \frac{1-e^{-\alpha(x-c)}}{\alpha}$ and $m(dx) = \frac{2e^{\alpha x}}{\sigma^2}dx$, where $\alpha = \frac{2\mu}{\sigma^2}$, respectively. If $\mu > 0$, then $\lim_{x \to \infty} S(x) = \frac{1-e^{\alpha c}}{2\alpha}$ and $\lim_{x \to -\infty} S(x) = -\infty$. The process is not recurrent and $P[\inf_{0 \leq t < \infty} X_t > -\infty] = 1$. If $\mu < 0$, then $\lim_{x \to \infty} S(x) = \infty$ and $\lim_{x \to -\infty} S(x) = \frac{1-e^{\alpha c}}{2\alpha}$. The process is not recurrent and $P[\sup_{0 \leq t < \infty} X_t < \infty] = 1$. In the former case X_t has an attracting boundary at ∞ ($P[\lim_{t \to \infty} X_t = \infty] = 1$, that is). In the latter case X_t has an attracting boundary at $-\infty$ ($P[\lim_{t \to \infty} X_t = -\infty] = 1$, that is). In both cases, it is easy to show that the boundary is unattainable, i.e., it cannot be reached in finite time with positive probability (c.f., Karatzas and Shreve, 1991, and Karlin and Taylor, 1981).*

Example 3 (Geometric Brownian motion) *Assume X_t is the solution to $dX_t = \mu X_t dt + \sigma X_t dB_t$, with $\mu, \sigma > 0$ and $\overline{X} > 0$. Then, $S(x) = c^\alpha \left[\frac{x^{-\alpha+1}}{-\alpha+1} - \frac{c^{-\alpha+1}}{-\alpha+1}\right]$, where $\alpha = \frac{2\mu}{\sigma^2}$ provided $\alpha < 1$ or $\alpha > 1$. The process is not recurrent for these choices of α. Specifically, if $\alpha < 1$, then $\lim_{x \to 0} S(x) = \frac{-c}{-\alpha+1}$ and $\lim_{x \to \infty} S(x) = \infty$ implying $P[\sup_{0 \leq t < \infty} X_t < \infty] = 1 = P[\lim_{t \to \infty} X_t = 0]$. If $\alpha > 1$, then $\lim_{x \to \infty} S(x) = \frac{-c}{-\alpha+1}$ and $\lim_{x \to 0} S(x) = -\infty$ implying $P[\inf_{0 \leq t < \infty} X_t > 0] = 1 = P[\lim_{t \to \infty} X_t = \infty]$. If $\alpha = 1$, then $S(x) = c[\log x - c]$ which implies $\lim_{x \to 0} S(x) = -\infty$ and $\lim_{x \to \infty} S(x) = \infty$, giving recurrence. In addition $m(dx) = \frac{2dx}{c\sigma^2 x}$ and is not integrable. Therefore, geometric Brownian motion is null recurrent when $2\mu = \sigma^2$.*

Of course, the same implications could have been derived by noticing that monotone transformations, such as exponentiation, preserve recurrence. By Ito's lemma, $d\log(X_t) = \theta dt + \sigma dB_t$, where $\theta = \mu - \frac{1}{2}\sigma^2$, thereby yielding Example 2, again, for the log transformation. Using Example 2, one could readily obtain that $\log(X_t)$ (X_t) is not recurrent and has an unattainable, attracting boundary at ∞ (∞) if $\theta > 0$ (or if $2\mu > \sigma^2$). Similarly, $\log(X_t)$

(X_t) is not recurrent and has an unattainable, attracting boundary at $-\infty$ (0) if $\theta < 0$ (or if $2\mu < \sigma^2$). If $\theta = 0$ (or if $2\mu = \sigma^2$), $\log(X_t)$ and X_t are null recurrent.

Example 4 (Bessel process) *Assume $dX_t = \frac{d-1}{2X_s}dt + dB_t$ with $d \geq 2$ and $\overline{X} > 0$. If $d > 2$, we obtain $S(x) = c^{d-1}\left[\frac{c^{2-d}-x^{2-d}}{d-2}\right]$ giving $\lim_{x\to\infty} S(x) = \frac{c}{d-2}$ and $\lim_{x\to 0} S(x) = -\infty$. In consequence, the process is not recurrent and $P[\lim_{t\to\infty} X_t = \infty] = 1$. If $d = 2$, then $S(x) = c[\log x - \log c]$ implying recurrence. Furthermore, the speed measure (i.e., $m(dx) = \frac{2xdx}{c}$) is not integrable between 0 and ∞ giving null recurrence.*

Example 5 (Affine models) *Assume both the drift and the infinitesimal variance are linear functions of the state (i.e., $\mu(x) = c_0 + c_1 x$ and $\sigma^2(x) = c_2 + c_3 x$ with $c_2, c_3 \geq 0$). The well-known Vasicek (Ornestein-Uhlenbeck) and CIR processes belong to this general class and are obtained by setting c_3 and c_2 equal to zero, respectively (c.f., Vasicek, 1977, and Cox et al., 1985). Under standard assumptions on the parameters (see Piazzesi, 2009, in this volume, for a discussion of scalar and multivariate affine models and related estimation procedures), affine diffusions are strongly ergodic (positive recurrent). Should c_0 and c_1 be equal to zero and $\sigma^2(x) = c_2 + c_3|x|$ with $c_2 > 0$ and $c_3 \geq 0$, then the invariant measure would not be integrable over $\mathfrak{R}$ and the resulting process would be null recurrent (c.f. Example 1).*

Prior to describing the estimation strategy, we wish to discuss descriptive tools that have been recently introduced to characterize recurrent SDPs. These tools rely on the notion of local time (Protter, 1995, and Revuz and Yor, 1998, are classical references). Local time is a random quantity which measures the amount of time that the process spends in the vicinity of a point. As a consequence, it might be interpreted as a spatial density and might be used to analyse the locational features of a possibly nonstationary process (for which it is defined, of course) in just the same way as a stationary probability density may be used to study stationary processes (Phillips, 2001, 2004). The next subsection defines local time and introduces a simple estimation strategy to identify it based on a discrete sample of observations. We will also discuss the role that estimated local time can play as a descriptive statistic for recurrent SDPs and its importance in designing robust (to deviations from stationarity) identification procedures for processes whose dynamics are driven by (14) (Bandi, 2002). The terms "local time," "spatial density," and "sojourn time" will be used interchangeably in what follows.

3.1 Generalized Density Estimation for SDPs

The local time of a continuous semimartingale is defined as the random quantity $L_X(t, a)$ satisfying

$$L_X(t,a) = \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \int_0^t \mathbf{1}_{[a,\ a+\varepsilon[}(X_s) d[X]_s \ \forall a, t, \tag{28}$$

where $[X]_t$ is the quadratic variation process of the underlying continuous semimartingale at t. Formula (28) clarifies the sense in which $L_X(t,a)$ measures time in information units or, more rigorously, in units of the quadratic variation process. Interestingly, one could write

$$[X]_t = \int_{-\infty}^{\infty} L_X(t,a) da, \tag{29}$$

thus expressing the quadratic variation process in terms of contributions coming from fluctuations in the process that occur in the vicinity of different spatial points $a \in (-\infty, \infty)$. Equation (29) is a spatial decomposition of variation. Readers familiar with stationary time series analysis will recognize the similarity of (29) to the decomposition of the variance σ_X^2 of a process X in terms of its spectral density at different frequencies, i.e.,

$$\sigma_X^2 = \int_{-\pi}^{\pi} f_{xx}(w) dw, \tag{30}$$

where $f_{xx}(w)$ is the spectral density of X.

We can specialize the analysis to the case of SDPs and consider a re-scaled version of the standard notion of sojourn time defined as

$$\overline{L}_X(t,a) = \frac{L_X(t,a)}{\sigma^2(a)}. \tag{31}$$

Since, for an SDP as in (23), $d[X]_s = \sigma^2(X_s)ds$, then formula (31) can be interpreted as representing time in real time units rather than in units of the non-decreasing quadratic variation process. In other words, $\overline{L}_X(t,a)$ records the amount of calendar time spent by the process in the neighborhood of a and can be defined as "chronological local time" (Bosq, 1998, and Phillips and Park, 1998). Such a notion has an interesting interpretation. Consider the occupation measure that was introduced in Section 2. As pointed out earlier, the quantity η_A^T represents the amount of time spent by the process in a certain spatial set of nonzero Lebesgue measure. Chronological local time is nothing but the density of the occupation measure of the process. Put differently, chronological local time is a version of the Radon-Nikodym derivative of the occupation measure with respect to the Lebesgue measure and is an occupation density (Geman and Horowitz, 1980). In fact, we can write

$$\eta_A^T = \int_0^T \mathbf{1}_{\{X_s \in A\}} ds = \int_A \overline{L}_X(T,a) da, \qquad \forall A \in \mathfrak{B}(\mathfrak{D}) \tag{32}$$

and, by linearity and monotone convergence,

$$\int_0^T \Phi(X_s) ds = \int \Phi(a) \overline{L}_X(T,a) da, \tag{33}$$

where Φ is a Borel measurable, non-negative, function (Bosq, 1998, *inter alia*). Eq. (33) is typically called the "occupation time formula" and may be regarded as the analogue of a more classical expectation (i.e., the integral with respect to a time-invariant probability measure) in the analysis of times series which are not necessarily endowed with a time-invariant probability measure.

From an applied standpoint, the notion of chronological local time is relevant for at least three, mutually reinforcing, reasons. First, chronological local time has an appealing interpretation in terms of calendar time spent by the process in the vicinity of values in its range. Second, chronological local time arises naturally as the limiting process to which density-like kernel estimators converge provided the underlying process is a scalar semimartingale and suitable conditions on the relevant bandwidth are met. Third, as shown in the next subsection, this is the notion of local time which will play a crucial role in understanding the limiting distributions of the infinitesimal moments $\mu(.)$ and $\sigma^2(.)$. In what follows, we will use the convention of referring to it simply as "local time."

The first two reasons together suggest the usefulness of local time as a new method for the descriptive analysis of data that might not be stationary, so that the techniques may be used in situations where estimated probability density functions do not make theoretical sense. Consistent with this logic, recent work has proposed nonparametric estimates of the local time process and has interpreted them in terms of generalized densities to be employed as new descriptive tools for studying the spatial characteristics of time series which might be nonstationary. The original intuition is due to Phillips (2001, 2004) in the context of nonstationary discrete-time series embeddable in Brownian motion (namely, discrete time series of the unit-root type). In continuous-time finance models, local time was first used as a descriptive tool for possibly nonstationary (recurrent) SDPs of the form analyzed here by Bandi (2002).

As pointed out earlier, a natural way to identify local time is to use density-like kernel estimators. Based on the same sampling scheme as in Section 2 with $T = \overline{T}$, we define an estimate of $\overline{L}_X(\overline{T}, a)$ as

$$\widehat{\overline{L}}_X(\overline{T}, a) = \frac{\Delta_{n,\overline{T}}}{h_{n,\overline{T}}} \sum_{i=1}^{n} \mathbf{K}\left(\frac{X_{i\Delta_{n,\overline{T}}} - a}{h_{n,\overline{T}}}\right), \tag{34}$$

where $h_{n,\overline{T}}$ is bandwidth sequence depending on n and $\mathbf{K}(.)$ is a conventional kernel function satisfying the assumptions in Section 2. Theorems 3.1 and 3.2 below show consistency of the local time estimator for its theoretical counterpart and provide a limiting distribution.

Theorem 3.1 *Assume X_t is the solution to (14). If $h_{n,\overline{T}} \to 0$ as $n \to \infty$ for a fixed time span T $(= \overline{T})$ in such a way that $\frac{1}{h_{n,\overline{T}}}(\Delta_{n,\overline{T}} \log(1/\Delta_{n,\overline{T}}))^{1/2} = o(1)$, then*

$$\widehat{\overline{L}}_X(\overline{T}, a) \stackrel{a.s.}{\to} \overline{L}_X(\overline{T}, a) \qquad \forall a \in \mathfrak{D}. \tag{35}$$

Proof *See Florens-Zmirou (1993). For a proof that allows for more general kernel functions than the indicator kernel used in Florens-Zmirou (1993) and utilizes different statistical tools, such as the occupation time formula in (33), see BP (2003). Park (2006) studies uniform L^1-consistency.*

We now turn to the asymptotic distribution. Here, and in what follows, the notation **MN** denotes the mixed Gaussian density.

Theorem 3.2 *Assume X_t is the solution to (14). If $h_{n,\overline{T}} \to 0$ as $n \to \infty$ for a fixed time span T $(= \overline{T})$ in such a way that $\frac{1}{h^{3/2}_{n,\overline{T}}}(\Delta_{n,\overline{T}} \log(1/\Delta_{n,\overline{T}}))^{1/2} = o(1)$, then*

$$\frac{1}{\sqrt{h_{n,\overline{T}}}}\left(\widehat{\overline{L}}_X(\overline{T},a) - \overline{L}_X(\overline{T},a)\right) \Rightarrow \mathbf{MN}\left(0, 8\mathbf{k}\frac{1}{\sigma^2(a)}\overline{L}_X(\overline{T},a)\right) \qquad \forall a \in \mathfrak{D}, \tag{36}$$

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

Proof *See Bandi (2002) for a proof in the case of an underlying SDP that is assumed to be the unique and strong solution to a stochastic differential equation like (14), in agreement with the statement of the theorem. See Phillips (2001, 2004) for a proof in the case of the estimated local time of linear, nonstationary, discrete-time series embeddable in Brownian motion.*

Theorem 3.1 justifies estimating the calendar time that an SDP spends in the local vicinity of a point by using a density-like kernel estimator. Theorem 3.2 enables us to construct asymptotic confidence intervals which closely resemble conventional intervals for probability densities obtained from kernel estimates (Bandi, 2002, and Phillips, 2004). The (estimated) asymptotic 95% confidence interval of $\overline{L}_X(\overline{T},a)$ is, in fact, given by

$$\widehat{\overline{L}}_X(\overline{T},a) \pm 1.96\left(8\mathbf{k}\frac{h_{n,\overline{T}}}{\widehat{\sigma}^2(a)}\widehat{\overline{L}}_X(\overline{T},a)\right)^{1/2}. \tag{37}$$

It is worth recalling here that the limiting process $\overline{L}_X(\overline{T},a)$ is random. As opposed to standard probability densities, spatial densities have a time dimension which can be fruitfully explored by changing the span of data used in the implementation of (34). In other words, $\widehat{\overline{L}}_X(\overline{T}_1,a)$ and $\widehat{\overline{L}}_X(\overline{T}_2,a)$ measure the time spent by the SDP of interest at a in the time intervals $[0,\overline{T}_1]$ and $[0,\overline{T}_2]$, respectively, and can be used as robust (to deviations from stationarity) descriptive statistics to summarize the spatial evolution of the SDP over time.

Some additional observations are in order. Given the interpretation of local time, the following result should come as no surprise.

Theorem 3.3 *Assume X_t the solution to (14) and $m(\mathfrak{D}) < \infty$ (as implied by Assumption (3b)). If $h_{n,T} \to 0$ as $n, T \to \infty$ in such a way that $\frac{1}{h_{n,T}}(\Delta_{n,T}\log(1/\Delta_{n,T}))^{1/2} = o(1)$, then*

$$\frac{\widehat{\overline{L}}_X(T,a)}{T} \xrightarrow{a.s.} f(a) = \frac{m(a)}{m(\mathfrak{D})} \qquad \forall a \in \mathfrak{D}. \tag{38}$$

Proof *See BP (2003) and Moloche (2004a). The interested reader is also referred to Moloche (2004a) for a discussion of the limiting properties of the expected local time process. An asymptotic theory for estimates of the expected local time is given in Park (2006).*

This result simply tells us that the standardized local time estimator of a strictly stationary (or positive recurrent) SDP converges point-wise to the stationary density of the process with probability one. Loosely speaking, if we divide the estimated time spent by the process at a between 0 and T by T, by appealing to the conventional frequentist notion of probability, we expect the ratio to converge to the probability mass at a when letting T diverge to infinity. Equivalently, we can say that the local time of a stationary, or positive recurrent, process diverges to infinity linearly with T (see, also, Bosq, 1997, and Bosq and Davidov, 1998). Naturally, we expect nonstationary, but recurrent, processes to have local times which diverge at speeds that are slower than T. Such speeds are generally not quantifiable. Nonetheless, Brownian motion is known to have a local time that diverges at speed $\sqrt{T}$. The following result can be easily proved for a standard Brownian motion (see BP, 2003, and, for a more general method of proof, Moloche, 2004a):

$$\frac{\widehat{L}_B(T,a)}{\sqrt{T}} = \frac{\sqrt{T}L_B\left(1, \frac{a}{\sqrt{T}}\right)}{\sqrt{T}} + o_{a.s.}(1) \xrightarrow{a.s.} L_B(1,0). \tag{39}$$

We will come back to a discussion of the divergence rates of local time when describing the estimation procedures for drift and diffusion function. For the time being, it suffices to stress that the class of SDPs that we are studying, namely the class of recurrent SDPs, has local times that diverge to infinity with probability one when the time span does so. The reason is easy to explain. Local time measures the time spent by the process at a point between 0 and T, say. Scalar recurrent processes visit every point an infinite number of times as T goes to infinity with probability one. Necessarily, therefore, the local time of a recurrent process diverges to infinity almost surely as T diverges to infinity.

We complete this subsection by pointing out that functions of spatial densities can be used as descriptive tools for possibly nonstationary SDPs just like functions of probability densities are employed as descriptive statistics in the context of stationary time series (Phillips, 2001, 2004). For example, Phillips (2004) defines a new kind of hazard function for discrete-time nonstationary time series as

$$\overline{H}_X(T,a) = \frac{\overline{L}_X(T,a)}{\int_a^\infty \overline{L}_X(T,s)ds}, \tag{40}$$

where, as usual, $\overline{L}_X(T,a)$ is the standard sojourn time. Such a function can be interpreted as the spatial counterpart of conventional hazard functions (see, e.g., Silverman, 1986, and Prakasa-Rao, 1983), where probability densities replace local times, and might be used to quantify hazards of certain financial time series such as inflation rates or interest rates. When applied to interest rates (as done in Bandi, 2002), for instance, formula (40) gives the conditional risk over the period $[0,T]$ of an interest rate level of a, given that interest rates are at least as large as a. An asymptotic theory for kernel estimates of spatial hazard rates is available to assist statistical inference.

Theorem 3.4 *Assume X_t is the solution to (14). If $h_{n,\overline{T}} \to 0$ as $n \to \infty$ for a fixed time span T $(=\overline{T})$ in such a way that $\frac{1}{h_{n,\overline{T}}}(\Delta_{n,\overline{T}} \log(1/\Delta_{n,T}))^{1/2} = o(1)$, then*

$$\widehat{\overline{H}}_X(\overline{T},a) = \frac{\widehat{\overline{L}}_X(\overline{T},a)}{\int_a^\infty \widehat{\overline{L}}_X(\overline{T},s)ds} \overset{a.s.}{\to} \overline{H}_X(\overline{T},a) \qquad \forall a \in \mathfrak{D}. \tag{41}$$

Furthermore, if $h_{n,\overline{T}} \to 0$ as $n \to \infty$ for a fixed time span T $(=\overline{T})$ and

$$\frac{1}{h^{3/2}_{n,\overline{T}}}(\Delta_{n,\overline{T}} \log(1/\Delta_{n,T}))^{1/2} = o(1), \tag{42}$$

then

$$\frac{1}{\sqrt{h_{n,\overline{T}}}}\left(\widehat{\overline{H}}_X(\overline{T},a) - \overline{H}_X(\overline{T},a)\right) \Rightarrow \mathbf{MN}\left(0, \frac{8\mathbf{k}(\overline{H}_X(\overline{T},a))^2}{\sigma^2(a)\overline{L}_X(\overline{T},a)}\right) \qquad \forall a \in \mathfrak{D}, \tag{43}$$

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

Proof *See Bandi (2002) for a proof in the case of an underlying SDP that is assumed to be the unique and strong solution to a stochastic differential equation like (14), consistent with the statement of the theorem. See Phillips (2001, 2004) for a proof in the case of the estimated local time of linear, nonstationary, discrete-time series embeddable in Brownian motion.*

For an introduction to descriptive methods for nonstationary discrete-time series using local time and related notions, the reader is referred to Phillips (2001, 2004). Estimated spatial densities and spatial hazard rates have been used by Bandi (2002) in studying the temporal dynamics of a nonparametric continuous-time specification (as in (14) above) for the short-term interest rate process. We refer the interested reader to that paper for a discussion about empirical implementation of the methodology in the case of SDPs. Park (2006) provides an estimation theory for functionals of spatial densities and spatial distributions while introducing novel (spatial) notions of value-at-risk and stochastic dominance, among other concepts. We now turn to the estimation of the infinitesimal conditional moments $\mu(.)$ and $\sigma^2(.)$.

3.2 N-W Kernel Estimation of the Infinitesimal Moments of an SDP

It is well-known that the transition density of the unique solution to (14) is completely characterized by the functions $\mu(.)$ and $\sigma^2(.)$. In other words, understanding the temporal evolution of a general SDP amounts to identifying the drift and the diffusion function. As discussed in Section 2, such functions have (infinitesimal) conditional moment definitions, i.e.

$$\mathbf{E}^a[X_t - a] = t\mu(a) + o(t) \tag{44}$$

$$\mathbf{E}^a[(X_t - a)^2] = t\sigma^2(a) + o(t), \tag{45}$$

as $t \downarrow 0$, or

$$\mu(a) = \lim_{t \to 0} \frac{1}{t} \mathbf{E}^a[X_t - a], \tag{46}$$

$$\sigma^2(a) = \lim_{t \to 0} \frac{1}{t} \mathbf{E}^a \left[(X_t - a)^2 \right], \tag{47}$$

where a is a generic initial condition and $\mathbf{E}^a$ is, as earlier in the Introduction, the expectation operator associated with the process started at a. Loosely speaking, (44) and (45) can be interpreted as representing the "instantaneous" conditional mean and the "instantaneous" conditional variance of the process when $X_0 = a$.

Our previous, informal arguments, combined with the definitions of $\mu(.)$ and $\sigma^2(.)$ in (44) and (45) above, suggest that standard functional techniques for conditional expectations based on local averages may be natural tools to estimate the two functions driving the evolution of a general SDP. This is the intuition in BP (2003) where sample analogues to infinitesimal conditional expectations are used to estimate *both* the drift and the diffusive volatility. Consider the same sampling scheme as in Section 2 (i.e., $n, T \to \infty$ with $\Delta_{n,T} = \frac{T}{n} \to 0$). Define

$$\widehat{\mu}_{(n,T)}(a) = \frac{1}{\Delta_{n,T}} \frac{\sum_{i=1}^{n-1} \mathbf{K}\left(\frac{X_{i\Delta_{n,T}} - a}{h_{n,T}}\right)(X_{(i+1)\Delta_{n,T}} - X_{i\Delta_{n,T}})}{\sum_{i=1}^{n} \mathbf{K}\left(\frac{X_{i\Delta_{n,T}} - a}{h_{n,T}}\right)}, \tag{48}$$

$$\widehat{\sigma}^2_{(n,T)}(a) = \frac{1}{\Delta_{n,T}} \frac{\sum_{i=1}^{n-1} \mathbf{K}\left(\frac{X_{i\Delta_{n,T}} - a}{h_{n,T}}\right)(X_{(i+1)\Delta_{n,T}} - X_{i\Delta_{n,T}})^2}{\sum_{i=1}^{n} \mathbf{K}\left(\frac{X_{i\Delta_{n,T}} - a}{h_{n,T}}\right)}, \tag{49}$$

where $\mathbf{K}(.)$ is a kernel function satisfying the assumptions in Section 2. Formulae (48) and (49) can be interpreted as the N-W kernel estimates corresponding to (46) and (47) above and they belong to the more general class of functional estimates suggested in BP (2003). We now consider some aspects of the asymptotic theory in that paper. We begin with (48).

Theorem 3.5 *Assume X_t is the solution to (14). Also, assume $h_{n,T}$ is such that*

$$\frac{\overline{L}_X(T,a)}{h_{n,T}}(\Delta_{n,T}\log(1/\Delta_{n,T}))^{1/2} = o_{a.s.}(1) \tag{50}$$

and $h_{n,T}\overline{L}_X(T,a) \stackrel{a.s.}{\to} \infty$ as $n, T \to \infty$ with $\frac{T}{n} \to 0$. Then,

$$\widehat{\mu}_{(n,T)}(a) \stackrel{a.s.}{\to} \mu(a). \tag{51}$$

Furthermore, if $h^5_{n,T}\overline{L}_X(T,a) = O_{a.s.}(1)$, then

$$\sqrt{h_{n,T}\widehat{\overline{L}}_X(T,a)}\left\{\widehat{\mu}_{(n,T)}(a) - \mu(a) - \Gamma_\mu(a)\right\} \Rightarrow \mathbf{N}\left(0, \mathbf{K}_2\sigma^2(a)\right), \tag{52}$$

where

$$\Gamma_\mu(a) = h^2_{n,T}\mathbf{K}_1\left[\mu'(a)\frac{m'(a)}{m(a)} + \frac{1}{2}\mu''(a)\right], \tag{53}$$

and $m(a)$ is the speed function of the process X at a, i.e., $m(a) = \frac{2}{S'(a)\sigma^2(a)}$.

Proof *See BP (2003).*

We now turn to (49).

Theorem 3.6 *Assume X_t is the solution to (14). Also, assume $h_{n,T}$ is such that*

$$\frac{\overline{L}_X(T,a)}{h_{n,T}}(\Delta_{n,T}\log(1/\Delta_{n,T}))^{1/2} = o_{a.s.}(1) \tag{54}$$

as $n, T \to \infty$ with $\frac{T}{n} \to 0$. Then,

$$\widehat{\sigma}^2_{(n,T)}(a) \stackrel{a.s.}{\to} \sigma^2(a). \tag{55}$$

Furthermore, if $\frac{h^5_{n,T}\overline{L}_X(T,a)}{\Delta_{n,T}} = O_{a.s.}(1)$, then

$$\sqrt{\frac{h_{n,T}\widehat{\overline{L}}_X(T,a)}{\Delta_{n,T}}}\left\{\widehat{\sigma}^2_{(n,T)}(a) - \sigma^2(a) - \Gamma_{\sigma^2}(a)\right\} \Rightarrow \mathbf{N}\left(0, 2\mathbf{K}_2\sigma^4(a)\right), \tag{56}$$ [8]

where

$$\Gamma_{\sigma^2}(a) = h^2_{n,T}\mathbf{K}_1\left[\left(\sigma^2(a)\right)'\frac{m'(a)}{m(a)} + \frac{1}{2}\left(\sigma^2(a)\right)''\right], \tag{57}$$

and $m(a)$ is the speed function of the process X at a, i.e., $m(a) = \frac{2}{S'(a)\sigma^2(a)}$.

[8] The proportionality factor reported in BP (2003), i.e., 4 rather than 2, is of course a typo (see, e.g., the related findings in BP, 2007, and BM, 2004).

Proof *See BP (2003).*

Under appropriate conditions on the bandwidths, the estimators converge to the true functions with probability one. The asymptotic distributions are normal and centered at the relevant functions provided the bandwidth sequences converge to zero sufficiently fast. If this is not the case, non-random bias terms affect the limiting distributions. The diffusion estimator converges to its theoretical counterpart at a faster rate $\left(\sqrt{\frac{h_{n,T}\widehat{\overline{L}}_X(T,a)}{\Delta_{n,T}}}\right)$ than the drift estimator $\left(\sqrt{h_{n,T}\widehat{\overline{L}}_X(T,a)}\right)$. We will now be more specific about the drift case but similar arguments apply to the diffusion function. A discussion of the difference between the two cases will follow.

When dealing with the drift, local time plays the same role which is played by the number of observations in the more standard estimation of conditional moments in discrete time. What matters to identify the drift at a point a is not the rate of divergence of the number of data points n but the rate of divergence of the number of calendar time units spent by the process in the vicinity of the level a (c.f. Section 2). Not surprisingly, therefore, the standard condition $nh_n \to \infty$ (or $Th_T \to \infty$) is replaced in our case by $h_{n,T}\overline{L}_X(T,a) \overset{a.s.}{\to} \infty$ as $n, T \to \infty$ with $\frac{T}{n} \to 0$. Equivalently, the point-wise condition which needs to be imposed on the bandwidth to prevent the insurgence of a bias term in the limit is $h^5_{n,T}\overline{L}_X(T,a) \overset{a.s.}{\to} 0$ as opposed to the more conventional condition $h^5_n n \to 0$ (or $h^5_T T \to 0$). In other words, the smoothing parameter has to converge to zero slowly enough as to guarantee that $h_{n,T}\overline{L}_X(T,a) \overset{a.s.}{\to} \infty$ (rather than $nh_n \to \infty$) but sufficiently fast as to satisfy $h^5_{n,T}\overline{L}_X(T,a) \overset{a.s.}{\to} 0$ (rather than $h^5_n n \to 0$).

Correspondingly, the rate of convergence of the estimator is random and equal to $\sqrt{h_{n,T}\widehat{\overline{L}}_X(T,a)}$ rather than $\sqrt{nh_n}$. Let us now consider the asymptotic variance and bias. These are given by

$$\frac{\left(\int_{-\infty}^{\infty}\mathbf{K}^2(s)ds\right)\sigma^2(a)}{h_{n,T}\widehat{\overline{L}}_X(T,a)} \tag{58}$$

and

$$h^2_{n,T}\left(\int_{-\infty}^{\infty}s^2\mathbf{K}(s)ds\right)\left[\mu'(a)\frac{m'(a)}{m(a)}+\frac{1}{2}\mu''(a)\right], \tag{59}$$

respectively. Their interpretation is clear when considering well-known findings about the asymptotic bias and variance of the standard N-W estimator of conditional moments in discrete time (see, for instance, formula (3.60) and Theorem 3.5 in Pagan and Ullah, 1999). In particular, the spatial density estimate $\widehat{\overline{L}}_X(T,a)$ and the ratio between the derivative of the speed function and the speed function itself play the same role as that played by the term $Tf(a)$, where $f(a)$ is the stationary density at a, and the ratio between the derivative of the density function and the density function itself in conventional nonparametric time-series analysis under stationarity. The features of the theory we discuss in this review are a reflection of the mildness of the assumptions imposed on the underlying process. As pointed out earlier, recurrence is all that is required.

This said, the theory is specializable to the positive recurrent and stationary cases. The following theorems mirror more conventional results in the functional estimation of conditional expectations for stationary, discrete-time series, and are immediate after noticing that, under conditions laid out earlier,

$$\frac{\widehat{\overline{L}}_X(T,a)}{T} \overset{a.s.}{\to} f(a) \tag{60}$$

and[9]

$$\frac{m'(a)}{m(a)} = \frac{f'(a)}{f(a)} = (\log f(a))' = \frac{2\mu(a) - \left(\sigma^2(a)\right)'}{\sigma^2(a)} \tag{61}$$

under positive recurrence or stationarity (c.f. the discussion in the previous subsection).

Theorem 3.7 *Assume X_t is the solution to (14) and $m(\mathfrak{D}) < \infty$ (as implied by Assumption (3b)). Furthermore, assume $h_{n,T} \to 0$ as $n, T \to \infty$ with $\Delta_{n,T} \to 0$ so that $\frac{T}{h_{n,T}}(\Delta_{n,T}\log(1/\Delta_{n,T}))^{1/2} = o(1)$ and $h_{n,T}T \to \infty$. Then,*

$$\widehat{\mu}_{(n,T)}(a) \overset{a.s.}{\to} \mu(a). \tag{62}$$

Additionally,

$$\sqrt{h_{n,T}T}\left\{\widehat{\mu}_{(n,T)}(a) - \mu(a) - \Gamma_\mu(a)\right\} \Rightarrow \mathbf{N}\left(0, \mathbf{K}_2\frac{\sigma^2(a)}{f(a)}\right), \tag{63}$$

if $h_{n,T} = O(T^{-1/5})$, where

$$\Gamma_\mu(a) = h_{n,T}^2\mathbf{K}_1\left[\mu'(a)\frac{f'(a)}{f(a)} + \frac{1}{2}\mu''(a)\right], \tag{64}$$

and $f(a)$ is the stationary density function of the process at a.

Proof *See BP(2003).*

[9] The last equality in (61) can be proved by solving the equation

$$\int_{\mathfrak{D}} \mathsf{L}\varphi(x)f(x)dx = 0$$

for a function $\varphi(.)$ in the domain of the infinitesimal generator L (c.f., Aït-Sahalia *et al.* (2009) in this volume). Such an equation holds by stationarity (c.f., Hansen and Scheinkman, 1995).

Theorem 3.8 *Assume X_t is the solution to (14) and $m(\mathfrak{D}) < \infty$ (as implied by Assumption (3b)). Furthermore, assume $h_{n,T} \to 0$ as $n, T \to \infty$ with $\Delta_{n,T} \to 0$ so that $\frac{T}{h_{n,T}}(\Delta_{n,T}\log(1/\Delta_{n,T}))^{1/2} = o(1)$. Then,*

$$\widehat{\sigma}^2_{(n,T)}(a) \overset{a.s.}{\to} \sigma^2(a). \tag{65}$$

Additionally,

$$\sqrt{nh_{n,T}}\left\{\widehat{\sigma}^2_{(n,T)}(a) - \sigma^2(a) - \Gamma_{\sigma^2}(a)\right\} \Rightarrow \mathbf{N}\left(0, 2\mathbf{K}_2\frac{\sigma^4(a)}{f(a)}\right), \tag{66}$$

if $h_{n,T} = O(n^{-1/5})$, where

$$\Gamma_{\sigma^2}(a) = h^2_{n,T}\mathbf{K}_1\left[\left(\sigma^2(a)\right)'\frac{f'(a)}{f(a)} + \frac{1}{2}\left(\sigma^2(a)\right)''\right], \tag{67}$$

and $f(a)$ is the stationary density function of the process at a.

Proof *See BP (2003).*

It was noted earlier that the rate of convergence of the drift estimator is $\sqrt{h_{n,T}\widehat{\overline{L}}_X(T,a)}$. This rate clarifies the sense in which identification of the infinitesimal first moment of an SDP requires an enlarging span of data (see, e.g., Geman, 1979). In effect, should T be fixed, then $\overline{L}_X(T,a)$ would be bounded in probability and the drift estimator would diverge at the rate $\sqrt{h_{n,T}}$ (Bandi, 2002). On the contrary, $\overline{L}_X(T,a)$ diverges to infinity as $T \to \infty$ by the assumption of recurrence, thereby ensuring that $\sqrt{h_{n,T}\widehat{\overline{L}}_X(T,a)} \overset{a.s.}{\to} \infty$, provided the bandwidth $h_{n,T}$ converges to zero slowly enough. The intuition why an enlarging span of data for drift estimation is needed was put forward in Section 2 but it is worth repeating here for clarity. In order to achieve consistency of the drift estimator at a certain spatial level, say a, we need the process to visit that level an infinite number of times over time. In this case we can take averages of first differences between observations on the continuous path of the process occurring in the local neighborhood of a (as suggested in Section 2) and hope to apply an appropriate law of large number. Recurrence guarantees that every level will be visited an infinite number of times over time provided $T \to \infty$. A diverging local time at a as $T \to \infty$ is simply the manifestation of the fact that, with probability one, the level a is crossed an infinite number of times, as the time span grows indefinitely.

Importantly, the diffusion function can be identified over a fixed span of data as shown by Florens-Zmirou (1993), Brugière (1991, 1993) and Jacod (1997) in important early work. In other words, one can employ (49) above and fix $T = \overline{T}$ to derive limiting results.

Theorem 3.9 *Assume X_t is the solution to (14). Given $n \to \infty$, $T = \overline{T}$ and $h_{n,\overline{T}} \to 0$ as $n \to \infty$ so that $\frac{1}{h_{n,\overline{T}}}(\Delta_{n,\overline{T}} \log(1/\Delta_{n,\overline{T}}))^{1/2} = o(1)$, the estimator (49) converges to the true function with probability one.*

If $nh^4_{n,\overline{T}} \to 0$, then the asymptotic distribution of (49) is driven by a "martingale" effect and has the form

$$\sqrt{nh_{n,\overline{T}}}\left\{\widehat{\sigma}^2_{(n,\overline{T})}(a) - \sigma^2(a)\right\} \Rightarrow \mathbf{MN}\left(0, \frac{2\mathbf{K}_2\sigma^4(a)}{\overline{L}_X(\overline{T},a)/\overline{T}}\right). \tag{68}$$

If $nh^4_{n,\overline{T}} \to \infty$, then the asymptotic distribution of (49) is driven by a "bias" effect and has the form

$$\frac{1}{h^{3/2}_{n,\overline{T}}}\left\{\widehat{\sigma}^2_{(n,\overline{T})}(a) - \sigma^2(a)\right\} \Rightarrow \mathbf{MN}\left(0, 16\varphi^{ind}\frac{\left(\sigma^{'}(a)\right)^2}{\overline{L}_X(\overline{T},a)}\right), \tag{69}$$

where $\varphi^{ind} = 2\int_0^\infty \int_0^\infty \mathbf{K}(a)\mathbf{K}(b)\min(a,b)dadb$.

Proof *Florens-Zmirou (1993) was the first to use the estimator (49) above to identify the diffusion function. The kernel used in Florens-Zmirou (1993) is a discontinuous indicator kernel and the consistency proof is based on mean-squared deviations (see also Jacod, 1997, for a interesting refinement of this approach). The asymptotic distribution (68) is provided in Florens-Zmirou (1993) with $2\mathbf{K}_2 = 2$ due the nature of the kernel used. Jiang and Knight (1997) modify the Florens-Zmirou estimator and define it using a continuous kernel. Their consistency proof follows Florens-Zmirou (1993) and is also in mean-squared. The statement of the theorem above is based on BP (2003) where almost sure convergence is shown. The treatment in BP (2003) highlights the potential for a random bias term (i.e., in Eq. (69) above) which might dominate the asymptotic distribution should the bandwidth sequence not converge to zero at a fast enough pace.*

We now turn to a brief discussion of bandwidth selection.

3.2.1 The Choice of Bandwidth

Optimal bandwidth selection is technically very demanding in these models and represents an open field of research, no rigorous treatment being available at present, at least to the authors' knowledge. Based on Theorem 3.5, in the drift case one can write

$$h^{drift}_{n,T} = c^{drift}_{n,T}\frac{1}{\log \widehat{\overline{L}}_X(T,a)}\widehat{\overline{L}}_X(T,a)^{-1/5}, \tag{70}$$

where $\widehat{\overline{L}}_X(T,a)$ is the estimated local time at a and $c^{drift}_{n,T}$ is a constant of proportionality. Such an expression derives from the fact that the asymptotic mean-squared-error (MSE) at a generic level a is of order

$$O_{a.s.}\left(\left(h_{n,T}^{drift}\right)^4\right)+O_{a.s.}\left(\frac{1}{h_{n,T}^{drift}\widehat{\overline{L}}_X(T,a)}\right) \tag{71}$$

and the best rate is obtained by taking $h_{n,T}^{drift}\propto\widehat{\overline{L}}_X(T,a)^{-1/5}$ in which case the limiting MSE is of order $\widehat{\overline{L}}_X(T,a)^{-4/5}$. Pre-multiplication by $\frac{1}{\log\widehat{\overline{L}}_X(T,a)}$ is, of course, somewhat ad-hoc but useful to achieve a close-to-optimal rate of convergence and undersmooth slightly, thereby eliminating the influence of the non-random bias term from the asymptotic distribution of the drift estimates.

As for the diffusion function, Theorem 3.6, and a similar argument to that above, suggest the expression

$$h_{n,T}^{diff}=c_{n,T}^{diff}\frac{1}{\log\left(\widehat{\overline{L}}_X(T,a)/\Delta_{n,T}\right)}\left(\widehat{\overline{L}}_X(T,a)/\Delta_{n,T}\right)^{-1/5} \tag{72}$$

and, in consequence, the approximation

$$h_{n,T}^{diff}\approx c_{n,T}^{diff}\frac{1}{\log n}n^{-1/5}, \tag{73}$$

for a T diverging to infinity sufficiently slowly (or for the case of local time diverging at speed T, i.e., the stationary case). When T is fixed, as in Theorem 3.9, the previous condition becomes

$$h_{n,\overline{T}}^{diff}=c_{n,\overline{T}}^{diff}\frac{1}{\log n}n^{-1/4}. \tag{74}$$

Again, both (72) and (74) imply close-to-optimal rates, namely rates that almost maximize the speed of convergence of the proposed estimators to the functions of interest while preventing the insurgence of a (deterministic or random) bias term in the limit.

Some observations are in order. First, Eq. (70) suggests that there is explicit scope for local adaptation of the drift bandwidth sequence to the number of visits to the point at which estimation is performed. In fact, it appears that the optimal bandwidth for the drift should be smaller at levels that are often visited.[10] In light of the approximation (73) and the result in (74) such an effect is more pronounced in the drift case than in the diffusion case. Second, Theorems 3.5 and 3.6 suggest that the optimal drift bandwidth is generally larger than the optimal diffusion bandwidth. Both observations are of course reflections of the fact that the local dynamics of an SDP contain more information about the diffusion function than about the drift, thereby rendering consistent estimation of the infinitesimal second moment possible over a fixed span of data.

[10] Bias reduction is the standard justification for suggesting variable bandwidths whose magnitude is inversely related to the estimated density function (c.f., Pagan and Ullah, 1999, page 31).

Unfortunately, despite being widely used in empirical work, standard automatic technologies (as discussed in Pagan and Ullah, 1999, among others) to select the constants $c^{drift}_{n,T}$, $c^{diff}_{n,T}$ and $c^{diff}_{n,\overline{T}}$, such as least squares cross-validation, have not been justified in the case of SDPs. Nonetheless, contrary to drift estimation, conventional cross-validation procedures appear to perform reasonably well when dealing with diffusion function estimation (c.f. Bandi and Nguyen, 1999). Future work should usefully focus on the development of convincing criteria for determining the constants $c^{drift}_{n,T}$, $c^{diff}_{n,T}$ and $c^{diff}_{n,\overline{T}}$ in (70), (72), (73) and (74) above and for selecting a preliminary smoothing sequence to define $\widehat{\overline{L}}_X(T,a)$, a fundamental quantity in the theory we are reviewing.

3.3 Extensions in Kernel Estimation for SDPs

3.3.1 Double-Smoothing

The estimators (48) and (49) belong to the general class of estimators suggested by BP (2003). Consistently with the discussion in BP (2003), one could envisage a more involved two-step procedure with further smoothing. First, one could define sample analogs to the values that drift and diffusion take on at the sampled points, i.e.,

$$\widetilde{\mu}_{(n,T)}(X_{i\Delta_{n,T}}) = \frac{1}{\Delta_{n,T}} \frac{\sum_{j=1}^{n-1} \mathbf{K}\left(\frac{X_{j\Delta_{n,T}}-X_{i\Delta_{n,T}}}{h_{n,T}}\right)(X_{(j+1)\Delta_{n,T}} - X_{j\Delta_{n,T}})}{\sum_{j=1}^{n} \mathbf{K}\left(\frac{X_{j\Delta_{n,T}}-X_{i\Delta_{n,T}}}{h_{n,T}}\right)}, \tag{75}$$

$$\widetilde{\sigma}^2_{(n,T)}(X_{i\Delta_{n,T}}) = \frac{1}{\Delta_{n,T}} \frac{\sum_{j=1}^{n-1} \mathbf{K}\left(\frac{X_{j\Delta_{n,T}}-X_{i\Delta_{n,T}}}{h_{n,T}}\right)(X_{(j+1)\Delta_{n,T}} - X_{j\Delta_{n,T}})^2}{\sum_{j=1}^{n} \mathbf{K}\left(\frac{X_{j\Delta_{n,T}}-X_{i\Delta_{n,T}}}{h_{n,T}}\right)}. \tag{76}$$

Second, estimated drift and diffusion values at the sampled points could be averaged using weights based on smooth kernels to recover the theoretical functions at levels that the sampled process does not visit, i.e.,

$$\overline{\mu}_{(n,T)}(a) = \frac{\sum_{i=1}^{n} \overline{\mathbf{K}}\left(\frac{X_{i\Delta_{n,T}}-a}{\varepsilon_{n,T}}\right)\widetilde{\mu}_{(n,T)}(X_{i\Delta_{n,T}})}{\sum_{i=1}^{n} \overline{\mathbf{K}}\left(\frac{X_{i\Delta_{n,T}}-a}{\varepsilon_{n,T}}\right)}, \tag{77}$$

$$\overline{\sigma}^2_{(n,T)}(a) = \frac{\sum_{i=1}^{n} \overline{\mathbf{K}}\left(\frac{X_{i\Delta_{n,T}}-a}{\varepsilon_{n,T}}\right)\widetilde{\sigma}^2_{(n,T)}(X_{i\Delta_{n,T}})}{\sum_{i=1}^{n} \overline{\mathbf{K}}\left(\frac{X_{i\Delta_{n,T}}-a}{\varepsilon_{n,T}}\right)}. \tag{78}$$

with $\overline{\mathbf{K}}(.)$ possibly different from $\mathbf{K}(.)$ and $h_{n,T}$ possibly different from $\varepsilon_{n,T}$. Both $\overline{\mathbf{K}}(.)$ and $\mathbf{K}(.)$ satisfy the assumptions in Section 2.[11]

[11] An interesting, alternative approach is studied in Reno' (2008) where the preliminary spot variance estimates $\widetilde{\sigma}^2_{(n,T)}$ at times $i\Delta_{n,T}$ are Fourier estimates as in Malliavin and Mancino (2002).

Asymptotically, the doubly-smoothed estimates (77) and (78) offer additional flexibility over their simple counterparts in (48) and (49) above. In effect, they can improve the asymptotic trade-off between bias and variance effects, thereby delivering smaller limiting MSEs than simple smoothing. As usual, let us focus on the drift while keeping in mind that the intuition extends to the diffusion function.

If $h_{n,T}$ satisfies the conditions in Theorem 3.5 and $\varepsilon_{n,T}/h_{n,T} \to \phi$, then the limiting bias and variance of the drift estimator (77) at a generic point a are, from BP (2003),

$$\frac{\theta_\phi \sigma^2(a)}{h_{n,T}\widehat{\overline{L}}_X(T,a)}, \tag{79}$$

with

$$\theta_\phi = \int\int\int \overline{\mathbf{K}}(a)\overline{\mathbf{K}}(e)\mathbf{K}(z-\phi e)\mathbf{K}(z-\phi a)dzdeda, \tag{80}$$

and

$$h_{n,T}^2 \mathbf{K}^\phi \left[\mu'(a)\frac{m'(a)}{m(a)} + \frac{1}{2}\mu''(a)\right], \tag{81}$$

with

$$\mathbf{K}^\phi = \int s^2 \mathbf{K}(s)ds + \phi \int s^2 \overline{\mathbf{K}}(s)ds. \tag{82}$$

Clearly, if $\varepsilon_{n,T}/h_{n,T} \to \phi = 0$, then double-smoothing coincides asymptotically with single-smoothing by a straightforward comparison of (79) and (81) with (58) and (59) above, respectively. Nonetheless, since θ_ϕ is a decreasing function of ϕ and $\mathbf{K}^\phi$ is an increasing function of it, there is scope for using convoluted kernels in order to achieve asymptotic MSEs which are optimized at values ϕ which are strictly larger than zero.

In finite samples, the extra level of smoothing which is implied by the use of convoluted kernels might be particularly beneficial, especially in the drift case. The intuition is as follows. We stressed earlier that the optimal smoothing parameter for the drift is generally larger than the corresponding choice for diffusion estimation. Nonetheless, there appears to be a fundamental difficulty in choosing the optimal drift bandwidth as a function of the estimated local time (as implied by Eq. (70)) and, consequently, as a function of the recurrence properties of the underlying Markov process. The use of convoluted kernels may achieve, in finite samples, the level of smoothing for the drift that weighted averages based on simple kernels would guarantee with relatively more appropriate (larger) choices of the bandwidth (BN, 1999).

3.3.2 Local Linear and Polynomial Estimation

It is immediate to see that the estimators (48) and (49) can be written as

$$\widehat{\mu}_{n,T}(a) = \arg\min_{\theta^{\mu}} \sum_{i=1}^{n-1} \mathbf{K}\left(\frac{X_{i\Delta_{n,T}} - a}{h_{n,T}}\right)\left\{\frac{1}{\Delta_{n,T}}(X_{(i+1)\Delta_{n,T}} - X_{i\Delta_{n,T}}) - \theta^{\mu}\right\}^2 \tag{83}$$

and

$$\widehat{\sigma}^2_{n,T}(a) = \arg\min_{\theta^{\sigma^2}} \sum_{i=1}^{n-1} \mathbf{K}\left(\frac{X_{i\Delta_{n,T}} - a}{h_{n,T}}\right)\left\{\frac{1}{\Delta_{n,T}}(X_{(i+1)\Delta_{n,T}} - X_{i\Delta_{n,T}})^2 - \theta^{\sigma^2}\right\}^2, \tag{84}$$

respectively. Specifically, as always, the N-W estimates of drift and diffusion function fit a constant line to data in vicinity of the level a. Alternatively, one might fit a polynomial locally and minimize the criteria

$$\sum_{i=1}^{n-1} \mathbf{K}\left(\frac{X_{i\Delta_{n,T}} - a}{h_{n,T}}\right)\left\{\frac{1}{\Delta_{n,T}}(X_{(i+1)\Delta_{n,T}} - X_{i\Delta n,T}) - \sum_{s=0}^{r} \theta^{\mu}_s (X_{i\Delta_{n,T}} - a)^s\right\}^2 \tag{85}$$

and

$$\sum_{i=1}^{n-1} \mathbf{K}\left(\frac{X_{i\Delta_{n,T}} - a}{h_{n,T}}\right)\left\{\frac{1}{\Delta_{n,T}}(X_{(i+1)\Delta_{n,T}} - X_{i\Delta n,T})^2 - \sum_{s=0}^{r} \theta^{\sigma^2}_s (X_{i\Delta_{n,T}} - a)^s\right\}^2 \tag{86}$$

with respect to $\boldsymbol{\theta}^{\mu} = (\theta^{\mu}_0, \theta^{\mu}_1, ..., \theta^{\mu}_r)^{\top}$ and $\boldsymbol{\theta}^{\sigma^2} = \left(\theta^{\sigma^2}_0, \theta^{\sigma^2}_1, ..., \theta^{\sigma^2}_r\right)^{\top}$ for all levels a. A simple argument based on Taylor expansions around a suggests that the proper estimates of $\mu(a)$ and $\sigma^2(a)$ are now the first components ($\widehat{\theta}^{\mu}_0$ and $\widehat{\theta}^{\sigma^2}_0$) of the estimated vectors $\widehat{\boldsymbol{\theta}}^{\mu}$ and $\widehat{\boldsymbol{\theta}}^{\sigma^2}$. The remaining components are estimates of the (standardized) derivatives of the functions of interest (provided these derivatives exist, of course).

In the case of recurrent SDPs of the kind analyzed in this review, this approach was suggested by Moloche (2004a) following classical work by Fan (1992, 1993) and Fan and Gijbels (1996) in nonparametric regression analysis for discrete-time series (see, also, Pagan and Ullah, 1999, page 93, for discussions). In the stationary case, local polynomial methods are employed by Fan and Zhang (2003).[12] Importantly, the estimated vectors $\widehat{\boldsymbol{\theta}}^{\mu}$ and $\widehat{\boldsymbol{\theta}}^{\sigma^2}$

[12] In particular, Fan and Zhang (2003) apply local polynomial methods to Stanton's kth-order approximations to drift and diffusion function (Stanton, 1997). Write

$$\begin{aligned}\mathbf{E}_t\{f(X_{t+\Delta}, t+\Delta)\} &= f(X_t,t) + \mathsf{L} f(X_t,t)\Delta + \frac{1}{2}\mathsf{L}^2 f(X_t,t)\Delta^2 + ... + \\ &\quad + \frac{1}{k!}\mathsf{L}^k f(X_t,t)\Delta^k + O(\Delta^{k+1}),\end{aligned}$$

where L is the infinitesimal generator in Eq. (20). Given discrete observations sampled at multiples of Δ (and weights $c_{k,j}$ with $j = 1, ..., k$), we can write can be expressed in the form of regression estimates since the criteria (85) and (86) may be readily interpreted in terms of classical weighted least-squares problems. As always, we are explicitly only about the drift case but similar observations apply to the diffusion function. Write

$$\mathbf{X}_{n,T}(a) = \begin{bmatrix} 1 & (X_{\Delta_{n,T}} - a) & ... & (X_{\Delta_{n,T}} - a)^r \\ ... & ... & ... & ... \\ 1 & (X_{(n-1)\Delta_{n,T}} - a) & ... & (X_{(n-1)\Delta_{n,T}} - a)^r \end{bmatrix}, \tag{87}$$

$$\mathbf{y}_{n,T} = \begin{bmatrix} \frac{1}{\Delta_{n,T}}(X_{2\Delta_{n,T}} - X_{\Delta_{n,T}}) \\ ... \\ \frac{1}{\Delta_{n,T}}(X_{n\Delta_{n,T}} - X_{(n-1)\Delta_{n,T}}) \end{bmatrix}, \tag{88}$$

and

$$\mathbf{W}_{n,T}(a) = diag\left(\frac{\Delta_{n,T}}{h_{n,T}}\mathbf{K}\left(\frac{X_{\Delta_{n,T}} - a}{h_{n,T}}\right), ..., \frac{\Delta_{n,T}}{h_{n,T}}\mathbf{K}\left(\frac{X_{(n-1)\Delta_{n,T}} - a}{h_{n,T}}\right)\right). \tag{89}$$

The point-wise drift estimator $\widehat{\theta}_0^{\mu}$ is the first component of the $(r+1)$-vector

$$\widehat{\boldsymbol{\theta}}^{\mu} = (\mathbf{X}_{n,T}^{\top}(a)\mathbf{W}_{n,T}(a)\mathbf{X}_{n,T}(a))^{-1}\mathbf{X}_{n,T}^{\top}(a)\mathbf{W}_{n,T}(a)\mathbf{y}_{n,T}. \tag{90}$$

Moloche (2004a) shows that under the same conditions on the bandwidth as in Theorem 3.5 and provided that the same sampling scheme as in Section 2 is adopted, the estimate $\widehat{\theta}_0^{\mu}$ converges to the true function with probability one and is (mixed) normally distributed in the limit. The local linear case ($r = 1$) is particularly relevant. The drift estimator can be conveniently expressed as a weighted N-W kernel estimator

$$\widehat{\mu}_{(n,T)}^{ll}(a) = \frac{1}{\Delta_{n,T}} \frac{\sum_{i=1}^{n-1} \mathbf{w}_i^{ll}(a, h_{n,T})\mathbf{K}\left(\frac{X_{i\Delta_{n,T}} - a}{h_{n,T}}\right)(X_{(i+1)\Delta_{n,T}} - X_{i\Delta_{n,T}})}{\sum_{i=1}^{n-1} \mathbf{w}_i^{ll}(a, h_{n,T})\mathbf{K}\left(\frac{X_{i\Delta_{n,T}} - a}{h_{n,T}}\right)} \tag{91}$$

$$\begin{aligned} \frac{1}{\Delta}\sum_{j=1}^{k} c_{k,j}\mathbf{E}_t\left\{f(X_{t+j\Delta}, t+j\Delta) - f(X_t, t)\right\} &= \left\{\sum_{j=1}^{k} jc_{k,j}\right\}\mathsf{L}f(X_t,t) + \left\{\sum_{j=1}^{k} j^2 c_{k,j}\right\}\frac{\mathsf{L}^2 f(X_t,t)}{2}\Delta \\ &+ ... + \left\{\sum_{j=1}^{k} j^{k+1}c_{k,j}\right\}\frac{\mathsf{L}^{k+1}f(X_t,t)}{(k+1)!}\Delta^k + O(\Delta^{k+1}). \end{aligned}$$

Now consider the drift case, i.e., $f(X_t, t) = X_t$. Clearly, $\mathsf{L}f(X_t, t) = \mu(X_t)$. Thus, if the weights $c_{k,j}$ are chosen so that $\sum_{j=1}^{k} jc_{k,j} = 1$ and $\sum_{j=1}^{k} j^p c_{k,j} = 0$ for all $2 \leq p \leq k$, then

$$\frac{1}{\Delta}\sum_{j=1}^{k} c_{k,j}\mathbf{E}_t\left\{X_{t+j\Delta} - X_t\right\} = \mu(X_t) + O(\Delta^k).$$

N-W kernel estimates can now be applied to data sampled at different frequencies (controlled by j) and can be appropriately weighed in order to estimate $\mu(.)$ (Stanton, 1997). Clearly, the case $k = 1$ corresponds to Eq. (46) above. Fan and Zhang (2003) use local polynomial estimates to show that a large k can be beneficial in terms of asymptotic bias but may translate into an exponentially-growing limiting variance.

with $\mathbf{w}_i^{ll} = \Xi_{n,2} - (X_{i\Delta_{n,T}} - a)\Xi_{n,1}$, where $\Xi_{n,k} = \frac{1}{h_{n,T}} \sum_{i=1}^{n-1} (X_{i\Delta_{n,T}} - a)^k \mathbf{K}\left(\frac{X_{i\Delta_{n,T}}-a}{h_{n,T}}\right)$ with $k = 1, 2$. Its asymptotic variance and bias have the form

$$\frac{\mathbf{K}_2 \sigma^2(a)}{h_{n,T} \widehat{\widehat{L}}_X(T,a)}, \tag{92}$$

and

$$h_{n,T}^2 \mathbf{K}_1 \frac{1}{2} \mu^{''}(a). \tag{93}$$

Expressions (92) and (93) should now be compared to the corresponding quantities for the N-W kernel estimates discussed earlier, namely (58) and (59) above. The comparison is standard and we refer the interested reader to the original work of Fan (1992) and the review of Pagan and Ullah (1999, pages 104-106) for details. Here we simply stress that the variances are the same, but the biases are different. In particular, the bias of the local linear estimator does not depend on the ratio $\frac{m^{'}(.)}{m(.)}$ and on the first derivative of μ at a and is "design adaptive" in the sense of Fan (1992). Similar expressions hold for the diffusion function estimator $\widehat{\theta}_0^{\sigma^2}$ (Moloche, 2004a).

3.3.3 Finite Sample Refinements

Local linear methods have favorable bias properties but may not guarantee positivity when positivity is required, as is the case for diffusion function estimation. To this extent, Xu (2008) suggests a *re-weighted* N-W diffusion estimator which is asymptotically equivalent (in terms of MSE properties) to the local linear estimator while retaining the non-negativity features of the classical local constant (or N-W) kernel estimator. Write

$$\widehat{\sigma}_{(n,T)}^{2(rNW)}(a) = \frac{1}{\Delta_{n,T}} \frac{\sum_{i=1}^{n-1} \mathbf{w}_i^{rNW}(a, h_{n,T}) \mathbf{K}\left(\frac{X_{i\Delta_{n,T}}-a}{h_{n,T}}\right) (X_{(i+1)\Delta_{n,T}} - X_{i\Delta_{n,T}})^2}{\sum_{i=1}^{n-1} \mathbf{w}_i^{rNW}(a, h_{n,T}) \mathbf{K}\left(\frac{X_{i\Delta_{n,T}}-a}{h_{n,T}}\right)}, \tag{94}$$

where the weights $\left\{\mathbf{w}_i^{rNW}(a, h_{n,T})\right\}$ are such that $\mathbf{w}_i \geq 0$, $\sum_{i=1}^{n-1} \mathbf{w}_i = 1$, and solve

$$\left\{\mathbf{w}_i^{rNW}(a, h_{n,T})\right\} = \max_{\{\mathbf{w}_i\}} \sum_{i=1}^{n-1} \log((n-1)\mathbf{w}_i), \tag{95}$$

under

$$\frac{1}{h_{n,T}} \sum_{i=1}^{n-1} \mathbf{w}_i (X_{i\Delta_{n,T}} - a) \mathbf{K}\left(\frac{X_{i\Delta_{n,T}} - a}{h_{n,T}}\right) = 0. \tag{96}$$

As emphasized by Xu (2008), the restriction in Eq. (96) is motivated by local linear estimation. It is, in fact, easily satisfied by $\mathbf{w}_i^{ll}$. Just like in local linear estimation, this is the restriction which yields a "design adaptive" bias component

$$h_{n,T}^2 \mathbf{K}_1 \frac{1}{2} \left(\sigma^2(a)\right)^{''} \tag{97}$$

and the same asymptotic variance as earlier. The positivity of the weights, of course, guarantees positivity of the final local estimates. For further discussions about the form of the weights and their derivation, we refer to Xu (2008). In particular, Xu (2008) discusses the interpretation of the criterion in terms of empirical likelihood. Inference for SDPs based on empirical likelihood is studied, e.g., in Chen *et al.* (2007) and Xu (2007) and we do not expand on it here.

While for the data routinely used in continuous-time econometrics the assumption of a limiting, vanishing distance between discretely-sampled observations ($\Delta_{n,T} \to 0$) represents an empirically valid asymptotic design, finite sample adjustments might sometime be important. Nicolau (2003) studies the finite-sample bias properties of the N-W diffusion estimator (in Eq. (49)) for a fixed distance between adjacent observations. He shows that if $\Delta_{n,T} = \Delta$ fixed, $T \to \infty$, $h_{n,T} \to 0$ and $nh_{n,T} \to \infty$ as $n \to \infty$,

$$\widehat{\sigma}^2_{(n,T)}(a) \xrightarrow{p} \sigma^2(a) + \mu^2(a)\Delta + \Pi(a)\Delta + O(\Delta^2), \tag{98}$$

with

$$\Pi(a) = \sigma^2(a)\mu^{'}(a) + \mu(a)\sigma(a)\sigma^{'}(a) + \frac{1}{2}\sigma^2(a)\left(\sigma^{'}(a)\right)^2 + \frac{1}{2}\sigma^3(a)\sigma^{''}(a). \tag{99}$$

Hence, of course, the estimator is biased for a fixed Δ. However, the first bias component ($\mu^2(a)\Delta$) may be eliminated asymptotically by using a (feasible) bias correction. One could, for instance, compute $\widehat{\sigma}^2_{(n,T)}(a)$ by averaging terms $(X_{(i+1)\Delta_{n,T}} - X_{i\Delta_{n,T}} - \widehat{\mu}_{(n,T)}(a)\Delta_{n,T})^2$, where $\widehat{\mu}_{(n,T)}(a)$ is the drift estimator in Eq. (48), rather than the traditional terms $(X_{(i+1)\Delta_{n,T}} - X_{i\Delta_{n,T}})^2$. While the asymptotic distribution of the resulting estimator for an increasing sampling frequency ($\Delta_{n,T} = \Delta \to 0$) is of course identical to that laid out in Theorem 3.6, the finite sample adjustment might be useful. In effect, it might work particularly well at levels a corresponding to large (in absolute terms) drift values (in regions where the degree of drift-induced mean-reversion is substantial). Reno' (2006) provides further discussions.

3.4 Using Nonparametric Information to Estimate and Test Parametric Models for SDPs

It is natural to use the information contained in the nonparametric estimates to design more accurate parametric models and test parametric assumptions. BP (2007) discuss a simple (semi-)parametric procedure to estimate potentially nonstationary diffusions which overcomes the usual inference problems posed by the unavailability of a closed-form expression for the transition density of the underlying process and does not require simulations.

They consider a parametric class $(\boldsymbol{\theta}^{\mu}, \boldsymbol{\theta}^{\sigma}) = \boldsymbol{\theta} \in \Theta$ for the underlying SDP and compute the parameters of interest as

$$\begin{aligned} \widehat{\boldsymbol{\theta}}^{\mu}_{n,T} \quad : &= \underset{\boldsymbol{\theta}^{\mu} \in \Theta^{\mu} \subset \Theta}{\arg\min}\; Q^{\mu}_{n,T} \\ &= \underset{\boldsymbol{\theta}^{\mu} \in \Theta^{\mu} \subset \Theta}{\arg\min} \frac{\overline{T}}{n} \sum_{i=1}^{n} \left(\widehat{\mu}_{(n,T)} \left(X_{i\Delta_{n,\overline{T}}} \right) - \mu \left(X_{i\Delta_{n,\overline{T}}}, \boldsymbol{\theta}^{\mu} \right) \right)^2, \end{aligned} \tag{100}$$

and

$$\begin{aligned} \widehat{\boldsymbol{\theta}}^{\sigma^2}_{n,T} \quad : &= \underset{\boldsymbol{\theta}^{\sigma^2} \in \Theta^{\sigma^2} \subset \Theta}{\arg\min}\; Q^{\sigma^2}_{n,T} \\ &= \underset{\boldsymbol{\theta}^{\sigma^2} \in \Theta^{\sigma^2} \subset \Theta}{\arg\min} \frac{\overline{T}}{n} \sum_{i=1}^{n} \left(\widehat{\sigma}^2_{(n,T)} \left(X_{i\Delta_{n,\overline{T}}} \right) - \sigma^2 \left(X_{i\Delta_{n,\overline{T}}}, \boldsymbol{\theta}^{\sigma^2} \right) \right)^2, \end{aligned} \tag{101}$$

where $\widehat{\mu}_{(n,T)} \left(X_{i\Delta_{n,\overline{T}}} \right)$ and $\widehat{\sigma}^2_{(n,T)} \left(X_{i\Delta_{n,\overline{T}}} \right)$ are functional estimates (defined over an enlarging span of data for consistency – see Subsection 3.2) of the N-W type (c.f. (48) and (49) above) at the i-th observation. The parameter values are chosen so that the average squared distance between the nonparametric curves at the sampled points and the adopted parametric specification is minimized. The asymptotic distributions of the parameter estimates are (variance mixtures of) normals and can be readily interpreted on the basis of well-known results for conventional nonlinear least-squares problems. Nonetheless, the integrals that appear in the limiting variances are not integrals with respect to probability measures (i.e. expectations) but integrals with respect to local times (i.e., occupation integrals, c.f. (33)) due to the generality of the approach in the present context. By virtue of the averaging, the rates of convergence of the parameter estimates are faster than the rates of convergence of the functional estimators used to define (100) and (101) above. This is, of course, a typical result in semiparametric problems (see, e.g., Andrews, 1989).

Apparently, the above criteria can be employed to test alternative parametric assumptions about the functions of interest. Consider the drift case. Assume one wishes to test the hypotheses $H_0 : \mu_0(x) = \mu(x, \boldsymbol{\theta}^{\mu})$ against $H_1 : \mu_0(x) \neq \mu(x, \boldsymbol{\theta}^{\mu})$. Provided a consistent (under the null) parametric estimate of $\boldsymbol{\theta}^{\mu}$, $\widetilde{\boldsymbol{\theta}}^{\mu}_{n,T}$ say, is obtained (the value $\widehat{\boldsymbol{\theta}}^{\mu}_{n,T}$ which minimizes (100) is, of course, a viable option) and the distribution of $\widehat{Q}^{\mu}_{n,T}(\widetilde{\boldsymbol{\theta}}^{\mu}_{n,T})$ is derived under the null, standard methods can be utilized to construct a consistent test. The use of nonparametric information to test parametric models based on the minimization of average squared errors like (100) and (101) above has a long history in hypothesis testing about density functions. Important early references in discrete-time are Bickel and Rosenblatt (1973) and Rosenblatt (1975). More recently, Aït-Sahalia (1996) has applied the idea to the study of stationary scalar diffusion models for the short-term interest rate process. Corradi and

White (1999) focus on the infinitesimal second moment over a fixed span of data and can, therefore, allow for transient dynamics. Relying on the informational content of the process' transition density, Hong and Li (2003) provide specification tests for both the drift and the diffusion of a stationary diffusion process. Empirical distribution function-based tests for stationary scalar and multivariate diffusion processes are proposed in Corradi and Swanson (2005).

To the authors' knowledge little work exists on parametric inference for null-recurrent diffusions. In addition to BP(2007), a recent contribution is the work by Höpfner and Kutoyants (2001) who discuss a method of inference for the parameter θ in the SDP

$$dX_t = \theta \frac{X_t}{1+X_t^2} dt + \sigma dB_t, \tag{102}$$

where $\theta \in \Theta = \left(-\frac{\sigma^2}{2}, \frac{\sigma^2}{2}\right)$. As they show, θ is the parameter determining the speed of divergence of additive integrable functionals of the process (in the sense discussed in Section 5 - Theorem 5.1) and Θ is the maximal open interval over which the process is null recurrent. Given knowledge of the diffusion function σ and availability of a continuum of observations, parametric estimation is conducted by maximum likelihood through arguments based on measure changes.

3.5 Time-Inhomogeneous SDPs

Allowing for time-dependence in the infinitesimal first and second moment may be empirically useful. However, while time-inhomogeneous diffusions have received some emphasis in the mathematical finance literature (see, e.g., Black *et al.*, 1990, Heath *et al.*, 1992, and the references therein), little work exists on their inference. Re-write Eq. (14) as

$$dX_t = \mu(X_t, t)dt + \sigma(X_t, t)dB_t. \tag{103}$$

The drift and diffusion function now depend on the state as well as on time. Fan *et al.* (2003) parametrize them as functions of X_t with time-varying parameters before localizing in time. Assume, as they do, that $\mu(X_t, t) = \alpha_0(t) + \alpha_1(t)X_t$. Given discretely-sampled observations X_{t_i} for $i = 1, ..., n$ with $\Delta_i = t_{i+1} - t_i$, a natural local least-squares criterion minimizes

$$\frac{1}{h}\sum_{i=1}^{n-1} \mathbf{K}\left(\frac{t_i - \widetilde{t}}{h}\right)\left\{\frac{X_{t_{i+1}} - X_{t_i}}{\Delta_i} - (\theta_0 + \theta_1 X_{t_i})\right\} \tag{104}$$

with respect to θ_0 and θ_1. Clearly, both estimates depend on $\widetilde{t}$, thereby giving $\widehat{\theta}_0 = \widehat{\alpha}_0(\widetilde{t})$ and $\widehat{\theta}_1 = \widehat{\alpha}_1(\widetilde{t})$. An analogous procedure can be applied to diffusion estimation under a similar parametrization. Fan *et al.* (2003), for instance, assume a constant elasticity-of-variance diffusion with time-varying parameters and write $\sigma(X_t, t) = \beta_0(t)X^{\beta_1(t)}$. Noting that the ratio

$$Y_{t_i} = \frac{X_{t_{i+1}} - X_{t_i} - (\widehat{\alpha}_0(t_i) + \widehat{\alpha}_1(t_i)X_{t_i})\,\Delta_i}{\sqrt{\Delta_i}} \approx \beta_0(t) X_{t_i}^{\beta_1(t)} \varepsilon_{t_i} \tag{105}$$

is approximately (conditionally) Gaussian, they optimize (for all $\widetilde{t}$) the local pseudo log-likelihood

$$-\frac{1}{2h}\sum_{i=1}^{n-1} \mathbf{K}\left(\frac{t_i - \widetilde{t}}{h}\right)\left\{\left(\log(\vartheta_0^2 X_{t_i}^{2\vartheta_1}\right) + \frac{Y_{t_i}^2}{\vartheta_0^2 X_{t_i}^{2\vartheta_1}}\right\} \tag{106}$$

to obtain $\widehat{\vartheta}_0 = \widehat{\beta}_0(\widetilde{t})$ and $\widehat{\vartheta}_1 = \widehat{\beta}_1(\widetilde{t})$. Alternatively, writing

$$\log(Y_{t_i}^2) \approx \log(\beta_0^2(t_i)) + \beta_1(t_i)\log(X_{t_i}^2) + \log(\varepsilon_{t_i}^2), \tag{107}$$

a similar (local) least-squares procedure as in Eq. (104) may be applied to the diffusion estimator. While notions of consistency for the resulting estimates under suitable sampling schemes have not been established yet, the methods are of course suggestive of the usefulness of kernel-based approaches to capture time-varying dynamics in parameters.

A fundamental class of models allowing for stochastic time-variation in the infinitesimal second moment is the family of stochastic volatility models. Consider Eq. (103) and write $\sigma(X_t, t) = \sigma(t)$. Assume the dynamics of $\sigma(t)$ are driven by an homogeneous stochastic differential equation and are recurrent. The next section provides a concise application of the ideas laid out in this section to stochastic volatility modelling in continuous time. For interesting, recent work focused on testing the null hypothesis $H_0 : \sigma^2(X_t, t) = \sigma^2(X_t)$ against the alternative $H_A : \sigma^2(X_t, t) =$ SDP not measurable with respect to the filtration generated by X_t, we refer the reader to Corradi and Distaso (2007). In what follows, we work under Corradi and Distaso's alternative hypothesis.

3.6 An Empirical Application: Stochastic Volatility

The recent literature on volatility estimation by virtue of high-frequency (intra-daily) asset price data has provided a set of tools to identify daily variance without the need for filtering using low-frequency asset returns. These high-frequency variance estimates may be put to work to understand variance dynamics from a new perspective. Using (i) intra-daily asset price data to generate *spot* variance estimates and (ii) N-W kernel estimates of the spot variance drift and diffusion (as suggested by BP, 2003), Bandi and Reno' (2008) (BR, henceforth) and Kanaya and Kristensen (2008) discuss a nonparametric theory of (continuous-time) stochastic volatility estimation (see, also, Comte *et al.*, 2007, for an alternative approach). The preliminary spot variance estimates in Kanaya and Kristensen (2008) are local (in time) averages of realized variance estimates as suggested in Kristensen (2007). In BR (2008) they are local (in time) averages of a family of robust (to market microstructure noise or jumps in returns) integrated variance estimates (for which an asymptotic theory of inference is provided in BR, 2008). Both BR (2008) and Kanaya and

Kristensen (2008) discuss conditions (to be added to those in Theorem 3.5 and Theorem 3.6) under which the estimation error introduced by the preliminary spot variance estimates is asymptotically negligible.

In order to briefly illustrate the methods in the SDP case, we follow BR (2008) in this review. Specifically, we estimate the spot variance of the S&P 500 index returns for all days between January 2, 1998, and March 31, 2006, by applying the two-scale estimator of Zhang et al. (2005) to intra-daily SPY returns (see BR, 2008, for details).[13] Figure 1 (a) represents the sojourn time of the spot variance estimates along with its asymptotic confidence bands (c.f. Theorem 3.2). Spot variance is expressed (on the horizontal axis, for instance) in daily terms and is multiplied by 10,000 for consistency with S&P 500 returns expressed in percentage terms. The sojourn time has a peak around 1 (corresponding to a volatility of annual S&P 500 returns equal to 15%). More generally, the stochastic variance process makes most of its visit at levels between about 0.3 and 1.5, namely for a volatility of annual market returns between about 8.5% and 19.5%. In this range, we expect the drift and diffusion estimates (in Figure 1 (b) and (c)) to be more precisely estimated as implied by point-wise asymptotic confidence bands whose width is inversely related to the number of visit to each spatial point. The drift is largely positive over the relevant variance range. The diffusion is a monotonically increasing and nonlinear function of the variance level. We do not dwell on these two functions here. We simply point out that the magnitude of the estimated drift in this subsection may be largely induced by the presence of positive jumps in the variance process affecting the infinitesimal first moment's estimates (see Eq. (147) below). Similarly, the magnitude of the diffusion estimates might be induced by infinitesimal second moment estimates which, in the presence of variance jumps, comprise genuinely diffusive volatility as well as the second moment of the discontinuous variance component (see Eq. (148) below). Consistent with these observations and the analysis in BR (2008), in Subsection 4.3 we will discuss a functional jump-diffusion model with exponential jump sizes for the market spot variance. This specification will represent a considerably superior modeling alternative to the SDP discussed here.

4 Scalar Jump-Diffusion Processes (SJDPs)

Throughout this section we model a time-series X_t as the solution of a stochastic differential equation with infrequent Poisson jumps. We assume that the jumps are bounded (i.e., $\sup_t |\Delta X_t| \leq C < \infty$ almost surely, where C is a non-random constant) and occur with conditional intensity $\lambda(.)$ (i.e., $\lambda(a)dt$ is the infinitesimal probability of a jump at the level a). The impact of a jump is given by the function $g(., y)$ whose arguments are the level

[13] SPY is the ticker symbol for the Standard and Poor's depository receipts (also known as Spiders). SPYs are shares in a trust which owns stocks in the same proportion as that found in the S&P 500 index. Importantly, they trade like a stock at approximately one-tenth of the level of the index. Thus, because changes in SPY value reflect changes in market value, SPY volatility reflects market volatility.

of the process and a generic random variable y which we assume to be endowed with the stationary probability measure $\Gamma(.)$. Specifically,

$$\Delta X_t = X_t - X_{t-} = \int_Y g(X_{t-}, y) N(dt, dy) = dJ_t, \tag{108}$$

where

$$N_t^{\Phi} = \sum_{j=1} \mathbf{1}_{[\tau_j \leq t, y_{\tau_j} \in \Phi]} \tag{109}$$

is, given a set Φ, a Poisson counting measure with stationary and independent increments (see, e.g., Protter, 1995). Write

$$\begin{aligned} dX_t &= \left[\mu(X_{t-}) - \lambda(X_{t-}) \int_Y g(X_{t-}, y)\Gamma(dy)\right] dt + \sigma(X_{t-})dB_t + dJ_t \\ &= \left[\mu(X_{t-}) - \lambda(X_{t-})\mathbf{E}_Y[g(X_{t-}, y)]\right] dt + \sigma(X_{t-})dB_t + dJ_t, \end{aligned} \tag{110}$$

where the standard Brownian motion $\{B_t : t \geq 0\}$ and the jump process $\{J_t : t \geq 0\}$ are assumed to be independent. The initial condition $X_0 = \overline{X}$ belongs to L^2 and is taken to be independent of both B_t and J_t.

The functions $\mu(.)$ and $\sigma(.)$ have a similar interpretation as in scalar diffusion models (c.f. Section 3). Nonetheless, due to the presence of a discontinuous jump-component $dJ_t = \Delta X_t$, the sample path of X_t fails to be continuous in the state space as in the case of standard SDPs despite being right continuous with left limits (*càdlàg*).

Conditions 1 through 4 (1 through 4b) below are imposed on the model (BN, 2003).

(1) *The functions $\mu(.)$, $\sigma(.)$, $g(., y)$ and $\lambda(.)$ are time-homogeneous and $\mathfrak{B}$-measurable on $\mathfrak{D} = (l, u)$ with $-\infty \leq l < u \leq \infty$, where $\mathfrak{B}$ is the σ-field generated by Borel sets on $\mathfrak{D}$. They satisfy local Lipschitz and growth conditions. Thus, for every compact subset J of the domain of the process, there exist constants C_1 and C_2 so that, for all x and z in J,*

$$|\mu(x) - \mu(z)| + |\sigma(x) - \sigma(z)| + \lambda(x) \int_Y |g(x, y) - g(z, y)|\Gamma(dy) \leq C_1|x - z|, \tag{111}$$

and

$$|\mu(x)| + |\sigma(x)| + \lambda(x) \int_Y |g(x, y)|\Gamma(dy) \leq C_2\{1 + |x|\}. \tag{112}$$

(2) *For a given $\alpha > 2$, there exists a constant C_3 such that*

$$\lambda(x) \int_Y |g(x, y)|^{\alpha}\Gamma(dy) \leq C_3\{1 + |x|^{\alpha}\}. \tag{113}$$

(3) $\lambda(.) > 0$ *and* $\sigma^2(.) > 0$ *on* $\mathfrak{D}$.

(4) *(Null recurrence) Define the second order elliptic operator* $\mathfrak{L}$ *and the integro-differential operator* $\mathfrak{A}$ *of the continuous and discontinuous portions of the solution to (110) above as*

$$\mathfrak{L}\varphi(.) = \varphi'(.)\mu(.) + \frac{1}{2}\varphi''(.)\sigma^2(.) \tag{114}$$

and

$$\mathfrak{A}\varphi(.) = \lambda(.)\int_Y \left[\varphi(. + g(., y)) - \varphi(.) - \varphi'(.)g(., y)\right]\Gamma(dy), \tag{115}$$

respectively. Assume Φ *is a Borel measurable and bounded function on the closure* $\overline{A}$. *The exterior Dirichlet problem, i.e.*

$$(\mathfrak{L} + \mathfrak{A})e \;=\; 0 \qquad a.e. \qquad in\ \mathfrak{D} \setminus \overline{A} \tag{116}$$

$$e \;=\; \Phi \qquad a.e. \qquad in\ \overline{A} \tag{117}$$

admits a unique bounded solution $e(x)$ *(see, e.g., Menaldi and Robin, 1999).*

(4bis) *(Positive recurrence) The exterior Dirichlet problem, i.e.*

$$-(\mathfrak{L} + \mathfrak{A})e \;=\; f \qquad a.e. \qquad in\ \mathfrak{D} \setminus \overline{A} \tag{118}$$

$$e \;=\; 0 \qquad a.e. \qquad in\ \overline{A} \tag{119}$$

admits a unique bounded solution $e(x)$ *(see, e.g., Menaldi and Robin, 1999).*

Under Assumptions 1 through 4 (4b) the SJDP (110) has a strong solution which is unique and null recurrent (positive recurrent). In particular, the càdlàg process X_t satisfies

$$X_t = \overline{X} + \int_0^t \mu(X_{s-})ds + \int_0^t \sigma(X_{s-})dB_s + \int_{0+}^t \int_Y g(X_{s-}, y)\overline{\nu}(ds, dy) \tag{120}$$

where

$$\overline{\nu}(ds, dy) \;=\; N(ds, dy) - \nu(X_{s-}, dy)ds \tag{121}$$

$$\;=\; N(ds, dy) - \lambda(X_{s-})\Gamma(dy)ds \tag{122}$$

is a compensated random measure and the notation $\int_{0+}^t = \int_{(0,t]}$ denotes the integral over the half open interval. It is noted that

$$\int_{0+}^t \int_Y g(X_{s-}, y)\overline{\nu}(ds, dy) \tag{123}$$

represents the conditional variation between $0+$ and t in the path of the process due to discontinuous jumps of random size y (with impact $g(., y)$) net of its expected conditional magnitude at $0+$. The model is defined as "compensated" by virtue of the presence of the term $\lambda(X_t)\mathbf{E}_Y[g(X_t, y)]$ denoting the conditional mean of the jump part. Its presence ensures that the jump component is a martingale, thereby making the solution to Eq. (110) a semimartingale. The semimartingale property, which is trivially satisfied by standard SDPs of the types analyzed earlier, makes SJDPs of the kind reviewed here attractive for modelling purposes in continuous-time finance. As is well-known, in the case of price processes, this property implies the existence of an equivalent martingale measure under which the process is a (local) martingale and absence of arbitrage in the spaces that preclude doubling strategies (Duffie, 1996).

Given Assumptions (1), (2), and (3), the infinitesimal conditional moments of the changes in the solution to (110) above can be written in terms of the functions $\mu(.)$, $\sigma(.)$, $g(.,.)$ and $\lambda(.)$ (Gikhman and Skorohod, 1972), i.e.,

$$M^1(a) = \lim_{t \to 0} \frac{1}{t}\mathbf{E}^a[X_t - a] = \mu(a), \tag{124}$$

$$M^2(a) = \lim_{t \to 0} \frac{1}{t}\mathbf{E}[(X_t - a)^2] = \sigma^2(a) + \lambda(a)\mathbf{E}_Y[g^2(a, y)], \tag{125}$$

$$M^k(a) = \lim_{t \to 0} \frac{1}{t}\mathbf{E}[(X_t - a)^k] = \lambda(a)\mathbf{E}_Y[g^k(a, y)] \ \forall k > 2, \tag{126}$$

for a generic $a \in \mathfrak{D}$. Specifically, formulae (124) through (126) suggest that the conditional infinitesimal moments of order higher than two contain important information about the intensity of the jumps and the distribution of the jump component. These moments are of course zero in the case of SDPs. Similarly to the case of SDPs, however, the first and second infinitesimal moments may be used to identify the drift and the diffusive volatility (given the estimated features of the jumps). These observations led Johannes (2004) to suggest nonparametric estimates of the infinitesimal moments and a procedure to extract the parameters and functions of interest from the estimated moments. We will be more accurate in the sequel. For now it suffices to point out that the $M^k(.)$s (for $k \geq 1$) will be our object of econometric interest.

We now turn to generalized density estimation for processes that are possibly nonstationary solutions to stochastic differential equations with infrequent jumps such as (110) above.

4.1 Generalized Density Estimation for SJDPs

The theory of local times for càdlàg semimartingales is well established in the stochastic process literature. We refer the reader to Protter (1995) for a thorough treatment. Here, consonant with our discussion in Section 3, we review some basic notions that will serve the purpose of illustrating the role that estimated local time may play as a descriptive tool for recurrent SJDPs. Assume X_t is a càdlàg semimartingale, then its sojourn time at T and a can be written as

$$L_X(T,a) = \lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\int_0^T \mathbf{1}_{[a,\ a+\varepsilon[}(X_s)d[X]_s^c, \tag{127}$$

where $[X]_t^c$ is the continuous part of the quadratic variation process of X_t, namely the non-decreasing process defined as

$$\begin{aligned}[X]_t^c &= [X]_t - \sum_{0<s\leq t}(\Delta X_s)^2 - X_0^2 &(128)\\ &= [X]_t - \sum_{0\leq s\leq t}(\Delta X_s)^2 &(129)\end{aligned}$$

with

$$[X]_0^c = 0, \tag{130}$$

(see, e.g., Yor, 1978). The interpretation is standard. Formula (127) represents the amount of time, in information units, that the càdlàg semimartingale spends in an arbitrarily small right neighborhood of a between time 0 and time T. Differently put, local time is an occupation density relative to the random clock $d[X]_s^c$.

A corresponding notion in calendar units can be easily obtained after noticing that $d[X]_t^c = \sigma^2(X_{t-})dt$ in the presence of a process whose dynamics are driven by (110) above. In fact,

$$\overline{L}_X(T,a) = \frac{1}{\sigma^2(a)}L_X(T,a) \tag{131}$$

is the chronological counterpart to (127) for SJDPs of the type analyzed here. As in the case of standard SDPs, the chronological sojourn time (131) can be interpreted as a version of the Radon-Nikodym derivative of the occupation measure with respect to the Lebesgue measure, i.e.,

$$\eta_A^T = \int_0^T \mathbf{1}_{\{X_s\in A\}}ds = \int_A \overline{L}_X(T,a)da, \qquad \forall A \in \mathfrak{B}(\mathfrak{D}), \tag{132}$$

and, as pointed out above, is an occupation density. Additionally, formula (132) readily leads to

$$[X]_t^c = \int_{-\infty}^{\infty} L_X(t,a)da \tag{133}$$

which can be interpreted as a decomposition of variance (c.f. (29)), coherently with our remarks in Section 3.

We now turn to estimation. As earlier when dealing with SDPs, there is a natural way to identify (131) using a sample of observations generated from (110), namely we can perform density-like kernel estimation as in (34). We proceed under the same sampling scheme as in Section 2.

Theorem 4.1 *Assume X_t is the solution to (110). If $h_{n,\overline{T}} \to 0$ as $n \to \infty$ with $T = \overline{T}$ in such a way as to guarantee that $\frac{1}{h_{n,\overline{T}}}(\Delta_{n,\overline{T}} \log(1/\Delta_{n,\overline{T}})^{1/2} = o(1)$, then*

$$\widehat{\overline{L}}_X(\overline{T}, a) = \frac{\Delta_{n,\overline{T}}}{h_{n,\overline{T}}} \sum_{i=1}^{n} \mathbf{K}\left(\frac{X_{i\Delta_{n,\overline{T}}} - a}{h_{n,\overline{T}}}\right) \overset{a.s.}{\rightarrow} \overline{L}_X(\overline{T}, a) \qquad \forall a \in \mathfrak{D}. \tag{134}$$

Proof *See BN (2003).*

Coherently with Theorem 3.1 in Section 3, Theorem 4.1 justifies using density-like kernel estimates as descriptive tools for SJDPs even in the presence of processes which might not possess a time-invariant stationary distribution. All we need to do is modify their interpretation since, in general, they cannot be regarded as estimates of the process' stationary density and recognize instead the role played by local time in characterizing the locational features of the process.

We conclude this subsection with two observations. First, recurrence implies divergence of the local time process as $T \to \infty$, just as when dealing with standard SDPs. In general, the rate of divergence cannot be quantified, although we expect positive recurrent and stationary processes to have local times that diverge at the fastest rate T. As in the case of the estimation of the infinitesimal moments of an SDP, the divergence properties of the local time factor affect the convergence properties of the infinitesimal moment estimators of (110) above, as shown in the next subsection. Second, functions of spatial densities, such as spatial hazard rates, can be readily defined as in Section 3. The intuition is immediate and follows from our discussion in the previous section. We do not dwell on it here.

4.2 N-W Kernel Estimation of the Infinitesimal Moments of an SJDP

As pointed out earlier, identification of a recurrent solution to (110) above essentially entails estimation of four quantities: the drift $\mu(.)$, the diffusive variance $\sigma^2(.)$, the intensity (or probability) of a jump $\lambda(.)$, and the distribution of the jump component. Importantly, such quantities can be identified from the estimated infinitesimal moments, as we show below by virtue of two examples.

To this extent, we first turn to infinitesimal moment estimation by virtue of N-W kernel estimates. Assume the same sampling mechanism as in Section 2 is adopted and write

$$\widehat{M}^k_{(n,T)}(a) = \frac{1}{\Delta_{n,T}} \frac{\sum_{i=1}^{n-1} \mathbf{K}\left(\frac{X_{i\Delta_{n,T}} - a}{h_{n,T}}\right)(X_{(i+1)\Delta_{n,T}} - X_{i\Delta_{n,T}})^k}{\sum_{i=1}^{n} \mathbf{K}\left(\frac{X_{i\Delta_{n,T}} - a}{h_{n,T}}\right)}, \tag{135}$$

$\forall k \geq 1$, where $\mathbf{K}(.)$ is a kernel function satisfying the assumptions in Section 2. The limiting properties of $\widehat{M}^k_{(n,T)}(a)$ are laid out in Theorem 4.2.

Theorem 4.2 *Assume X_t is the solution to (110). If $n \to \infty$, $T \to \infty$, $\frac{T}{n} \to 0$ and $h_{n,T} \to 0$ so that $\frac{\overline{L}_X(T,a)}{h_{n,T}}(\Delta_{n,T}\log(1/\Delta_{n,T}))^{1/2} = o_{a.s.}(1)$ and $h_{n,T}\overline{L}_X(T,a) \overset{a.s.}{\to} \infty$, then*

$$\widehat{M}^k_{(n,T)}(a) \overset{a.s.}{\to} M^k(a) \qquad \forall k \geq 1. \tag{136}$$

Furthermore, if $h^5_{n,T}\overline{L}_X(T,a) = O_{a.s.}(1)$, then

$$\sqrt{h_{n,T}\widehat{\overline{L}}_X(T,a)}\left(\widehat{M}^k_{(n,T)}(a) - M^k(a) - \Gamma_{M^k}(a)\right) \Rightarrow \mathbf{N}\left(0, \mathbf{K}_2 M^{2k}(a)\right), \tag{137}$$

$\forall k \geq 1$, where

$$\Gamma_{M^k}(a) = h^2_{n,T}\mathbf{K}_1\left(\left(M^k(a)\right)' \frac{m'(a)}{m(a)} + \frac{1}{2}\left(M^k(a)\right)''\right) \qquad \forall k \geq 1, \tag{138}$$

and $m(dx) = m(x)dx$ is the invariant measure of the process.

Proof *See BN (2003).*

By virtue of our discussion in the case of SDPs, the implications of Theorem 4.2 should be clear. Here we focus on the main differences between the case with discontinuities and the case without discontinuities examined in the previous section. Contrary to diffusion estimation, all of the infinitesimal moment estimators converge at the same rate, namely $\sqrt{h_{n,T}\widehat{\overline{L}}_X(T,a)}$. The intuition is as follows. The family of estimators in (135) above hinges on averages of terms of order $\sqrt{dt}$ for every $k \geq 1$. In the case of standard SDPs, the drift estimator (48) is an average of terms of order $\sqrt{dt}$, whereas the diffusion estimator (49) averages terms of order dt, leading to a faster rate of convergence. Apparently, no infinitesimal moment can be identified over a fixed span of data in the presence of an underlying SJDP. If T were fixed, then $\widehat{\overline{L}}_X(T,a)$ would be bounded in probability and $\sqrt{h_{n,T}}$ would not be a proper convergence rate. Again, the intuition is simple. In jump-diffusion models like (110) above, the function $\mu(.)$ has the same interpretation as in the standard set-up without jumps. As a consequence, an enlarging span of data is expected to be necessary for the consistency of $\widehat{M}^1_{(n,T)}(.)$. As for the higher moments, formulae (124), (125), and (126) illustrate their dependence on the characteristics of the discontinuous jump component. A fixed span of data can not possibly contain sufficient information for the identification of the features of infrequent Poisson jumps since the number of jumps of this type on any fixed time span is finite with probability one.

We now turn to the identification of the functions and parameters of interest by virtue of two examples.

Example 1 Assume $g(x,y) = y$, where y is normally distributed with mean 0 and variance σ_y^2. In other words, $dJ_t = ydN_t$, with $y \sim N(0, \sigma_y^2)$. Then, we can rewrite (124), (125), and (126) with $k = 1$, 2, 4, and 6 as

$$\begin{aligned} M^1(a) &= \mu(a), &(139)\\ M^2(a) &= \sigma^2(a) + \lambda(a)\sigma_y^2, &(140)\\ M^4(a) &= 3\lambda(a)\left(\sigma_y^2\right)^2, &(141)\\ M^6(a) &= 15\lambda(a)\left(\sigma_y^2\right)^3, &(142)\end{aligned}$$

$\forall a \in \mathfrak{D}$. Given (139) through (142), Johannes (2004) suggested the following identification scheme:

$$\begin{aligned} \left(\widehat{\sigma}_y^2\right)_{(n,T)} &= \frac{1}{n}\sum_{i=1}^{n} \frac{\widehat{M}^6_{(n,T)}(X_{i\Delta_{n,T}})}{5\widehat{M}^4_{(n,T)}(X_{i\Delta_{n,T}})}, &(143)\\ \widehat{\lambda}_{(n,T)}(a) &= \frac{\widehat{M}^4_{(n,T)}(a)}{3\left(\widehat{\sigma}_y^4\right)_{(n,T)}}, &(144)\\ \widehat{\sigma}^2_{(n,T)}(a) &= \widehat{M}^2_{(n,T)}(a) - \widehat{\lambda}_{(n,T)}(a)\left(\widehat{\sigma}_y^2\right)_{(n,T)}, &(145)\\ \widehat{\mu}_{(n,T)}(a) &= \widehat{M}^1_{(n,T)}(a). &(146)\end{aligned}$$

This simple scheme was successfully applied to the analysis of the continuous-time dynamics of the short end of the term structure of interest rates (Johannes, 2004). For an interesting, alternative approach to identification in this class of discontinuous processes, we refer the reader to Mancini and Reno' (2006).

Example 2 Assume $g(x,y) = y$, where y is exponentially distributed with mean β. In other words, $dJ_t = ydN_t$, with $y \sim \exp(\beta)$. Also, assume the process X_t is not compensated, thereby implying that the first infinitesimal moment also contains a component equal to the first conditional moment of the jump part. We can thus re-write (124), (125), and (126) with $k = 1$, 2, 3, and 4 as

$$\begin{aligned} M^1(a) &= \mu(a) + \lambda(a)\beta, &(147)\\ M^2(a) &= \sigma^2(a) + 2\lambda(a)\beta^2, &(148)\\ M^3(a) &= 6\lambda(a)\beta^3, &(149)\\ M^4(a) &= 24\lambda(a)\beta^4, &(150)\end{aligned}$$

$\forall a \in \mathfrak{D}$. Given (147) through (150), BR (2008) suggested the following identification scheme:

$$\widehat{\beta}_{(n,T)} = \frac{1}{n}\sum_{i=1}^{n}\frac{\widehat{M}^4_{(n,T)}(X_{i\Delta_{n,T}})}{4\widehat{M}^3_{(n,T)}(X_{i\Delta_{n,T}})}, \tag{151}$$

$$\widehat{\lambda}_{(n,T)}(a) = \frac{\widehat{M}^4_{(n,T)}(a)}{24\widehat{\beta}^4_{(n,T)}}, \tag{152}$$

$$\widehat{\sigma}^2_{(n,T)}(a) = \widehat{M}^2_{(n,T)}(a) - 2\widehat{\lambda}_{(n,T)}(a)\widehat{\beta}^2_{(n,T)}, \tag{153}$$

$$\widehat{\mu}_{(n,T)}(a) = \widehat{M}^1_{(n,T)}(a) - \widehat{\lambda}_{(n,T)}(a)\widehat{\beta}_{(n,T)}. \tag{154}$$

Given preliminary spot variance estimates (as in Subsection 3.6 above, for instance), BR (2008) apply this scheme to the analysis of the continuous-time dynamics of the spot variance process. We will follow their treatment in the empirical application below.

Some observations are in order. First, reasonable parametric assumptions on the jump component are necessary for identification. While the identification methods in the examples above have been shown to be empirically successful in a variety of contexts in continuous-time finance, more involved (and potentially more efficient) methodologies making use of alternative moments may have been suggested instead. This said, any sensible identification scheme entails averages of nonparametric estimates for the purpose of the estimation of the model's parameters (c.f. (143) and (151) above). Hence, while the estimates of the functions of interest (namely, $\mu(.)$, $\sigma^2(.)$ and $\lambda(.)$) converge at the nonparametric rate $\sqrt{h_{n,T}\widehat{\overline{L}}_X(T,a)}$ (by a simple application of the delta method), the parameter estimates $\widehat{\sigma}^2_y$ and $\widehat{\beta}$ converge at a faster rate. Differently put, while causing a loss in terms of generality of the model, the important (for identification) imposition of parametric assumptions is, not surprisingly, beneficial for estimation (BR, 2008). Second, the methodology is flexible. Alternative parametric assumptions could have been imposed. Obviously, the functions $\mu(.)$, $\sigma^2(.)$ and $\lambda(.)$ are left fairly unrestricted, thereby allowing for non-linearities which, as shown below, might prove useful in continuous-time finance modeling.

4.3 An Empirical Application: Stochastic Volatility

The estimation scheme in Eq. (151) through Eq. (154) has been applied by BR (2008) to identify the features of exponential jumps in spot variance while allowing drift, diffusive variance, and intensity of the jumps to be nonlinear functions of the state. Here we consider two-scale estimates of spot market variance relying on SPY data (for the period between January 2, 1998, and March 31, 2006), as in Subsection 3.6. In addition to drift and diffusive variance, Figure 2 reports the probability of the exponential jumps (expressed in terms of the number of annual jumps per variance level) and the (constant) expected jump size.[14] For all plots, we display the parametric estimates (dotted lines) reported in

[14] The form of the corresponding asymptotic bands is derived in BR, 2008.

Eraker *et al.* (2003) for a model with linear drift ($\mu(a) = c_0 + c_1 a$), square-root volatility ($\sigma^2(a) = c_3 a$), and constant jump intensity ($\lambda(a) = c_4$) - c.f., Eraker *et al.*, 2008, Table III. As indicated in Subsection 3.6, allowing for jumps reduces the magnitude of the estimated drift and diffusion vis-à-vis the SDP case reported earlier. The functional drift implies a bit more mean-reversion than the parametric model. The nonparametric diffusive function suggests more variability associated with the continuous component of the process than in the parametric specification of Eraker *et al.* (2003). While this difference may be visually small and is statistically insignificant, model diagnostics reported in BR (2008) indicate that it is important. In the more informative, empirical range of the process (as highlighted by the local time estimates in Figure 1 (a)), the average annual number of jumps is about 1, which is consistent with the parametric model. The expected jump size is about 5 with the parametric estimate (about 1.8) within the 95% asymptotic bands. Importantly, the obtained value for the expected jump size may be influenced by spot variance estimates (constructed using the two-scale estimator) which are affected by the presence of jumps in the market return process. The use of spot variance estimates robust to discontinuities in market returns would yield an expected jump size of about 2 and roughly unchanged jump intensities (BR, 2008). These values would therefore lend some support to the variance jump sizes and the frequency of jumps obtained by Eraker *et al.* (2003) by virtue of (*i*) a fully parametric model allowing for jumps in returns and (*ii*) a more classical volatility-filtering method relying on MCMC methods (for a discussion of MCMC methods in finance, see Johannes and Polson, 2009, in this volume).

5 Multivariate Diffusion Processes (MDPs)

In this section we focus on a vector process X_t expressed as the d-dimensional solution of the multivariate stochastic differential equation

$$dX_t = \boldsymbol{\mu}\left(X_t\right) dt + \boldsymbol{\sigma}\left(X_t\right) d\mathbf{B}_s, \tag{155}$$

where $\mathbf{B} = \left\{\mathbf{B}_t, \Im_t^B; 0 \leq t < \infty\right\}$ is an m-dimensional standard Brownian motion, $\boldsymbol{\mu}(.) = \{\mu_i(.)\}_{1 \leq i \leq d}$ is a $d \times 1$ Borel measurable vector, $\boldsymbol{\sigma}(.) = \{\sigma_{ij}(.)\}_{\substack{1 \leq i \leq d \\ 1 \leq j \leq m}}$ is a $d \times m$ Borel measurable matrix, and $X_0 = \overline{X} \in \mathfrak{D} \subseteq \Re^d$ is a given initial condition taken to be independent of $\Im_\infty^B$ and with finite second moment, i.e. $\mathbf{E}||\overline{X}|| < \infty$. Define the left-continuous filtration

$$\overline{\Im}_t := \sigma(\overline{X}) \vee \Im_t^B = \sigma(\overline{X}, \mathbf{B}_s; 0 \leq s \leq t) \qquad 0 \leq t < \infty \tag{156}$$

and the collection of null sets

$$\aleph := \{N \subseteq \Omega; \exists G \in \overline{\Im}_\infty \; with \; N \subseteq G \; and \; P(G) = 0\}. \tag{157}$$

Now create the augmented filtration

$$\widetilde{\Im}_t^X := \sigma(\overline{\Im}_t \cup \aleph) \qquad 0 \le t < \infty. \tag{158}$$

Assumption 1 and 2 (1 and 2b) below are imposed on (155) to guarantee the existence of a unique and null recurrent (positive recurrent) solution to (155).

(1) *$\boldsymbol{\mu}(.)$ and $\boldsymbol{\sigma}(.)$ are time-homogeneous, $\mathfrak{B}$-measurable functions on $\mathfrak{D} \subseteq \Re^d$ where $\mathfrak{B}$ is the σ-field generated by Borel sets on $\mathfrak{D}$. Both functions satisfy local Lipschitz and linear growth conditions. Thus, for every compact subset J of the range of the process, there exist constants C_1 and C_2 such that, for all x and y in J,*

$$||\boldsymbol{\mu}(x) - \boldsymbol{\mu}(y)|| + ||\boldsymbol{\sigma}(x) - \boldsymbol{\sigma}(y)|| \le C_1||x - y||, \tag{159}$$

and

$$||\boldsymbol{\mu}(x)|| + ||\boldsymbol{\sigma}(x)|| \le C_2\{1 + ||x||\}, \tag{160}$$

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

(2) *(Null recurrence) Define the positive definite matrix $\mathbf{s}(x) = \boldsymbol{\sigma}(x)\boldsymbol{\sigma}(x)^\top$ so that $s_{ik}(x) = \sum_{g=1}^{m}\sigma_{ig}(x)\sigma_{gk}(x)\ \forall x \in \mathfrak{D} \subseteq \Re^d$ and assume that every open and bounded set $A \in \mathfrak{D}$ satisfies*

$$\min_{x\in\overline{A}} s_{ii}(x) > 0, \tag{161}$$

for some $1 \le i \le d$. Write the second-order elliptic operator

$$\mathfrak{L}\varphi(.) = \sum_{i=1}^{d}\mu_i(.)\frac{\partial\varphi(.)}{\partial x_i} + \frac{1}{2}\sum_{i=1}^{d}\sum_{k=1}^{d}s_{ik}(.)\frac{\partial\varphi(.)}{\partial x_i\partial x_k}. \tag{162}$$

There is a function $\varphi(.) : \Re^d\setminus\{0\} \to \Re$ of class C^2 in the domain of the operator which satisfies

$$\mathfrak{L}\varphi(.) \le 0 \qquad \text{on} \qquad \Re^d\setminus\{0\}, \tag{163}$$

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

(2b) *(Positive recurrence) There is a function $\varphi(.) : \Re^d\setminus\{0\} \to \Re$ of class C^2 in the domain of the operator which satisfies*

$$\mathfrak{L}\varphi(.) \le -1 \qquad \text{on} \qquad \Re^d\setminus\{0\}, \tag{164}$$

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

Under Assumptions 1 and 2 (2 b), the stochastic differential equation (155) has a strong solution X_t which is unique and null recurrent (positive recurrent). Specifically, the process X_t satisfies

$$X_t = \overline{X} + \int_0^t \boldsymbol{\mu}(X_s)\, ds + \int_0^t \boldsymbol{\sigma}(X_s)\, d\mathbf{B}_s, \tag{165}$$

and is square integrable, i.e. $\mathbf{E}||X_t||^2 < \infty \quad \forall t$. Equivalently, we can write each coordinate X_t^j as

$$X_t^j = \overline{X}^j + \int_0^t \mu_j(X_s)\, ds + \sum_{g=1}^m \int_0^t \sigma_{jg}(X_s)\, dB_s^g, \qquad 0 \le t < \infty, 1 \le j \le d. \tag{166}$$

In agreement with the scalar model in Section 3, the dynamics of X_t are driven by Brownian shocks and determined by the functional forms of the matrices $\boldsymbol{\mu}(.)$ and $\mathbf{s}(.)$ (recall from Assumption (2) above that $\mathbf{s}(x) = \boldsymbol{\sigma}(x)\boldsymbol{\sigma}(x)^\top$). Such matrices will be the object of econometric interest in the present section. As in the scalar case, they both have straightforward representations in terms of infinitesimal conditional moments. In particular,

$$\mathbf{E}^a\left[X_t^i - a_i\right] = t\mu_i(a) + o(t) \tag{167}$$

$$\mathbf{E}^a\left[\left(X_t^i - a_i\right)\left(X_t^j - a_j\right)\right] = ts_{ij}(a) + o(t) \tag{168}$$

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

The notions of recurrence implied by Assumption 2 are standard (c.f. Section 2). Namely, the multidimensional process X_t is Harris recurrent if there is a $\sigma-$finite measure $m(dx)$ so that $m(A) > 0$ implies $\lim_{T\to\infty} \eta_A^T = \infty$ with probability one $\forall A \in \mathfrak{B}(\mathfrak{D})$, where $\eta_A^T = \int_0^T \mathbf{1}_{\{X_s \in A\}} ds$ is, as earlier, the occupation measure of A. The following result gives us the (implicit) rate at which η_A^T diverges to infinity and a weak convergence result for η_A^T.

Theorem 5.1 *Assume X_t is the solution to (155) above.*[15] *Consider a non-negative function $\delta(.)$. If there exists a constant $\alpha \in [0,1]$ and a slowly-varying function at infinity $L(T)$ such that*

$$\lim_{T\to\infty} \mathbf{E}^a\left[\int_0^\infty e^{-\frac{s}{T}}\delta(X_s)ds\right]/(T^\alpha L(T)) = C_X > 0 \qquad \forall a \in \mathfrak{D}, \tag{169}$$

then

$$\lim_{T\to\infty} P^a\left\{\frac{1}{C_X\left(T^\alpha L(T)\right)}\int_0^T \delta(X_s)ds < x\right\} = G_\alpha(x), \tag{170}$$

[15] The theorem, in its form stated here, applies to more general continuous-time Markov processes than MDPs (Darling and Kac, 1956).

where

$$G_\alpha(x) = \frac{1}{\pi\alpha}\int_0^x \sum_{j=1}^{\infty} \frac{(-1)^{j-1}}{j!} \sin(\pi\alpha j)\,\Gamma(\alpha j+1) y^{j-1} dy, \tag{171}$$

$\Gamma(.)$ *is the Gamma function,* $C_X = C_X^* \int_{-\infty}^{\infty} \delta(x) m(dx)$, $m(dx)$ *is the invariant measure, and* C_X^* *is a process-specific constant.*

Proof *See Darling and Kac (1956) for the original statement of the theorem. Bingham (1971) contains a functional version of the same finding. Several papers discuss limit results for slowly-increasing occupation times associated with null-recurrent Markov processes, see Höpfner and Löcherbach (2001) for a complete, recent survey of the literature on the subject. The interested reader is referred to Khasminskii (2001) and Khasminskii and Yin (2000) for a detailed treatment of the one-dimensional null-recurrent diffusion case.*

We can rewrite (170) as follows:

$$\frac{\int_0^T \delta(X_s) ds}{C_X u(T)} \Rightarrow g_\alpha, \tag{172}$$

where g_α is the Mittag-Leffler density with parameter α,[16] i.e.,

$$g_\alpha(x) = \frac{1}{\pi\alpha} \sum_{j=1}^{\infty} \frac{(-1)^{j-1}}{j!} \sin(\pi\alpha j)\Gamma(\alpha j+1) x^{j-1}, \tag{173}$$

and $u(T) = T^\alpha L(T)$. Theorem 5.1 shows that additive functionals of the process ($\int_0^T \delta(X_s)ds$) converge weakly to a random variable endowed with the Mittag-Leffler density g_α when standardized appropriately (by $u(T)$). The rate of divergence to infinity of the standardizing factor (and, as a consequence, the rate of divergence to infinity of the continuous averages) depends on the statistical features of the process through the constant α. Clearly, α defines the nature of the Mittag-Leffler density as well.

Some observations are in order. First, the theorem readily applies to the occupation measures since we can take $\delta(.) = \mathbf{1}_A$ (the characteristic function of the generic set A) giving $\int_0^T \delta(X_s)ds = \eta_A^T \; \forall A \in \mathfrak{B}(\mathfrak{D})$. Second, it extends to a large class of continuous-time Harris recurrent Markov processes, provided assumption (169) is satisfied (Darling and Kac, 1956). In general, we can apply it to SDPs and SDJPs of the type analyzed in the present review. Importantly, its implications appear to be particularly interesting when studying processes for which a standard notion of local time cannot be defined, as is the case with multivariate diffusions, for example. We do not dwell on this idea here (and refer the reader to BM, 2004) but the meaning of the statement will become clear in the next subsection.

[16] We abuse notation somewhat to signify convergence to a limiting random variable endowed with the Mittag-Leffler density.

We now briefly consider some interesting special cases. We noticed earlier that the characteristics of the underlying recurrent process affect the weak convergence result through the constant α which modifies both the rate of divergence of the occupation measure and the resulting limiting distribution. This constant is known only for a few processes. Specifically, if X_t is Brownian motion on the plane, then $\alpha = 0$ and the Mittag-Leffler distribution coincides with the exponential distribution, i.e.

$$\frac{\int_0^T \delta(X_s)ds}{C_X \log T} \Rightarrow e^{-x} \qquad x \geq 0. \tag{174}$$

If X_t is a scalar Brownian motion, then $\alpha = \frac{1}{2}$ and the Mittag-Leffler density is equal to the truncated normal, i.e.

$$\frac{\int_0^T \delta(X_s)ds}{C_X \sqrt{T}} \Rightarrow \frac{2}{\sqrt{2\pi}} e^{-x^2/2} \qquad x \geq 0, \tag{175}$$

implying, importantly, that the dimensionality of the system has, in general, an impact on the rate of divergence of continuous averages of the process. It is noted, in fact, that we go from a $\sqrt{T}$-rate to a $\log T$-rate when moving from the scalar Brownian case to its bivariate counterpart. Importantly, the dimensionality of the process does not influence the rate of divergence of the continuous averages (or the rate of divergence of the occupation measures) if stationarity is satisfied. Under stationarity $\alpha = 1$ and

$$\frac{\int_0^T \delta(X_s)ds}{T} \xrightarrow{p} \int_{-\infty}^{\infty} \delta(x) f(dx), \tag{176}$$

where $f(dx) = f(x)dx$ is the stationary probability measure of X_t, which is a form of the classical ergodic theorem. We will return to Theorem 5.1 and its implications in the sequel. We now consider generalized density estimation for multivariate solutions to (155).

5.1 Generalized Density Estimation for MDPs

Just as it is natural to estimate multivariate density functions using multidimensional extensions of kernel estimates for scalar densities (see, e.g., Pagan and Ullah, 1999), it might appear natural to estimate the local time of a vector process using a multivariate counterpart of the standard estimator from Section 3, i.e.

$$\widehat{\widehat{\mathbb{L}}}_{(n,T)}(\overline{T}, a) = \frac{\Delta_{n,\overline{T}}}{\mathbf{h}_{n,\overline{T}}} \sum_{i=1}^{n} \left(\prod_{j=1}^{d} \mathbf{K} \left(\frac{X^j_{i\Delta_{n,\overline{T}}} - a_j}{h_{n,\overline{T}}} \right) \right), \tag{177}$$

where $\mathbf{h}_{n,\overline{T}} = h^d_{n,\overline{T}}$ and $a = (a_1, a_2, .., a_d) \in \Re^d$.[17] As it happens, local time is not generally defined for multidimensional semimartingales (see Brugière, 1991, among others).

[17] Solely for notational simplicity, in order to focus on the main ideas, we assume here that all bandwidths are the same (i.e., $h_{n,\overline{T},j} = h_{n,\overline{T}}$ for all j) in which case $\mathbf{h}_{n,\overline{T}} = \prod_{j=1}^{d} h_{n,\overline{T},j} = h^d_{n,\overline{T}}$. It is of course straight-

In consequence, we cannot build a notion of (spatial) density for multivariate, potentially nonstationary, continuous-time processes based on local time as suggested in Section 3 for SDPs and in Section 4 for SJDPs. Consistently, over a fixed span of data $\overline{T}$, the quantity $\widehat{\widehat{\mathbb{L}}}_{(n,\overline{T})}(\overline{T},a)$ cannot be interpreted as a multivariate sojourn time estimator despite being a local time estimator for $d=1$. Nonetheless, its asymptotic features as $n\to\infty$ for a fixed $T=\overline{T}$ can still be characterized. Using a multivariate indicator kernel (but we expect the results not to change in the presence of a continuous kernel function), Brugière (1993) shows that

$$\frac{1}{\mathbf{h}_{n,\overline{T}}\log\left(1/\mathbf{h}^2_{n,\overline{T}}\right)}\widehat{\widehat{\mathbb{L}}}_{(n,\overline{T})}(\overline{T},a) \tag{178}$$

converges weakly (as $n\to\infty$) to an exponentially distributed random variable when the dimension of the system d is equal to two. Furthermore, the quantity

$$\frac{1}{\mathbf{h}_{n,\overline{T}}}\widehat{\widehat{\mathbb{L}}}_{(n,\overline{T})}(\overline{T},a) \tag{179}$$

converges weakly to

$$\int_0^\infty \mathbf{1}_{\{\sigma(0)\mathbf{B}_s<1\}}ds \tag{180}$$

when $d\geq 3$.

Interestingly, while preventing us from constructing appealing descriptive statistics for multidimensional (potentially nonstationary) semimartingales based on (177), the nonexistence of a notion of local time is not prohibitive when it comes to dealing with the estimation of the infinitesimal moments of (155). This result might at first appear surprising since the local time estimates are known to play a fundamental role in the scalar limit theory for recurrent SDPs and SJDPs. On the other hand, it is well know that simple matrix functionals of multivariate processes, such as the functional $\int_0^t \mathbf{B}(s)\mathbf{B}(s)^\top ds$ of the vector Brownian motion $\mathbf{B}$, are well defined and sample functions converge weakly to them, even though these functionals may not have a representation in terms of local time, as they do from the occupation time formula in the scalar case.

Coherently, we now discuss a finding which is important in building an estimation theory for recurrent, multivariate, diffusions without resorting to a notion of local time. The following result, which heavily hinges on Theorem 5.1. above, characterizes the behavior of $\widehat{\widehat{\mathbb{L}}}_{(n,T)}(T,a)$ over an enlarging span of observations, namely as $T\to\infty$ (with $n\to\infty$). In the next subsection we will show that the asymptotic behavior of $\widehat{\widehat{\mathbb{L}}}_{(n,T)}(T,a)$ as $T,n\to\infty$ is crucial in interpreting the limit theory of kernel estimates of the infinitesimal moments

forward to extend the framework and allow for process-specific smoothing parameters. We refer the reader to BM (2004) for discussions.

of the solution to (155). We continue to use the sampling scheme which was laid out in Section 2.

Theorem 5.2 *Assume X_t is the solution to (155). If the vanishing bandwidth $h_{n,T}$ satisfies*

$$\frac{(\Delta_{n,T}\log(1/\Delta_{n,T}))^{1/2}}{\mathbf{h}_{n,T}} = o(1), \tag{181}$$

as $n, T \to \infty$ with $\Delta_{n,T} \to 0$, then

$$\frac{\widehat{\widehat{\mathbb{L}}}_{(n,T)}(T,a)}{C_X u(T)} \Rightarrow m(a)\, g_\alpha \quad \forall a \in \mathfrak{D} \subseteq \mathfrak{R}^d \tag{182}$$

for some function $u(T) = T^\alpha L(T)$, with $0 \leq \alpha \leq 1$ and $L(T)$ slowly-varying, where g_α is the Mittag-Leffler random variable with the same parameter α, and $m(dx) = m(x)dx$ is the invariant measure of the process. C_X is a process-specific constant.

Proof *See BM (2004).*

Theorem 5.2 links the divergence properties of $\widehat{\widehat{\mathbb{L}}}_{(n,T)}(T,a)$ to those of the occupation time measure η_A^T. This result is hardly surprising, being that $\widehat{\widehat{\mathbb{L}}}_{(n,T)}(T,a)$ is an estimate of the time spent by the process in the vicinity of the spatial point a, even though the dimensionality of the system prevents us from interpreting it as a consistent estimate of the local time of the process at a.

Two observations are in order. First, Theorem 5.2 applies to SDPs of the type analyzed in Section 3. Previously, we pointed out that the local time estimates of stationary processes and standard scalar Brownian motion diverge at rate T and $\sqrt{T}$, respectively. The same result can be deduced from Theorem 5.2 as a subcase of the more general theory laid out in this section. Second, in the presence of stationary processes of any dimension $\widehat{\widehat{\mathbb{L}}}_{(n,T)}(T,a)$ represents a well-defined density estimator. In fact, if $\alpha = 1$, then $g_\alpha = 1$ and $C_X = \frac{1}{m(\mathfrak{D})}$. Thus,

$$\frac{\widehat{\widehat{\mathbb{L}}}_{(n,T)}(T,a)}{T} = \frac{1}{n\mathbf{h}_{n,T}} \sum_{i=1}^{n} \left(\prod_{j=1}^{d} \mathbf{K}\left(\frac{X^j_{i\Delta_{n,T}} - a_j}{h_{n,T}} \right) \right) \tag{183}$$

converges to $\frac{m(a)}{m(\mathfrak{D})} = f(a)$, which is a standard finding in multivariate density estimation for both discrete and continuous-time stationary processes (see, e.g., Prakasa-Rao, 1983, and Silverman, 1986).[18] We now turn to the estimation of the infinitesimal moments.

[18] For invariant density estimation in the case of multidimensional diffusions, we refer the reader to Bianchi (2007) and Dalalyan and Reiss (2006).

5.2 N-W Kernel Estimation of the Infinitesimal Moments of an MDP

Following our discussion in the previous sections, it is natural to estimate the matrices $\boldsymbol{\mu}(.)$ and $\mathbf{s}(.) = \boldsymbol{\sigma}(.)\boldsymbol{\sigma}(.)^\top$ using nonparametric kernel estimates of the N-W type, i.e.,

$$\widehat{\boldsymbol{\mu}}_{(n,T)}(a) = \frac{\frac{1}{\Delta_{n,T}} \sum_{i=1}^{n-1} \mathbf{K}_i^a \left(X_{(i+1)\Delta_{n,T}} - X_{i\Delta_{n,T}} \right)}{\sum_{i=1}^{n} \mathbf{K}_i^a}, \tag{184}$$

and

$$\widehat{\mathbf{s}}_{(n,T)}(a) = \frac{\frac{1}{\Delta_{n,T}} \sum_{i=1}^{n-1} \mathbf{K}_i^a \left(X_{(i+1)\Delta_{n,T}} - X_{i\Delta_{n,T}} \right) \left(X_{(i+1)\Delta_{n,T}} - X_{i\Delta_{n,T}} \right)^\top}{\sum_{i=1}^{n} \mathbf{K}_i^a} \tag{185}$$

where

$$\mathbf{K}_i^a = \prod_{j=1}^{d} \mathbf{K} \left(\frac{X_{i\Delta_{n,T}}^j - a_j}{h_{n,T}} \right). \tag{186}$$

As pointed out earlier, the nonexistence of local time for multivariate solutions to (155) above does not represent an impossible obstacle when deriving an estimation theory based on (184) and (185). The intuition relies on the following observations: any limit results for (184) and (185) in the $d \geq 1$ case should collapse in the findings that we illustrated in Section 3, i.e. (52) and (56), when reducing the dimensionality of the system to $d = 1$. Let us focus on the drift for illustration purposes. Based on (52), our best guess of a weak convergence result for (184) is:

$$\sqrt{\mathbf{h}_{n,T} \widehat{\widehat{\mathbb{L}}}_{(n,T)}(T,a)} \left\{ \widehat{\boldsymbol{\mu}}_{(n,T)}(a) - \boldsymbol{\mu}(a) \right\} \Rightarrow \mathbf{N} \left(0, \mathbf{K}_2^d \mathbf{s}(a) \right) \tag{187}$$

where $\mathbf{h}_{n,T} = h_{n,T}^d$. First, (187) reduces to (52) when $d = 1$, thereby satisfying our requirement. Second, the impossibility of interpreting $\widehat{\widehat{\mathbb{L}}}_{(n,T)}(T,a)$ as a local time estimator for $d > 1$ does not have an impact on the credibility of the intuition leading to (187). In fact, as shown earlier, $\widehat{\widehat{\mathbb{L}}}_{(n,T)}(T,a)$ converges (as $n, T \to \infty$ and if standardized appropriately) to a well-defined random variable for dimensions higher than one while also being a local time estimator when $d = 1$. The following theorem confirms our intuition. As usual, we adopt the same sampling scheme as in Section 2.

Theorem 5.3 *Assume X_t is the solution to (155). Also, assume the vanishing sequence $h_{n,T}$ satisfies*

$$\frac{(\Delta_{n,T} \log(1/\Delta_{n,T}))^{1/2} u(T)}{\mathbf{h}_{n,T}} = o(1) \tag{188}$$

and

$$\mathbf{h}_{n,T} u(T) \to \infty, \tag{189}$$

as $n,T\to\infty$ with $\Delta_{n,T}\to 0$, for some function $u(T)=T^{\alpha}L(T)$ with $L(T)$ slowly-varying and a process-specific parameter α so that $0\le\alpha\le 1$. Then,

$$\widehat{\boldsymbol{\mu}}_{(n,T)}(a)\overset{a.s.}{\to}\boldsymbol{\mu}(a)\qquad\forall a\in\mathfrak{D}\subseteq\Re^{d}. \tag{190}$$

Furthermore, if $h_{n,T}=O\left(u(T)^{-\frac{1}{d+4}}\right)$, *then*

$$\begin{aligned}&\sqrt{\mathbf{h}_{n,T}\widehat{\widehat{\mathbb{L}}}_{(n,T)}(T,a)}\left(\widehat{\boldsymbol{\mu}}_{(n,T)}(a)-\boldsymbol{\mu}(a)-\boldsymbol{\Gamma}_{\boldsymbol{\mu}}(a)\right)\\ \Rightarrow\quad&(\mathbf{s}(a))^{1/2}\,\mathbf{N}\left(0,\mathbf{K}_2^{d}\mathbf{I}\right)\qquad\qquad\forall a\in\mathfrak{D}\subseteq\Re^{d},\end{aligned} \tag{191}$$

where

$$\boldsymbol{\Gamma}_{\boldsymbol{\mu}}(a)=(bias_1,bias_2,...,bias_d)(a), \tag{192}$$

$$bias_i(a)\;=\;h_{n,T}^2\mathbf{K}_1\left(\sum_{k=1}^{d}\frac{\partial\mu_i\left(a\right)}{\partial a_k}\frac{\frac{\partial m(a)}{\partial a_k}}{m(a)}+\frac{1}{2}\sum_{k=1}^{d}\frac{\partial^2\mu_i\left(a\right)}{\partial a_k\partial a_k}\right)\qquad i=1,...,d, \tag{193}$$

and $m\left(dx\right)=m(x)dx$ is the invariant measure of the process.

Proof *See BM (2004).*

All our comments in the scalar case apply to the multivariate set-up examined here up to some minor modifications. We will therefore not be as detailed as in Section 3. Nonetheless, it should be noted that the asymptotic bias is $O(h_{n,T}^2)$, as in the scalar case, whereas the asymptotic variance is of order $\widehat{\widehat{\mathbb{L}}}_{(n,T)}(T,a)^{-1}h_{n,T}^{-d}$ rather than $\widehat{\widehat{\mathbb{L}}}_{(n,T)}(T,a)^{-1}h_{n,T}$. In the standard estimation of conditional first moments in the discrete-time, stationary, framework, the limiting bias is $O(h_{n,T}^2)$ while the limiting variance is $n^{-1}h_{n,T}^{-d}$, rather than $(nh_{n,T})^{-1}$. In other words, the variance increases with the dimensionality of the system. This effect is commonly known as *the curse of dimensionality* (see, e.g., Silverman, 1986). Importantly, here we have a curse of dimensionality which mirrors the classical result in conventional non-parametric estimation of conditional moments in discrete time and manifests itself through the factor h^{-d}, as well as an additional curse of dimensionality operating via the quantity $\widehat{\widehat{\mathbb{L}}}_{(n,T)}(T,a)$. The latter effect is a genuine by-product of the generality of this theory and, in particular, is due to robustness to deviations from stationarity. In fact, should the system be stationary (or positive recurrent), then $\widehat{\widehat{\mathbb{L}}}_{(n,T)}(T,a)$ would diverge at speed T (c.f. the previous subsection) regardless of the number of equations and the order of the variance term would simply be $T^{-1}h_{n,T}^{-d}$. Hence, we would be in the presence of a rather ordinary dimensionality problem since only the power d would be affected by the number of equations in the system. By contrast, consider the null recurrent situation. We pointed out earlier that scalar Brownian motion and Brownian motion on the plane imply divergence rates for $\widehat{\overline{\mathbb{L}}}_{(n,T)}(T,a)$ that are equal to $\sqrt{T}$ and $\log T$, respectively (see the previous subsection). This result has broader implications. We expect the dimensionality of the system to have a negative influence on the rate of divergence of the factor $\widehat{\overline{\mathbb{L}}}_{(n,T)}(T,a)$ for null recurrent processes that are more general than Brownian motion, thereby reinforcing the conventional effect that comes into play through the term $h_{n,T}^d$ and leading to a slower rate of convergence of the nonparametric estimates to the theoretical vector $\boldsymbol{\mu}(.)$. The optimal bandwidth sequence, i.e.

$$h_{n,T} \propto \widehat{\overline{\mathbb{L}}}_{(n,T)}^{-\frac{1}{d+4}}(T,a) \tag{194}$$

accounts for both effects or, in other words, for *the two curses of dimensionality,* in the terminology of BM (2004).

We now turn to diffusion estimation. The symbol $\otimes$ in the statement of Theorem 5.4 denotes the standard Kronecker product. When applied to a generic matrix A the operator *vec* stacks the columns of A. The operator *vech* selects the non-redundant elements of *vec*.

Theorem 5.4 *Assume X_t is the solution to (155). Also, assume the vanishing sequence $h_{n,T}$ satisfies*

$$\frac{(\Delta_{n,T}\log(1/\Delta_{n,T}))^{1/2}u(T)}{\mathbf{h}_{n,T}} = o(1), \tag{195}$$

as $n,T\to\infty$ with $\Delta_{n,T}\to 0$, for some function $u(T)=T^{\alpha}L(T)$ with $L(T)$ slowly-varying and a process-specific parameter α so that $0\leq\alpha\leq 1$. Then,

$$\widehat{\mathbf{s}}_{(n,T)}(a) \stackrel{a.s.}{\to} \mathbf{s}(a) \qquad \forall a\in\mathfrak{D}\subseteq\mathfrak{R}^d. \tag{196}$$

Furthermore, if $h_{n,T}^2\sqrt{\frac{\mathbf{h}_{n,T}u(T)}{\Delta_{n,T}}} = O(1)$, *then*

$$\sqrt{\frac{\mathbf{h}_{n,T}\widehat{\overline{\mathbb{L}}}_{(n,T)}(T,a)}{\Delta_{n,T}}}\left(\text{vech}\widehat{\mathbf{s}}_{(n,T)}(a) - \text{vech}\mathbf{s}(a) - \boldsymbol{\Gamma}_{\boldsymbol{\sigma}^2}(a)\right)$$
$$\Rightarrow \quad (\boldsymbol{\Xi}(a))^{1/2}\,\mathbf{N}\left(0,\mathbf{K}_2^d\mathbf{I}\right), \qquad \forall a\in\mathfrak{D}\subseteq\mathfrak{R}^d, \tag{197}$$

where

$$\boldsymbol{\Gamma}_{\boldsymbol{\sigma}^2}(a) = (bias_{1,1}, bias_{2,1}, ..., bias_{d,d})(a), \tag{198}$$

$$bias_{i,j}(a) = h_{n,T}^2\mathbf{K}_1\left(\sum_{k=1}^{d}\frac{\partial s_{i,j}(a)}{\partial a_k}\frac{\frac{\partial m(a)}{\partial a_k}}{m(a)} + \frac{1}{2}\sum_{k=1}^{d}\frac{\partial^2 s_{i,j}(a)}{\partial a_k\partial a_k}\right) \qquad i,j=(1,1),...,(d,d), \tag{199}$$

$$\boldsymbol{\Xi}(a) = L_D\left(2\mathbf{s}(a)\otimes\mathbf{s}(a)\right)L_D^{\top}, \tag{200}$$

$$L_D = (D^\top D)^{-1} D^\top, \tag{201}$$

D is the standard duplication matrix, i.e., the unique $d^2 \times (d(d+1))/2$ matrix so that L_D eliminates redundant elements, viz.,

$$\text{vech}\mathbf{s}(a) = L_D \text{vec}\mathbf{s}(a) = \begin{bmatrix} s_{1,1} \\ s_{2,1} \\ s_{2,2} \\ s_{3,1} \\ \cdots \\ s_{d,d} \end{bmatrix}, \tag{202}$$

and $m(dx) = m(x)dx$ is the invariant measure of the process.

Proof *See BM (2004).*

Our comments in the scalar diffusion case (c.f. Section 3) and in the multivariate drift case should suffice to interpret the results in Theorem 5.4 above. Here we simply note that, as in the scalar case, the local properties of the process contain sufficient information to identify the diffusion matrix, i.e. $\mathbf{s}(a)$ can be estimated consistently over a fixed span of data $T = \overline{T}$. The interested reader is referred to the work of Brugiére (c.f. Brugière, 1991, 1993) for a thorough treatment in the fixed T case. In particular, Brugière (1991) discusses weak consistency of (185) for the matrix of interest, while Brugière (1993) proves the asymptotic normality of the diffusion matrix estimator. The kernel used in both papers is the discontinuous indicator kernel. Extending the results in Brugière to the use of continuous kernels should be immediate. Genon-Catalot and Jacod (1993) offer interesting, related methods.

In recent work, Jeffrey *et al.* (2004) propose and study a multivariate nonparametric volatility estimator in the context of the successful HJM term-structure model (Heath *et al.*, 1992). An alternative approach to functional multivariate volatility estimation for the purpose of fixed income pricing is contained in Knight *et al.* (2006). We refer the reader to both papers for multidimensional kernel methods for diffusions tailored to flexible continuous-time pricing issues. The consistency and limiting distribution of derivative prices obtained by virtue of parametric, semiparametric, and nonparametric estimators for diffusions is studied in Kristensen (2008).

6 Concluding Remarks

In surveying the tools that have been recently introduced to describe and study the formulation and estimation of classes of continuous-time Markov models, this chapter illustrates the important role that is played by *local nonparametric* methods along with the assumption of *recurrence*. Our focus has been on estimation procedures which are general both in terms of model specification and in terms of statistical assumptions needed for identification. Local nonparametric methods achieve the former by being robust (at the cost of an efficiency loss) to model mispecifications. Recurrence is a promising avenue to achieve the latter.

Similar arguments in favor of minimal conditions on the underlying statistical structure of the process of interest may, however, be put forward when dealing with parametric models and discrete-time series. Sometimes empirical researchers may be a lot more comfortable avoiding restrictions like stationarity or arbitrary mixing conditions on the processes they are modeling. In the same circumstances, it might also seem inappropriate to impose explicit nonstationary behaviour (often of the random walk or $\frac{1}{2}$-null recurrent type) in the specification. Indeed, many practical situations arise where neither stationarity nor nonstationarity can be safely ruled out in advance and, in such situations, the assumption of recurrence appears to be a suitable alternative condition that permits a wide range of plausible sample behaviors and includes both stationary and nonstationary processes. Interestingly, statistical inference can often be carried out in recurrent models using limiting laws defined in terms of random norming (the averaged kernel in the definition of the estimated local time being an example, c.f. (52) and (56) for instance). Such random norming captures the divergence features of time series with various degrees of recurrence and allows the user to be agnostic about the recurrence features of the processes of interest. The practical advantage of this fact is apparent. While standard asymptotic theory treats stationary and nonstationary models differently in deriving implications for statistical inference, reliance on recurrence permits one to consider both cases as subcases of a more general theory of inference. Additionally, even when the existence of a stationary density appears to be an unquestionable feature of the data and/or is dictated by economic theory, the dynamic structure of Markov processes renders conventional forms of mixing not crucial to derive limiting results and, consequently, vital tools for statistical analysis.

Having made these points, we should add the qualification that the use of recurrence as an identifying condition is still in its infancy in the econometrics literature. Harris recurrence is the identifying assumption in Yakovitz (1989), but the treatment in that paper only focuses on the discrete-time ergodic case.[19] More recently, kernel density estimation for real-valued positive Harris recurrent Markov chains is discussed in Athreya and Atuncar (1998). Karlsen and Tjøstheim (2001) provide a theory of inference for nonparametric (auto-)regressions of β-null recurrent discrete-time Markov chains. As was discussed earlier, should the rate of divergence of the occupation density of the process be known in closed form, then the optimal bandwidth choice would depend on it. However, this is not the case in general and suitable data-driven methods should deliver bandwidths capable of capturing

[19] In his 1989 paper, Yakovitz conjectured that "...in the Markov case the mixing assumptions are not essential...Even in the absence of a stationary distribution, under conditions general enough to include unbounded random walks and ARMA processes, [nonlinear] regression estimation is possible. We require only stationarity of the transition law, not of the process."

the recurrence properties of the covariates. This is the proposal formulated in recent work by Guerre (2004). Karlsen *et al.* (2007) and Schienle (2008) focus on the functional estimation of co-integrating relations between β-null recurrent discrete-time Markov chains. Moloche (2004b) tackles the nonparametric estimation of (potentially co-integrating) regressions between (either null or positive) Harris recurrent discrete-time Markov processes. An alternative approach based on Skorohod embedding and nonlinear transformations of the embedded process was initiated by Phillips and Park (1998). They study nonparametric kernel estimation of nonstationary time series embeddable in Brownian motion. Wang and Phillips (2008, 2009), and the references therein, expand on that approach and make it applicable more generally, including cases of fractional limit processes.

The methods reviewed in the present chapter, along with the recent treatments mentioned above, have helped to lay some foundations for econometric inference with continuous- and discrete-time series under mild assumptions on their parametric form and statistical evolution. The field is a new one, however, and, as this chapter has suggested, much more needs to be done.

References

[1] AÏT-SAHALIA, Y. (1996). Testing Continuous-Time Models of the Spot Interest Rate. *Review of Financial Studies* 2, 385–426.

[2] AÏT-SAHALIA, Y., L.P. HANSEN and J. SCHEINKMAN (2009). Operator Methods for Continuous-Time Markov Processes. *This volume.*

[3] ANDREWS, D.W.K. (1989). Asymptotics for Semiparametric Econometric Models: III. Testing and Examples. *Cowles Foundation Discussion Paper N. 910, Yale University.*

[4] ATHREYA, K.B., and G.S. ATUNCAR (1998). Kernel Estimation for Real-Valued Markov Chains. *Sankhyā* 60, 1-17.

[5] AZÉMA, J., M. KAPLAN-DUFLO and D. REVUZ (1967). Mesure Invariante Sur Les Classes Récurrentes des Processus de Markov. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 8, 157–181.

[6] BANDI, F.M. (2002). Short-Term Interest Rate Dynamics: A Spatial Approach. *Journal of Financial Economics* 65, 73-110.

[7] BANDI, F.M., and G. MOLOCHE (2004). On the Functional Estimation of Multivariate Diffusion Processes. *Working paper.*

[8] BANDI, F.M., and T. NGUYEN (1999). Fully Nonparametric Estimators for Diffusions: a Small Sample Analysis. *Working paper.*

[9] BANDI, F.M., and T. NGUYEN (2003). On the Functional Estimation of Jump-Diffusion Models. *Journal of Econometrics* 116, 293-328.

[10] BANDI, F.M., and P.C.B. PHILLIPS (2003). Fully Nonparametric Estimation of Scalar Diffusion Models. *Econometrica* 71, 241-283.

[11] BANDI, F.M., and P.C.B. PHILLIPS (2007). A Simple Approach to the Parametric Estimation of Potentially Nonstationary Diffusions. *Journal of Econometrics* 137, 354-395.

[12] BANDI, F.M., and R. RENO' (2008). Nonparametric Stochastic Volatility. *Working paper.*

[13] BANDI, F.M., and J.R. RUSSELL (2007). Volatility. In J. Birge and V. Linetski (Eds.), *Handbook of Financial Engineering*, Elsevier, 183-222.

[14] BIANCHI, A. (2007). Invariant Density Estimation for Multidimensional Diffusions. *Math everywhere.* Springer.

[15] BICKEL, P.J., and M. ROSENBLATT (1973). On Some Global Measures of The Deviations of Density Function Estimates. *Annals of Statistics* 1, 1071-1095.

[16] BLACK, F., DERMAN, E. and W. TOY (1990). A One-Factor Model of Interest Rates and its Applications to Treasury Bond Options. *Financial Analysts Journal* 46, 33-39.

[17] BOSQ, D. (1997). Temps Local et Estimation Sans Biais de la Densite' en Temps Continu*. C.R. Acad. Sci.,* T. 325, 1, 527-530.

[18] BOSQ, D. (1998). *Nonparametric Statistics for Stochastic Processes.* Springer-Verlag.

[19] BOSQ, D., and Y. DAVYDOV (1998). Local Time and Density Estimation in Continuous Time*. Publ. IRMA Univ. Lille 1, Vol. 44, III.*

[20] BRUGIÈRE, P. (1991). Estimation de la Variance d'un Processus de Diffusion dans le Cas Multidimensionnel. *C. R. Acad. Sci.,* T. 312, 1, 999-1005.

[21] BRUGIÈRE, P. (1993). Théorém de Limite Centrale Pour un Estimateur Non Paramétrique de la Variance d'un Processus de Diffusion Multidimensionnelle. *Ann. Inst. Henri Poincaré* 29, 357-389.

[22] CAI, Z., and Y. HONG (2003). Nonparametric Methods in Continuous-Time Finance: A Selective Review. In M.G. Akritas and D.N. Politis (Eds.), *Recent Advances and Trends in Nonparametric Statistics*. Elsevier, 283-302.

[23] CAMPBELL, J.Y., A. LO and A.C. MACKINLAY (1997). *The Econometrics of Financial Markets.* Princeton University Press.

[24] CHEN, X., L.P. HANSEN, and M. CARRASCO (1999). Nonlinearity and Temporal Dependence. *Working paper.*

[25] CHEN, S.X., J. GAO and C.Y. TANG (2008). A Test for Model Specification of Diffusion Processes. *Annals of Statistics*, forthcoming.

[26] COMTE, F., GENON-CATALOT, V. and Y. ROZENHOLC (2007). Nonparametric Estimation for a Stochastic Volatility Model. *Working paper.*

[27] CONLEY, T., L.P. HANSEN, E. LUTTMER, and J. SCHEINKMAN (1997). Short-term Interest Rates as Subordinated Diffusions. *Review of Financial Studies* 10, 525-577.

[28] CORRADI, V., and W. DISTASO (2007). Diagnostic tests for volatility models. *Working paper.*

[29] CORRADI, V., and H. WHITE (1999). Specification Tests for the Variance of a Diffusion. *Journal of Time Series Analysis* 20, 253-270.

[30] CORRADI, V., and N.R. SWANSON (2005). Bootstrap Specification Tests for Diffusion Processes. *Journal of Econometrics* 124, 117-148.

[31] COX, J.C., J.E. INGERSOLL and S.A. ROSS (1985). A Theory of the Term Structure of Interest Rates. *Econometrica* 53, 385-408.

[32] DALALYAN, A., and M. REISS (2006). Asymptotic Statistical Equivalence for Ergodic Diffusions: the Multidimensional Case. *Probability Theory and Related Fields* 137, 25-47.

[33] DARLING, D.A., and M. KAC (1957). On Occupation Times for Markoff Processes. *Transactions of the American Mathematical Society* 84, 444–458.

[34] DUFFIE, D. (1996). *Dynamic Asset Pricing Theory*. Princeton University Press.

[35] ERAKER, B., M. JOHANNES and N. POLSON (2003). The Impact of Jumps in Volatility and Returns. *Journal of Finance,* 1269-1300.

[36] ETHIER, S.N., and T.G. KURTZ (1986). *Markov Processes: Characterization and Convergence*. John Wiley & Sons.

[37] FAN, J. (1992). Design-Adaptive Nonparametric Regression. *Journal of the American Statistical Association* 87, 998-1004.

[38] FAN, J. (1993). Local Linear Regression Smoothers and their Minimax Efficiencies. *Annals of Statistics* 21, 196-216.

[39] FAN, J. (2005). A Selective Overview of Nonparametric Methods in Financial Econometrics. *Statistical Science* 20, 317-357.

[40] FAN, J., and I. GIJBELS (1996). *Local Polynomial Modelling and Its Applications.* Chapman and Hall.

[41] FAN, J., and Q. YAO (2003). *Nonlinear Time Series: Parametric and Nonparametric Methods*. Springer-Verlag.

[42] FAN, J., and C. ZHANG (2003). A Re-Examination of Diffusion Estimators with Applications to Financial Model Validation. *Journal of the American Statistical Association* 98, 118-134.

[43] FAN, J., J. JIANG, C. ZHANG, and Z. ZHOU (2003). Time-Dependent Diffusion Models for Term Structure Dynamics. *Statistica Sinica* 13, 965-992.

[44] FLORENS-ZMIROU, D. (1993). On Estimating the Diffusion Coefficient from Discrete Observations. *Journal of Applied Probability* 30, 790-804.

[45] GALLANT, R., and G. TAUCHEN (2009). Simulated Methods and Indirect Inference for Continuous-Time Models. *This volume.*

[46] GEMAN, D., and J. HOROWITZ (1978). Occupation Densities. *Annals of Probability* 1, 1-67.

[47] GEMAN, S.A. (1979). On a Common Sense Estimator for the Drift of a Diffusion. *Working Paper*.

[48] GENON-CATALOT, V., and J. JACOD (1993). On the Estimation of the Diffusion Coefficient for Multi-Dimensional Diffusion Processes. *Ann. Inst. H. Poincaré Probab. Statist.* 29, 119–151.

[49] GUERRE, E., 2007. Design-Adaptive Point-Wise Nonparametric Regression Estimation for Recurrent Markov Time Series. *Working paper.*

[50] GIKHMAN, I.I., and A.V. SKOROHOD (1972). *Stochastic Differential Equations*. Springer-Verlag.

[51] GOURIEROUX, C., and J. JASIAK (2001). *Financial Econometrics: Problems, Models and Methods.* Princeton University Press.

[52] HANSEN, L.P., and T.J. SARGENT (1983). The Dimensionality of the Aliasing Problem in Models with Rational Spectral Densities. *Econometrica* 50, 377-387.

[53] HANSEN, L.P., and J.A. SCHEINKMAN (1995). Back to the Future: Generating Moment Implications for Continuous-Time Markov Processes. *Econometrica* 63, 767-804.

[54] HÄRDLE, W. (1990). *Applied Nonparametric Regression*. Cambridge University Press.

[55] HÄRDLE, W., and O. LINTON (1994). Applied Nonparametric Methods. In D. McFadden and R.F. Engle (Eds.), *Handbook of Econometrics,* Vol. IV, Elsevier, 2295-2339.

[56] HEATH, D., R. JARROW and A. MORTON (1992). Bond Pricing and the Term Structure of Interest Rates: A New Methodology for Contingent Claims Valuation. *Econometrica* 60, 77-105.

[57] HONG, Y., and H. LI (2003). Nonparametric Specification Testing for Continuous-time Models with Application to Spot Interest Rates. *Review of Financial Studies* 18, 37-84.

[58] HÖPFNER, R., and E. LÖCHERBACH (2001). Limit Theorems for Null-Recurrent Markov Processes. *Preprint N. 22, Fachbereich Mathematik, Universität Mainz.*

[59] HÖPFNER, R., and Y. KUTOYANTS (2001). On a Problem of Statistical Inference in Null Recurrent Diffusions. *Preprint N. 7, Fachbereich Mathematik, Universität Mainz.*

[60] 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.*

[61] JACOD, J. (2009). Inference for Stochastic Processes. *This volume.*

[62] JEFFREY, A., D. KRISTENSEN, O. LINTON, T. NGUYEN, and P.C.B. PHILLIPS (2004). Nonparametric Estimation of a Multifactor Heath-Jarrow-Morton Model: An Integrated Approach. *Journal of Financial Econometrics* 2, 251-289.

[63] JIANG, G.J., and J. KNIGHT (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.

[64] JIANG, G.J., and J. KNIGHT (1999). Finite Sample Comparison of Alternative Estimators of Itô Diffusion Processes: A Monte-Carlo Study. *Journal of Computational Finance* 2, 5-38.

[65] JOHANNES, M. (2004). The Statistical and Economic Role of Jumps in Continuous-Time Interest Rate Models. *Journal of Finance* 59, 227-260.

[66] JOHANNES, M., and N. POLSON (2009). Numerical Bayesian Methods for Estimating Continuous-Time Models. *This volume.*

[67] KANAYA, S., and D. KRISTENSEN (2008). Estimation of Stochastic Volatility Models by Nonparametric Filtering. *Working paper.*

[68] KARATZAS, I., and S.E. SHREVE (1991). *Brownian Motion and Stochastic Calculus.* Springer-Verlag.

[69] KARLIN, S., and H.M. TAYLOR (1981). *A Second Course in Stochastic Processes.* Academic Press.

[70] KARLSEN, H.A., and D. TJØSTHEIM (2001). Nonparametric Estimation in Null Recurrent Time Series. *Annals of Statistics* 29, 372-416.

[71] KARLSEN, H.A., T. MYKLEBUST, and D. TJØSTHEIM (2007). Nonparametric Estimation in a Nonlinear Cointegrating Type Model. *Annals of Statistics* 35, 252-299.

[72] KHASMINSKII, R.Z. (1980). *Stochastic Stability of Differential Equations.* Sijthoff & Noordhoff.

[73] KHASMINSKII, R.Z. (2001). Limit Distributions of Some Integral Functionals for Null-Recurrent Diffusions. *Stochastic Processes and their Applications* 92, 1-9.

[74] KHASMINSKII, R.Z., and G. YIN (2000). Asymptotic Behavior of Parabolic Equations Arising from One-Dimensional Null-Recurrent Diffusions. *Journal of Differential Equations* 161, 154-173.

[75] KNIGHT, J., F. LI, and M. YUAN (2006). A semiparametric Two-Factor Term Structure Model. *Journal of Financial Econometrics* 4, 204-237.

[76] KRISTENSEN, D. (2007). Nonparametric Filtering of the Realized Spot Volatility: a Kernel-Based Approach. *Econometric Theory*, forthcoming.

[77] KRISTENSEN, D. (2008). Estimation of Partial Differential Equations with Applications in Finance. *Journal of Econometrics* 144, 392-408.

[78] LI, Q., and J. S. RACINE. (2006). *Nonparametric Econometrics: Theory and Practice.* Princeton.

[79] McCRORIE, J.R., (2009). Estimating Continuous-Time Models on the Basis of Discrete Data via an Exact Discrete Analog. *Econometric Theory*, forthcoming.

[80] MANCINI, C., and R. RENO' (2006). Threshold Estimation of Jump-Diffusion Models and Interest Rate Modeling. *Working paper.*

[81] MALLIAVIN, P., and M. MANCINO (2002). Fourier Series Method for Measurement of Multivariate Volatilities. *Finance and Stochastics* 6, 49-61.

[82] MENALDI, J.L., and M. ROBIN (1999). Invariant Measure for Diffusions with Jumps. *Applied Mathematics and Optimization* 40, 105-140.

[83] MEYN, S.P., and R.L. TWEEDIE (1993). *Markov Chains and Stochastic Stability.* Springer-Verlag.

[84] MERTON, R.C. (1990). *Continuous-Time Finance.* Blackwell.

[85] MOLOCHE, G. (2004a). Local Nonparametric Estimation of Scalar Diffusions. *Working paper.*

[86] MOLOCHE, G. (2004b). Kernel Regression for Non-Stationary Harris Recurrent Processes. *Working paper.*

[87] NICOLAU, J. (2003). Bias-Reduction in Nonparametric Diffusion Coefficient Estimation. *Econometric Theory* 19, 754-777.

[88] PAGAN, A., and A. ULLAH (1999). *Nonparametric Statistics.* Cambridge University Press.

[89] PARK, J. (2006). The Spatial Analysis of Time Series. *Working paper.*

[90] PIAZZESI, M. (2009). Affine Term-Structure Models. *This volume.*

[91] PHILLIPS, P.C.B. (1973). The Problem of Identification in Finite Parameter Continuous-Time Models. *Journal of Econometrics* 4, 351-362.

[92] PHILLIPS, P.C.B. (2001). Descriptive Econometrics for Nonstationary Time Series with Empirical Illustrations. *Journal of Applied Econometrics* 16, 389-413.

[93] PHILLIPS, P.C.B. (2004). Econometric Analysis of Fisher's Equation. In J. Geanakoplos and W.C. Brainard (eds.). Essays in Memory of Irving Fisher.

[94] PHILLIPS, P.C.B., and J. PARK (1998). Nonstationary Density Estimation and Kernel Autoregression. *Cowles Foundation Discussion Paper, No. 1181, Yale University.*

[95] PHILLIPS, P.C.B, and J. YU (2008). Maximum Likelihood and Gaussian Estimation of Continuous-Time Models in Finance. In R. Davis (Ed.), *Handbook of Financial Time Series*, forthcoming.

[96] POLLACK, M., and D. SIEGMUND (1985). A Diffusion Process and its Applications to Detecting a Change in the Drift of Brownian Motion. *Biometrika* 72, 267-280.

[97] PRAKASA-RAO, B.L.S. (1983). *Nonparametric Functional Estimation.* Academic Press.

[98] PROTTER, P. (1995). *Stochastic Integration and Differential Equations.* Springer-Verlag.

[99] RENO', R. (2006). Nonparametric Estimation of Stochastic Volatility Models. *Economic Letters* 90, 390-395.

[100] RENO', R. (2008). Nonparametric Estimation of the Diffusion Coefficient of Stochastic Volatility Models. *Econometric theory*, forthcoming.

[101] REVUZ, D., and M. YOR (1998). *Continuous Martingales and Brownian Motion.* Springer-Verlag.

[102] ROSENBLATT, M. (1975). Quadratic Measure of Deviation of Two-Dimensional Density Estimates and a Test of Independence. *Annals of Statistics* 3, 1-14.

[103] SCHIENLE, M. (2008). Nonparametric Nonstationary Regression. Doctoral Thesis, Mannheim University.

[104] SILVERMAN, B. (1986). *Density Estimation for Statistics and Data Analysis.* Chapman and Hall.

[105] STANTON, R. (1997). A Nonparametric Model of Term Structure Dynamics and the Market Price of Interest Rate Risk. *Journal of Finance* 52, 1973-2002.

[106] SUNDARESAN, S.M. (2000). Continuous-Time Methods in Finance: A Review and an Assessment. *Journal of Finance* 55, 1569-1622.

[107] TSAY, R.S. (2005). *Analysis of Financial Time Series.* John Wiley & Sons.

[108] VASICEK, O. (1977). An Equilibrium Characterization of the Term Structure. *Journal of Financial Economics* 5, 177-188.

[109] WANG, Q., and P.C.B. PHILLIPS (2008). Structural Nonparametric Cointegrating Regression. *Working paper.*

[110] WANG, Q., and P.C.B. PHILLIPS (2009). Asymptotic Theory for Local Time Density Estimation and Nonparametric Cointegrating Regression. *Econometric Theory,* forthcoming.

[111] XU, K.-L. (2007). Empirical Likelihood-Based Inference for Nonlinear Diffusions. *Working paper.*

[112] XU, K.-L. (2008). Re-Weighted Functional Estimation of Diffusion Models. *Working paper.*

[113] YAKOWITZ, S. (1989). Nonparametric Density and Regression Estimation for Markov Sequences Without Mixing Assumptions. *Journal of Multivariate Analysis* 30, 124-136.

[114] YOR, M. (1978). Sur la Continuité des Temps Locaux Associés à Certaines Semi-Martingales. *Astérisque* 52-53, 219-221.

[115] ZHANG, L., P. MYKLAND, and Y. AÏT-SAHALIA (2005). A Tale of Two Time Scales: Determining Integrated Volatility with Noisy High-Frequency Data. *Journal of the American Statistical Association* 100, 1394-1411.

Figure 1(a)
Spot variance local time

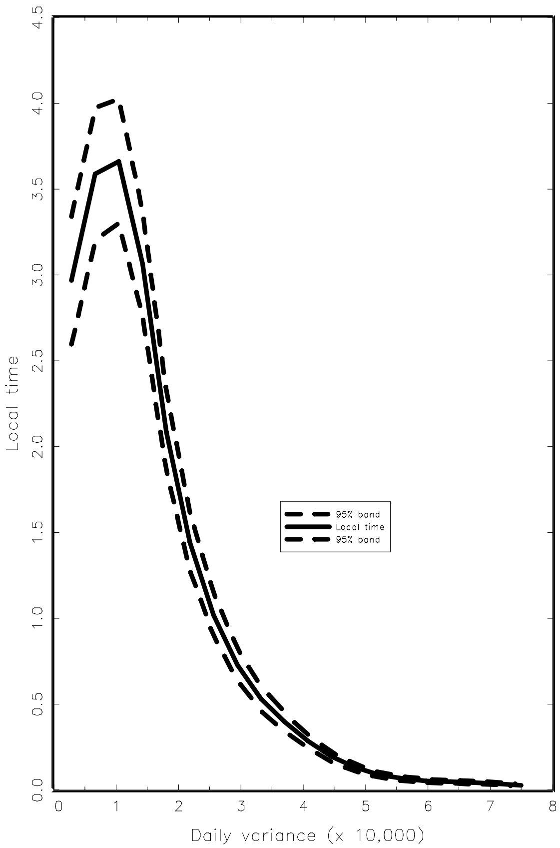

Figure 1(b)
Spot variance drift estimates

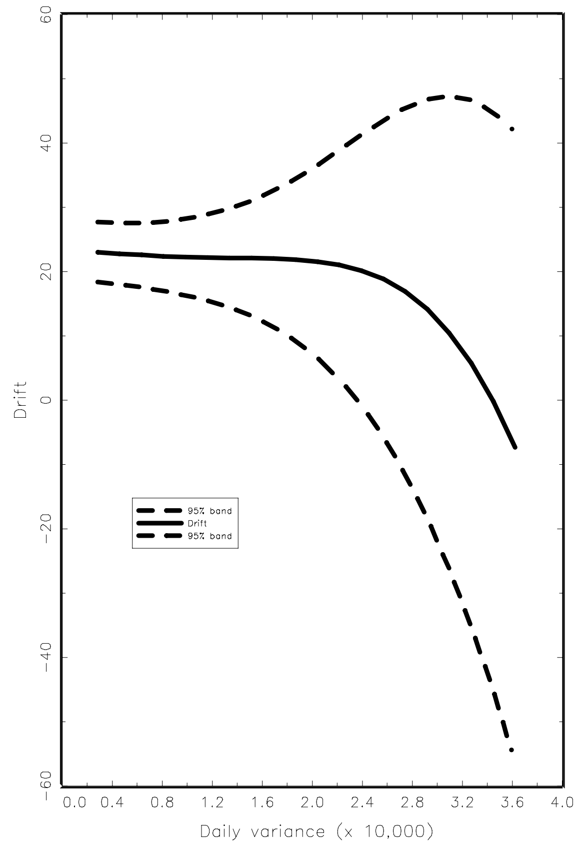

Figure 1(c)
Spot variance diffusion estimates

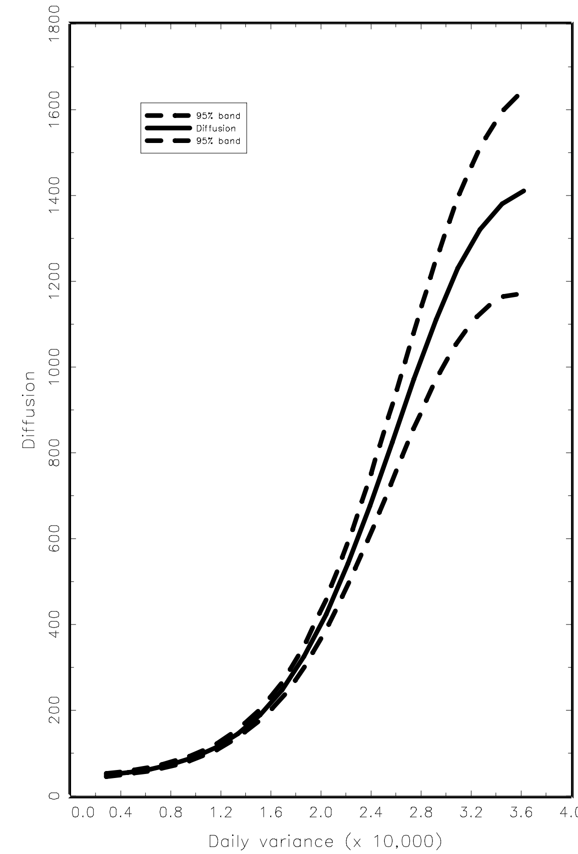

Figure 2(a)
Spot variance drift estimates

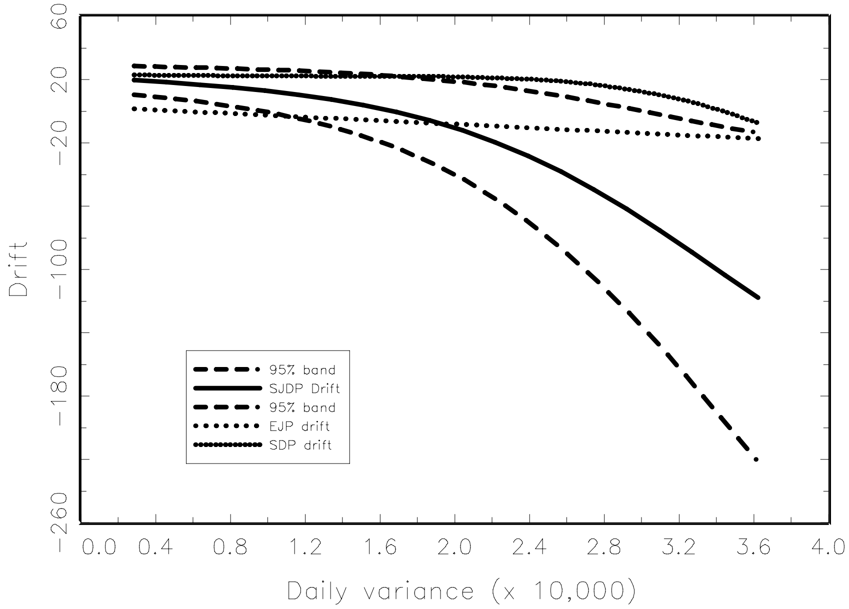

Figure 2(b)
Spot variance diffusion estimates

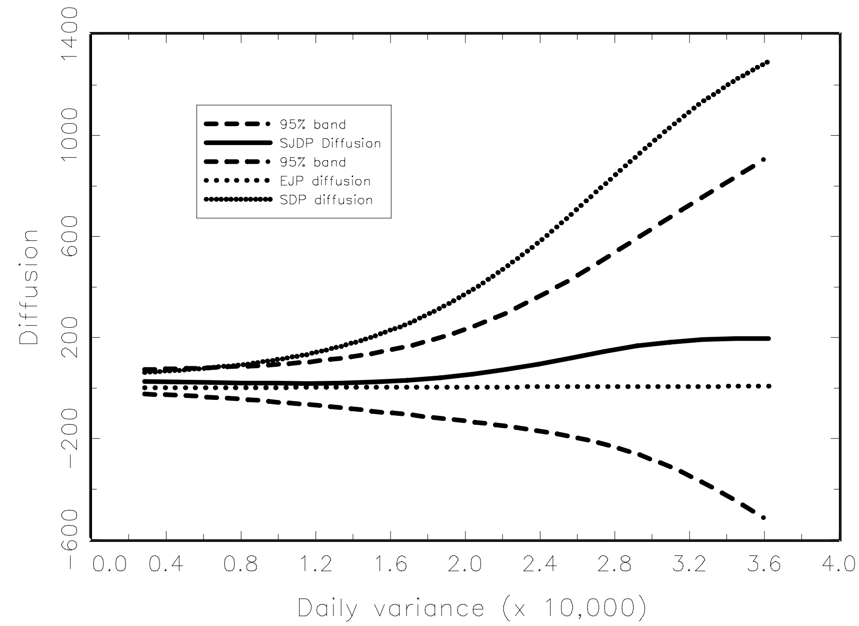

Figure 2(c)
Spot variance jump intensity

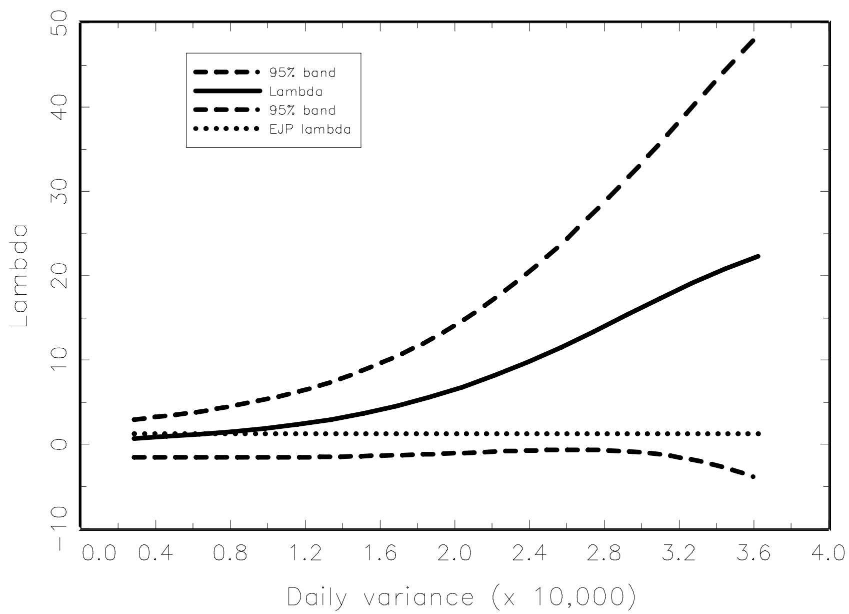

Figure 2(d)
Spot variance expected jump size