NONPARAMETRIC NONSTATIONARY REGRESSIONS IN CONTINUOUS TIME

SAMUELE CENTORRINO

Toulouse School of Economics

Abstract. This paper extends non parametric estimation to time homogeneous nonstationary diffusion processes where the drift and the diffusion coefficients are function of a multivariate exogenous time dependent variable Z. We base our estimation framework on a discrete sampling of data, following a recent stream of literature. We prove almost sure convergence and normal asymptotic distribution using the concept of multivariate occupation densities, in order to make the multivariate kernel estimation meaningful in the context of nonstationary time processes. We also provide an extension of this model when Z is a long memory process of dimension 1.

JEL Codes: C13, C14, C22.

Keywords: Nonstationary time series, Harris recurrence, Nonparametric estimation, Occupation density, Brownian motion, fractional Brownian motion.

E-mail address: `samuele.centorrino@univ-tlse1.fr`.

Date: September 30, 2011.

This paper has been presented at the 2011 Asian Meeting of the Econometric Society and at the 65th European Meeting of the Econometric Society. I thank Jean-Pierre Florens, Nour Meddahi, Jeffrey Racine and Dennis Kristensen for useful comments and remarks. I also thank participants to the 2^{nd} Humboldt-Copenhagen Conference in Financial Econometrics (Copenhagen, May 14-15, 2011) and to the XII Workshop in Quantitative Finance (Padova, January 27-28, 2011). Funding from MIUR (Italian Ministry of Education, Universities and Research) is gratefully acknowledged. All remaining errors are mine.

1. Introduction

In economics, a time homogeneous diffusion process in dimension one is often used to characterize the behaviour of a given variable Y_t, called the state variable (e.g., a stock price, the interest or the exchange rate). The structural model is written under the form:

$$dY_t = \mu(Y_t)dt + \sigma(Y_t)dB_t^* \tag{1.1}$$

where dB_t^* is the time increment of a standard Brownian motion, that is normally distributed with zero mean and variance equal to the time increment dt[1]. The two functions $\mu(Y_t)$ and $\sigma(Y_t)$ are called the drift and the diffusion coefficient, respectively.

This paper copes with a more general structural form of the model, where the drift and the diffusion coefficients can possibly be function of a time dependent variable Z_t. Our data generating process (DGP) can therefore be written in the following way:

$$dY_t = \mu(Y_t, Z_t)dt + \sigma(Y_t, Z_t)dB_t \tag{1.2}$$

where dB_t is the Brownian motion associated to the covariate depending process. This model can be interpreted as a general location scale model in continuous time. In particular, in this regression model, the objects of interest are both the location and the scale function.

This structural model is interesting in different respects. First of all, it generalizes to continuous Markov processes the economic idea that a given phenomenon may not be *self-explanatory*. Other factors may intervene in determining the outcome of the state today. This may be summarized in the concept of causality, which is central in econometrics but which has not yet been extended, to the best of our knowledge, to continous time diffusions. Furthermore, Z_t may be thought as a set of parameters which varies over time. The latent stochastic volatility model in continuous time can be therefore encompassed in this more general framework (e.g. see Bandi and Reno, 2009). Finally, the model is not reducible to a simplified multivariate diffusion. We

[1]For a review of the properties of a standard Brownian motion, see Karatzas and Shreve (1991) and Øksendal (2003)

are not assuming a covariance structure between two diffusion processes, but that a given process Z_t may directly intervene into the realization of $Y_{t+\delta}$. In that sense, we also allow for greater flexibility of the covariate process and we discuss a particular case in which Z_t exhibits long memory.

This structural model is not completely new to applied literature. Creedy and Martin (1994) and Creedy et al. (1996) develop a framework in which the variable Z represents market fundamentals that influence the behaviour of prices and US/UK exchange rate respectively[2]. In a more recent paper, Fernandes (2006) generalizes the same framework in order to supply a model for forecasting financial crashes. However, these papers use parametric method (e.g., maximum likelihood) solely to estimate and analyse the behaviour of the stationary conditional distribution of the state variable.

The novelty of this work is twofold. On the one hand, it extends the model used in applied literature to encompass exogenous time dependent covariates and provides clear assumptions on the covariate process Z which enable us to make inference on the scale and location functions. On the other hand, it focuses on nonparametric estimation of the structural models while relaxing the assumption of stationarity, following a recent stream of literature (Bandi and Phillips, 2003; Bandi and Nguyen, 2003, among others)[3].

Nonparametric estimation of stochastic diffusion processes hinges on a considerably rich literature. The main objects of interest being the drift and the diffusion coefficients, it may be difficult to identify them without further assumptions when the data are discretely sampled, because of the so-called *aliasing problem* (Phillips, 1973; Hansen and Sargent, 1983). Furthermore, while the drift term is of order dt, the diffusion term is of order $\sqrt{dt}$, which means that much of the infinitesimal variation in the process reflects the latter more than the former. This entails the impossibility to show consistency of the drift estimator as the sample frequency increases, i.e. $dt \to 0$ (so-called *infill* asymptotics).

[2]For a more recent application see also Jäger and Kostina (2005)

[3]Interested readers are referred to Bandi and Phillips (2010), for a complete review of the existing econometric literature on Nonparametric Estimation for Nonstationary Processes in Continuous Time.

A possible way to correctly identify both the diffusion and the drift coefficient is to assume that the process is time stationary, so that a time invariant density $\pi(y)$ exists. The *backward* and the *forward* Kolmogorov equations allow then to specify a relation between this density, the drift and the diffusion coefficients.

Nevertheless, the assumption of stationarity seems somehow too restrictive and it does not take into account many interesting phenomena in economics. Relaxing the assumption of stationarity requires careful handling of kernel estimators, which is not meaningful any more as an estimator of the invariant density. An interpretation of the kernel estimator in time series, both in the univariate and multivariate case, may be given in terms of occupation densities (Geman and Horowitz, 1980). Namely, in the univariate case, Phillips and Park (1998) show the convergence of the nonparametric kernel estimator to the *chronological* local time of the stochastic process (see, e.g. Revuz and Yor, 1999, Ch. VI, for a review of the properties of local time).

Bandi and Phillips (2003) are then able to overcome the identification issues without assuming stationarity. *Harris recurrence*, which is a substantially milder assumption, is required instead. To ensure consistency of the drift term, they couple *infill* asymptotics with lengthening time span of observations, i.e. $T \to \infty$ (so-called *long span* asymptotics).

In related papers, Löcherbach and Loukianova (2008) and Bandi and Moloche (2008) use the same framework under the assumption of Harris recurrence for the joint process to prove convergence of such an estimator in the multivariate case.

In this paper, we show that their convergence results can be extended to the nonparametric estimator of the drift and the diffusion in model (1.2).

However, while we show the properties of our estimation for any dimension d of the covariate process, we run simulations for the case in which $d = 1$. As pointed out by Schienle (2011), Harris recurrence is a property which is rarely satisfied when the dimension of the process increases. We do not tackle this question here, as it goes beyond the scope of the present paper. We therefore acknowledge the limited applicability of this framework that may be a topic for further research.

The paper is structured as follows. Section 2 set up the general framework. Section 3 overviews the theoretical foundations on which this work is based upon. Section

4 provides the main estimation framework and the asymptotic properties. Section 5 discusses an extension to long memory processes. Finally, section 6 includes a simulation study which draws the finite sample properties of the estimator.

2. Motivations and theoretical foundations

The possibility to meaningfully define conditional moments for continuous time processes is a necessary condition to perform statistical inference based on sample analogues. Diffusion type processes are very useful in this sense, as the definition of conditional moments is straightforward under the Markov property. Moreover, the behaviour of a diffusion is fully described by its location and scale parameters, which are the objects of interest. The goal of this section is therefore to show that, under suitable assumptions on the conditional and the marginal process, we can make our data generating process being a diffusion process.

We suppose here to observe a multivariate continuous time process $\{Z_t : t \geq 0\}$ of given dimension d; and a scalar process $\{Y_t : t \geq 0\}$ which is Markov conditionally on Z_t. We denote by X_t the joint process $\{Y_t, Z_t\}$ which takes value in a Polish space $(E, \mathcal{E})$.

Define $(\Omega_z, \mathcal{Z}, \mathcal{P}_z)$ and $\{\mathcal{Z}_t\}_{t\geq 0}$ the probability space and the natural filtration associated to the process Z_t, respectively.

We further consider a univariate Brownian motion $\{B_t : t \geq 0\}$ defined on the probability space $(\Omega_B, \mathcal{F}^B, \mathcal{P}_B)$ and adapted to a filtration $\{\mathcal{F}_t^B\}_{t\geq 0}$. We assume B_t to be a $\mathcal{Z}_t$−adapted martingale, so that $\mathbb{E}\left[dB_t|\mathcal{Z}_t\right] = 0$.

The joint filtration, generated by the process $\{X_t : t \geq 0\}$ is set as follows:

$$\mathcal{X}_t := \mathcal{Y}_t \vee \mathcal{Z}_t = \sigma(y) \vee \mathcal{Z}_t \vee \mathcal{F}_t^B = \sigma\left(y, Z_s, B_s; 0 \leq s \leq t\right) \tag{2.1}$$

We assume all filtrations satisfy the *usual conditions* (or hypotheses), i.e. they contain all the sets of zero measure for $t = 0$ and they are *right-continuous*.

In our framework, the filtration generated by the process Z_t enters the construction of the filtration under which the process Y_t is defined. To ensure exogeneity of the joint process, we apply the following definition:

Definition 2.1 (STRONG GLOBAL NONCAUSALITY, Florens and Fougere, 1996)**.** $\mathcal{X}_t$ does not strongly cause Z_t given $\mathcal{Z}_s$ if:

$$\mathcal{Z}_t \perp\!\!\!\perp \mathcal{X}_s | \mathcal{Z}_s \qquad \forall s, t \in [0, T] \qquad \blacksquare$$

This properties is trivially satisfied if $t \leq s$. Nevertheless, if $\mathcal{X}_t$ does not strongly cause Z_t, every $\mathcal{Z}_t$-adapted martingale is also a $\{\mathcal{X}_t\}$-martingale (Florens and Fougere, 1996, Theorem 2.2).

The assumption of strong global noncausality is simply stating that, conditionally on the observation of the process Z at time s, the joint process is not delivering any additional information about the marginal process Z_t, $\forall t$. However, the most important implication of this hypothesis is that it immediately entails the preservation of the martingale property of B_t under the joint filtration.

It is also important to notice that, in this context, the assumption of global noncausality is equivalent to the assumption of instantaneous noncausality (in a Granger sense) and to any other noncausality assumption, as $\mathcal{Z}$ is also the conditioning filtration (see, Comte and Renault, 1996; Florens and Fougere, 1996). Therefore, using the most restrictive assumption of noncausality only serves maintaining the martingale property.

Remark 1. To clarify the meaning of our assumption, consider a simple linear autoregressive model in discrete time.

$$Y_t = \alpha Y_{t-1} + \beta Z_{t-1} + \varepsilon_t$$

If $X = (Y, Z)$, strong global noncausality is equivalent to:

$$\mathbb{E}(\varepsilon_t | X_{t-1}) = 0 \quad \forall t \geq 0$$

i.e. a strict exogeneity assumption on the two regressors. $\blacksquare$

Under the conditional markovianity of Y_t and noncausality, we can give to our regression model the attribute of a stochastic differential equation (Karatzas and Shreve, 1991). The *conditional* diffusion process is thus defined as::

$$dY_t = \mu(Y_t, Z_t)dt + \sigma(Y_t, Z_t)dB_t \tag{2.2}$$

where $\mu(\cdot,\cdot) : \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}$ and $\sigma(\cdot,\cdot) : \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}$, which are our objects of interest in what follows. This model can be considered as an extension of the conditional mean model studied in Park (2008). We extend his model in two respects. First of all, we allow the volatility term to also depend on Z_t. Second, we allow for any (possibly nonlinear) specification of the drift and the diffusion term[4].

For ease of notations, we write our DGP as follows:

$$dY_t = \mu(X_t)dt + \sigma(X_t)dB_t \tag{2.3}$$

where X_t denotes the joint process.

Remark 2. As an example, consider a Ornstein-Uhlembeck process, where the drift function μ is linear and the diffusion function is a constant:

$$dY_t = (\theta_1(Z_t) - \theta_2 Y_t)\, dt + \theta_3 dB_t$$

where $\theta_2 > 0$ (so that the process is mean reverting) and any function θ_1 of Z_t. ■

Remark 3 (Stochastic volatility model). We consider the stochastic volatility model without jumps (Bandi and Reno, 2009):

$$dY_t = \mu(\sigma_t^2)dt + \sigma_t^2 dB_t^r$$
$$df(\sigma_t^2) = m_f(\sigma_t^2)dt + \Lambda_f(\sigma_t^2)dB_t^\sigma$$

where $\{B_t^r, B_t^\sigma\}$ are possibly correlated Brownian motions.

The assumption of noncausality in such a model, when $\{B_t^r, B_t^\sigma\}$ are independent Wiener processes, is discussed in Comte and Renault (1996). In this context, the assumption serves to generate some instantaneous noncausality between the latent variable σ_t and the log return Y_t.

Nevertheless, when we introduce some dependence between the two Brownian terms, local instantaneous noncausality disappears. As a consequence, any stronger assumption of noncausality would fail. However, the mild assumption that

$$\mathbb{E}\left(dB_t^r | \mathcal{Y}_t \vee \sigma(\sigma_t)\right) = 0$$

[4]Park (2008) considers as an error term in his model any continuous martingale with bounded variations. However, up to a time change, any continuous martingale can be rewritten as a *Dambis-Dubins-Schwarz* Brownian motion.

would still make the model fit into our more general framework. ■

We assume the following conditions to hold in studying (2.3).

Assumption 1. *The functions $\mu(\cdot)$ and $\sigma(\cdot)$ satisfy the following assumptions:*

(i) *They are measurable on the σ-field generated by all the Borel sets on $\mathcal{E}$ and they are at least twice continuously differentiable with respect to both their arguments;*

(ii) *They satisfy local Lipschitz and growth conditions in X_t, i.e. for every compact set $B \in \mathcal{E}$, there exists a constant, C, such that, for any realization x_1 and x_2 in B,*

$$\|\mu(x_1) - \mu(x_2)\| + \|\sigma(x_1) - \sigma(x_2)\| \leq C\|x_1 - x_2\| \tag{2.5}$$

and

$$\|\mu(x_1)\|^2 + \|\sigma(x_1)\|^2 \leq C^2\left(1 + \|x_1\|^2\right) \tag{2.6}$$

(iii) *Nondegeneracy (ND) - $\sigma^2(\cdot) > 0$ on $\mathcal{E}$*

(iv) *Local Integrability (LI) with respect to Y_t, for any realization of the process $Z_t = z$:*

$$\forall (y_1, z) \in E, \exists \delta > 0 \quad \textit{such that} \quad \int_{y_1-\delta}^{y_1+\delta} \frac{|\mu(\zeta, z)| d\zeta}{\sigma^2(\zeta, z)} < \infty \qquad ■ \tag{2.7}$$

Conditions (ii) and (iii) (Karatzas and Shreve, 1991, Theorem 2.2, p. 289) ensure the existence of a strong solution to equation (2.3). We can therefore write the usual Itô's stochastic differential equation, which is the solution of our DGP in the following form:

$$Y_t = y + \int_0^t \mu(X_s) ds + \int_0^t \sigma(X_s) dB_s \tag{2.8}$$

where y is an initial condition independent of the Brownian motion B_t and Y_t is adapted to the filtration $\mathcal{Y}_t \vee \mathcal{Z}_t$.

The drift and the diffusion coefficients can be thus defined as in the standard framework. Take any function $f \in \mathbb{C}^2$ of Y_t, so to preserve the semimartingale properties of our solution (see Protter, 2003, Theorem 32, p. 174). Using Itô's lemma and taking expectation over any couple of realizations (y,z), the infinitesimal generator $\mathcal{L}$ of equation (1.2) can be defined as:

$$\lim_{t\to 0}\frac{1}{t}\mathbb{E}^x\left[f(Y_t)-f(y)\right]=(\mathcal{L}f)(y) \tag{2.9}$$

Taking $f(Y_t) = Y_t$, we obtain the drift coefficient as the conditional instantaneous change in the process:

$$\mathbb{E}^x\left[Y_t-y\right]=t\mu(x)+o(t) \tag{2.10}$$

while, taking $f(Y_t) = (Y_t - y)^2$, we obtain the diffusion coefficient as the conditional instantaneous change in the volatility of the process,

$$\mathbb{E}^x\left[(Y_t-y)^2\right]=t\sigma^2(x)+o(t) \tag{2.11}$$

We can then proceed as in any standard nonparametric inference problem for conditional moments, using sample analogues to identify conditional expectations over infinitesimal time distances. In practise, the exogenous case is encompassed in the existing literature for stochastic processes. In the next sessions, we show that the asymptotic properties of the drift and the diffusion term are equivalent to those of a multivariate diffusion when the dimension is equal to $d+1$.

3. Additive Functionals and Occupation Density

Before to explicitly derive the nonparametric estimators of the drift and the diffusion coefficient, we need to set up the main definitions and theorems which allow us to meaningfully define a standard kernel estimator in such a nonstationary context.

We assume the following conditions about the joint process to hold.

Assumption 2. (i) *X_t is Harris recurrent;*

(ii) *Under $\mathcal{X}_t$, X_t is a special semi-martingale and it admits a Doob-Meyer decomposition of the type:*

$$X_t = H_t + M_t \qquad \forall t\in(0,T]$$

where H_t is a $\mathcal{X}_t$-predictable process and M_t is a $\mathcal{X}_t$-local martingale such that $\mathbb{E}(M_t|\mathcal{X}_s)=0,\forall s<t$. ■

In particular, since every $\mathcal{X}_t$-martingale can be written as a time changed *Dambis-Dubins-Schwarz* Brownian motion (Revuz and Yor, 1999, Ch. V, Theorem 1.6), X_t is a Brownian semimartingale.

Condition (i) is the minimal requirement to perform nonparametric inference on the joint process. It is possible to show that conditional stationarity of Y given Z and Harris recurrence of Z are sufficient conditions to obtain Harris recurrence of the joint process (see the Appendix). However, it is not possible to assume a more general structure on the conditional process and still to obtain condition (i).

Remark 4. The conditional Ornstein-Uhlembeck process has a stationary distribution given $Z = z$. The stationary density of the conditional process is Gaussian with mean equal to $\frac{\theta_1(z)}{\theta_2}$ and variance $\frac{\theta_3^2}{2\theta_2}$. Therefore, for any process Z_t which satisfies the assumption of Harris recurrence, the joint process would fit assumption (2), (i). ■

Remark 5. It is worth noticing as well that, conditions (i) and (ii) together entail the markovianity of X_t. These assumptions, coupled with the definition of noncausality 2.1, imply that Z_t is a Markov process (Florens et al., 1990). In section 5, we slightly relax this assumption and propose a solution to extend our main framework to long memory processes. ■

For any measurable Borel set $B \subset \mathcal{E}$, we choose a measure m. This measure is invariant if and only if,

$$m(B) = \int_E \mathbb{P}\left(X_t^{(x)} \in B\right) m(dx) \tag{3.1}$$

where $X_t^{(x)}$ denotes the realization of the joint process at time t for a given initial condition x. In particular, Harris recurrence is a sufficient condition for the existence of an invariant measure, unique up to multiplication by a constant and absolutely continuous with respect to the Lebesgue measure λ on $\mathbb{R}^{d+1}$ (i.e. $m \ll \lambda$). The absolute continuity of m further implies that m admits a density with respect to the Lebesgue measure, i.e. a random function $p_t(\cdot)$ such that $m(dx) = p_t(x)\lambda(dx)$.

Definition 3.1 (Höpfner and Löcherbach, 2003)**.** An additive functional of X is a process $A = (A_t)_{t\geq 0}$, such that:

(i) A is $\mathcal{X}$−adapted, $A_0 = 0$;

(ii) All paths of A are nondecreasing and right-continuous;

(iii) For all $s, t \geq 0$, we have $A_{t+s} = A_t + A_s * \theta_t$, where θ_t is a family of shift operators for X. ■

We focus our attention here to integrable additive functionals. For every Borel set B, the measure ν defined by the functional A for each t is equal to:

$$\nu_A(B) = \mathbb{E}_m\left(\int_0^1 \mathbb{1}_B(X_s)dA_s\right) = \frac{1}{t}\mathbb{E}_m\left(\int_0^t \mathbb{1}_B(X_s)dA_s\right)$$

A functional is termed *integrable* when:

$$\| \nu_A \| = \nu_A(E) = \mathbb{E}_m(A_1) < \infty$$

In particular, when the functional $A_t = t$, for each Borel set B, we can define:

$$\eta_t^B = \int_0^t \mathbb{1}_B(X_s)ds \quad , \quad t \geq 0$$

which heuristically counts the amount of times for which X_s belong to B, for $T \to \infty$. In this particular case, we obtain that:

$$\mathbb{E}_m\left(\int_0^1 \mathbb{1}_B(X_s)ds\right) = m(B)$$

which defines the *occupation measure* for the set B (Geman and Horowitz, 1980), i.e. the time spent by the process in the set B up to time t. Therefore, the measure defined by the constant functional on each subset of $\mathcal{E}$ is equivalent to the invariant measure of X_t. Since the invariant measure admits a density with respect to the Lebesgue measure, there exists a random function $p_t(\cdot)$, such that:

$$m(B) = \int_B p_t(x)\lambda(dx)$$

We define , following this terminology, $p_t(\cdot)$ to be the *occupation density* of X. In dimension 1, the invariant measure is defined to be the *sojourn* time of a given process X (Park, 2005), while the random function $p_t(x)$ corresponds to the local time of the process (Borodin, 1989). This is formally defined as the Radon-Nykodim derivative of the sojourn time with respect to the Lebesgue measure. Our approach can be thus considered a generalization of the univariate case.

Remark 6. For the stationary case we have that:

$$\int p_t(x)\lambda(dx) = 1$$

so that $p_t(x) = \pi(x)$ is the invariant stationary density of X_t. ■

The following theorem gives the condition for weak convergence of additive functionals of a Harris recurrent process X:

Theorem 3.2 (Höpfner and Löcherbach, 2003). *For a given constant $\alpha \in (0,1]$ and a function $l(\cdot)$ slowly varying at infinity*[5]*, the following are equivalent:*

(i) *For every nonnegative measurable function $g(\cdot)$ with $0 < m(g) < \infty$, one has regular variation at 0 of resolvants*[6] *in X if*

$$(R_{1/t}g)(x) = \mathbb{E}_x\left(\int_0^\infty e^{-\frac{1}{t}s}g(X_s)ds\right) \quad \sim \quad \frac{t^\alpha}{l(t)}m(g) \quad , \quad t \to \infty \tag{3.2}$$

(ii) *every additive functional A of X with $0 < \mathbb{E}_m(A_1) < \infty$, one has:*

$$\frac{(A_t)_{t\geq 0}}{t^\alpha/l(t)} \to \mathbb{E}_m(A_1)W^\alpha \qquad as \quad t \to \infty \tag{3.3}$$

under the Skorohod topology, where W^α is the Mittag-Leffler process of index α[7]. ■

Remark 7. Equation 3.2 simply states that we are restricting our attention to null recurrent diffusions with regular variation of the resolvent at 0. In the more general case, one should define the kernel estimator for any function $v_t = \mathbb{E}_m\left[\int_0^t g(X_s)ds\right]$ (Löcherbach and Loukianova, 2008). In our case we take $v_t = t^\alpha/l(t)$. Moreover, equation 3.2 is equivalent to the condition given by Bandi and Moloche (2008, Theorem 2), where $C_X = m(g) < \infty$. ■

---

[5] A function $f : [a,\infty) \to (0,\infty)$, $a > 0$ is said to be slowly-varying at infinity in the sense of Karamata if $\lim_{x\to\infty} f(\lambda x)/f(x) \to 1$, for $\lambda > 0$.

[6] For $\alpha > 0$ and a continuously differentiable function $g(\cdot)$, we define the resolvent operator R_α, by $(R_\alpha g)(x) = \mathbb{E}_x\left(\int_0^\infty e^{-\alpha s}g(X_s)ds\right)$. $R_\alpha g$ is a bounded continuous function (Øksendal, 2003, Definition 8.1.2 and Lemma 8.1.3, pg. 135).

[7] Interested readers are referred to Höpfner and Löcherbach (2003), for general definition and properties of Mittag-Leffler processes.

Remark 8. For stationary processes, we simply set $\alpha = 1$, $l(t) = 1$ and the Mittag-Leffler process $W^1 = Id$ (the deterministic process) by definition. Thus, for any measurable bounded function $f(\cdot)$, we obtain convergence by equation 3.3, i.e.:

$$\frac{1}{T}\int_0^T f(X_s)ds \xrightarrow{p} \int f(x)\pi(x)dx = \mathbb{E}(f(x))$$

where $\pi(\cdot)$ is the invariant stationary probability density. ■

4. Estimation and Asymptotic Properties

For simplicity, we suppose that the process $\{X_t, t \geq 0\}$ is sampled at equispaced times in the interval $[0, T]$, where T is a strictly positive number. If n is the sample size in $[0, T]$, we obtain that the time lag between two observations is equal to $\Delta_{n,T} = \frac{T}{n}$. The observed sample is therefore denoted as $X_{i\Delta_{n,T}}$ for all $i = 1, \cdots, n$.

Under these hypotheses and following the definitions given in equations (2.10) and (2.11), we can estimate the drift and the diffusion coefficients as follows:

$$\hat{\mu}_{n,T}(x) = \frac{1}{\Delta_{n,T}} \frac{\frac{1}{h_{n,T}^{d+1}} \sum_{i=1}^{n} \mathbf{K}_{h_{n,T}}\left(X_{i\Delta_{n,T}} - x\right)\left(Y_{(i+1)\Delta_{n,T}} - Y_{i\Delta_{n,T}}\right)}{\frac{1}{h_{n,T}^{d+1}} \sum_{i=1}^{n} \mathbf{K}_{h_{n,T}}\left(X_{i\Delta_{n,T}} - x\right)} \tag{4.1}$$

$$\hat{\sigma}^2_{n,T}(x) = \frac{1}{\Delta_{n,T}} \frac{\frac{1}{h_{n,T}^{d+1}} \sum_{i=1}^{n} \mathbf{K}_{h_{n,T}}\left(X_{i\Delta_{n,T}} - x\right)\left(Y_{(i+1)\Delta_{n,T}} - Y_{i\Delta_{n,T}}\right)^2}{\frac{1}{h_{n,T}^{d+1}} \sum_{i=1}^{n} \mathbf{K}_{h_{n,T}}\left(X_{i\Delta_{n,T}} - x\right)} \tag{4.2}$$

where,

$$\mathbf{K}_{h_{n,T}}\left(X_{i\Delta_{n,T}} - x\right) = K\left(\frac{Y_{i\Delta_{n,T}} - y}{h_{n,T}^{(y)}}\right)\prod_{j=1}^{d} K\left(\frac{Z_{j,i\Delta_{n,T}} - z_j}{h_{n,T}^{(z)}}\right)$$

where $h_{n,T}^{(y)}$ and $h_{n,T}^{(z)}$ are two bandwidths parameters for the process Y_t and Z_t respectively. For notational brevity, we also suppose that $h_{n,T}^{(y)} = h_{n,T}^{(z)}$. For further ease of notations, we denote $x = (y, z)$.

The kernel functions $\mathbf{K}(\cdot)$ and $K(\cdot)$ satisfy the following conditions.

Assumption 3. - *(Pagan and Ullah, 1999; Bandi and Moloche, 2008; Ruppert and Wand, 1994)*

(i) *The function $K(\cdot)$ is a non negative, bounded, continuous, and symmetric function such that:*

$$\int_{-\infty}^{\infty} K(u)du = 1 \qquad \int_{-\infty}^{\infty} K^2(u)du < \infty \quad \text{and} \quad \int_{-\infty}^{\infty} u^2 K(u)du < \infty$$

(ii) *The function $\mathbf{K}(\cdot)$ is a compactly supported, bounded kernel, such that $\int uu'\mathbf{K}(u)du = \rho_2(\mathbf{K})\mathbb{I}$, where $\rho_2(\mathbf{K}) \neq 0$ is a scalar and $\mathbb{I}$ is the identity matrix of dimension $d+1$.*

(iii) *Additionally, there exists a non negative function $D(v,\varepsilon)$ such that:*

$$|\mathbf{K}(x) - \mathbf{K}(v)| \leq D(v,\varepsilon)\|x - v\| \tag{4.3}$$

$\forall x, v \in \mathbb{R}^{d+1}$ so that $\|x - v\| < \varepsilon$. Furthermore,

$$\lim_{\varepsilon \to 0} \int D(v,\varepsilon)dv < \infty \tag{4.4}$$

and

$$\lim_{\varepsilon \to 0} \int D(v,\varepsilon)m(dv) < \infty \qquad \forall \varepsilon < \infty \qquad \blacksquare \tag{4.5}$$

While many of these assumptions are standard in the nonparametric literature, assumption *(iii)* deserves some additional discussions. The multivariate kernel function is often supposed to satisfy a regularity condition, e.g. some Hölder type of continuity. However, in the nonstationary case, any function which satisfies such a kind of uniform continuity will explode as $T \to \infty$, when it is integrated with respect to time. Therefore, we need to bound the kernel function by a function whose integral is defined when it is integrated with respect to the invariant measure.

Under assumption (3), we can thus define the kernel estimator of the occupation density of X:

$$\hat{L}^X(T,x) = \frac{\Delta_{n,T}}{h_{n,T}^{d+1}} \sum_{i=1}^{n} \mathbf{K}_{h_{n,T}}\left(X_{i\Delta_{n,T}} - x\right) \tag{4.6}$$

Using theorem 3.2, it is possible to show the weak convergence of this estimator towards the Radon-Nikodym derivative of m with respect to the Lebesgue measure on R^{d+1}.

Corollary 4.1. *Consider the following additive functional of X_s:*

$$\Phi_t = \int_0^t \frac{1}{h_{n,T}^{d+1}} \mathbf{K}_{h_{n,T}} \left(X_s - x\right) ds$$

which is strictly positive and integrable $\forall t \geq 0$. The kernel estimator (4.6) converges almost surely to Φ_t for $n, T \to \infty$, provided that:

$$\frac{\hat{L}^X(T,x)}{h_{n,T}^{d+1}} \left(\Delta_{n,T} \log(1/\Delta_{n,T})\right)^{1/2} \xrightarrow{a.s.} 0$$

Moreover, when $h_{n,T} \to 0$, we obtain:

$$\frac{\Phi_t}{t^\alpha / l(t)} \to C p_\infty(x) W^\alpha \qquad as \quad t \to \infty$$

by theorem 3.2, where C is a process specific constant.

Proof. See the Appendix. ∎

Remark 9. Under stationarity, (4.6) is a well defined estimator of the stationary density, as $\frac{\hat{L}^X(T,x)}{T} \xrightarrow{p} \pi(x)$. ∎

Remark 10. The estimator presented here has been firstly proposed by Bandi and Moloche (2008) and it is a generalization to multivariate processes of the local time estimator for scalar diffusion process presented in Florens-Zmirou (1993).

4.1. **Estimation and asymptotic distribution of the drift coefficient.** In this section we report the convergence properties of the drift estimator.

Theorem 4.2. Almost sure convergence of the drift estimator.
Suppose that:

$$\frac{\hat{L}^X(T,x)}{h_{n,T}^{d+1}} \left(\Delta_{n,T} \log(1/\Delta_{n,T})\right)^{1/2} \xrightarrow{a.s.} 0$$

with $\hat{L}_X(T,x) h_{n,T}^{d+1} \to \infty$ with $\Delta_{n,T} \to 0$, $h_{n,T} \to 0$ and $n, T \to \infty$, then the estimator of equation (4.1) converges almost surely to the drift coefficient. I.e.:

$$\hat{\mu}_{n,T}(x) \xrightarrow{a.s.} \mu(x) \tag{4.7}$$

Proof. See the Appendix. ∎

Theorem 4.3. Asymptotic distribution of the drift estimator.
Suppose that:

$$\frac{\hat{L}^X(T,x)}{h_{n,T}^{d+1}}\left(\Delta_{n,T}\log(1/\Delta_{n,T})\right)^{1/2} \xrightarrow{a.s.} 0$$
$$\hat{L}^X(T,x)h_{n,T}^{d+1} \xrightarrow{a.s.} \infty$$

with $h_{n,T} = O_{a.s}\left(\hat{L}^X(T,x)^{-\frac{1}{d+1}}\right)$, $\Delta_{n,T} \to 0$, $h_{n,T} \to 0$ *and* $n, T \to \infty$, *then the estimator described in equation (4.1) converges in distribution to a Gaussian random variable.*

$$\begin{aligned}\sqrt{\hat{L}^X(T,x)h_{n,T}^{d+1}}\left(\hat{\mu}_{n,T}(x) - \mu(x) - \Gamma^\mu(x)\right)\\ \xrightarrow{d} \sigma(x)\mathcal{N}\left(0, \left(\int \mathbf{K}^2(u)du\right)\right)\end{aligned} \tag{4.8}$$

where $\Gamma^\mu(x)$ *is a bias term, equal to:*

$$\Gamma^\mu(x) = h_{n,T}^2\rho_2(\mathbf{K})\left(tr\left\{\mathcal{D}_{\mu,p}(x)\right\} + \frac{1}{2}tr\left\{\mathcal{H}_\mu(x)\right\}\right) \tag{4.9}$$

where,

$$\mathcal{H}_\mu(x) = \left(\frac{\partial^2\mu(x)}{\partial x_j \partial x_l}\right)_{j,l=1}^{d+1} \qquad \mathcal{D}_{\mu,p}(x) = \left(\frac{\partial\mu(x)}{\partial x_j}\frac{\partial p_t(x)}{\partial x_l}\right)_{j,l=1}^{d+1}$$

Instead if, everything being equal:

$$\hat{L}^X(T,x)h_{n,T}^{d+5} \xrightarrow{a.s.} 0$$

the bias term disappears asymptotically.

Proof. See the Appendix. ∎

Remark 11. The random speed of covergence of the drift estimator depends on the occupation density of the joint process. This is a natural consequence of considering the occupation density as the number of visits of the process in a small set which diverges to infinity as the time span grows. Therefore, the higher the dimension d of the covariate process, the slower the speed of convergence. Together with the standard dimensionality problem in nonparametric statistics, Bandi and Moloche (2008) refer to it as *double curse of dimensionality.*

Remark 12 (Bandwidth choice). The asymptotic mean squared error (AMSE) is equal to:

$$O\left(h_{n,T}^4\right) + O\left(\frac{1}{h_{n,T}^{d+1}\hat{L}^X(T,x)}\right)$$

This suggests the bandwidth parameter for the drift term being set proportionally to $\hat{L}^X(T,x)^{-\frac{1}{d+5}}$. As already pointed out in related papers, drift bandwidth selection is locally adapted in order to account for the number of visits to the point in which the estimation is performed.

Remark 13 (Stationary case). In the stationary case, we showed that $\hat{L}^X(T,x) \xrightarrow{p} T\pi(x)$. Therefore, our result can be restated as follows:

$$\sqrt{Th_{n,T}^{d+1}}\left(\hat{\mu}_{n,T}(x) - \mu(x) - \Gamma^{\mu}(x)\right) \xrightarrow{d} \sigma(x)\mathcal{N}\left(0, \left(\frac{\int \mathbf{K}^2(u)du}{\pi(x)}\right)\right)$$

as $Th_{n,T}^{d+1} \to \infty$. The bias term is now equal to:

$$\frac{h_{n,T}^2}{\pi(x)}\rho_2(\mathbf{K})\left(tr\left\{\left(\frac{\partial\mu(x)}{\partial x_j}\frac{\partial\pi(x)}{\partial x_l}\right)_{j,l=1}^{d+1}\right\} + \frac{1}{2}tr\left\{\left(\frac{\partial^2\mu(x)}{\partial x_j\partial x_l}\right)_{j,l=1}^{d+1}\right\}\right)$$

This is a standard results in conditional moments estimation (see, e.g. Pagan and Ullah, 1999, p. 101).

4.2. **Estimation and asymptotic distribution of the diffusion coefficient.** In this section we report the convergence properties of the diffusion estimator.

Theorem 4.4. Almost sure convergence of the diffusion estimator. *Suppose that:*

$$\frac{\hat{L}^X(T,x)}{h_{n,T}^{d+1}}\left(\Delta_{n,T}\log(1/\Delta_{n,T})\right)^{1/2} \xrightarrow{a.s.} 0$$

with $\Delta_{n,T} \to 0$, $h_{n,T} \to 0$ *and* $n, T \to \infty$, *then the estimator of equation (4.2) converges almost surely to the diffusion coefficient. I.e.:*

$$\hat{\sigma}_{n,T}^2(x) \xrightarrow{a.s.} \sigma^2(x) \tag{4.10}$$

Proof. See the Appendix. ■

Theorem 4.5. Asymptotic distribution of the diffusion estimator. *Suppose that:*

$$\frac{\hat{L}^X(T,x)}{h_{n,T}^{d+1}}\left(\Delta_{n,T}\log(1/\Delta_{n,T})\right)^{1/2} \xrightarrow{a.s.} 0$$

$$\hat{L}^X(T,x)h_{n,T}^{d+1} \xrightarrow{a.s.} \infty$$

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

$$\sqrt{\frac{h_{n,T}^{d+5}\hat{L}^X(T,x)}{\Delta_{n,T}}} \xrightarrow{a.s.} 0$$

then the estimator described in equation (4.2) converges in distribution to a Gaussian random variable.

$$\sqrt{\frac{\hat{L}^X(T,x)h_{n,T}^{d+1}}{\Delta_{n,T}}}\left(\hat{\sigma}_{n,T}^2(x) - \sigma^2(x)\right) \xrightarrow{d} 2\sigma^2(x)\mathcal{N}\left(0, \left(\int \mathbf{K}^2(u)du\right)\right) \tag{4.11}$$

If, instead,

$$\sqrt{\frac{h_{n,T}^{d+5}\hat{L}^X(T,x)}{\Delta_{n,T}}} = O_{a.s.}(1)$$

then, there is an asymptotic bias term $\Gamma^{\sigma^2}(x)$*, equal to:*

$$\Gamma^{\sigma^2}(x) = h_{n,T}^2\rho_2(\mathbf{K})\left(tr\left\{\mathcal{D}_{\sigma^2,p}(x)\right\} + \frac{1}{2}tr\left\{\mathcal{H}_{\sigma^2}(x)\right\}\right) \tag{4.12}$$

where,

$$\mathcal{H}_{\sigma^2}(x) = \left(\frac{\partial^2\sigma^2(x)}{\partial x_j \partial x_l}\right)_{j,l=1}^d \qquad \mathcal{D}_{\sigma^2,p}(x) = \left(\frac{\partial\sigma^2(x)}{\partial x_j}\frac{\partial p_t(x)}{\partial x_l}\right)_{j,l=1}^d$$

Proof. See the Appendix. ∎

Remark 14. It is also possible to identify the diffusion term for any fixed time horizon T. This has been already pointed out in Bandi and Moloche (2008) and goes back to a result first shown in Brugière (1993). The general results can also be applied to our setting. In the fixed T case, if one is ready to assume that:

$$h_{n,T}^{d+1} = O_{a.s.}\left(\sqrt{\Delta_{n,T}\log(1/\Delta_{n,T})}\right)$$

it is possible to show the consistency and asymptotic normality of the diffusion estimator. In particular, for $\Delta_{n,T}, h_{n,T}^{d+1} \to 0$ and $n \to \infty$, it is possible to show that:

$$\sqrt{\frac{h_{n,T}^{d+1}}{\Delta_{n,T}}}\left(\hat{\sigma}_{n,T}^2(x) - \sigma^2(x)\right) \sim MN\left(0, \frac{4\sigma^4(x)}{\hat{L}^X(T,x)}\right)$$

where, MN denotes a mixed normal distribution, with mixing factor $\hat{L}^X(T,x)$. ∎

Remark 15. The asymptotic mean squared error (AMSE) is equal to:

$$O\left(h_{n,T}^{4}\right) + O\left(\frac{\Delta_{n,T}}{h_{n,T}^{d+1}\hat{L}^{X}(T,x)}\right)$$

This suggests to use again an adaptive scheme to set the bandwidth for the diffusion term. In particular, we oversmooth in areas that are less visited by the process and undersmooth in areas that are often visited. The diffusion bandwidth is therefore set proportionally to $\left(\frac{\hat{L}^{X}(T,x)}{\Delta_{n,T}}\right)^{-\frac{1}{d+5}}$. However, as long as the diffusion term can be identified for fixed T, we can also choose a constant bandwidth which is going to be proportional to $n^{-1/(d+5)}$. ■

5. An extension to long memory processes

The results presented so far are obtained under the assumption that the joint process X_t is a Markov process. However, it is possible to extend this model to allow for the marginal process Z_t to be a long memory process (e.g. fractional Brownian motion, fBM, or stochastic differential equations driven by a fBM), at least when Z_t is defined on the real line.

The problem which arises in this case is that processes driven by fBM are not semi-martingales and are not Markov[8]. Therefore our assumption 2 would completely fail.

Let $\{B_t^H, t \geq 0\}$ to be a fBM, with Hurst parameter equal to $H \in (0,1)$ and suppose that Z_t follows a stochastic differential equation driven by a B_t^H,

$$Z_t = \int_0^t \psi(s)ds + \int_0^t \xi(s)dB_t^H$$

where $\{\psi(t), t \geq 0\}$ is a $\mathcal{Z}_t$-adapted process and $\xi(t)$ is a non-vanishing deterministic function. Although Z_t is not a semimartingale in this case, one can associate to it a semi-martingale $\{J_t, t \geq 0\}$, called the *fundamental semi-martingale* such that the natural filtration $\mathcal{J}_t$ of the process J coincides with $\mathcal{Z}_t$ (Kleptsyna et al., 2000). Therefore, one can perform inference on Y_t in model 2.3 using J_t instead of Z_t without losing any information.

---

[8]For an extensive review of the properties of fBM and stochastic diffusions driven by fBM (see, e.g. Biagini et al., 2008; Rao, 2010)

Define, for $0 < s < t$:

$$\begin{aligned} k_H(t,s) &= \kappa_H^{-1} s^{\frac{1}{2}-H}(t-s)^{\frac{1}{2}-H}, \qquad \kappa_H = 2H\Gamma\left(\frac{3}{2}-H\right)\Gamma\left(H+\frac{1}{2}\right) \\ w_t^H &= \lambda_H^{-1} t^{2-2H}, \qquad \lambda_H = \frac{2H\Gamma(3-2H)\Gamma\left(H+\frac{1}{2}\right)}{\Gamma\left(\frac{3}{2}-H\right)} \\ M_t^H &= \int_0^t k_H(t,s)\, dB_t^H \end{aligned}$$

where M_t^H is referred to as the *fundamental martingale* associated to the fBM B_t^H, whose quadratic variation is nothing but the function w_t^H (Norros et al., 1999).

Finally suppose that the sample paths of the function $\xi^{-1}(t)\psi(t)$ are smooth enough and define:

$$Q_t^H = \frac{d}{dw_t^H}\int_0^t k_H(t,s)\,\xi^{-1}(s)\psi(s)ds, \quad t \in [0,T]$$

We can therefore define the process J_t as:

$$J_t = \int_0^t k_H(t,s)\,\xi^{-1}(s)dZ_s$$

such that (see Kleptsyna et al., 2000):

(i) J_t is a semi-martingale which admits the following decomposition:

$$J_t = \int_0^t Q_t^H(s)dw_s^H + M_t^H$$

(ii) Z_t admits a representation as a stochastic integral with respect to J_t.

(iii) the natural filtrations $\mathcal{Z}_t$ and $\mathcal{J}_t$ coincide.

We can therefore define the joint process $X_t^* = (Y_t, J_t)$ defined on the natural filtration $\mathcal{X}_t^*$. Under the *fundamental semi-martingale* result and definition 2.1 of noncausality, the filtrations $\mathcal{X}_t$ and $\mathcal{X}_t^*$ coincide.

This equivalence between the two filtrations allows us to perform inference on Y_t by means of the process X_t^*, as long as the information carried by Z_t and J_t is the same. We can therefore restate assumption 2 as follows:

Assumption 2a. (i) *$X_t^* \in \mathbb{R}^2$ is Harris recurrent.*

(ii) *Under $\mathcal{X}_t^*$, X_t^* is a special semi-martingale and it admits a Doob-Meyer decomposition of the type:*

$$X_t^* = H_t^* + M_t^* \qquad \forall t \in (0,T]$$

where H_t^* *is a* $\mathcal{X}_t^*$*-predictable process and* M_t^* *is a* $\mathcal{X}_t^*$*-local martingale such that* $\mathbb{E}(M_t^*|\mathcal{X}_s^*) = 0, \forall s < t$. ∎

Under this assumption, our inference results can be used to deal with the case of Z_t being a long memory process in $\mathbb{R}$.

The two following equations would be used to theoretically identify the drift and the diffusion coefficient:

$$\mathbb{E}^{x^*}\left[Y_t - y\right] = t\mu(x) + o(t) \tag{5.1}$$

$$\mathbb{E}^{x^*}\left[(Y_t - y)^2\right] = t\sigma^2(x) + o(t) \tag{5.2}$$

where $x^* = (y, j)$. Under assumption 2a, we can apply the same estimation technique and asymptotic theory presented in previous sections.

6. Simulations

Notwithstanding the curse of dimensionality problem which is common to nonparametric inference and which can be even more severe in the case of nonstationary diffusion processes, because of the random divergence of the occupation density, we provide here a simulation study in which the diffusion process is a function of a scalar covariate Z. This is the minimal framework that can be use to prove the reliability of our estimation procedure in finite samples. Programming has been conducted in Matlab and codes are available upon request.

We consider the following true data generating processes:

$$dY_t^{(1)} = \left(\theta_1(Z_t) - \theta_2 Y_t^{(1)}\right) dt + dB_t^{(1)} \tag{6.1a}$$

$$dY_t^{(2)} = \left(\theta_1(Z_t) - \theta_2 Y_t^{(2)}\right) dt + \zeta \left(Y_t^{(2)} + Z_t\right) dB_t^{(2)} \tag{6.1b}$$

where $\theta_2 = 2$ and $\zeta = 0.4$. The former process is a generalization of a Ornstein-Uhlenbeck process, where the drift only is function of Z and the diffusion is a constant (taken equal to one for simplicity); while the latter is a CKLS model (Chan et al., 1992), generalized to encompass the dependence on the covariate. The process Z has been taken as follows:

$$Z_t^{(1)} = E_t \tag{6.2a}$$

$$Z_t^{(2)} = B_t^{H=0.2} \tag{6.2b}$$

$$Z_t^{(3)} = B_t^{H=0.7} \tag{6.2c}$$

where $\{E_t\}_{t\geq 0}$ is a standard Wiener process and $\{B_t^H\}_{t\geq 0}$ is a fractional Brownian motion, with Hurst index equal to 0.2 and 0.7, respectively. Namely, the latter numerical schemes have been chosen to assess the performance of our estimate where Z is a long memory process. For the sake of simplicity, we consider $\theta_1(Z_t) = Z_t^2$ in all replications. We draw 250 paths of the processes in (6.1a) and (6.1b), using a Milstein scheme which reaches an order of approximation equal to one (Iacus, 2008).

Remark 16. Following Phillips (1973), because of the *aliasing problem* in the estimation of stochastic diffusions, when data are discretely sampled, it is not possible to identify a nonlinear drift without imposing any structural restrictions on the model. In our simulating equations, structural restrictions are coming both from the additive form of the drift and from the dependence on Z. ■

The goal of this exercise is to recover an estimate of the functional form of $\theta_1(\cdot)$.

If we hope to correctly identify both the drift and the diffusion term, we have to construct a finite sample in which dt is sufficiently small and T is sufficiently large. We therefore set $\Delta_{n,T} = 1/52$ and $n = 4800$. In practical application, this would imply weekly observations over roughly 100 years time span. However, the scope of this exercise is to check that our estimators have desirable properties. Research on the applicability of this method is in progress.

To the best of our knowledge, there is not a general theory for choosing a bandwidth parameter to estimate the occupation density of multidimensional nonstationary processes in continuous time. Moreover, the bandwidth parameter depends on the recurrence properties of the underlying stochastic process which are difficult to assess. Following Schienle (2011), we set the bandwidth according to an adaptive scheme. For each evaluation point, we count the number of neighbours in a small interval around that point. That is, for a fixed interval I_j around the point x_j:

$$h_{n,T}(x_j) = \left(\sum_{i=0}^{n} \mathbb{1}(X_{i\Delta_{n,T}} \in I_j) \right)^{-\frac{1}{q+4}} \tag{6.3}$$

where q is the dimension of the joint process (Y_t, Z_t). The estimators for the drift and the diffusion coefficient have been computed using (2.10) and (2.11), respectively. In order to recover the functional form of $\theta_1(\cdot)$, a semiparametric method has been applied. In particular, we first project the estimated drift on Y_t and Z_t using a simple linear regression model. We obtain a first estimate of θ_2, say $\hat{\theta}_2^{(1)}$. We then use this estimate to compute:

$$\hat{\theta}_1^{(1)}(z) = \frac{\sum_{i=1}^{n-1} K_h\left(Z_{i\Delta_{n,T}} - z\right)\left(\hat{\mu}(Z_{i\Delta_{n,T}}, Y_{i\Delta_{n,T}}) - \hat{\theta}_2^{(1)} Y_{i\Delta_{n,T}}\right)}{\sum_{i=1}^{n} K_h\left(Z_{i\Delta_{n,T}} - z\right)}$$

We then plug the nonparametric estimate into the first step regression in order to get a new value of θ_2, say $\hat{\theta}_2^{(2)}$, and we iterate until convergence.

The drift bandwidth parameter has been set according to the theoretical proportionality rule i.e.:

$$h_{n,T}^{dr} = \hat{L}^X(T, x)^{-\frac{1}{d+5}}$$

Remark 17. Bandi and Moloche (2008) suggest applying a correction factor in order to undersmooth and center at zero the asymptotic distribution. However, we do not find this correction factor having any impact in our simulation study. ■

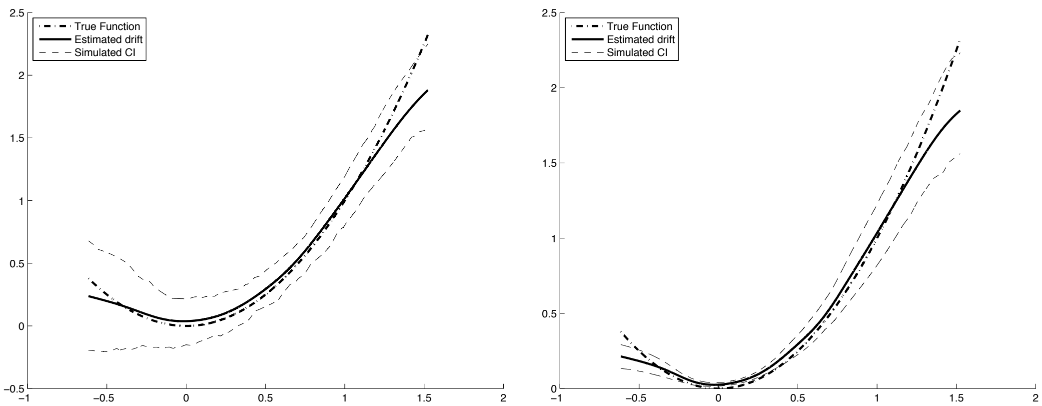
The diffusion bandwidth has instead been taken constant and proportional to the sample size. That is:

$$h_{n,T}^{df} = n^{-\frac{1}{d+5}}$$

We report separately the results for the estimation of the drift, for models 6.1a and 6.1b. We also draw simulated confidence bands over the interval 2.5% – 97.5%.

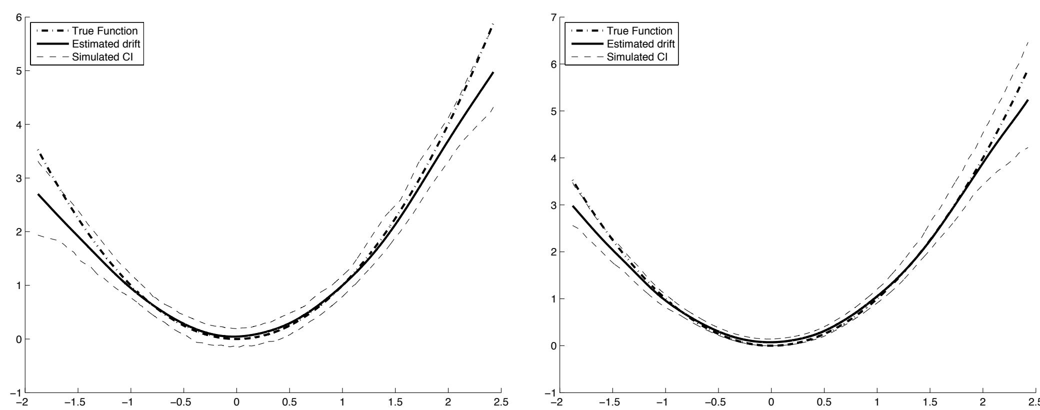
As it can be seen from figures 1 and 2, the drift estimation is rather satisfactory, despite a poorer behaviour at the boundaries.

FIGURE 1. Estimation of $\theta_1(\cdot)$ when Z_t is drawn from 6.2a, with 100 simulated paths.



(a) Model 6.1a (b) Model 6.1b

FIGURE 2. Estimation of $\theta_1(\cdot)$ when Z_t is drawn from 6.2b, with 100 simulated paths.



(a) Model 6.1a (b) Model 6.1b

7. CONCLUSIONS

This paper delivers a new structural model, where the causal relation between an endogenous variable and a set of covariates is brought to continuous time diffusions. Our main asymptotic results refine the current literature on the topic. We also show that it is possible to extend this framework in the case when Z_t is a long memory process of dimension 1. This may have potential applications in finance, especially for stochastic volatility model but it can also be a new explorative tool for macroeconomic variables, such as the interest and the exchange rates.

Current research is focusing on relaxing the assumption of joint Harris recurrence, by imposing an additivity property on the drift and the diffusion coefficients. In the same way, we are trying to extend nonparametric inference in models where the assumption of strict exogeneity of Z may be dropped (Florens and Simon, 2010).

Another interesting line of research that has been set in a paper by Phillips and Tyurin (1999) would be to explore the properties of nonparametric estimators of the local time of fractional Brownian motion. This is an open question both in economics and statistical mathematics which can have huge impacts on the theory presented in the current paper and, more generally, to explore nonparametric inference in stochastic systems fully driven by fBM.

References

Azéma, J., M. Duflo, and D. Revuz (1969): "Mesure Invariante des Processus de Markov récurrents," *Séminaires de probabilité (Strasbourg)*, 3, 24–33.

Bandi, F. M. and G. Moloche (2008): "On the Functional Estimation of Multivariate Diffusion Processes," *Working Paper, University of Chicago.*

Bandi, F. M. and T. H. Nguyen (2003): "On the functional estimation of jump-diffusion models," *Journal of Econometrics*, 116, 293 – 328.

Bandi, F. M. and P. C. B. Phillips (2003): "Fully Nonparametric Estimation of Scalar Diffusion Models," *Econometrica*, 71, 241–283.

——— (2010): "Nonstationary Continuous-Time Processes," in *Handbook of Financial Econometrics*, ed. by Y. Ait-Sahalia and L. P. Hansen, Elsevier.

Bandi, F. M. and R. Reno (2009): "Nonparametric Stochastic Volatility," Global coe hi-stat discussion paper series, Institute of Economic Research, Hitotsubashi University.

Biagini, F., Y. Hu, B. Øksendal, and T. Zhang (2008): *Stochastic calculus for fractional Brownian motion and applications*, Probability and its applications, Springer.

BILLINGSLEY, P. (1979): *Probability and Measure*, Wiley series in Probability and mathematical statistics., Wiley.

BORODIN, A. (1989): "Brownian Local Time," *Russ. Math. Surv.*, 44.

BRUGIÈRE, P. (1993): "Théorème de limite centrale pour un estimateur non paramétrique de la variance d'un processus de diffusion multidimensionnelle." *Probabilités et Statistiques*, 3, 357–389.

CHAN, K. C., G. A. KAROLYI, F. A. LONGSTAFF, AND A. B. SANDERS (1992): "An Empirical Comparison of Alternative Models of the Short-Term Interest Rate," *The Journal of Finance*, 47, pp. 1209–1227.

COMTE, F. AND E. RENAULT (1996): "Noncausality in Continuous Time Models," *Econometric Theory*, 12, 215–256.

CREEDY, J., J. LYE, AND V. L. MARTIN (1996): "A Non-Linear Model of the Real US/UK Exchange Rate," *Journal of Applied Econometrics*, 11, 669–686.

CREEDY, J. AND V. L. MARTIN (1994): "A Model of The Distribution of Prices," *Oxford Bulletin of Economics and Statistics*, 56, 67–76.

FERNANDES, M. (2006): "Financial Crashes as endogenous jumps: estimation, testing and forecasting," *Journal of Economic Dynamics and Control*, 30, 111–141.

FLORENS, J., M. MOUCHART, AND J. ROLIN (1990): *Elements of Bayesian statistics*, Pure and Applied Mathematics, M. Dekker.

FLORENS, J. P. AND D. FOUGERE (1996): "Noncausality in Continuous Time," *Econometrica*, 64, 1195–1212.

FLORENS, J.-P. AND G. SIMON (2010): "Endogeneity and Instrumental Variables In Dynamic Models," *Working Paper.*

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

GEMAN, D. AND J. HOROWITZ (1980): "Occupation Densities," *The Annals of Probability*, 8, 1–67.

HANSEN, L. P. AND T. J. SARGENT (1983): "The Dimensionality of the Aliasing Problem in Models with Rational Spectral Densities," *Econometrica*, 51, 377–387.

HÖPFNER, R. AND E. LÖCHERBACH (2003): *Limit Theorems for Null Recurrent Markov Processes*, American Mathematical Society.

Iacus, S. (2008): *Simulation and inference for Stochastic Differential Equations: with R examples*, Springer Series in Statistics, Springer.

Jäger, S. and E. Kostina (2005): "Parameter estimation for forward Kolmogorov equation with application to nonlinear exchange rate dynamics," *PAMM*, 5, 745–746.

Karatzas, I. and S. Shreve (1991): *Brownian Motion and Stochastic Calculus*, Springer-Verlag.

Kleptsyna, M., A. Le Breton, and M.-C. Roubaud (2000): "Parameter Estimation and Optimal Filtering for Fractional Type Stochastic Systems," *Statistical Inference for Stochastic Processes*, 3, 173–182.

Löcherbach, E. and D. Loukianova (2008): "On Nummelin Splitting for Continuous Time Harris Recurrent Markov Processes and Application to Kernel Estimation for Multi-dimensional Diffusions," *Stochastic Processes and their Applications*, 118, 1301 – 1321.

McKean, H. P. (1969): *Stochastic Integrals*, Academic Press, Inc.

Norros, I., E. Valkeila, and J. Virtamo (1999): "An Elementary Approach to a Girsanov Formula and Other Analytical Results on Fractional Brownian Motions," *Bernoulli*, 5, pp. 571–587.

Øksendal, B. (2003): *Stochastic Differential Equations: an Introduction with Applications*, Universitext (1979), Springer.

Pagan, A. and A. Ullah (1999): *Nonparametric Econometrics*, Cambridge University Press.

Park, J. Y. (2005): "The Spatial Analysis of Time Series," Discussion papers, Indiana University.

——— (2008): "Martingale Regression and Time Change," *Working Paper*.

Phillips, P. C. and J. Y. Park (1998): "Nonstationary Density Estimation and Kernel Autoregression," Cowles Foundation Discussion Papers 1181, Cowles Foundation, Yale University.

Phillips, P. C. B. (1973): "The problem of Identification in Finite Parameter Continuous Time Models," *Journal of Econometrics*, 1, 351 – 362.

Phillips, P. C. B. and W. Ploberger (1996): "An Asymptotic Theory of Bayesian Inference for Time Series," *Econometrica*, 64, 381–412.

Phillips, P. C. B. and K. Tyurin (1999): "The Occupation Density of fractional Brownian Motion and some of its Applications," *Working Paper, Cowles Foundation.*

Protter, P. E. (2003): *Stochastic Integration and Differential Equations, 2nd ed.*, Springer-Verlag.

Rao, B. (2010): *Statistical Inference for Fractional Diffusion Processes*, Wiley Series in Probability and Statistics, John Wiley & Sons.

Revuz, D. (1984): *Markov chains*, North-Holland mathematical library, North-Holland.

Revuz, D. and M. Yor (1999): *Continous Martingale and Brownian Motion*, Springer-Verlag.

Ruppert, D. and M. P. Wand (1994): "Multivariate Locally Weighted Least Squares Regression," *The Annals of Statistics*, 22, 1346–1370.

Schienle, M. (2011): "Nonparametric Nonstationary Regression with Many Covariates," Discussion papers, Humboldt University.

8. Appendix

8.1. General Definitions, Corollaries and Theorems.

Definition 8.1 (Harris Recurrence Azéma et al., 1969). A strongly Markov process X taking values in a Polish space $(E,\mathcal{E})$ is Harris recurrent, if there exists some $\sigma-$finite measure m on $(E,\mathcal{E})$, such that:

$$m(A) > 0 \Rightarrow \forall x \in E \quad : \qquad P_x\left(\int_0^\infty \mathbb{1}_A(X_s)ds = \infty\right) = 1$$

This process is also called $m - irreducible$. ■

Definition 8.2 (Höpfner and Löcherbach, 2003). A Harris recurrent process X, taking values in a Polish space $(E,\mathcal{E})$, with invariant measure m is called positive recurrent (or ergodic) if $m(E) < \infty$, null recurrent if $m(E) = \infty$. ■

Theorem 8.3 (Ratio Limit Theorem Azéma et al., 1969). *If a process X is Harris recurrent with invariant measure m and A and B are two integrable additive functionals and if $\| \nu_B \| > 0$, then:*

(i) $\lim_{t\to\infty} \frac{\mathbb{E}_x(A_t)}{\mathbb{E}_x(B_t)} = \frac{\|\nu_A\|}{\|\nu_B\|} \qquad m - a.s.,$

(ii) $\lim_{t\to\infty} \frac{A_t}{B_t} = \frac{\|\nu_A\|}{\|\nu_B\|} \qquad P_x - a.s.,\ \forall x.$

■

Definition 8.4 (Modulus of Continuity of Multivariate Brownian Semimartingales). Suppose X is a special multivariate Brownian semimartingale, and denote:

$$\kappa_{n,T} = \sup_{|t-s|<\Delta_{n,T},[0\leq s<t\leq T]} |X_t - X_s|$$

to be its modulus of continuity. We can then write (McKean, 1969):

$$\mathbb{P}\left[\lim \sup_{\Delta_{n,T}\to 0} \frac{\kappa_{n,T}}{\sqrt{\Delta_{n,T}\left(1/\Delta_{n,T}\right)}} = \max_{t\leq T} \sqrt{2\gamma(X_t)}\right] = 1$$

where $\gamma(X_t)$ is the biggest eigenvalue of the covariance matrix of the process X. ■

8.2. **Proof of Lemma (4.1).** We want to prove that:

$$\frac{\Delta_{n,T}}{h_{n,T}^{d+1}}\sum_{i=1}^{n}\mathbf{K}_{h_{n,T}}\left(X_{i\Delta_{n,T}}-x\right)\xrightarrow{a.s.}\frac{1}{h_{n,T}^{d+1}}\int_0^T\mathbf{K}_{h_{n,T}}\left(X_s-x\right)ds$$

We start by writing:

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

by the triangle inequality and assumption (3). Finally using the *Ratio Limit theorem*, we have that:

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

By theorem (3.2), we now have that, for $n,T\to\infty$:

$$\frac{\frac{1}{h_{n,T}^{d+1}}\int_0^T\mathbf{K}_{h_{n,T}}\left(X_s-x\right)ds}{t^\alpha/l(t)}\to\mathbb{E}_m\left(\frac{1}{h_{n,T}^{d+1}}\int_0^T\mathbf{K}_{h_{n,T}}\left(X_s-x\right)ds\right)W^\alpha$$

Therefore, to prove our final result, we only need to prove that:

$$\mathbb{E}_m\left(\frac{1}{h_{n,T}^{d+1}}\int_0^T\mathbf{K}_{h_{n,T}}\left(X_s-x\right)ds\right)=Cp_t(x)\tag{8.1}$$

By the strong version of the Ratio Limit Theorem, for any couple of integrable functions $f(\cdot)$ and $g(\cdot)$, we have that:

$$\frac{\mathbb{E}_m(f)}{\mathbb{E}_m(g)}=\frac{m(f)}{m(g)}$$

which implies:

$$\mathbb{E}_m(f)=Cm(f)\qquad\text{where}\qquad C=\frac{m(g)}{\mathbb{E}_m(g)}$$

We can then write:

$$\mathbb{E}_m\left(\frac{1}{h_{n,T}^{d+1}}\int_0^T \mathbf{K}_{h_{n,T}}\left(X_s-x\right)ds\right)=C\int_{\mathcal{E}}\frac{1}{h_{n,T}^{d+1}}\mathbf{K}_{h_{n,T}}\left(X_s-x\right)m(dX_s)$$
$$=\int_{\mathcal{E}}\frac{1}{h_{n,T}^{d+1}}\mathbf{K}_{h_{n,T}}\left(X_s-x\right)p_\infty(X_s)\lambda(dX_s)=\int_{\mathcal{E}}\frac{1}{h_{n,T}^{d+1}}\mathbf{K}(u)p_\infty(uh_{n,T}+x)\lambda(h_{n,T}du)$$
$$=\int_{\mathcal{E}}\mathbf{K}(u)p_\infty(uh_{n,T}+x)\lambda(du)$$

where we use the continuity of m wrt λ and the properties of the Lebesgue measure (Billingsley, 1979, Theorem 12.2, p.172). Finally, as $h_{n,T}^{d+1}\to 0$:

$$\int_{\mathcal{E}}\mathbf{K}(u)p_t(uh_{n,T}^{d+1}+x)\lambda(du)\to p_\infty(x)\int_{\mathcal{E}}\mathbf{K}(u)\lambda(du)=p_\infty(x)$$

By the relation between Riemann and Lebesgue integration and assumption (3). This concludes the proof.

8.3. **Proof of Theorem (4.2).** We want to prove that:

$$\hat{\mu}_{n,T}(x)\xrightarrow{a.s.}\mu(x)$$

We start by writing the drift estimator of equation (4.1) as follows:

$$\hat{\mu}_{n,T}(x)$$
$$=\frac{1}{\Delta_{n,T}}\frac{\frac{\Delta_{n,T}}{h_{n,T}^{d+1}}\sum_{i=1}^{n}\mathbf{K}_{h_{n,T}}\left(X_{i\Delta_{n,T}}-x\right)\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}\mu(X_s)ds}{\frac{\Delta_{n,T}}{h_{n,T}^{d+1}}\sum_{i=1}^{n}\mathbf{K}_{h_{n,T}}\left(X_{i\Delta_{n,T}}-x\right)} \tag{8.2}$$

$$+\frac{1}{\Delta_{n,T}}\frac{\frac{\Delta_{n,T}}{h_{n,T}^{d+1}}\sum_{i=1}^{n}\mathbf{K}_{h_{n,T}}\left(X_{i\Delta_{n,T}}-x\right)\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}\sigma(X_s)dB_s}{\frac{\Delta_{n,T}}{h_{n,T}^{d+1}}\sum_{i=1}^{n}\mathbf{K}_{h_{n,T}}\left(X_{i\Delta_{n,T}}-x\right)} \tag{8.3}$$

We start with the numerator of equation (8.2). We want to prove that:

$$\frac{1}{h_{n,T}^{d+1}}\sum_{i=1}^{n-1}\mathbf{K}_{h_{n,T}}\left(X_{i\Delta_{n,T}}-x\right)\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}\mu(X_s)ds$$
$$\xrightarrow{a.s.}\frac{1}{h_{n,T}^{d+1}}\int_0^T\mathbf{K}_{h_{n,T}}\left(X_s-x\right)\mu(X_s)ds \tag{8.4}$$

We start by writing:

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

by the triangle inequality and assumption (3). Finally using the *Ratio Limit theorem*, we have that:

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

which ensures that the statement in equation (8.4) to be true. We are now left with the following expression:

$$
\frac{\frac{1}{h_{n,T}^{d+1}} \int_0^T \mathbf{K}_{h_{n,T}} \left(X_s - x \right) \mu(X_s) ds + O_{a.s.} \left(\frac{\left(\Delta_{n,T} \log(1/\Delta_{n,T}) \right)^{1/2} \hat{L}^X(T,x)}{h_{n,T}^{d+1}} \right)}{\frac{1}{h_{n,T}^{d+1}} \int_0^T \mathbf{K}_{h_{n,T}} \left(X_s - x \right) ds + O_{a.s.} \left(\frac{\left(\Delta_{n,T} \log(1/\Delta_{n,T}) \right)^{1/2} \hat{L}^X(T,x)}{h_{n,T}^{d+1}} \right)}
$$

We have now to prove that this converges to the true functional form of the drift coefficient. We denote the true functional as $\mu(x)$ and write the following equation:

$$
\frac{\frac{1}{h_{n,T}^{d+1}} \int_0^T \mathbf{K}_{h_{n,T}} \left(X_s - x \right) \left(\mu(X_s) - \mu(x) \right) ds}{\frac{1}{h_{n,T}^{d+1}} \int_0^T \mathbf{K}_{h_{n,T}} \left(X_s - x \right) ds}
$$

We want to show that the numerator converges almost surely to 0. To do so, we exploit the Lipschitz continuity property of the drift function. Write:

$$\begin{aligned}
&\left|\frac{1}{h_{n,T}^{d+1}}\int_0^T \mathbf{K}_{h_{n,T}}\left(X_s-x\right)\left(\mu(X_s)-\mu(x)\right)ds\right| \\
&\leq \frac{1}{h_{n,T}^{d+1}}\int_0^T \left|\mathbf{K}_{h_{n,T}}\left(X_s-x\right)\right|\left|\mu(X_s)-\mu(x)\right|ds \\
&\leq \frac{C}{h_{n,T}^{d+1}}\int_0^T \left|\mathbf{K}_{h_{n,T}}\left(X_s-x\right)\right|\left|X_s-x\right|ds \leq C(\kappa_{n,T})\frac{1}{h_{n,T}^{d+1}}\int_0^T \mathbf{K}_{h_{n,T}}\left(X_s-x\right)ds \\
&= C(\kappa_{n,T})O_{a.s.}\left(\frac{1}{h_{n,T}^{d+1}}\int_0^T \mathbf{K}_{h_{n,T}}\left(X_s-x\right)ds\right)
\end{aligned}$$

which gives the desired result.

In order to prove that equation (8.3) converges to zero almost surely, we proceed as follows. We notice that, as in Bandi and Phillips (2003), the numerator of the equation can be embedded in a continuous time martingale for any value of $X_{i\Delta_{n,T}}$. As a matter of fact we have:

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

is a stochastic integral which is $\mathcal{Y}_{(i+1)\Delta_{n,T}} \vee \mathcal{Z}_{(i+1)\Delta_{n,T}}$-measurable and such that $\mathbb{E}\left[\beta_{(i+1)\Delta_{n,T}}\right] = 0$. Moreover by Itô isometry (see Øksendal, 2003, Lemma 3.15, p. 26):

$$var(\beta_{(i+1)\Delta_{n,T}}) = \mathbb{E}\left[\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}} \sigma(X_s)dB_s\right]^2 = \mathbb{E}\left[\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}} \sigma^2(X_s)ds\right] < \infty$$

We can therefore construct the following continuous martingale:

$$\begin{aligned}
M^{X_{i\Delta_{n,T}}}(r) &= \sqrt{h_{n,T}^{d+1}}\left(\frac{1}{h_{n,T}^{d+1}}\sum_{i=1}^{[nr]}\mathbf{K}_{h_{n,T}}\left(X_{i\Delta_{n,T}}-x\right)\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}\sigma(X_s)dB_s\right) \\
&= \frac{1}{\sqrt{h_{n,T}^{d+1}}}\sum_{i=1}^{[nr]}\mathbf{K}_{h_{n,T}}\left(X_{i\Delta_{n,T}}-x\right)\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}\sigma(X_s)dB_s
\end{aligned} \tag{8.5}$$

whose quadratic variation is equal to:

$$\left[M^{X_{i\Delta_{n,T}}}(r)\right] = \frac{1}{h_{n,T}^{d+1}}\sum_{i=1}^{[nr]}\mathbf{K}_h^2\left(X_{i\Delta_{n,T}}-x\right)\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}\sigma^2(X_s)ds \tag{8.6}$$

Using the same method applied for equation (8.2) and using the *Ratio Limit theorem*, we can show that:

$$\left[M^{X_{i\Delta_{n,T}}}(1)\right] = O_{a.s.}\left(\frac{1}{h_{n,T}^{d+1}}\int_0^T \mathbf{K}_{h_{n,T}}(X_s - x)\,ds\right) \tag{8.7}$$

Finally, as in Phillips and Ploberger (1996), expanding the probability space as needed:

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

which gives:

$$\sqrt{\hat{L}^X(T,x)h_{n,T}^{d+1}}\left(\frac{\frac{1}{\Delta_{n,T}}\frac{\Delta_{n,T}}{h_{n,T}^{d+1}}\sum_{i=1}^{n}\mathbf{K}_{h_{n,T}}(X_s - x)\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}\sigma(X_s)dB_s}{\frac{\Delta_{n,T}}{h_{n,T}^{d+1}}\sum_{i=1}^{n}\mathbf{K}_{h_{n,T}}(X_s - x)}\right) = O_{a.s.}(1)$$

Therefore, the term in equation (8.3) converges almost surely to zero, provided that $\hat{L}^X(T,x)h_{n,T}^{d+1} \xrightarrow{a.s.} \infty$. This completes the proof.

8.4. **Proof of Theorem (4.3).** We start by decomposing the estimator into a bias and a variance component:

$$\underbrace{(8.2) - \mu(x)}_{\text{BIAS}} + \underbrace{(8.3)}_{\text{VARIANCE}}$$

We start by analyzing the variance term. We use again the fact that this term can be written as a sequence of martingale components. Namely, we know that every martingale array can be written as a time changed *Dambis, Dubins-Schwartz* Brownian motion. We call τ, the time change associated to $M^{X_{i\Delta_{n,T}}}(1)$. This implies:

$$\frac{M_\tau^{X_{i\Delta_{n,T}}}(1)}{\sqrt{\frac{\Delta_{n,T}}{h_{n,T}^{d+1}}\sum_{i=1}^{n}\mathbf{K}_{h_{n,T}}(X_s - x)}} \xrightarrow{d} \mathcal{N}\left(0, \frac{\left[M_\tau^{X_{i\Delta_{n,T}}}(1)\right]}{\frac{\Delta_{n,T}}{h_{n,T}^{d+1}}\sum_{i=1}^{n}\mathbf{K}_{h_{n,T}}(X_s - x)}\right)$$

Using dominated convergence and the *Ratio Limit Theorem*, we can show that the numerator of the variance of $M_\tau^{X_{i\Delta_{n,T}}}(1)$ converges to:

$$\begin{aligned} &\frac{1}{h_{n,T}^{d+1}}\sum_{i=1}^{n}\mathbf{K}_h^2\left(X_{i\Delta_{n,T}} - x\right)\int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}}\sigma^2(X_s)ds \\ &\xrightarrow{a.s.} \sigma^2(x)\left(\int \mathbf{K}^2(u)du\right) \end{aligned} \tag{8.8}$$

Now, we turn to the bias term. Write the bias term in the following way:

$$\frac{\frac{1}{h_{n,T}^{d+1}}\int_0^T \mathbf{K}_{h_{n,T}}\left(X_s - x\right)\left(\mu(X_s) - \mu(x)\right)ds}{\frac{1}{h_{n,T}^{d+1}}\int_0^T \mathbf{K}_{h_{n,T}}\left(X_s - x\right)ds}$$

$$\xrightarrow{a.s.} \frac{\frac{1}{h_{n,T}^{d+1}}\int \mathbf{K}(u)\left(\mu(x+uh_{n,T}) - \mu(x)\right)p_t(x+uh_{n,T})\lambda(du)}{\frac{1}{h_{n,T}}\int \mathbf{K}(u)p(x+uh_{n,T}^{d+1})\lambda(du)}$$

We therefore compute the Taylor expansion of this function around x.

$$\frac{\int \mathbf{K}(u)\left[h_{n,T}\sum_{j=1}^{d+1}\frac{\partial\mu(x)}{\partial x_j}u_j + \frac{h_{n,T}^2}{2}\sum_{j,l=1}^{d+1}\frac{\partial^2\mu(x)}{\partial x_j\partial x_l}u_ju_l\right]\left[p_t(x) + h_{n,T}\sum_{j=1}^{d+1}\frac{\partial p_t(x)}{\partial x_j}u_j\right]\lambda(du)}{\int \mathbf{K}(u)\left[p_t(x) + h_{n,T}\sum_{j=1}^{d+1}\frac{\partial p_t(x)}{\partial x_j}u_j + o(h_{n,T})\right]\lambda(du)}$$

Using the symmetry of kernels and neglecting terms of order higher than $h_{n,T}^2$ leads to:

$$\int \mathbf{K}(u)\left[h_{n,T}^2\left(\sum_{j,l=1}^{d+1}\frac{\partial\mu(x)}{\partial x_j}\frac{\frac{\partial p_t(x)}{\partial x_l}}{p_t(x)}u_ju_l\right) + \frac{h_{n,T}^2}{2}\left(\sum_{j,l=1}^{d+1}\frac{\partial^2\mu(x)}{\partial x_j\partial x_l}u_ju_l\right)\right]\lambda(du)$$

We define

$$\mathcal{H}_\mu(x) = \left(\frac{\partial^2\mu(x)}{\partial x_j\partial x_l}\right)_{j,l=1}^d \qquad \mathcal{D}_{\mu,p}(x) = \left(\frac{\partial\mu(x)}{\partial x_j}\frac{\partial p_t(x)}{\partial x_l}\right)_{j,l=1}^d$$

where $\mathcal{H}_\mu(x)$ is the symmetric hessian matrix of the function μ and we rewrite the bias term as follows:

$$\begin{aligned} & h_{n,T}^2 tr\left\{\int \mathbf{K}(u)u'\left(\mathcal{D}_{\mu,p}(x) + \frac{1}{2}\mathcal{H}_\mu(x)\right)u\lambda(du)\right\} \\ =& h_{n,T}^2 tr\left\{\left(\mathcal{D}_{\mu,p}(x) + \frac{1}{2}\mathcal{H}_\mu(x)\right)\int \mathbf{K}(u)uu'\lambda(du)\right\} \\ =& h_{n,T}^2\rho_2(\mathbf{K})\left(tr\left\{\mathcal{D}_{\mu,\lambda(du)p}(x)\right\} + \frac{1}{2}tr\left\{\mathcal{H}_\mu(x)\right\}\right)\end{aligned}$$

using the properties of the trace operator, the relation between Lebesgue and Riemann integration and assumption (3).

8.5. **Proof of Theorem (4.4).** We want to prove that:

$$\hat{\sigma}_{n,T}^2(x) \xrightarrow{a.s.} \sigma^2(x)$$

Using Itô's lemma, we can show that $(Y_{(i+1)\Delta_{n,T}} - Y_{i\Delta_{n,T}})^2$ satisfies the following SDP:

$$(Y_{(i+1)\Delta_{n,T}} - Y_{i\Delta_{n,T}})^2 = \int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}} \left(2(Y_s - Y_{i\Delta_{n,T}})\mu(X_s) + \sigma^2(X_s)\right) ds + \int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}} 2(Y_s - Y_{i\Delta_{n,T}})\sigma(X_s) dB_s$$

This leads us to decompose equation (4.2) as follows:

$$\hat{\sigma}^2_{n,T}(x) = \frac{1}{\Delta_{n,T}} \frac{\frac{\Delta_{n,T}}{h^{d+1}_{n,T}} \sum_{i=1}^{n} \mathbf{K}_{h_{n,T}} \left(X_{i\Delta_{n,T}} - x\right) \int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}} \sigma^2(X_s) ds}{\frac{\Delta_{n,T}}{h^{d+1}_{n,T}} \sum_{i=1}^{n} \mathbf{K}_{h_{n,T}} \left(X_{i\Delta_{n,T}} - x\right)} \tag{8.9}$$

$$+ \frac{1}{\Delta_{n,T}} \frac{\frac{\Delta_{n,T}}{h^{d+1}_{n,T}} \sum_{i=1}^{n} \mathbf{K}_{h_{n,T}} \left(X_{i\Delta_{n,T}} - x\right) \int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}} 2(Y_s - Y_{i\Delta_{n,T}})\sigma(X_s) dB_s}{\frac{\Delta_{n,T}}{h^{d+1}_{n,T}} \sum_{i=1}^{n} \mathbf{K}_{h_{n,T}} \left(X_{i\Delta_{n,T}} - x\right)} \tag{8.10}$$

$$+ \frac{1}{\Delta_{n,T}} \frac{\frac{\Delta_{n,T}}{h^{d+1}_{n,T}} \sum_{i=1}^{n} \mathbf{K}_{h_{n,T}} \left(X_{i\Delta_{n,T}} - x\right) \int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}} 2(Y_s - Y_{i\Delta_{n,T}})\mu(X_s) ds}{\frac{\Delta_{n,T}}{h^{d+1}_{n,T}} \sum_{i=1}^{n} \mathbf{K}_{h_{n,T}} \left(X_{i\Delta_{n,T}} - x\right)} \tag{8.11}$$

In order to prove consistency of the diffusion term, we treat the drift as a nuisance parameter. As in the proof of theorem (4.2), using dominated convergence, the properties of the diffusion function and the *Ratio Limit Theorem*, we can prove that equation (8.9) almost surely converges to the true value of the diffusion term, as long as $\frac{\hat{L}^X(T,x)}{h^{d+1}_{n,T}} \left(\Delta_{n,T} \log(1/\Delta_{n,T})\right)^{1/2} = O_{a.s.}(1)$.

For equation (8.10) and equation (8.11), we follow Florens-Zmirou (1993) and Bandi and Phillips (2003). The term in $(Y_s - Y_{i\Delta_{n,T}})$ is a semi-martingale, so that we can use Burkholder-Davis-Gundy inequality (see,e.g. Protter, 2003, Theorem 48, p. 193) to show that its expectation can be bounded by the square root of its quadratic variation which converges at a rate equal to $\sqrt{\Delta_{n,T}}$. Therefore, following the proof of theorem (4.2), the component in equation (8.10) can be embedded in a continuous martingale whose expectation converges to zero as long as $\sqrt{\frac{\hat{L}^X(T,x) h^{d+1}_{n,T}}{\Delta_{n,T}}}$ diverges to infinity. In the same way, the term in (8.11) is bounded as long as $\sqrt{\frac{\hat{L}^X(T,x)h_{n,T}^{d+1}}{\Delta_{n,T}}}$ diverges (Bandi and Phillips, 2003; Bandi and Moloche, 2008).

8.6. **Proof of theorem (4.5).** Using the same procedure as in theorem (4.3), we decompose our estimator into a bias and a variance component:

$$\underbrace{(8.9) - \sigma^2(x)}_{\text{BIAS}} + \underbrace{(8.10) + (8.11)}_{\text{VARIANCE}}$$

For the variance, the component in equation (8.11) converges to zero almost surely as noted in the previous proof. Using the Ratio Limit theorem we can prove that equation (8.10) converges in distribution to a normal with variance equal to:

$$4\sigma^4(x)\left(\int \mathbf{K}^2(u)du\right) \tag{8.12}$$

We then turn to the bias term. We can follow the same procedure that for theorem (4.3). Define:

$$\mathcal{H}_{\sigma^2}(x) = \left(\frac{\partial^2\sigma^2(x)}{\partial x_j \partial x_l}\right)_{j,l=1}^{d} \qquad \mathcal{D}_{\sigma^2,p}(x) = \left(\frac{\partial\sigma^2(x)}{\partial x_j}\frac{\partial p_t(x)}{\partial x_l}\right)_{j,l=1}^{d}$$

where $\mathcal{H}_\sigma(x)$ is the symmetric hessian matrix of the function σ. Then the bias term is equal to:

$$h_{n,T}^2\rho_2(\mathbf{K})\left(tr\left\{\mathcal{D}_{\sigma^2,p}(x)\right\} + \frac{1}{2}tr\left\{\mathcal{H}_{\sigma^2}(x)\right\}\right)$$

8.7. **Additional Proofs.**

Theorem 8.5. *Suppose Y_t is a stationary process conditionally on Z_t and Z_t is Harris Recurrent. Then $X_t = (Y_t, Z_t)$ is a joint Harris Recurrent process.*

Proof. Remember that X_t lies in a Polish space $(E, \mathcal{E})$. We have to show that there exists a measure m, such that:

$$0 < m(A) < \infty \qquad \forall A \subset \mathcal{E}$$

i.e. a $\sigma-$finite measure on E, such that X is m-irreducible (see Definition 8.1).

We start to show that, for every set A and $t \to \infty$, if a measure exists, it is $\sigma-$finite. Take any set $A \subset \mathcal{E}$, such that $A = B \times C$, where B and C are compact, with $Z_{s+1} \in B$ and $Y_{s+1} \in C$. We denote by ϕ_z the invariant measure of the process Z_t and by $\pi(y|z)$

the stationary probability measure of Y given Z. We can write down the transition probability for the joint process, under the markovianity of X, as:

$$\begin{aligned}
& \int_0^\infty \mathbb{P}\left(X_{s+1} \in A | X_s\right) ds \\
= & \int_0^\infty \mathbb{P}\left(Z_{s+1} \in B, Y_{s+1} \in C | Z_s, Y_s\right) ds \\
= & \int_0^\infty \mathbb{P}\left(Z_{s+1} \in B | Z_s\right) \mathbb{P}\left(Y_{s+1} \in C | Z_s, Y_s, Z_{s+1} \in B\right) ds \\
\leq & \left(\int_0^\infty \mathbb{P}\left(Z_{s+1} \in B | Z_s\right) ds\right)\left(\int_0^\infty \mathbb{P}\left(Y_{s+1} \in C | Z_s, Y_s, Z_{s+1} \in B\right) ds\right) \\
= & \left(\int \mathbb{P}\left(Z_{s+1} \in B\right) \phi_z(dz)\right)\left(\int_0^\infty \mathbb{P}\left(Y_{s+1} \in C | Z_s, Y_s, Z_{s+1} \in B\right) ds\right) \\
= & \left(\int \mathbb{P}\left(Z_{s+1} \in B\right) \phi_z(dz)\right)\left(\int \mathbb{P}\left(Y_{s+1} \in C | Z_{s+1} \in B\right) \pi(dy|z)\right)
\end{aligned}$$

with a straightforward application of Bayes' theorem. Finally:

$$\phi_z(B) = \int \mathbb{P}\left(Z_{s+1} \in B\right) \phi_z(dz) < \infty$$

since A is bounded, and:

$$\pi(y \in C | z \in B) = \int \mathbb{P}\left(Y_{s+1} \in C | Z_{s+1} \in B\right) \pi(dy|z) \in (0, 1]$$

This implies:

$$\int_0^\infty \mathbb{P}\left(X_{s+1} \in A | X_s\right) ds < \infty \tag{8.13}$$

Therefore, for every set A, there exists a $\sigma-$finite measure for X. This concludes the first part of the proof.

Now, denote $\tau_A = inf\{t \geq 0, X_t \in A\}$, the hitting time of set A, for a given realization of X_t, $x = (z, y) \notin A$. For any arbitrary measure m:

$$\mathbb{P}^x(\tau_A < \infty) = 1 \tag{8.14}$$

implies $m(A) > 0$ (Revuz, 1984). We set $\tau_B^z = inf\{t \geq 0, Z_t \in B\}$ and $\tau_C^y = inf\{t \geq 0, Y_t \in C\}$. Then define:

$$\begin{aligned}
& \mathbb{P}^x(\tau_A < \infty) = \mathbb{P}^x(\tau_B^z < \infty, \tau_C^y < \infty) \\
= \; & \mathbb{P}^x(\tau_B^z < \infty) \mathbb{P}^x(\tau_C^y < \infty | \tau_B^z < \infty)
\end{aligned}$$

where the conditional probability is well defined since τ_B^z is a stopping time and $\{\tau_B^z < \infty\} \in \mathcal{Z}_\infty$ (Protter, 2003). Since Y is stationary conditional on Z, we have that:

$$\mathbb{E}^x\big(\tau_C^y | \tau_B^z < \infty\big) < \infty$$

which implies:

$$\Big\{ \sup_{t \geq 0, \tau_B^z < \infty} \tau_C^y \Big\} < \infty \quad \rightarrow \quad \mathbb{P}^x\big(\tau_C^y < \infty | \tau_B^z < \infty\big) = 1$$

We then obtain (8.14), from the Harris recurrence of Z.

Therefore, for every set A, X is m-irreducible and m is a $\sigma-$finite measure by (8.13). By definition (8.1), X is Harris recurrent. This concludes the proof. ∎