ON THE FUNCTIONAL ESTIMATION OF JUMP-DIFFUSION MODELS*

Federico M. Bandi†
Graduate School of Business
The University of Chicago

and

Thong H. Nguyen
Bank of America Securities

First draft: September 17, 2000
This version: February 5, 2001

Abstract

We provide a general asymptotic theory for the fully functional estimates of the infinitesimal moments of continuous-time models with discontinuous sample paths of the jump-diffusion type. Minimal requirements are placed on the dynamic properties of the underlying jump-diffusion process, i.e. stationarity is not required.

Our theoretical framework justifies consistent (in a statistical sense) nonparametric extraction of the parameters and functions that drive the dynamic evolution of the process of interest (i.e. the potentially non-affine and level dependent intensity of the jump arrival being an example) from the estimated infinitesimal conditional moments as suggested by Johannes (2000).

*We thank Jean Bertoin, Michael Johannes, Guillermo Moloche, Benoit Perron, George Tauchen (the co-editor), two anonymous referees and, especially, Darrell Duffie, for helpful comments. We are grateful to the seminar participants at the NBER/NSF Time Series Conference (Fort Collins, CO, September 23-24, 2000), the Econometric Society Winter Meetings 2001 (New Orleans, LA, January 5-7) and the CRDE Workshop "Modeling, Estimating and Forecasting Volatility" (Montreal, February 28, 2001) for discussions. Bandi thanks the Graduate School of Business of the University of Chicago for financial support.

†*Corresponding author. Address: Graduate School of Business, The University of Chicago, 1101 East 58th Street, Chicago, IL, 60637. Comments are welcome and may be sent to federico.bandi@gsb.uchicago.edu*

1 Introduction

Growing evidence demonstrates that sensible continuous-time models for several financial time series should account for the presence of discontinuous jump components (see Bakshi, Cao and Chen (1997), Duffie, Pan, and Singleton (2000) and the references therein for discussions regarding the equity market; see Das (1998), Piazzesi (2000), Johannes (2000) and the references therein for descriptions of jump-diffusion behavior in the fixed-income market). Unfortunately, econometric estimation of the parameters representing the jump arrival intensity and the distribution of the jump size is particularly cumbersome when using data sampled at discrete time intervals. In particular, it is empirically difficult to discriminate between variation caused by the continuous Brownian motion shocks and genuine discontinuities in the path of the process.[1] Even from a theoretical standpoint, the second (conditional) infinitesimal moment of the process can be expressed as the sum of the conditional volatility of the diffusion component (i.e. the so-called diffusive volatility) and the conditional second moment of the jump part (c.f. formula (11) in Section 2).

Nonetheless, should jumps play a role, then all the infinitesimal moments of order higher than one (c.f. formulae (11) and (12) in Section 2) would carry information about the probability of arrival and the features of the distribution of the jump size (c.f. Gikhman and Skorohod (1972) for a classical treatment). As a consequence, coherently with Bandi and Phillips (1998) (BP, hereafter) in the context of scalar diffusion processes and Johannes (2000) in the case of jump-diffusion models, in this work we pursue identification by considering estimators that can be readily interpreted as functional sample analogues to the instantaneous conditional moments of the underlying process.

The procedure is nonparametric. In contrast to the previous work on the functional estimation of continuous-time processes with discontinuous sample paths (c.f. Johannes (2000)), a complete asymptotic theory for the estimates is derived. In particular, we provide a theoretical framework that justifies the nonparametric extraction of the parameters and functions controlling the arrival of a jump and the distribution of the jump size from the estimated infinitesimal conditional moments as suggested by Johannes (2000).[2] The estimation procedure is sufficiently flexible to allow for potential nonlinearities in the drift, in the diffusive volatility and in the intensity of the discontinuous jump component.

Our results apply to both stationary and nonstationary jump-diffusions under mild assumptions. Technically, we only require the underlying process to be Harris recurrent (see Section 2 for a formal

[1] There have been recent advances in dealing with this problem. *Inter alia,* Andersen *et al.* (1998) rely on the efficient method of moment, Singleton (2001) and Chacko and Viceira (1999) employ moment conditions in time and/or frequency domain based on the characteristic function of the sampled data, Johannes (2000) advocates nonparametric methods.

[2] Our discussion complements the existing theoretical treatments on the estimation of structural breaks in discrete-time (c.f. Yin (1998), Müller (1992), Chu and Wu (1993), Delgado and Hidalgo (2000) and Perron (1999), *inter alia*)

definition) for the consistency and weak convergence results in this paper to be valid. Intuitively, Harris recurrence guarantees infinite returns of the continuous (in time) sample path of the process to every set of non zero Lebesgue measure in its range with probability one (c.f. (13) below). By recurrence we are able to identify the relevant functions at every admissible level in the range of the process through the contemporaneous implementation of *infill* and *long* span asymptotics (c.f. BP (1998)). The former (asymptotically decreasing distance between adjacent discretely sampled observations, that is) allows us to approximate the discrete sampled path of the process with its underlying continuous counterpart. The later (asymptotically increasing length of data, that is) permits us to exploit the properties of the path of the system for the purpose of the (pointwise) identification of the conditional infinitesimal moments of the process through infinite visits to every spatial set, as implied by recurrence.

We expect our theory to be particularly useful to study and model discontinuous processes for which standard parametrizations (often of the affine type in the finance literature[3]) are likely to be misspecified and for which the ubiquitous stationarity assumption appears to be excessively restrictive, as sometimes the case when dealing with financial time series in continuous-time models for asset pricing (c.f. Bandi (2001) for a discussion of potentially nonstationary behavior, and its implications, in a nonparametric diffusion model for the short-term interest rate process).

The plan of the paper is as follows. Section 2 introduces the model along with some useful preliminary results about the *cádlág* local time (or sojourn time, as it is sometimes referred to) of semimartingales. Consistently with the pure diffusion case discussed by BP (1998), local time plays an important role in affecting the convergence rates of the nonparametric estimates of the infinitesimal moments. In consequence, local time estimation is a necessary step for proper inference. Additionally, estimated local time is known to represent a valuable descriptive tool for nonstationary discrete-time series and recurrent continuous-time processes as discussed by Phillips (2001) and Bandi (2001), respectively. The focus in this paper is on the identification of a discontinuous process through the estimation of its infinitesimal moments. Hence, we do not dwell on the use (and logic) of local time as a descriptive statistic (the interested reader is referred to the papers cited above). In Section 3 we present the functional estimation scheme for the infinitesimal moments of the process and the local time factor. Section 4 contains the limit results. Identification of the features of the jump component through nonparametric extraction from the estimated infinitesimal moments is discussed in Section 5 with the aid of two examples. A simple Monte Carlo exercise showing the finite sample accuracy of our asymptotic theory and the empirical plausibility of our sampling scheme relying on twofold asymptotics is presented in Section 6. Section 7 concludes. All

[3]The reader is referred to Singleton (2001) and the references therein for a discussion of affine asset pricing models and related estimation methods.

technical proofs are confined to the Appendix.

2 The model

The model we analyze is time-homogeneous Markov and described by the equation

$$\begin{aligned} dX_t &= \left[\mu(X_{t-}) - \lambda(X_{t-})\int_Y c(X_{t-},y)\Pi(dy)\right]dt + \sigma(X_{t-})dW_t + dJ_t \\ &= \left[\mu(X_{t-}) - \lambda(X_{t-})\mathbf{E}_Y[c(X_{t-},y)]\right]dt + \sigma(X_{t-})dW_t + dJ_t \end{aligned} \tag{1}$$

where $\{W_t \,: t \geq 1\}$ and $\{J_t \,: t \geq 1\}$ are a standard scalar Brownian motion and an independent jump process, respectively. The initial condition $\overline{X}$ belongs to $\mathbf{L}^{\beta}$ for some $\beta > 0$ and is taken to independent of both W_t and J_t. The functions $\mu(.)$ and $\sigma(.)$ have the conventional interpretation in diffusion models. The jumps are bounded (i.e. $\sup_t |\Delta X_t| \leq \overline{C} < \infty$ almost surely where $\overline{C}$ is a non-random constant[4]) and occur with conditional (on the level of the process) intensity $\lambda(.)$.[5] The conditional impact of a jump is given by the function $c(., y)$ where y is a stationary random variable with probability distribution function represented by $\Pi(.)$. In consequence,

$$dJ_t = \Delta X_t = X_t - X_{t-} = \int_Y c(X_{t-},y)N(dt,dy) \tag{2}$$

where

$$N_t^{\Phi} = \sum_{j=1} \mathbf{1}_{[\tau_j \leq t, y_{\tau_j} \in \Phi]} \tag{3}$$

is a Poisson counting measure with stationary and independent increments. In the integral form, write

$$X_{t+\Delta} = X_t + \int_t^{t+\Delta} \mu(X_{s-})ds + \int_t^{t+\Delta} \sigma(X_{s-})dW_s + \int_{t+}^{t+\Delta}\int_Y c(X_{s-},y)\overline{\nu}(ds,dy) \tag{4}$$

where

$$\begin{aligned} \overline{\nu}(dt,dy) \;\; &: \;= N(dt,dy) - E(N(dt,dy)) \\ &: \;= N(dt,dy) - \lambda(X_{t-})\Pi(dy)dt \end{aligned} \tag{5}$$

[4]If V_t is a *Lévy* process with bounded jumps, then $\mathbf{E}\{|V_t^n|\} < \infty \; \forall n$. In other words, V_t has bounded moments of all orders (c.f. Protter (1995), Theorem 34, page 25).

[5]Contrary to most current estimation methodologies, we allow the drift, the diffusive volatility and the intensity of the jump ($\mu(.)$, $\sigma^2(.)$ and $\lambda(.)$, that is) to be fairly general (c.f. Assumption 1), potentially non-affine, functions.

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

$$\int_{t+}^{t+\Delta} \int_{Y} c(X_{s-}, y)\overline{\nu}(ds, dy) = \int_{t+}^{t+\Delta} dJ_t - \lambda(X_{s-})\mathbf{E}_Y[c(X_{s-}, y)]ds \tag{6}$$

represents the conditional variation between $t+$ and $t+\Delta$ in the path of the process due to discontinuous jumps of random impact $c(., y)$ net of its expected conditional magnitude at $t+$. The model is defined "compensated" by virtue of the presence of the term $\lambda(X_t)\mathbf{E}_Y[c(X_t, y)]dt$ denoting the conditional mean of the jump component. Its presence ensures that the component (6) is a martingale.[6] The martingale nature of the jump part will be heavily exploited in the proofs.

We impose the following conditions on the model. They guarantee existence of a *càdlàg* strong solution to (1).

Assumption 1.

(i) *The functions $\mu(.)$, $\sigma(.)$, $c(.,.)$ 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 are at least twice continuously differentiable. They satisfy local Lipschitz and growth conditions. Thus, for every compact subset Ψ of the domain of the process, there exists a constant C_1 such that, for all x and z in Ψ,*

$$|\mu(x) - \mu(z)| + |\sigma(x) - \sigma(z)| + \lambda(x) \int_{Y} |c(x, y) - c(z, y)|\Pi(dy) \leq C_1|x - z|. \tag{7}$$

Furthermore, there exists a constant C_2 so that for any $x \in \mathfrak{D}$,

$$|\mu(x)| + |\sigma(x)| + \lambda(x) \int_{Y} |c(x, y)|\Pi(dy) \leq C_2\{1 + |x|\}. \tag{8}$$

(ii) *For a given $\alpha > 2$, there exists a constant C_3 such that for any $x \in \mathfrak{D}$,*

$$\lambda(x) \int_{Y} |c(x, y)|^{\alpha}\Pi(dy) \leq C_3\{1 + |x|^{\alpha}\}. \tag{9}$$

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

Given (i), (ii) and (iii), the infinitesimal conditional moments of the changes in the solution to (1) can be written in terms of the functions $\mu(.)$, $\sigma(.)$, $c(.,.)$ and $\lambda(.)$ (c.f. Gikhman and Skorohod (1972)). In particular,

[6] In particular, the solution to (1) is a semimartingale. It is known that the semimartingale property implies the existance of an equivalent martingale measure under which the process is a (local) martingale. In consequence, should (1) be a price process, then absence of arbitrage in the spaces that preclude *doubling strategies* would be guaranteed by the semimartingale property of the price process itself (c.f. Duffie (1990)).

$$\mathbf{M}^1(x) = lim_{\Delta \to 0} \frac{1}{\Delta} \mathbf{E}[X_{t+\Delta} - X_t | X_t = x] = \mu(x), \tag{10}$$

$$\mathbf{M}^2(x) = lim_{\Delta \to 0} \frac{1}{\Delta} \mathbf{E}[(X_{t+\Delta} - X_t)^2 | X_t = x] = \sigma^2(x) + \lambda(x) \mathbf{E}_Y[c^2(x,y)], \tag{11}$$

$$\mathbf{M}^k(x) = lim_{\Delta \to 0} \frac{1}{\Delta} \mathbf{E}[(X_{t+\Delta} - X_t)^k | X_t = x] = \lambda(x) \mathbf{E}_Y[c^k(x,y)] \ \forall k > 2. \tag{12}$$

Equations (10), (11) and (12) will form the basis for our estimation procedure. It is noted that the generic $\mathbf{M}^k(x)$ is defined as an infinitesimal conditional expectation. We will show that every $\mathbf{M}^k(x)$ is identifiable, for every sample path, using standard functional methods for conditional expectations (c.f. Section 3). Consistent estimates of the objects of interest, i.e. $\mu(.)$, $\sigma(.)$, $\mathbf{E}_Y[c^k(x,y)]$ and $\lambda(.)$, can then be obtained, through nonparametric extraction from the estimated moments, provided appropriate identifying conditions are imposed on the underlying system (c.f. Section 5).

We now discuss the main identifying assumption in this paper: Harris recurrence (c.f. Meyn and Tweedie (1993) for a standard treatment). Let A be a measurable set of the range $\mathfrak{D}$ of the process of interest. Define the first hitting time of A as $\tau_A = \inf \{t \geq 0 : X_t \in A\}$. A generic Markov process X_t is null Harris recurrent if there is an invariant measure $\phi(dx)$ (see below for a definition) such that $\phi(A) > 0$ implies $\mathbf{P}^a [\tau_A < \infty] = 1$ for every $a \in \mathfrak{D}/\overline{A}$. Positive Harris recurrence holds if there is an invariant measure $\phi(dx)$ such that $\phi(A) > 0$ implies $\mathbf{E}^a[\tau_A] < \infty$ for every $a \in \mathfrak{D}/\overline{A}$. Under both notions of recurrence, the process returns to A an infinite number of time over times, i.e.

$$P_x \left(\int_0^\infty \mathbf{1}_{\{X_s \in A\}} ds = \infty \right) = 1, \tag{13}$$

for any $x \in \mathfrak{D}$.

Null Harris recurrence guarantees the existance of an invariant measure $\phi(dx)$ so that

$$\phi(A) = \int_{\mathfrak{D}} P(X_t^{(x)} \in A) \phi(dx) \qquad \forall A \in \mathfrak{B}(\mathfrak{D}). \tag{14}$$

Under positive recurrence the invariant measure is finite on $\mathfrak{D}$ and the process possesses an invariant probability measure given by $f(dx) = \frac{\phi(dx)}{\phi(\mathfrak{D})}$. Stationary processes are positive Harris recurrent processes started at the invariant probability measure $f(dx)$.

Either Assumption 2 or Assumption 3 below will be imposed on the dynamic properties of the jump-diffusion process of interest.

Assumption 2. *The solution to (1) is Harris recurrent.*

Assumption 3. *The solution to (1) is positive Harris recurrent.*[7]

Our estimation procedure only requires infinite returns of the path of the underlying process to every measurable set in its range (c.f. (13)). Assumption 2 is therefore sufficient. Nonetheless, consistently with the pure diffusion case discussed elsewhere (c.f. BP (1998)), the existence of a stationary probability measure (as implied by Assumption 3) increases the rate of convergence of the functional estimates to their theoretical counterparts.

2.1 Some useful preliminaries

We now report results about the local times of *cádlág* semimartingales that will be useful in the development of our limit theory. All what is needed below is contained in standard treatments like Protter (1995) and Revuz and Yor (1998).

Lemma 1. (The Tanaka formula) *Let X be a semimartingale and let $L_X(.,a)$ be its local time at a. Then,*

$$\begin{aligned}(X_t-a)^+-(X_0-a)^+ &= \int_{0+}^{t}\mathbf{1}_{(X_{s-}>a)}dX_s+\sum_{0<s\leq t}\mathbf{1}_{(X_{s-}>a)}(X_s-a)^- \\ &\quad +\sum_{0<s\leq t}\mathbf{1}_{(X_{s-}\leq a)}(X_s-a)^+ +\frac{1}{2}L_X(t,a), \end{aligned} \tag{15}$$

and

$$\begin{aligned}(X_t-a)^--(X_0-a)^- &= -\int_{0+}^{t}\mathbf{1}_{(X_{s-}\leq a)}dX_s+\sum_{0<s\leq t}\mathbf{1}_{(X_{s-}>a)}(X_s-a)^- \\ &\quad +\sum_{0<s\leq t}\mathbf{1}_{(X_{s-}\leq a)}(X_s-a)^+ +\frac{1}{2}L_X(t,a). \end{aligned} \tag{16}$$

Lemma 2. (Continuity of semimartingale local time) *Let X be a semimartingale with $\sum_{0<s\leq t}|\Delta X_s|<\infty$ a.s. $\forall t>0$. Then, there exist a $\mathfrak{B}\otimes\mathfrak{P}$ measurable version of $(a,t,\varpi)\rightarrow L_X(t,a,\varpi)$ which is everywhere jointly right continuous in a and continuous in t. Moreover a.s. the limits $L_X(t,a-)=\lim_{b\rightarrow a,\, b<a}L_X(t,b)$ exist.*

[7] The interested reader is referred to the papers by Menaldi and Robin (1999) and Wee (2000) for necessary and sufficient conditions for null recurrence, positive recurrence and transience that make use of appropriately defined Lyapounov functions.

Lemma 3. (The occupation time formula) *Let X be a semimartingale with local time $(L_X(.,a))_{a\in\Re}$. Let g be a Borel measurable function. Then, a.s.*

$$\int_{-\infty}^{\infty} L_X(t,a)g(a)da = \int_0^t g(X_{s-})d[X]_s^c \tag{17}$$

where $[X]^c$ is the continuous part of the quadratic variation of X.

Lemma 4. (Local times) *Let X be a semimartingale satisfying $\sum_{0<s\leq t}|\Delta X_s| < \infty$ a.s. $\forall t > 0$. Then, for every (a,t) we have*

$$L_X(t,a+) = L_X(t,a) = \lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\int_0^t \mathbf{1}_{(a\leq X_s\leq a+\varepsilon)}d[X]_s^c, \text{ a.s.}, \tag{18}$$

and

$$L_X(t,a-) = \lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\int_0^t \mathbf{1}_{(a-\varepsilon\leq X_s\leq a)}d[X]_s^c \text{ a.s.} \tag{19}$$

Also,

$$\frac{L_X(t,a+) + L_X(t,a-)}{2} = \lim_{\varepsilon\to 0}\frac{1}{2\varepsilon}\int_0^t \mathbf{1}_{(|X_s-a|\leq\varepsilon)}d[X]_s^c := L_X^{\oplus}(t,a) \text{ a.s.} \tag{20}$$

is a symmetrized version of local time.

Consistently with the standard diffusion case, the local time (c.f. Lemma 1) of a semimartingale with discontinuous sample path measures the amount of time spent by the process in the local neighborohood of a point. Time is measured in units of the continuous part of the quadratic variation process ($[X]_s^c = \sigma^2(X_s)ds$ in our case), i.e. in information units (c.f. Lemma 4). For general semimartingales, the local time process is *càdlàg* (c.f. Lemma 2). This observation leads to the notions of symmetrized local time (20), local time from the left (19) and local time from the right (18), as Lemma 4 reveals. Chronological versions (where time is measured in real time units) of the various local time notions at the spatial point a, say, can be defined in the usual fashion by simply rescaling the corresponding expressions by $\sigma^2(a)$. It is noted that chronological local time from the right is a version of the Radon-Nykodym derivative of the occupation measure of the process (i.e. $\int_0^T \mathbf{1}_{\{X_s\in A\}}ds$, $\forall A\in\mathfrak{B}(\mathfrak{D})$). The result follows from the occupation time formula in Lemma 3 by simply replacing the function $g(.)$ with the indicator over the set A. In the case of the solution to (1) above, the three notions of local time coincide since

$$\int_0^{\infty} \mathbf{1}_{(X_{s-}=a)}|dV| = 0 \qquad \forall a\in\mathfrak{D} \tag{21}$$

where V is the continous finite variation component of X (c.f. Yor (1978)). As a result, the process of interest has a bicontinuous (in a and t) modification of its family of local times (c.f. Revuz and Yor (1998, Theorem 1.7) and Protter (1995, Theorem 56, Corollary 1)).

As briefly mentioned in the introduction, the role played by local time is twofold.

First, estimated local times are known to be valuable in defining descriptive statistics for potentially non-stationary discrete time series and recurrent continuous-time models (c.f. Phillips (2001) and Bandi (2001)) in just the same way as estimated probability densities assist in summarizing the information contained in stationary processes. In Section 3 we introduce a general methodology, which we specialize to the case of the jump-diffusion process analyzed here, to identify (c.f. Theorem 1 in Section 4) the three notions of local time discussed earlier by virtue of averaged kernel functions constructed using symmetric, left and right kernels ($\mathbf{K}^{\oplus}$, $\mathbf{K}^{-}$ and $\mathbf{K}^{+}$, respectively) whose properties are listed in Assumption 4 below. Being the three notions of local time equivalent in our framework, the use of a standard symmetric kernel is generally preferable in virtue of its superior stability properties and will be our choice in the sequel.

Second, coherently with the pure diffusion case discussed elsewhere (c.f. BP (1998)), local time affects the rate of convergence (and, in consequence, the asymptotic variances) of the functional estimates of the infinitesimal moments of recurrent jump-diffusions (c.f. Theorem 3 in Section 4). Consistent estimation of the local time factor will then be crucial to perform inference on the estimated moments.

We now present the properties of the kernel functions used in this paper before turning to a discussion of the estimation procedure and limit results. In what follows the notation $\mathbf{K}(.)^{\pm}$ signifies either $\mathbf{K}^{+}$ or $\mathbf{K}^{-}$, where $\mathbf{K}^{+}$ is a right kernel function (zero for negative values and nonnegative for positive values) and $\mathbf{K}^{-}$ is a left kernel function (zero for positive values and nonnegative for negative values), as shown in the assumption below.

Assumption 4.

(i) *(Symmetric kernel function) The kernel* $\mathbf{K}^{\oplus}(.)$ *is a continuously differentiable, symmetric and nonnegative function such that*

$$\int_{\mathfrak{R}} \mathbf{K}^{\oplus}(s)ds = 1, \quad \int_{\mathfrak{R}} \left(\mathbf{K}^{\oplus}(s)\right)^2 ds < \infty \tag{22}$$

and

$$\int_{\mathfrak{R}} |s^2 \mathbf{K}^{\oplus}_{(m)}(s)|ds < \infty, \;\; \int_{\mathfrak{R}} |\mathbf{K}^{\oplus}_{(m)}(s)|ds < \infty \tag{23}$$

for $m = 0, 1$ *where* $\mathbf{K}^{\oplus}_{(m)}(s) = \partial^m\ \mathbf{K}^{\oplus}(s)/\partial s^m$.

(ii) *(Asymmetric kernels) The kernels* $\mathbf{K}^{\pm}(.)$ *are continuously differentiable and asymmetric functions such that*

$$\mathbf{K}^{\pm}(.) : \Re^{\pm} \to \Re \quad where \int_{\Re^{\pm}} \mathbf{K}^{\pm}(s)ds = 1 \ and \int_{\Re^{\pm}} \left(\mathbf{K}^{\pm}(s)\right)^2 ds < \infty \tag{24}$$

In addition,

$$\int_{\Re^{\pm}} |s^2 \mathbf{K}^{\pm}_{(m)}(s)|ds < \infty \quad and \quad \int_{\Re^{\pm}} |\mathbf{K}^{\pm}_{(m)}(s)|ds < \infty \tag{25}$$

for $m = 0, 1$ *where* $\mathbf{K}^{\pm}_{(m)}(s) = \partial^m\ \mathbf{K}^{\pm}(s)/\partial s^m$.[8]

3 Econometric estimation

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

We start with the identification of the local time factors. Assume the time span is fixed, i.e. $T = \overline{T}$. Then, functional estimation of (18), (19) and (20) at the spatial point a and $\overline{T}$ can be performed based on

$$\widehat{\overline{L}}_X(\overline{T}, a+) = \widehat{\overline{L}}_X(\overline{T}, a) = \frac{\Delta_{n,\overline{T}}}{h_{n,\overline{T}}} \sum_{i=1}^{n} \mathbf{K}^{+}\left(\frac{X_{i\Delta_{n,\overline{T}}} - a}{h_{n,\overline{T}}}\right), \tag{26}$$

$$\widehat{\overline{L}}_X(\overline{T}, a-) = \frac{\Delta_{n,\overline{T}}}{h_{n,\overline{T}}} \sum_{i=1}^{n} \mathbf{K}^{-}\left(\frac{X_{i\Delta_{n,\overline{T}}} - a}{h_{n,\overline{T}}}\right) \tag{27}$$

and

$$\widehat{\overline{L}}^{\oplus}_X(\overline{T}, a) = \frac{\Delta_{n,\overline{T}}}{h_{n,\overline{T}}} \sum_{i=1}^{n} \mathbf{K}^{\oplus}\left(\frac{X_{i\Delta_{n,\overline{T}}} - a}{h_{n,\overline{T}}}\right), \tag{28}$$

respectively. We now turn to the nonparametric estimates of the infinitesimal moments (10) - (12). We study the estimators for $\mathbf{M}^1(x)$, $\mathbf{M}^2(x)$ and $\mathbf{M}^k(x)$ with $k > 2$ suggested by Johannes (2000). Johannes extends the nonparametric procedure developed by Stanton (1997) and BP (1998). The generality of functional procedures based on sample analogues to the infinitesimal moments of

[8] See Delgado and Hidalgo (2000) and Perron (1999) for similar assumptions on the kernel weights.

continuous-time processes with conditional moment definitions is discussed in Bandi (2000). We propose the following estimators:

$$\widehat{\mathbf{M}}^1_{n,T}(a) = \frac{\frac{1}{h_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}^{\oplus}\left(\frac{X_{i\Delta_{n,T}}-a}{h_{n,T}}\right)[X_{(i+1)\Delta_{n,T}} - X_{i\Delta_{n,T}}]}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}^{\oplus}\left(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}}\right)}, \tag{29}$$

$$\widehat{\mathbf{M}}^2_{n,T}(a) = \frac{\frac{1}{h_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}^{\oplus}\left(\frac{X_{i\Delta_{n,T}}-a}{h_{n,T}}\right)[X_{(i+1)\Delta_{n,T}} - X_{i\Delta_{n,T}}]^2}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}^{\oplus}\left(\frac{X_{i\Delta_{n,T}}-a}{h_{n,T}}\right)}, \tag{30}$$

and

$$\widehat{\mathbf{M}}^k_{n,T}(a) = \frac{\frac{1}{h_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}^{\oplus}\left(\frac{X_{i\Delta_{n,T}}-a}{h_{n,T}}\right)[X_{(i+1)\Delta_{n,T}} - X_{i\Delta_{n,T}}]^k}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}^{\oplus}\left(\frac{X_{i\Delta_{n,T}}-a}{h_{n,T}}\right)} \tag{31}$$

$\forall k > 2$. The expressions (29) - (31) are sample analogues to conditional expectations defined as weighted averages of differences between observations sampled discretely on the discontinuous path of the process raised to some power.

4 Limit theory

We begin with the consistency (Theorem 1) of the estimates of the local time process.

In Theorem 1 we assume a fixed time span $\overline{T}$ as in the definitions of (26) - (28). A fixed span of time does not permit to exploit the recurrence properties of the underlying process. In consequence, recurrence is not a necessary assumption for estimating the local time factors over a predefined period of time as sometimes the case when using local times to describe the locational features of a possibly nonstationary continuous-time process of interest (c.f. Bandi (2001)).

In Corollary 1 and Theorem 2 below we let T diverge to infinity. We will show that pointwise explosion (for every sampled path) of the local time factor as T grows to infinity is a necessary assumption for the consistency of the infinitesimal moment estimators (c.f. Theorem 2). As the time span increases asymptotically, almost sure explosion of local time is guaranteed by the recurrence of the underlying continuous-time process (c.f. Corollary 1).

Theorem 1. *If $n \to \infty$ (with $T = \overline{T}$) and $h_{n,\overline{T}}$ $(\to 0)$ is such that $\frac{1}{h_{n,\overline{T}}}(\Delta_{n,\overline{T}})^{\theta} = O(1)$ for some $\theta \in \left(0, \frac{1}{2}\right)$ $\forall a \in \mathfrak{D}$, then*

$$\widehat{\overline{L}}_X(\overline{T}, a\pm) \overset{a.s.}{\to} \overline{L}_X(\overline{T}, a\pm) \tag{32}$$

where

$$\overline{L}_X(\overline{T},a+)=\overline{L}_X(\overline{T},a)=\frac{1}{\sigma^2(a)}\lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\int_0^{\overline{T}}\mathbf{1}_{(a\leq X_s\leq a+\varepsilon)}d[X]_s^c,\ a.s. \tag{33}$$

and

$$\overline{L}_X(\overline{T},a-)=\frac{1}{\sigma^2(a)}\lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\int_0^{\overline{T}}\mathbf{1}_{(a-\varepsilon\leq X_s\leq a)}d[X]_s^c,\ a.s. \tag{34}$$

Furthermore,

$$\widehat{\overline{L}}_X^{\oplus}(\overline{T},a)\overset{a.s.}{\to}\overline{L}_X^{\oplus}(\overline{T},a) \tag{35}$$

with

$$\begin{aligned}\overline{L}_X^{\oplus}(\overline{T},a) &= \frac{\overline{L}_X(\overline{T},a+)+\overline{L}_X(\overline{T},a-)}{2}=\frac{\overline{L}_X(\overline{T},a)+\overline{L}_X(\overline{T},a-)}{2}\\ &= \frac{1}{\sigma^2(a)}\lim_{\varepsilon\to 0}\frac{1}{2\varepsilon}\int_0^{\overline{T}}\mathbf{1}_{(|X_s-a|\leq\varepsilon)}d[X]_s^c,\ a.s.\end{aligned} \tag{36}$$

Corollary 1. *If $T\to\infty$ with n but $\frac{T}{n}=\Delta_{n,T}\to 0$ and $h_{n,T}$ $(\to 0)$ in such a way that $\frac{\overline{L}_X(T,a)}{h_{n,T}}(\Delta_{n,T})^{\theta}=O_{a.s.}(1)$ for some $\theta\in\left(0,\frac{1}{2}\right)$ $\forall a\in\mathfrak{D}$, then*

$$\widehat{\overline{L}}_X^{\pm}(T,a)\overset{a.s.}{\to}\overline{L}_X\left(\sup\{s:X_s=X_{s-}=a\},a\right) \tag{37}$$

and

$$\widehat{\overline{L}}_X^{\oplus}(T,a)\overset{a.s.}{\to}\overline{L}_X\left(\sup\{s:X_s=X_{s-}=a\},a\right). \tag{38}$$

Under Assumption 2,

$$\overline{L}_X\left(\sup\{s:X_s=X_{s-}=a\},a\right)=\infty. \tag{39}$$

As pointed out earlier, depending on the choice of the kernel function (c.f. Assumption 4) we obtain almost sure convergence to the various notions of local time for *cádlág* semimartingales that were presented in Lemma 4. The three notions are equivalent in the case of the solution to (1), thereby rendering the use of the more stable symmetric kernel preferable. We employ a symmetric kernel in what follows.

We now discuss the asymptotic theory of the infinitesimal conditional moments.

Theorem 2. *Assume Assumption 2 is satisfied. If* $n \to \infty$, $T \to \infty$, $\frac{T}{n} \to 0$ *and* $h_{n,T}$ $(\to 0)$ *is such that* $\frac{\overline{L}_X^{\oplus}(T,a)}{h_{n,T}}(\Delta_{n,T})^{\theta} = O_{a.s.}(1)$ *for some* $\theta \in \left(0, \frac{1}{2}\right)$ *and* $h_{n,T}\overline{L}_X^{\oplus}(T,a) \stackrel{a.s.}{\to} \infty$ $\forall a \in \mathfrak{D}$, *then*

$$\widehat{\mathbf{M}}^p_{n,T}(a) \stackrel{a.s.}{\to} \mathbf{M}^p(a) \qquad \forall p. \tag{40}$$

Furthermore, if $h^5_{n,T}\overline{L}_X^{\oplus}(T,a) \stackrel{a.s.}{\to} 0$ $\forall a \in \mathfrak{D}$, *then*

$$\sqrt{h_{n,T}\overline{L}_X^{\oplus}(T,a)}\left(\widehat{\mathbf{M}}^p_{n,T}(a) - \mathbf{M}^p(a)\right) \Rightarrow \mathbf{N}\left(0, \mathbf{K}_2^{\oplus}\mathbf{M}^{2p}(a)\right) \qquad \forall p \tag{41}$$

where

$$\mathbf{K}_2^{\oplus} = \int_{\mathfrak{R}} \left(\mathbf{K}^{\oplus}(s)\right)^2 ds. \tag{42}$$

Theorem 2 shows that straight sample analogues to the infinitesimal moments converge to the true functions with probability one as the time span and frequency of observations increase asymptotically. As in the case of drift function estimation, but contrary to the case of diffusion function estimation, in the pure scalar diffusion context (c.f. BP (1998)), an enlarging time span ($T \to \infty$) is necessary to guarantee the consistency result for *all* the estimated moments. A fixed T (or a non-diverging local time in the presence of explosive processes) would make the functional estimates diverge at speed $\frac{1}{\sqrt{h_{n,T}}}$. The result is intuitive and reflects the common belief that a long span of observations is necessary to gather sufficient (for consistency) information about the features of the Lévy measure of the jump component (intensity of the jump and probability distribution of the jump size, that is).

The asymptotic distributions are normal and the rates of convergence are path-dependent and defined pointwise as $\sqrt{h_{n,T}\overline{L}_X^{\oplus}(T,a)}$ where $\overline{L}_X^{\oplus}(T,a)$ is the local time process of the underlying jump-diffusion process. Contrary to the standard diffusion case (c.f. BP (1998)), the second infinitesimal moment estimator has a rate of convergence that is the same as the rate of convergence of the first infinitesimal moment estimator. Apparently, this is due to the presence of discontinuous breaks that have an equal impact on all the functional estimates. In fact, all the estimated functions have the same convergence rate pointwise. As in the pure diffusion case (c.f. BP (1998)), we expect the rate of convergence to be maximized $\left(\sqrt{h_{n,T}T},\ \text{that is}\right)$ when the underlying process is endowed with a stationary probability measure (that is, when the process is positive Harris recurrent, as implied by Assumption 3, or stationary). In this case the local time process increases like T, i.e.

$$\frac{\overline{L}_X^{\oplus}(T,a)}{T} \stackrel{a.s.}{\to} f(a) = \frac{\phi(a)}{\phi(\mathfrak{D})} \qquad \forall a \in \mathfrak{D} \tag{43}$$

(c.f. Davydov (1976), for example). Slower divergence rates for the local time factor (and slower convergence rates for estimated moments) occur in the presence of null Harris recurrent jump-diffusion processes.

If $h_{n,T}$ satisfies the above conditions but $h_{n,T} = O_{a.s.}\left(\overline{L}_X^{\oplus}(T,a)^{-1/5}\right)$, then a non-random bias term plays a role in the limit (c.f. the proof of Theorem 2). Its form is

$$h_{n,T}^2 \left(\int_{\mathfrak{R}} s^2 \mathbf{K}^{\oplus}(s) ds \right) \left(\frac{1}{2} \left(\mathbf{M}^p(a)\right)'' + \left(\mathbf{M}^p(a)\right)' \frac{\phi'(a)}{\phi(a)} \right) \qquad \forall p \tag{44}$$

where $\phi(dx)$ is the invariant measure of the process. The features of the bias term imply an asymptotic mean-squared error of order $h_{n,T}^4 + \frac{1}{h_{n,T}\overline{L}_X^{\oplus}(T,a)}$ and, in consequence, optimal bandwidth sequences of order $\left(\overline{L}_X^{\oplus}(T,a)\right)^{-1/5}$. For all practical purposes, the bandwidth sequence for the generic moment $\mathbf{M}^p(a)$ can be set equal to

$$h_{n,T}^p(a) = \vartheta^p \log\left(\frac{1}{\widehat{\overline{L}}_X^{\oplus}(T,a)} \right) \left(\widehat{\overline{L}}_X^{\oplus}(T,a) \right)^{-1/5} \tag{45}$$

where ϑ^p is a moment specific constant of proportionality and $\log\left(\frac{1}{\widehat{\overline{L}}_X^{\oplus}(T,a)}\right)$ is a standard adjustment factor intended to guarantee slight undersmoothing and, consequently, absence of a limiting bias term (c.f. (41) above) from the asymptotic distributions of the functional estimates as in the statement of Theorem 2. Two additional observations are in order. First, the optimal bandwidth sequences have a local nature. Larger bandwidths are required in regions where the data is sparser as summarized by the information contained in the local time factor. Second, the proportionality factor ϑ^p and the optimal smoothing sequence for local time estimation should be determined based on data-driven criteria. The design of such criteria goes beyond the scope of the present paper and is left for future research.

5 Identification: two examples

We now discuss two examples that will serve the purpose of illustrating possible mechanisms to extract the functions of interest from the estimated nonparametric moments. For simplicity, we write $c(.,y) = y$. Then, equations (10), (11) and (12) reduce to

$$\mathbf{M}^1(x) = \mu(x), \tag{46}$$

$$\mathbf{M}^2(x) = \sigma^2(x) + \lambda(x)\mathbf{E}_Y[y^2], \tag{47}$$

and

$$\mathbf{M}^k(x) = \lambda(x)\mathbf{E}_Y[y^k] \ \forall k > 2. \tag{48}$$

Identification of the drift, diffusive volatility, intensity of the jump size and parameters of the distribution of the jump component simply requires choice of an appropriate parametric family for the probability measure of the jump part as well as use of the estimated moments. In this section we discuss two choices for the distribution of the jump component that accomodate different jump behaviors, namely normal and mixed normal jump sizes. Extensions to alternative specifications are straightforward based on our subsequent discussion.

5.1 Normal jumps

Assume $y \rightsquigarrow \mathbf{N}\left(0, \sigma_y^2\right)$ (c.f. Johannes (2000)). Thus,

$$\mathbf{E}_Y[y^{2r}] = \sigma_y^{2r} \prod_{n=1}^{r} (2n-1) \tag{49}$$

and

$$\mathbf{E}_Y[y^{2r-1}] = 0 \tag{50}$$

for $r = 1, 2, 3, ...$ A natural way to extract estimates of the underlying objects of interest from the moment restrictions (46)-(48) is to use the following sequential algorithm suggested by Johannes (2000):[9]

[1] Obtain an estimate of σ_y^2 via

$$\left(\widehat{\sigma}_y^2\right)_{n,T} = \frac{1}{n} \sum_{i=1}^{n} \frac{\widehat{\mathbf{M}}_{n,T}^6(X_{i\Delta_{n,T}})}{5\widehat{\mathbf{M}}_{n,T}^4(X_{i\Delta_{n,T}})}. \tag{51}$$

[2] Obtain an estimate of $\lambda(x)$ via

$$\widehat{\lambda}_{n,T}(x) = \frac{\widehat{\mathbf{M}}_{n,T}^4(x)}{3\left(\widehat{\sigma}_y^4\right)_{n,T}}. \tag{52}$$

[3] Obtain an estimate of $\sigma^2(x)$ via

$$\widehat{\sigma}_{n,T}^2(x) = \widehat{\mathbf{M}}_{n,T}^2(x) - \widehat{\lambda}_{n,T}(x)\left(\widehat{\sigma}_y^2\right)_{n,T}. \tag{53}$$

[4] Obtain an estimate of $\mu(x)$ via

$$\widehat{\mathbf{M}}_{n,T}^1(x). \tag{54}$$

[9] Johannes (1999) contains a thorough illustration of the empirical issues posed by this estimation procedure. We refer the interested reader to his work.

Due to the averaging, and coherently with standard semiparametric models, we expect the rate of convergence of the parameter estimate $\left(\widehat{\sigma}_y^2\right)_{n,T}$ to be faster than those of the functional estimates $\widehat{\lambda}_{n,T}(.)$, $\widehat{\sigma}^2_{n,T}(.)$ and $\widehat{\mu}_{n,T}(.)$. The same intuition applies to the parameter estimates in the mixed normal model discussed below.

5.2 Mixed normal jumps (The Variance Gamma model)

Assume $y \rightsquigarrow \mathbf{N}\left(0, \sigma_y^2 \mathbf{V}\right)$ and $\mathbf{V} \rightsquigarrow \mathbf{G}(v) = \frac{1}{\Gamma\left(\frac{1}{b}\right)} \frac{1}{b} \left(\frac{1}{b} v\right)^{\frac{1}{b}-1} e^{-\frac{1}{b} v} \mathbf{I}\left\{v \geq 0\right\}$ where $\Gamma\left(.\right)$ is the gamma function (c.f. Madan and Soneta (1990)).[10] $\mathbf{V}$ has expected value equal to 1 and variance equal to b, hence σ_y^2 serves as a scale parameter. Madan and Soneta (1990) introduced this model as a way to capture long taildeness in daily stock returns (the proportional excess kurtosis over the normal kurtosis is given by the positive parameter b).[11] It is easy to note that

$$\begin{aligned} \mathbf{E}_Y[y^{2r}] &= \mathbf{E}\left(\mathbf{E}[y^{2r}|\mathbf{V}]\right) = \mathbf{E}\left(\sigma_y^{2r}\left(\prod_{n=1}^{r}(2n-1)\right)\mathbf{V}^r\right) \\ &= \sigma_y^{2r}\left(\prod_{n=1}^{r}(2n-1)\right)\mathbf{E}[\mathbf{V}^r] = \sigma_y^{2r}\left(\prod_{n=1}^{r}(2n-1)\right)\frac{\Gamma\left(\frac{1}{b}+r\right)}{\Gamma\left(\frac{1}{b}\right)}b^r \end{aligned} \tag{55}$$

and

$$\mathbf{E}_Y[y^{2r-1}] = 0 \tag{56}$$

for $r = 1, 2, 3, ...$ Thus, a possible way to extract the functions and parameters of interest from the estimated moments is as follows:

[1] Define $\widehat{\Psi}^{6,4}_{n,T}$ and $\widehat{\Psi}^{8,6}_{n,T}$ as

$$\widehat{\Psi}^{6,4}_{n,T} = \frac{1}{n}\sum_{i=1}^{n}\frac{\widehat{\mathbf{M}}^6_{n,T}(X_{i\Delta_{n,T}})}{\widehat{\mathbf{M}}^4_{n,T}(X_{i\Delta_{n,T}})}, \tag{57}$$

$$\widehat{\Psi}^{8,6}_{n,T} = \frac{1}{n}\sum_{i=1}^{n}\frac{\widehat{\mathbf{M}}^8_{n,T}(X_{i\Delta_{n,T}})}{\widehat{\mathbf{M}}^6_{n,T}(X_{i\Delta_{n,T}})}. \tag{58}$$

Hence,

[10] In a Bayesian framework, one could interpret $\mathbf{G}$ as being a prior distribution on the parameter $\mathbf{V}$ (Praetz (1972)).

[11] The Lévy process which is consistent with the Variance Gamma model as a distribution for the unit period dynamics is a time changed Brownian motion with zero drift and variance σ^2 where the time change has gamma increments with mean 1 and variance b over unit intervals (c.f. Madan and Soneta (1990)). Option pricing with the variance gamma process is discussed in Madan *et al.* (1998).

$$\begin{aligned}\left(\widehat{\sigma}_y^2(b)\right)_{n,T} &= \frac{1}{5b}\widehat{\Psi}_{n,T}^{6,4}\left(\frac{\Gamma\left(\frac{1}{b}+2\right)}{\Gamma\left(\frac{1}{b}+3\right)}\right)\\ &= \frac{1}{5b}\widehat{\Psi}_{n,T}^{6,4}\left(\frac{1}{\frac{1}{b}+2}\right)\\ &= \frac{1}{5}\left(\frac{1}{1+2b}\right)\widehat{\Psi}_{n,T}^{6,4},\end{aligned}\tag{59}$$

and $\widehat{b}_{n,T}$ can be found by solving

$$\widehat{\Psi}_{n,T}^{8,6} - \frac{7}{5}\widehat{\Psi}_{n,T}^{6,4}\left(\frac{\Gamma\left(\frac{1}{b}+2\right)\Gamma\left(\frac{1}{b}+4\right)}{\Gamma^2\left(\frac{1}{b}+3\right)}\right) = 0 \tag{60}$$

which implies,

$$\widehat{b}_{n,T} = \frac{\widehat{\Psi}_{n,T}^{8,6} - \frac{7}{5}\widehat{\Psi}_{n,T}^{6,4}}{\frac{21}{5}\widehat{\Psi}_{n,T}^{6,4} - 2\widehat{\Psi}_{n,T}^{8,6}}. \tag{61}$$

It is now straightforward to compute $\left(\widehat{\sigma}_y^2\right)_{n,T}$ as

$$\left(\widehat{\sigma}_y^2\left(\widehat{b}_{n,T}\right)\right)_{n,T} = \frac{1}{5}\left(\frac{1}{1+2\widehat{b}_{n,T}}\right)\widehat{\Psi}_{n,T}^{6,4}. \tag{62}$$

[2] Obtain an estimate of $\lambda(x)$ via

$$\widehat{\lambda}_{n,T}(x) = \frac{\widehat{\mathbf{M}}_{n,T}^4(x)}{3\left(\widehat{\sigma}_y^4\right)_{n,T}\left(1+\widehat{b}_{n,T}\right)}. \tag{63}$$

[3] Obtain an estimate of $\sigma^2(x)$ via

$$\widehat{\sigma}_{n,T}^2(x) = \widehat{\mathbf{M}}_{n,T}^2(x) - \widehat{\lambda}_{n,T}(x)\left(\widehat{\sigma}_y^2\right)_{n,T}. \tag{64}$$

[4] Obtain an estimate of $\mu(x)$ via

$$\widehat{\mathbf{M}}_{n,T}^1(x). \tag{65}$$

6 A Monte Carlo exercise

In this section we conduct a simple simulation experiment to examine the finite sample properties of the estimators discussed in the paper and the coherence between asymptotic and finite sample distributions. An extended version of a process widely used in modeling interest rates, namely the Cox, Ingersoll, and Ross (1985) process, is our choice here. The model we analyze is

$$d\log(r_t) = \beta\left(\alpha - r_t\right)dt + \sigma\sqrt{r_t}dW_t + dJ_t \tag{66}$$

where J_t is a jump process with constant arrival intensity $\lambda(.) = \lambda$ and jump size $y \rightsquigarrow \mathbf{N}\left(0, \sigma_y^2\right)$. The underlying parameters are chosen as

$$\lambda(r_t) = \lambda = 20 \qquad \forall t \tag{67}$$

$$\beta = .85837 \tag{68}$$

$$\alpha = .089102 \tag{69}$$

$$\sigma = .3 \tag{70}$$

$$\sigma_y = .03630427. \tag{71}$$

We select values that determine sample trajectories that mimick the observed behavior of conventionally used short-term interest rate series in the US bond market. The process displays a mean-reverting drift and satisfies affine structures, i.e. both the drift and the diffusive volatility are linear functions of the state variable. The log specification guarantees that the series stays positive as required by nominal interest rates. Furthermore, consistently with the observation that interest rates are more volatile at higher levels, this model allows the diffusive volatility to be an increasing function of the underlying interest rate level.

We use the Euler's scheme to simulate $10,000$ daily observations for each path, which is equivalent to a daily data set spanning about 40 years of observations. $10,000$ paths are generated using the antithetic variate technique. The use of daily frequencies in continuous-time asset pricing accommodates the fact that the researcher often does not observe (or wishes to employ, due to spurious microstructure contaminations) higher frequency data. We will show that the method is robust to this "crude" frequency.

Our choice of the kernel is the ubiquitous, second order, symmetric Gaussian kernel, i.e. $\mathbf{K}^{\oplus}(x) = \frac{1}{h\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x}{h}\right)^2\right)$ and $\mathbf{K}_2^{\oplus} = \frac{1}{2\sqrt{\pi}}$. Choosing the optimal bandwidth is still largely an elusive question in the nonparametric literature. Moreover, the existing methods are generally designed for standard regression contexts. The need for theoretical treatments of bandwidth selection procedures in the case of continuous-time model estimation, especially when dealing with possibly nonstationary series, has been discussed by Bandi and Nguyen (1999). In this simple exercise we opt for employing a flat smoothing parameter equal to 1.5%. Such value is similar to bandwidth values that are reported in empirical studies on the dynamics of US short-term interest rates (for example, Aït-Sahalia (1996) utilizes a bandwidth equal to 1.6033%). We leave the use of more appropriate bandwidth selection methods for future research. In particular, as discussed earlier, the optimal bandwidth should accomodate the divergence properties of the local time factor and be level-specific. We estimate the functions and parameters of interest by employing the extraction scheme in Subsection 5.1 (c.f. Johannes (2000)).

The estimation results are reported in Figure 1, where the graphs are for drift, diffusive volatility and jump intensity. In each graph, the four curves represent the true value, the estimated value (pointwise sample means across the $10,000$ simulations) and the 25 and 75 percentiles. The standard deviation of the jump size (σ_y) is estimated quite accurately. Its mean value is 0.03582. The standard deviation of the estimates is 0.0037. The minimum and maximum values are 0.0267 and 0.0691, respectively. Despite the use of a simple extraction scheme and a rather naively-chosen smoothing parameter, the estimators appear to capture sufficiently well the true parameter value and functions in finite sample.

Figure 2 through 5 contain graphical comparisons between the limiting distributions of the first four infinitesimal moment estimators from Theorem 2 and their empirical counterparts. We report results for two levels that are often visited by the simulated trajectories (i.e. for which the mean estimated local time is large) such as 6% and 7%. The asymptotic approximations appear to be very satisfactory in finite sample. Similar results occur when considering different levels.

7 Conclusion

Since the fundamental paper by Merton (1976) which introduces jump-diffusion models as a possible solution to the problem stemmed from the fact that the Black and Scholes (1973) approach fails to fit observed option prices and stock price dynamics, continuous-time processes with discontinuous sample paths have incited an on-again, off-again interest in the finance and econometrics literature. The recent compelling evidence on the importance of discontinuous components in models for continuous-time financial series has determined a vigorously renewed attention to the econometrics of jump-diffusions.

This paper tackles the identification of the functions of interest under mild assumptions on the underlying process. More specifically, no parametric specification is assumed for drift, diffusive volatility and intensity of the jump size (c.f. Johannes (2000)). In addition, the process is not required to possess a stationary probability measure. Together with the extreme computational simplicity, these are features that should make the methodology particularly appealing to study the dynamics of time series for which stationarity is an issue, as sometimes the case in finance, and for which simple parametrizations appear to be too restrictive.

The asymptotic theory contained in this paper, along with the existing treatments in the standard diffusion case (c.f. BP (1998) and Bandi (2000)), represent useful theoretical tools to test for the presence of discontinuous breaks based on the properties of the infinitesimal moments of the underlying semimartingale when jumps play a role. Additionally, in light of the presence of a wide array of possible parametric forms for sensible jump-diffusion models, the asymptotics developed here can be fruitfully employed to construct functional tests of model (mis)specification. Research on these topics is being conducted and will be reported in later work.

Appendix: Proofs

Proof of Lemma1. See Protter (1995), Theorem 49, page 165.

Proof of Lemma 2. See Protter (1995), Theorem 56, page 176.

Proof of Lemma 3. See Protter (1995), Corollary 1, page 168.

Proof of Lemma 4. See Protter (1995), Corollary 3, page 178.

Proof of Theorem 1. Consider the quantity

$$\int_{0+}^{\overline{T}} \frac{1}{h_{n,\overline{T}}} \mathbf{K}^{\pm}\left(\frac{X_{s-}-x}{h_{n,\overline{T}}}\right) ds. \tag{72}$$

From Lemma 3 (i.e. the occupation time measure), we obtain

$$\int_{0+}^{\overline{T}} \frac{1}{h_{n,\overline{T}}} \mathbf{K}^{\pm}\left(\frac{X_{s-}-x}{h_{n,\overline{T}}}\right) \frac{d[X]_s^c}{\sigma^2(X_s)} = \int_{-\infty}^{\infty} \frac{1}{h_{n,\overline{T}}} \mathbf{K}^{\pm}\left(\frac{a-x}{h_{n,\overline{T}}}\right) \frac{1}{\sigma^2(a)} L_X(\overline{T},a)da. \tag{73}$$

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

$$\begin{aligned} & \int_{-\infty}^{\infty} \mathbf{K}^{\pm}(q) \frac{1}{\sigma^2(h_{n,\overline{T}}q+x)} L_X(\overline{T}, h_{n,\overline{T}}q+x)dq \\ = & \int_{\Re^+} \mathbf{K}^{\pm}(q) \frac{1}{\sigma^2(h_{n,\overline{T}}q+x)} L_X(\overline{T}, h_{n,\overline{T}}q+x)dq \\ & + \int_{\Re^-} \mathbf{K}^{\pm}(q) \frac{1}{\sigma^2(h_{n,\overline{T}}q+x)} L_X(\overline{T}, h_{n,\overline{T}}q+x)dq. \end{aligned} \tag{74}$$

For general semimartingales, the map $a \mapsto L_t^a$ is a.s. *cádlág* (i.e. right-continuous with left limits) for a fixed t (c.f. Lemma 2). Hence, as $n \to \infty$ and $h_{n,\overline{T}} \to 0$, and since $\int_{\Re^{\pm}} \mathbf{K}^{\pm}(q)dq = 1$ from Assumption 4 in Section 2, we have

$$\begin{aligned} & \int_{\Re} \mathbf{K}^{\pm}(q) \frac{1}{\sigma^2(h_{n,\overline{T}}q+x)} L_X(\overline{T}, h_{n,\overline{T}}q+x)dq \\ = & \int_{\Re^+} \mathbf{K}^{+}(q) \frac{1}{\sigma^2(h_{n,\overline{T}}q+x)} L_X(\overline{T}, h_{n,\overline{T}}q+x)dq \\ & \overset{a.s.}{\to} \frac{1}{\sigma^2(x)} L_X(\overline{T}, x+) = \frac{1}{\sigma^2(x)} L_X(\overline{T}, x) \\ = & \frac{1}{\sigma^2(x)} \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \int_0^{\overline{T}} \mathbf{1}_{(x \le X_s \le x+\varepsilon)} d[X]_s^c := \overline{L}_X(\overline{T}, x) \end{aligned} \tag{75}$$

by dominated convergence and provided $\mathbf{K}^{\pm}(.) = \mathbf{K}^{+}(.)$. If $\mathbf{K}^{\pm}(.) = \mathbf{K}^{-}(.)$, then

$$\begin{aligned} & \int_{\Re} \mathbf{K}^{-}(q) \frac{1}{\sigma^2(h_{n,\overline{T}}q+x)} L_X(\overline{T}, h_{n,\overline{T}}q+x)dq \\ = & \int_{\Re^-} \mathbf{K}^{-}(q) \frac{1}{\sigma^2(h_{n,\overline{T}}q+x)} L_X(\overline{T}, h_{n,\overline{T}}q+x)dq \\ & \overset{a.s.}{\to} \frac{1}{\sigma^2(x)} L_X(\overline{T}, x-) \\ = & \frac{1}{\sigma^2(x)} \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \int_0^{\overline{T}} \mathbf{1}_{(x-\varepsilon \le X_s \le x)} d[X]_s^c := \overline{L}_X(\overline{T}, x-) \end{aligned} \tag{76}$$

again by dominated convergence. To prove the result we have to show that

$$\frac{\Delta_{n,\overline{T}}}{h_{n,\overline{T}}}\sum_{i=1}^{n}\mathbf{K}^{\pm}\left(\frac{X_{i\Delta_{n,\overline{T}}}-x}{h_{n,\overline{T}}}\right)-\int_{0+}^{\overline{T}}\frac{1}{h_{n,\overline{T}}}\mathbf{K}^{\pm}\left(\frac{X_{s-}-x}{h_{n,\overline{T}}}\right)ds\overset{a.s.}{\to}0, \tag{77}$$

under the stated conditions. This is equivalent to proving that

$$\begin{aligned}&\frac{1}{h_{n,\overline{T}}}\sum_{i=0}^{n-1}\int_{i\overline{T}/n+}^{(i+1)\overline{T}/n}\left[\mathbf{K}^{\pm}\left(\frac{X_{i\Delta_{n,\overline{T}}}-x}{h_{n,\overline{T}}}\right)-\mathbf{K}^{\pm}\left(\frac{X_{s-}-x}{h_{n,\overline{T}}}\right)\right]ds\\&-\frac{\Delta_{n,\overline{T}}}{h_{n,\overline{T}}}\mathbf{K}^{\pm}\left(\frac{X_{0+}-x}{h_{n,\overline{T}}}\right)+\frac{\Delta_{n,\overline{T}}}{h_{n,\overline{T}}}\mathbf{K}^{\pm}\left(\frac{X_{n\Delta_{n,\overline{T}}}-x}{h_{n,\overline{T}}}\right)\overset{a.s.}{\to}0.\end{aligned} \tag{78}$$

Using Assumption 4 again, the left side of (78) is bounded by

$$\begin{aligned}&\left|\frac{1}{h_{n,\overline{T}}}\sum_{i=0}^{n-1}\int_{i\overline{T}/n+}^{(i+1)\overline{T}/n}\left[\mathbf{K}^{\pm}\left(\frac{X_{i\Delta_{n,\overline{T}}}-x}{h_{n,\overline{T}}}\right)-\mathbf{K}^{\pm}\left(\frac{X_{s-}-x}{h_{n,\overline{T}}}\right)\right]ds\right|\\&+\frac{\Delta_{n,\overline{T}}}{h_{n,\overline{T}}}\left|\mathbf{K}^{\pm}\left(\frac{X_{0}-x}{h_{n,\overline{T}}}\right)\right|+\frac{\Delta_{n,\overline{T}}}{h_{n,\overline{T}}}\left|\mathbf{K}^{\pm}\left(\frac{X_{n\Delta_{n,\overline{T}}}-x}{h_{n,\overline{T}}}\right)\right|\\ \leq\;&\frac{1}{h_{n,\overline{T}}}\left|\sum_{i=0}^{n-1}\int_{i\overline{T}/n+}^{(i+1)\overline{T}/n}\mathbf{K}_{(1)}^{\pm}\left(\frac{\widetilde{X}_{is-}-x}{h_{n,\overline{T}}}\right)\left(\frac{X_{s-}-X_{i\Delta_{n,\overline{T}}}}{h_{n,\overline{T}}}\right)ds\right|+2C_4\frac{\Delta_{n,\overline{T}}}{h_{n,\overline{T}}}\\ \leq\;&\left(\frac{1}{h_{n,\overline{T}}}\right)\frac{1}{h_{n,\overline{T}}}\sum_{i=0}^{n-1}\int_{i\overline{T}/n+}^{(i+1)\overline{T}/n}\left|\mathbf{K}_{(1)}^{\pm}\left(\frac{\widetilde{X}_{is-}-x}{h_{n,\overline{T}}}\right)\right|\left|\left(X_{s-}-X_{i\Delta_{n,\overline{T}}}\right)\right|ds+2C_4\frac{\Delta_{n,\overline{T}}}{h_{n,\overline{T}}}\end{aligned} \tag{79}$$

where $\widetilde{X}_{is-}$ is on the line segment connecting X_{s-} and $X_{i\Delta_{n,\overline{T}}}$ and C_4 is a suitable constant. Define

$$\delta_{n,\overline{T}}=\max_{i\leq n}\sup_{i\Delta_{n,\overline{T}}+\leq s\leq(i+1)\Delta_{n,\overline{T}}}|X_{s-}-X_{i\Delta_{n,\overline{T}}}|. \tag{80}$$

The increments $X_{t+\Delta}-X_t$ are of order $\sqrt{\Delta}$. The order of magnitude can be deduced from the Lévy-Khintchine representation (c.f. Protter (1990), Theorem 4.3, page 32). Furthermore, there is $\Omega_0\in\Im$ with $\mathbf{P}[\Omega_0]=1$ such that, for every $\varpi\in\Omega_0$, $X_t(\varpi)$ is right-continuous in $t\geq 0$ and has left limit in $t>0$ (c.f. Sato (1999), Definition 1.6(5)). Then,

$$\frac{\delta_{n,\overline{T}}}{\left(\Delta_{n,\overline{T}}\right)^{\theta}}=o_{a.s.}(1) \tag{81}$$

for $\forall\ \theta<\frac{1}{2}$. Thus, if $h_{n,\overline{T}}$ is such that $\frac{1}{h_{n,\overline{T}}}(\Delta_{n,\overline{T}})^{\theta}=O(1)$ for some $\theta\in\left(0,\frac{1}{2}\right)$, then

$$\frac{\delta_{n,\overline{T}}}{h_{n,\overline{T}}}=\frac{\delta_{n,\overline{T}}}{\left(\Delta_{n,\overline{T}}\right)^{\theta}}\frac{(\Delta_{n,\overline{T}})^{\theta}}{h_{n,\overline{T}}}=o_{a.s.}(1) \tag{82}$$

as $n\to\infty$. Using (82) we have

$$\mathbf{K}_{(1)}^{\pm}\left(\frac{\widetilde{X}_{is-}-x}{h_{n,\overline{T}}}\right)=\mathbf{K}_{(1)}^{\pm}\left(\frac{X_{s-}-x}{h_{n,\overline{T}}}+o_{a.s}(1)\right), \tag{83}$$

uniformly over $i=1,...,n$. It follows from (81) through (83) that (79) is bounded by

$$\left(\frac{\delta_{n,\overline{T}}}{h_{n,\overline{T}}}\right)\frac{1}{h_{n,\overline{T}}}\sum_{i=0}^{n-1}\int_{i\overline{T}/n+}^{(i+1)\overline{T}/n}\left|\mathbf{K}_{(1)}^{\pm}\left(\frac{X_{s-}-x}{h_{n,\overline{T}}}+o_{a.s}(1)\right)\right|ds+2C_4\frac{\Delta_{n,\overline{T}}}{h_{n,\overline{T}}}$$

$$
\begin{aligned}
&\leq \left(\frac{\delta_{n,\overline{T}}}{h_{n,\overline{T}}}\right)\frac{1}{h_{n,\overline{T}}}\int_{0+}^{\overline{T}}\left|\mathbf{K}_{(1)}^{\pm}\left(\frac{X_{s-}-x}{h_{n,\overline{T}}}+o_{a.s}(1)\right)\right|ds+2C_4\frac{\Delta_{n,\overline{T}}}{h_{n,\overline{T}}}\\
&= \left(\frac{\delta_{n,\overline{T}}}{h_{n,\overline{T}}}\right)\frac{1}{h_{n,\overline{T}}}\int_{-\infty}^{\infty}\left|\mathbf{K}_{(1)}^{\pm}\left(\frac{p-x}{h_{n,\overline{T}}}+o_{a.s}(1)\right)\right|\overline{L}_X\left(\overline{T},p\right)dp+2C_4\frac{\Delta_{n,\overline{T}}}{h_{n,\overline{T}}}\\
&= \left(\frac{\delta_{n,\overline{T}}}{h_{n,\overline{T}}}\right)\int_{-\infty}^{\infty}\left|\mathbf{K}_{(1)}^{\pm}\left(q+o_{a.s}(1)\right)\right|\overline{L}_X\left(\overline{T},qh_{n,\overline{T}}+x\right)dq+2C_4\frac{\Delta_{n,\overline{T}}}{h_{n,\overline{T}}}\\
&\leq C_5\left(\frac{\delta_{n,\overline{T}}}{h_{n,\overline{T}}}\right)O_{a.s}\left(\overline{L}_X^{\pm}(\overline{T},x)\right)+2C_4\frac{\Delta_{n,\overline{T}}}{h_{n,\overline{T}}}, \qquad (84)
\end{aligned}
$$

for some constant C_5, by virtue of the absolute integrability of $\mathbf{K}_{(1)}^{\pm}$ (from Assumption 4) and the continuity properties of $\overline{L}_X$ from Lemma 2. Since $\frac{\delta_{n,\overline{T}}}{h_{n,\overline{T}}} \to 0$ by (82), then the bound vanishes as $n \to \infty$. This proves the stated result in the case of asymmetric kernels $\mathbf{K}^{\pm}$. The derivation is similar when using a symmetric kernel $\mathbf{K}^{\oplus}$ as in Assumption 4 and is omitted here for brevity.

Proof of Corollary 1. Expressions (37) and (38) follow from Remark 2 of Proposition 1.3 in Revuz and Yor (1998) using the fact that the measure $dL_X(t,x)(w)$ is carried by the set $\{s : X_{s-}(w) = X_s(w) = x\}$ for $a.\ a.\ w$ (c.f. Protter (1995), Theorem 50, page 166). Expression (39) is immediate given recurrence (c.f. Corollary 1 in BP (1998)).

Proof of Theorem 2. We show limit results for the first and the second infinitesimal moment. Extensions to the higher moments are rather straighforward based on the analysis presented below and can be provided by the authors upon request.

We begin with the consistency of the first infinitesimal moment. Write

$$
\begin{aligned}
\widehat{\mathbf{M}}_{n,T}^1(x) &= \frac{\frac{1}{h_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}^{\oplus}\left(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}}\right)\left[X_{(i+1)\Delta_{n,T}}-X_{i\Delta_{n,T}}\right]}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}^{\oplus}\left(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}}\right)} \qquad (85)\\
&= \frac{\frac{1}{h_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}^{\oplus}\left(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}}\right)\int_{i\Delta_{n,T}+}^{(i+1)\Delta_{n,T}}\mu(X_{s-})ds}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}^{\oplus}\left(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}}\right)} \qquad (86)\\
&\quad+\frac{\frac{1}{h_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}^{\oplus}\left(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}}\right)\int_{i\Delta_{n,T}+}^{(i+1)\Delta_{n,T}}\sigma(X_{s-})dW_s}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}^{\oplus}\left(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}}\right)} \qquad (87)\\
&\quad+\frac{\frac{1}{h_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}^{\oplus}\left(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}}\right)\int_{i\Delta_{n,T}+}^{(i+1)\Delta_{n,T}}\int_Y c(X_{s-},y)\overline{\nu}(ds,dy)}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}^{\oplus}\left(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}}\right)}. \qquad (88)\\
&= \alpha_{n,T}(x)+\beta_{n,T}(x)+\gamma_{n,T}(x). \qquad (89)
\end{aligned}
$$

We start with the term $\alpha_{n,T}(x)$. Arguments contained in the proof of Theorem 1 allow us to obtain

$$
\begin{aligned}
&\alpha_{n,T}(x)\\
&= \frac{\frac{1}{h_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}^{\oplus}\left(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}}\right)\int_{i\Delta_{n,T}+}^{(i+1)\Delta_{n,T}}\mu(X_{s-})ds}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}^{\oplus}\left(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}}\right)} \qquad (90)
\end{aligned}
$$

$$= \frac{\frac{1}{h_{n,T}}\int_{T/n}^{T}\mathbf{K}^{\oplus}\left(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}}\right)\mu(X_{s-})ds}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}^{\oplus}\left(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}}\right)} \tag{91}$$

$$= \frac{\frac{1}{h_{n,T}}\int_{0}^{T}\mathbf{K}^{\oplus}\left(\frac{X_{s}-x}{h_{n,T}}\right)\mu(X_{s-})ds + O_{a.s.}\left(\frac{\overline{L}_X^{\oplus}(T,a)}{h_{n,T}}(\Delta_{n,T})^{\frac{1}{2}-\varepsilon}\right)}{\frac{1}{h_{n,T}}\int_{0}^{T}\mathbf{K}^{\oplus}\left(\frac{X_{s}-x}{h_{n,T}}\right)ds + O_{a.s.}\left(\frac{\overline{L}_X^{\oplus}(T,a)}{h_{n,T}}(\Delta_{n,T})^{\frac{1}{2}-\varepsilon}\right)}. \tag{92}$$

Using the Quotient limit Theorem for Harris recurrent Markov processes (c.f. Azema *et al.* (1966) and Revuz and Yor (1998), *inter alia*) for a fixed bandwidth $h_{n,T}$, we can write

$$\alpha_{n,T}(x) \overset{a.s.}{\to} \frac{\int_{\mathfrak{R}}\mathbf{K}^{\oplus}(a)\mu(x+ah_{n,T})\phi(x+ah_{n,T})da}{\int_{\mathfrak{R}}\mathbf{K}^{\oplus}(a)\phi(x+ah_{n,T})da} \tag{93}$$

where $\phi(dx)$ is the $\sigma-$finite invariant measure of the underlying discontinuous semimartingale. In the case of the solution to (1) above, such measure is absolutely continuous with respect to the Lebesgue measure (c.f. Menaldi and Robin (1999)), i.e. $\phi(dx) = \phi(x)dx$. Provided $h_{n,T}$ converges to zero slowly enough as to guarantee that $\frac{\overline{L}_X^{\oplus}(T,x)}{h_{n,T}}(\Delta_{n,T})^{\theta} = O_{a.s.}(1)$ for some $\theta \in \left(0, \frac{1}{2}\right)$ and $h_{n,T}\overline{L}_X^{\oplus}(T,x) \overset{a.s.}{\to} \infty \ \forall x \in \mathfrak{D}$ as in the statement of the theorem, then

$$\alpha_{n,T}(x) \overset{a.s.}{\to} \mu(x). \tag{94}$$

Now consider the term $\beta_{n,T}(x)$. By the strong law of large numbers for martingale difference sequences (MGDS's, henceforth) with zero first moments and finite second moments (c.f. Hall and Heyde (1986)), we can write

$$\beta_{n,T}(x) \overset{a.s.}{\to} 0. \tag{95}$$

The rate of convergence can be found invoking Knight's embedding theorem (c.f. Revuz and Yor (1998)), for example. Fix T $(=\overline{T})$, for simplicity. Define by $\beta_{n,\overline{T}}^{num}(x)$ the numerator of the term $\beta_{n,\overline{T}}(x)$ and write,

$$\begin{aligned}\left(\widehat{\beta}_{n,\overline{T}}^{num}(x)\right)_r &= \sqrt{h_{n,\overline{T}}}\left(\beta_{n,\overline{T}}^{num}(x)\right)_r \\ &= \frac{1}{\sqrt{h_{n,\overline{T}}}}\sum_{i=1}^{[nr]-1}\mathbf{K}^{\oplus}\left(\frac{X_{i\Delta_{n,\overline{T}}}-x}{h_{n,\overline{T}}}\right)\int_{i\Delta_{n,\overline{T}}}^{(i+1)\Delta_{n,\overline{T}}}\sigma(X_{s-})dW_s. \end{aligned}\tag{96}$$

Using the occupation time formula in Lemma 3, the quadratic variation process at r of $\left(\widehat{\beta}_{n,\overline{T}}^{num}(x)\right)_r$ can be expressed as

$$\begin{aligned}\left[\widehat{\beta}_{n,\overline{T}}^{num}(x)\right]_r &= \frac{1}{h_{n,\overline{T}}}\sum_{i=1}^{[nr]-1}\left(\mathbf{K}^{\oplus}\left(\frac{X_{i\Delta_{n,\overline{T}}}-x}{h_{n,\overline{T}}}\right)\right)^2\int_{i\Delta_{n,\overline{T}}}^{(i+1)\Delta_{n,\overline{T}}}\sigma^2(X_{s-})ds \\ &\overset{a.s.}{\to} \left(\int_{-\infty}^{\infty}\left(\mathbf{K}^{\oplus}(s)\right)^2 ds\right)\overline{L}_X^{\oplus}(r\overline{T},x)\sigma^2(x). \end{aligned}\tag{97}$$

Hence, the term $\left(\widehat{\beta}_{n,\overline{T}}^{num}(x)\right)_r$ converges (as $n \to \infty$) to the continuous local martingale $M_{r\overline{T}}(x)$ with increasing process

$$[M(x)]_{r\overline{T}} = \mathbf{K}_2^{\oplus}\overline{L}_X^{\oplus}(r\overline{T},x)\sigma^2(x) \tag{98}$$

where $\mathbf{K}_2^{\oplus} = \int_{-\infty}^{\infty}\left(\mathbf{K}^{\oplus}(s)\right)^2 ds$. In consequence, we can write

$$B_{r\overline{T}} = M_{\tau_{r\overline{T}}}(x) \tag{99}$$

with

$$\tau_{r\overline{T}} = \inf\left\{s : \mathbf{K}_2^{\oplus}\overline{L}_X^{\oplus}(s,x)\sigma^2(x) > r\overline{T}\right\}. \tag{100}$$

Equivalently,

$$M_{r\overline{T}}(x) = B_{\mathbf{K}_2^{\oplus}\overline{L}_X^{\oplus}(r\overline{T},x)\sigma^2(x)}$$

where B and W are independent Brownian motions (see Revuz and Yor (Theorem 2.6, 1998) for the independence property). It follows that

$$\frac{\left(\widehat{\beta}_{n,\overline{T}}^{num}(x)\right)_r}{\frac{\Delta_{n,\overline{T}}}{h_{n,\overline{T}}}\sum_{i=1}^{[rn]}\mathbf{K}^{\oplus}\left(\frac{X_{i\Delta_{n,\overline{T}}}-x}{h_{n,\overline{T}}}\right)} \Rightarrow \mathbf{MN}\left(0, \frac{\mathbf{K}_2^{\oplus}\sigma^2(x)}{\overline{L}_X^{\oplus}(r\overline{T},x)}\right). \tag{101}$$

Finally,

$$\sqrt{h_{n,T}\overline{L}_X^{\oplus}(T,x)}\left(\beta_{n,T}(x)\right) \Rightarrow \mathbf{N}\left(0, \mathbf{K}_2^{\oplus}\sigma^2(x)\right), \tag{102}$$

when $r = 1$ and $\overline{T} \to \infty$ as in the statement of the theorem. Combining (95) and (102), we obtain

$$\beta_{n,T}(x) = O_{a.s.}\left(\frac{1}{\sqrt{h_{n,T}\overline{L}_X^{\oplus}(T,x)}}\right). \tag{103}$$

Now consider the term $\gamma_{n,T}(x)$. Define

$$\begin{aligned} J_{(i+1)\Delta_{n,T}} &= \frac{1}{\sqrt{h_{n,T}}}\int_{i\Delta_{n,T}+}^{(i+1)\Delta_{n,T}}\mathbf{K}^{\oplus}\left(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}}\right)\int_Y c(X_{s-},y)\overline{\nu}(ds,dy) \\ &= \frac{1}{\sqrt{h_{n,T}}}\sum_{i\Delta_{n,T}+\leq s\leq(i+1)\Delta_{n,T}}\mathbf{K}^{\oplus}\left(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}}\right)\Delta X_s \\ &\quad -\frac{1}{\sqrt{h_{n,T}}}\int_{i\Delta_{n,T}+}^{(i+1)\Delta_{n,T}}\mathbf{K}^{\oplus}\left(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}}\right)\left(\lambda(X_{s-})\int_Y c(X_{s-},y)\Pi(y)dy\right)ds. \end{aligned} \tag{104}$$

Notice that $J_{(i+1)\Delta_{n,T}}$ is a martingale difference measurable with respect to $\Im_{(i+1)\Delta_{n,T}}$. Furthermore,

$$\mathbf{E}\left(J_{(i+1)\Delta_{n,T}}\right) = 0 \tag{105}$$

and

$$\begin{aligned} &\lambda_{(i+1)\Delta_{n,T}} \\ = \;&\mathbf{Var}(J_{(i+1)\Delta_{n,T}}) = \frac{1}{h_{n,T}}\mathbf{E}\left(\int_{i\Delta_{n,T}+}^{(i+1)\Delta_{n,T}}\left(\mathbf{K}^{\oplus}\left(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}}\right)\right)^2\left(\lambda(X_{s-})\int_Y c^2(X_{s-},y)\Pi(y)dy\right)ds\right) \\ < \;&\infty \end{aligned} \tag{106}$$

(c.f. Protter, Theorem 38, 1995). Hence, $\left(J_{(i+1)\Delta_{n,T}}, \Im_{(i+1)\Delta_{n,T}}\right)$ is a MGDS with zero mean and finite variance $\lambda_{(i+1)\Delta_{n,T}}$. As earlier, we invoke a standard strong law of large numbers for MGDS*'s* to prove that

$$\gamma_{n,T}(x) \overset{a.s.}{\to} 0. \tag{107}$$

Additionally (see Hall and Heyde (1986, Theorem 3.2, page 58), for instance), we can write

$$\frac{\sum_{i=1}^{n-1} J_{(i+1)\Delta_{n,T}}}{\sqrt{\sum_{i=1}^{n-1}\lambda_{i\Delta_{n,T}+,\ (i+1)\Delta_{n,T}}}} \Rightarrow \mathbf{N}(0,1), \tag{108}$$

where

$$
\begin{aligned}
&\sum_{i=1}^{n-1} \lambda_{i\Delta_{n,T}+,\ (i+1)\Delta_{n,T}} \\
= \quad &\sum_{i=1}^{n-1} \frac{1}{h_{n,T}} \left(\int_{i\Delta_{n,T}+}^{(i+1)\Delta_{n,T}} \left(\mathbf{K}^{\oplus} \left(\frac{X_{i\Delta_{n,T}} - x}{h_{n,T}} \right) \right)^2 \left(\lambda(X_{s-}) \int_Y c^2(X_{s-}, y)\Pi(y)dy \right) ds \right) \\
&\overset{a.s.}{\to} \mathbf{K}_2^{\oplus} \left(\int_Y c^2(x,y)\nu(dy) \right) \overline{L}_X^{\oplus}(T,x), \qquad (109)
\end{aligned}
$$

using standard arguments. In consequence,

$$
\sqrt{h_{n,T}\overline{L}_X^{\oplus}(T,x)} \left(\gamma_{n,T}(x)\right) \Rightarrow \mathbf{N}\left(0, \mathbf{K}_2^{\oplus}\left(\int_Y c^2(x,y)\nu(dy)\right)\right). \qquad (110)
$$

Finally, combining (110) and (107), we obtain

$$
\gamma_{n,T}(x) = O_{a.s.}\left(\frac{1}{\sqrt{h_{n,T}\overline{L}_X^{\oplus}(T,x)}}\right). \qquad (111)
$$

To conclude,

$$
\begin{aligned}
\widehat{\mathbf{M}}_{n,T}^1(x) \quad &= \quad \alpha_{n,T}(x) + \beta_{n,T}(x) + \gamma_{n,T}(x) \qquad (112) \\
&= \quad \mu(x) + O_{a.s.}\left(\frac{1}{\sqrt{h_{n,T}\overline{L}_X^{\oplus}(T,x)}}\right) \overset{a.s.}{\to} \mu(x) \qquad (113)
\end{aligned}
$$

as $h_{n,T}\overline{L}_X^{\oplus}(T,x) \overset{a.s.}{\to} \infty\ \forall x \in \mathfrak{D}$. We now turn to the asymptotic distribution. Write the estimation error decomposition as

$$
\begin{aligned}
&\widehat{\mathbf{M}}_{n,T}^1(x) - \mu(x) \\
= \quad &\frac{\frac{1}{h_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}^{\oplus}\left(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}}\right)\int_{i\Delta_{n,T}+}^{(i+1)\Delta_{n,T}}\mu(X_{s-})}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}^{\oplus}\left(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}}\right)} - \mu(x) \qquad (114) \\
&+\frac{\frac{1}{h_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}^{\oplus}\left(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}}\right)\int_{i\Delta_{n,T}+}^{(i+1)\Delta_{n,T}}\sigma(X_{s-})dW_s}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}^{\oplus}\left(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}}\right)} \qquad (115) \\
&+\frac{\frac{1}{h_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}^{\oplus}\left(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}}\right)\int_{i\Delta_{n,T}+}^{(i+1)\Delta_{n,T}}\int_Y c(X_{s-},y)\overline{\nu}(ds,dy)}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}^{\oplus}\left(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}}\right)}. \qquad (116) \\
= \quad &\alpha_{n,T}^{'}(x) + \beta_{n,T}(x) + \gamma_{n,T}(x) \qquad (117)
\end{aligned}
$$

where $\alpha_{n,T}^{'}(x) = (114)$. From (102) and (110) the terms $\beta_{n,T}(x)$ and $\gamma_{n,T}(x)$ are distributed as

$$
\sqrt{h_{n,T}\overline{L}_X^{\oplus}(T,x)} \left(\beta_{n,T}(x)\right) \Rightarrow \mathbf{N}\left(0, \mathbf{K}_2^{\oplus}\sigma^2(x)\right), \qquad (118)
$$

and

$$
\sqrt{h_{n,T}\overline{L}_X^{\oplus}(T,x)} \left(\gamma_{n,T}(x)\right) \Rightarrow \mathbf{N}\left(0, \mathbf{K}_2^{\oplus}\left(\int_Y c^2(x,y)\nu(dy)\right)\right) \qquad (119)
$$

where

$$\mathbf{K}_2^{\oplus} = \int_{\mathfrak{R}} \left(\mathbf{K}^{\oplus}(s)\right)^2 ds, \tag{120}$$

respectively. Now consider the bias term $\alpha^{'}_{n,T}(x)$. Write

$$\alpha^{'}_{n,T}(x) = \frac{\frac{1}{h_{n,T}}\int_0^T \mathbf{K}^{\oplus}\left(\frac{X_{s-}-x}{h_{n,T}}\right)(\mu(X_{s-})-\mu(x))\,ds + O_{a.s.}\left(\frac{\overline{L}_X^{\oplus}(T,x)\Delta_{n,T}^{1/2-\varepsilon}}{h_{n,T}}\right)}{\frac{1}{h_{n,T}}\int_0^T \mathbf{K}^{\oplus}\left(\frac{X_{s-}-x}{h_{n,T}}\right)ds + O_{a.s.}\left(\frac{\overline{L}_X^{\oplus}(T,x)\Delta_{n,T}^{1/2-\varepsilon}}{h_{n,T}}\right)} \tag{121}$$

We fix the bandwidth and, as earlier, use the Quotient limit Theorem for Harris recurrent processes to obtain

$$\alpha^{'}_{n,T}(x) = h_{n,T}^2 \mathbf{K}_1^{\oplus}\left(\frac{1}{2}\mu^{''}(x) + \mu^{'}(x)\frac{\phi^{'}(x)}{\phi(x)}\right) + o_{a.s.}(1). \tag{122}$$

where $\mathbf{K}_1^{\oplus} = \int_{\mathfrak{R}} u^2 \mathbf{K}^{\oplus}(u)\,du$. The result (122) applies when $h_{n,T} \to 0$ so that $\frac{\overline{L}_X^{\oplus}(T,x)}{h_{n,T}}(\Delta_{n,T})^{\theta} = O_{a.s.}(1)$ for some $\theta \in \left(0, \frac{1}{2}\right)$ and $h_{n,T}\overline{L}_X^{\oplus}(T,x) \overset{a.s.}{\to} \infty\ \forall x \in \mathfrak{D}$. We conclude by noticing that, in light of the independence between $\{W_t : t \geq 0\}$ and $\{N(dt,dy) : t \geq 0\}$ (c.f. Section 2), the quantities $\gamma_{n,T}(x)$ and $\beta_{n,T}(x)$ are also independent. This fact implies that

$$\begin{aligned} &\sqrt{h_{n,T}\overline{L}_X^{\oplus}(T,x)}\left(\widehat{\mathbf{M}}^1_{n,T}(x) - \mu(x)\right) \\ \Rightarrow\ &\mathbf{N}\left(0, \mathbf{K}_2^{\oplus}\left(\sigma^2(x) + \mathbf{E}_Y[c^2(x,y)]\lambda(x)\right)\right) \end{aligned} \tag{123}$$

provided $h_{n,T}^5\overline{L}_X^{\oplus}(T,x) \overset{a.s.}{\to} 0\ \forall x \in \mathfrak{D}$. This proves the stated result for the first infinitesimal moment estimator (c.f. (29)).

We now turn to the second infinitesimal moment and show consistency of the corresponding estimator (c.f. (30)). Consider a generic function $\varphi \in C^2$. A simple extension of Itô's formula to the jump-diffusion setting (c.f. Gikhman and Skorohod (1972) and Protter (Theorem 3.2, 1995)) permits us to write

$$\begin{aligned} d\varphi(X_t) &= \mathcal{L}\varphi(X_{t-})dt + \mathcal{A}\varphi(X_{t-})dt + \varphi_x(X_{t-})\sigma(X_{t-})dW_t \\ &\quad + \int_Y \left[\varphi(X_{t-} + c(X_{t-},y)) - \varphi(X_{t-})\right]\overline{\nu}(dt,dy) \end{aligned} \tag{124}$$

where $\mathcal{L}$ and $\mathcal{A}$ are the second order elliptic operator and the integro-differential operator corresponding to the continuous and discontinuous portions of the process, respectively. More precisely,

$$\mathcal{L}\varphi(.) = \varphi_x(.)\mu(.) + \frac{1}{2}\varphi_{xx}(.)\sigma^2(.) \tag{125}$$

and

$$\mathcal{A}\varphi(.) = \lambda(.)\int_Y \left[\varphi(. + c(.,y)) - \varphi(.) - \varphi_x(.)c(.,y)\right]\Pi(dy). \tag{126}$$

Then,

$$\begin{aligned} dX_s^2 &= 2X_{s-}\mu(X_{s-})ds + 2X_{s-}\sigma(X_{s-})dW_s + \sigma^2(X_{s-})ds + \int_Y \left((X_{s-}+c)^2 - X_{s-}^2\right)\overline{\nu}(dt,dy) \\ &\quad + \lambda(X_{s-})\int_Y \left((X_{s-}+c)^2 - X_{s-}^2 - 2X_{s-}c(X_{s-},y)\right)\Pi(dy)ds \\ &= 2X_{s-}\mu(X_{s-})ds + 2X_{s-}\sigma(X_{s-})dW_s + \sigma^2(X_{s-})ds + \int_Y \left((X_{s-}+c)^2 - X_{s-}^2\right)\overline{\nu}(dt,dy) \\ &\quad + \lambda(X_{s-})\int_Y c^2(X_{s-},y)\Pi(dy)ds, \end{aligned} \tag{127}$$

and

$$
\begin{aligned}
X^2_{(i+1)\Delta_{n,T}} - X^2_{i\Delta_{n,T}} &= 2\int_{i\Delta_{n,T}+}^{(i+1)\Delta_{n,T}} X_{s-}\mu(X_{s-})ds \\
&\quad +2\int_{i\Delta_{n,T}+}^{(i+1)\Delta_{n,T}} X_{s-}\sigma(X_{s-})dW_s \\
&\quad +\int_{i\Delta_{n,T}+}^{(i+1)\Delta_{n,T}} \sigma^2(X_{s-})ds \\
&\quad +\int_{i\Delta_{n,T}+}^{(i+1)\Delta_{n,T}} \left(\int_Y c^2(X_{s-},y)\Pi(dy)\right)\lambda(X_{s-})ds \\
&\quad +\int_{i\Delta_{n,T}+}^{(i+1)\Delta_{n,T}} \int_Y \left((X_{s-}+c)^2 - X^2_{s-}\right)\overline{\nu}(ds,dy). \qquad (128)
\end{aligned}
$$

Finally,

$$
\begin{aligned}
&\left(X_{(i+1)\Delta_{n,T}} - X_{i\Delta_{n,T}}\right)^2 \\
=\;& X^2_{(i+1)\Delta_{n,T}} - X^2_{i\Delta_{n,T}} - 2X_{i\Delta_{n,T}}[X_{(i+1)\Delta_{n,T}} - X_{i\Delta_{n,T}}] \\
=\;& 2\int_{i\Delta_{n,T}+}^{(i+1)\Delta_{n,T}} X_{s-}\mu(X_{s-})ds + 2\int_{i\Delta_{n,T}+}^{(i+1)\Delta_{n,T}} X_{s-}\sigma(X_{s-})dW_s + \int_{i\Delta_{n,T}+}^{(i+1)\Delta_{n,T}} \sigma^2(X_{s-})ds \\
&+\int_{i\Delta_{n,T}+}^{(i+1)\Delta_{n,T}} \left(\int_Y c^2(X_{s-},y)\Pi(dy)\right)\lambda(X_{s-})ds + \int_{i\Delta_{n,T}+}^{(i+1)\Delta_{n,T}} \int_Y \left((X_{s-}+c)^2 - X^2_{s-}\right)\overline{\nu}(ds,dy) \\
&-2\int_{i\Delta_{n,T}+}^{(i+1)\Delta_{n,T}} X_{i\Delta_{n,T}}\mu(X_{s-})ds - 2\int_{i\Delta_{n,T}+}^{(i+1)\Delta_{n,T}} X_{i\Delta_{n,T}}\sigma(X_{s-})dW_s \\
&-2\int_{i\Delta_{n,T}+}^{(i+1)\Delta_{n,T}} \int_Y X_{i\Delta_{n,T}}c(X_{s-},y)\overline{\nu}(ds,dy) \\
=\;& 2\int_{i\Delta_{n,T}+}^{(i+1)\Delta_{n,T}} \left(X_{s-} - X_{i\Delta_{n,T}}\right)\mu(X_{s-})ds + 2\int_{i\Delta_{n,T}+}^{(i+1)\Delta_{n,T}} \left(X_{s-} - X_{i\Delta_{n,T}}\right)\sigma(X_{s-})dW_s \\
&-2\int_{i\Delta_{n,T}+}^{(i+1)\Delta_{n,T}} \int_Y X_{i\Delta_{n,T}}c(X_{s-},y)\overline{\nu}(ds,dy) + \int_{i\Delta_{n,T}+}^{(i+1)\Delta_{n,T}} \sigma^2(X_{s-})ds \\
&+\int_{i\Delta_{n,T}+}^{(i+1)\Delta_{n,T}} \left(\int_Y c^2(X_{s-},y)\Pi(dy)\right)\lambda(X_{s-})ds + \int_{i\Delta_{n,T}+}^{(i+1)\Delta_{n,T}} \int_Y \left((X_{s-}+c)^2 - X^2_{s-}\right)\overline{\nu}(ds,dy). \\
&\qquad (129)
\end{aligned}
$$

Now write,

$$
\widehat{\mathbf{M}}^2_{n,T}(x) = \frac{\frac{1}{h_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}^{\oplus}\left(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}}\right)2\int_{i\Delta_{n,T}+}^{(i+1)\Delta_{n,T}}\left(X_{s-}-X_{i\Delta_{n,T}}\right)\mu(X_{s-})ds}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}^{\oplus}\left(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}}\right)} \qquad (130)
$$

$$
+\frac{\frac{1}{h_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}^{\oplus}\left(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}}\right)2\int_{i\Delta_{n,T}+}^{(i+1)\Delta_{n,T}}\left(X_{s-}-X_{i\Delta_{n,T}}\right)\sigma(X_{s-})dW_s}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}^{\oplus}\left(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}}\right)} \qquad (131)
$$

$$
-\frac{\frac{1}{h_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}^{\oplus}\left(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}}\right)2\int_{i\Delta_{n,T}+}^{(i+1)\Delta_{n,T}}\int_Y X_{i\Delta_{n,T}}c(X_{s-},y)\overline{\nu}(ds,dy)}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}^{\oplus}\left(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}}\right)} \qquad (132)
$$

$$+\frac{\frac{1}{h_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}^{\oplus}\left(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}}\right)\int_{i\Delta_{n,T}+}^{(i+1)\Delta_{n,T}}\left(\sigma^2(X_{s-})+\left(\int_Y c^2(X_{s-},y)\Pi(dy)\right)\lambda(X_{s-})\right)ds}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}^{\oplus}\left(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}}\right)} \tag{133}$$

$$+\frac{\frac{1}{h_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}^{\oplus}\left(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}}\right)\int_{i\Delta_{n,T}+}^{(i+1)\Delta_{n,T}}\int_Y\left((X_{s-}+c)^2-X_{s-}^2\right)\overline{\nu}(ds,dy)}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}^{\oplus}\left(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}}\right)} \tag{134}$$

$$= a_{n,T}(x)+b_{n,T}(x)+c_{n,T}(x)+d_{n,T}(x)+e_{n,T}(x). \tag{135}$$

Previous arguments suggest that

$$\begin{aligned} d_{n,T}(x) &= \frac{\frac{1}{h_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}^{\oplus}\left(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}}\right)\int_{i\Delta_{n,T}+}^{(i+1)\Delta_{n,T}}\left(\sigma^2(X_{s-})+\left(\int_Y c^2(X_{s-},y)\Pi(dy)\right)\lambda(X_{s-})\right)ds}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}^{\oplus}\left(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}}\right)} \\ &\overset{a.s.}{\to} \sigma^2(x)+\left(\int_Y c(x,y)\Pi(dy)\right)\lambda(x)=\sigma^2(x)+\mathbf{E}_Y\left[c^2(x,y)\right]\lambda(x). \end{aligned} \tag{136}$$

Some of the remaining quantities ($b_{n,T}(x), c_{n,T}(x)$ and $e_{n,T}(x)$, that is) are sample averages of MDGS*'s* converging to zero at some rate. As for $a_{n,T}(x)$ note that

$$\begin{aligned} a_{n,T}(x) &= \frac{\frac{1}{h_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}^{\oplus}\left(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}}\right)2\int_{i\Delta_{n,T}+}^{(i+1)\Delta_{n,T}}\left(X_{s-}-X_{i\Delta_{n,T}}\right)\mu(X_{s-})ds}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}^{\oplus}\left(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}}\right)} \\ &\leq O_{a.s.}\left(\Delta_{n,T}^{1/2}\right)\frac{\frac{1}{h_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}^{\oplus}\left(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}}\right)2\int_{i\Delta_{n,T}+}^{(i+1)\Delta_{n,T}}\mu(X_{s-})ds}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}^{\oplus}\left(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}}\right)} \\ &= O_{a.s.}\left(\Delta_{n,T}^{1/2}\right)\left(\mu(x)+o_{a.s.}(1)\right)\overset{a.s.}{\to}0. \end{aligned} \tag{137}$$

Finally,

$$\widehat{\mathbf{M}}_{n,T}^2(x)\overset{a.s.}{\to}\sigma^2(x)+\mathbf{E}_Y\left[c^2(x,y)\right]\lambda(x). \tag{138}$$

We now evaluate the limiting distribution. Write the estimation error decomposition as

$$\begin{aligned} &\left[\widehat{\mathbf{M}}_{n,T}^2(x)-\left(\sigma^2(x)+\left(\int_Y c^2(x,y)\Pi(dy)\right)\lambda(x)\right)\right] \\ &= \frac{\frac{1}{h_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}^{\oplus}\left(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}}\right)\int_{i\Delta_{n,T}+}^{(i+1)\Delta_{n,T}}\left(\sigma^2(X_{s-})+\left(\int_Y c(X_{s-},y)\Pi(dy)\right)\lambda(X_{s-})\right)ds}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}^{\oplus}\left(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}}\right)} \\ &\quad -\left(\sigma^2(x)+\left(\int_Y c^2(x,y)\Pi(dy)\right)\lambda(x)\right)+a_{n,T}(x)+b_{n,T}(x)+c_{n,T}(x)+e_{n,T}(x) \\ &= d_{n,T}^{'}(x)+a_{n,T}(x)+b_{n,T}(x)+c_{n,T}(x)+e_{n,T}(x) \end{aligned} \tag{139}$$

where

$$\begin{aligned} d_{n,T}^{'}(x) &= \frac{\frac{1}{h_{n,T}}\sum_{i=1}^{n-1}\mathbf{K}^{\oplus}\left(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}}\right)\int_{i\Delta_{n,T}+}^{(i+1)\Delta_{n,T}}\left(\sigma^2(X_{s-})+\left(\int_Y c(X_{s-},y)\Pi(dy)\right)\lambda(X_{s-})\right)ds}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}^{\oplus}\left(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}}\right)} \\ &\quad -\left(\sigma^2(x)+\left(\int_Y c^2(x,y)\Pi(dy)\right)\lambda(x)\right). \end{aligned} \tag{140}$$

Notice that

$$a_{n,T}(x) = o_{a.s.}(c_{n,T}(x)), \tag{141}$$

$$b_{n,T}(x) = o_{a.s.}(c_{n,T}(x)), \tag{142}$$

and

$$c_{n,T}(x) = O_{a.s.}(e_{n,T}(x)). \tag{143}$$

From (141), (142) and (143) the limit distribution depends on the relationship between the speeds of convergence of the bias term $d'_{n,T}(x)$ and $e_{n,T}(x) + c_{n,T}(x)$. Under the conditions that we assumed for the smoothing sequence $h_{n,T}$, we obtain

$$d'_{n,T}(x) = O(h^2_{n,T}) \tag{144}$$

as in the case of (122) above. In fact,

$$d'_{n,T}(x) = h^2_{n,T}\mathbf{K}^{\oplus}_1 \left(\frac{1}{2}\left(\mathbf{M}^2(x)\right)'' + \left(\mathbf{M}^2(x)\right)' \frac{\phi'(x)}{\phi(x)} \right) + o_{a.s.}(1), \tag{145}$$

where $\mathbf{K}^{\oplus}_1$ was defined earlier. Now write,

$$\begin{aligned} \left(\widehat{e}^{num}_{n,T}(x)\right) &= \sqrt{h_{n,T}}\left(e^{num}_{n,T}(x)\right) \\ &= \frac{1}{\sqrt{h_{n,T}}} \sum_{i=1}^{n-1} \mathbf{K}^{\oplus}\left(\frac{X_{i\Delta_{n,T}} - x}{h_{n,T}}\right) \int_{i\Delta_{n,T}+}^{(i+1)\Delta_{n,T}} \int_{Y} \left((X_{s-}+c)^2 - X^2_{s-}\right) \overline{\nu}(ds,dy). \end{aligned} \tag{146}$$

As before,

$$\begin{aligned} &\sqrt{h_{n,T}\overline{L}^{\oplus}_X(T,x)} \left(\frac{e^{num}_{n,T}(x)}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}^{\oplus}\left(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}}\right)} \right) \\ \Rightarrow\ & \mathbf{N}\left(0, \mathbf{K}^{\oplus}_2\left(\lambda(x)\mathbf{E}_Y\left[\left((x+c(x,y))^2 - x^2\right)^2\right]\right)\right). \end{aligned} \tag{147}$$

Also write,

$$\begin{aligned} \left(\widehat{c}^{num}_{n,T}(x)\right) &= \sqrt{h_{n,T}}\left(c^{num}_{n,T}(x)\right) \\ &= -\frac{1}{\sqrt{h_{n,T}}} \sum_{i=1}^{n-1} \mathbf{K}^{\oplus}\left(\frac{X_{i\Delta_{n,T}} - x}{h_{n,T}}\right) 2\int_{i\Delta_{n,T}+}^{(i+1)\Delta_{n,T}} \int_{Y} X_{i\Delta_{n,T}} c(X_{s-},y)\overline{\nu}(ds,dy). \end{aligned} \tag{148}$$

Then,

$$\begin{aligned} &\sqrt{h_{n,T}\overline{L}^{\oplus}_X(T,x)} \left(\frac{c^{num}_{n,T}(x)}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}^{\oplus}\left(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}}\right)} \right) \\ \Rightarrow\ & \mathbf{N}\left(0, \mathbf{K}^{\oplus}_2\left(4\lambda(x)\mathbf{E}_Y\left[c^2(x,y)x^2\right]\right)\right). \end{aligned} \tag{149}$$

Finally, the limiting covariance between $\sqrt{h_{n,T}}\left(c_{n,T}(x)\right)$ and $\sqrt{h_{n,T}}\left(e_{n,T}(x)\right)$ can be characterized as

$$
\begin{aligned}
& Asicov\left(\frac{\sqrt{h_{n,T}}c_{n,T}^{num}(x)}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}^{\oplus}\left(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}}\right)}, \frac{\sqrt{h_{n,T}}e_{n,T}^{num}(x)}{\frac{\Delta_{n,T}}{h_{n,T}}\sum_{i=1}^{n}\mathbf{K}^{\oplus}\left(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}}\right)}\right) \\
& \overset{a.s.}{\to} \frac{1}{\overline{L}_X^{\oplus}(T,x)}\mathbf{K}_2^{\oplus}\left(-2\lambda(x)\mathbf{E}_Y\left[xc(x,y)\left((x+c)^2-x^2\right)\right]\right). \qquad (150)
\end{aligned}
$$

Hence,

$$
\begin{aligned}
& \sqrt{h_{n,T}\overline{L}_X^{\oplus}(T,x)}\left(\widehat{\mathbf{M}}_{n,T}^2(x)-\left(\sigma^2(x)+\left(\int_Y c(x,y)\Pi(dy)\right)\lambda(x)\right)\right) \\
\Rightarrow \quad & \mathbf{N}\left(0,\mathbf{K}_2^{\oplus}\left(\lambda(x)\mathbf{E}_Y\left[c^4(x,y)\right]\right)\right), \qquad (151)
\end{aligned}
$$

with $\mathbf{K}_2^{\oplus}=\int_{-\infty}^{\infty}\left(\mathbf{K}^{\oplus}(s)\right)^2 ds$. This proves the stated result for the second infinitesimal moment estimator.

Notation

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

References

[1] Andersen, T., Benzoni, L., Lund J., 1998. Estimating jump-diffusions for equity returns. Working paper, Northwestern.

[2] Bakshi, G., Cao, C., Chen Z., 1997. Empirical performance of alternative option pricing models. Journal of Finance 52, 2003-49.

[3] Bandi, F.M., 2000. Nonparametric fixed-income pricing. Unpublished working paper, University of Chicago.

[4] Bandi, F.M., 2001. The Short-Term Interest Rate Dynamics: a Spatial Approach. *The Journal of Financial Economics, Forthcoming.*

[5] Bandi, F.M., Nguyen T., 1999. Fully nonparametric estimators for diffusions: a small sample analysis. Unpublished working paper, Yale University.

[6] Bandi, F.M., Phillips P.C.B., 1998. Fully nonparametric estimation of scalar diffusion models. Unpublished working paper, University of Chicago.

[7] Chacko, G., Viceira L.M., 1999. Spectral GMM estimation of continuous-time processes. Working paper, Harvard University.

[8] Chu, J.S., Wu C.K., 1993. Kernel-type estimators of jump points and values of a regression function. Annals of Statistics 21, 1545-1566.

[9] Das, S., 1998. Poisson-Gaussian processes and the bond market. NBER Working paper, #6631.

[10] Davydov Y., 1976. Local times of random processes. Th. Probab. and Appl. 21, 1, p. 172-179.

[11] Delgado M.A., Hidalgo J., 2000. Nonparametric inference on structural breaks. Journal of Econometrics 96, 113-143.

[12] Duffie, D., Pan, J., Singleton, K., 2000. Transform analysis and asset pricing for affine jump-diffusions. Econometrica 68, 1343-1376.

[13] Gikhman, I.I., Skorohod A.V., 1972. Stochastic Differential Equations. Springer-Verlag.

[14] Hall, P., Heyde C.C., 1986. Martingale Limit Theory and its Applications. New York: Academic Press.

[15] Johannes M., 2000. Jumps in interest rates: a nonparametric approach. Working paper, University of Chicago.

[16] Karatzas, I., Shreve, S.E., 1988. Brownian Motion and Stochastic Calculus. Springer-Verlag, New York.

[17] Karlin, S., Taylor, 1981. A Second Course in Stochastic Processes. Academic Press, New York.

[18] Madan, D., Carr, P., Chang E.C., 1998. The Variance Gamma process and option pricing. Working Paper, University of Maryland.

[19] Madan, D., Seneta, E., 1990. The Variance Gamma (V.G.) model for share market returns. Journal of Business, 511-524.

[20] Menaldi, J.L., Robin, M., 1999. Invariant measure for diffusions with jumps. Applied Mathematics and Optimization 40, 105-140.

[21] Meyn, S.R., Tweedie R.L., 1993. Markov Chains and Stochastic Stability. Springer Verlag.

[22] Müller, H.G., 1992. Change-points in nonparametric regression analysis. Annals of Statistics 20, 737-761.

[23] Pagan, A, Ullah A., 1999. Nonparametric Econonometrics. Cambridge University Press.

[24] Perron, B., 1999. Jumps in the volatility of financial markets. Working paper, University of Montreal.

[25] Piazzesi, M., 2000. Monetary policy and macroeconomic variables in a model of the term structure of interest rates. Working paper, Stanford University.

[26] Praetz, P., 1972. The distribution of share price changes. Journal of Business, 49-55.

[27] Protter, P., 1995. Stochastic Integration and Differential Equations. Springer-Verlag, New York.

[28] Revuz, D., Yor, M., 1998. Continuous Martingales and Brownian Motion. Second Edition, Springer-Verlag, New York.

[29] Singleton, K., 2001. Estimation of affine asset pricing models based on the empirical characteristic function. Journal of Econometrics 102, 111-141.

[30] Wee, I., 2000. Recurrence and transience for jump-diffusion processes. Forthcoming in Stochastic Analysis and Applications.

[31] Yin, Y.Q., 1988. Detection of the number, locations and magnitudes of jumps. Communications in Statistics: Stochastic Models 4, 445-455.

[32] Yor, M., 1978. Rappels et préliminaires généraux. Astérisque 52-53, 17-22.

[33] Yor, M., 1978. Sur la continuité des temps locaux associés à certaines semi-martingales. Astérisque 52-53, 23-35.

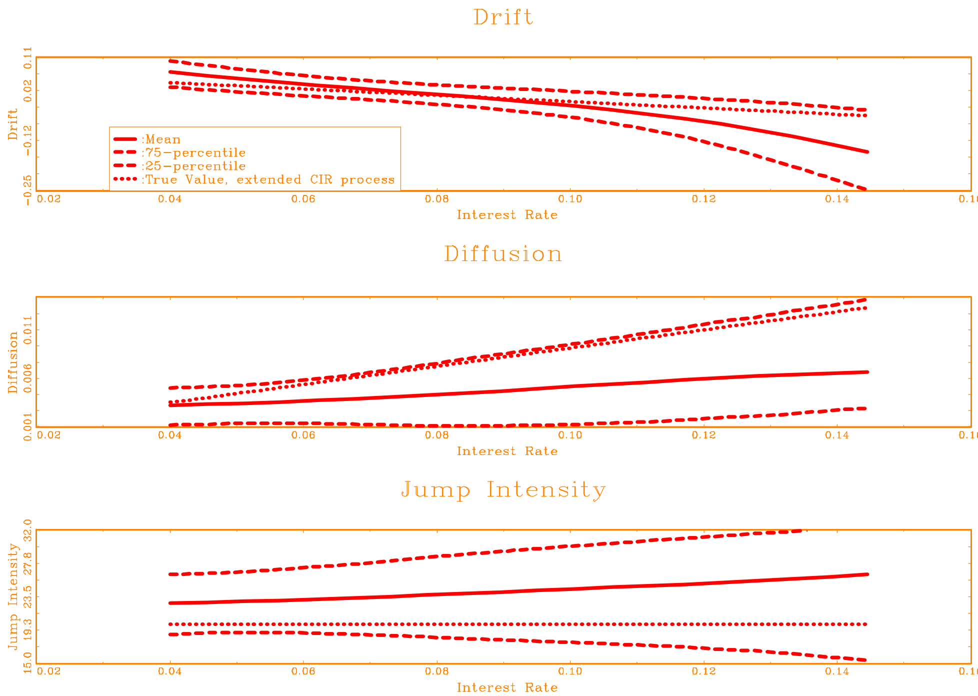

Figure 1: Mean, 75 percentile and 25 percentile of the estimates of drift, diffusive volatility and jump intensity. The results are obtained from simulating 10000 paths, and 10000 daily observations for every path, of the jump-diffusion process $d\ln r = \beta(\alpha - r) + \sigma\sqrt{r}dW + dJ$ with $y \sim N\left(0, \sigma_y^2\right)$, $\lambda = 20$, $\beta = .85837$, $\alpha = .089102$, $\sigma = .3$ and $\sigma_y = .03630427$. We employ a second order Gaussian kernel. The value of the bandwidth parameter is set equal to .015.

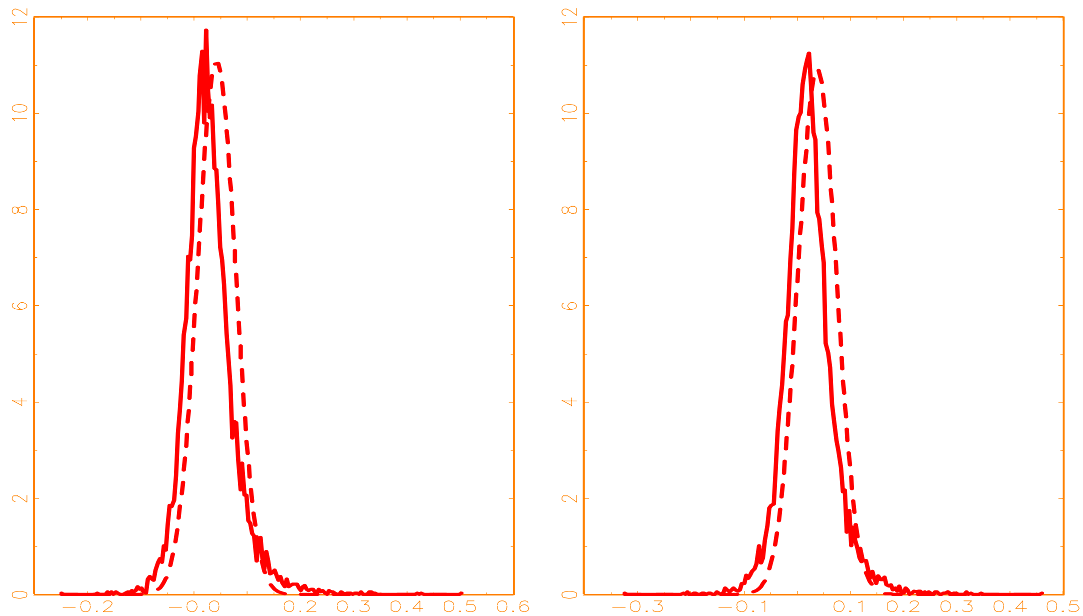

Figure 2: Distributions of the estimates of the first moment for interest rate levels equal to 6% and 7%. The dashed line is the normal distribution dictated from the asymptotic theory while the continuous line is the real distribution of the estimates. We employ a second order Gaussian kernel. The value of the bandwidth parameter is set equal to .015.

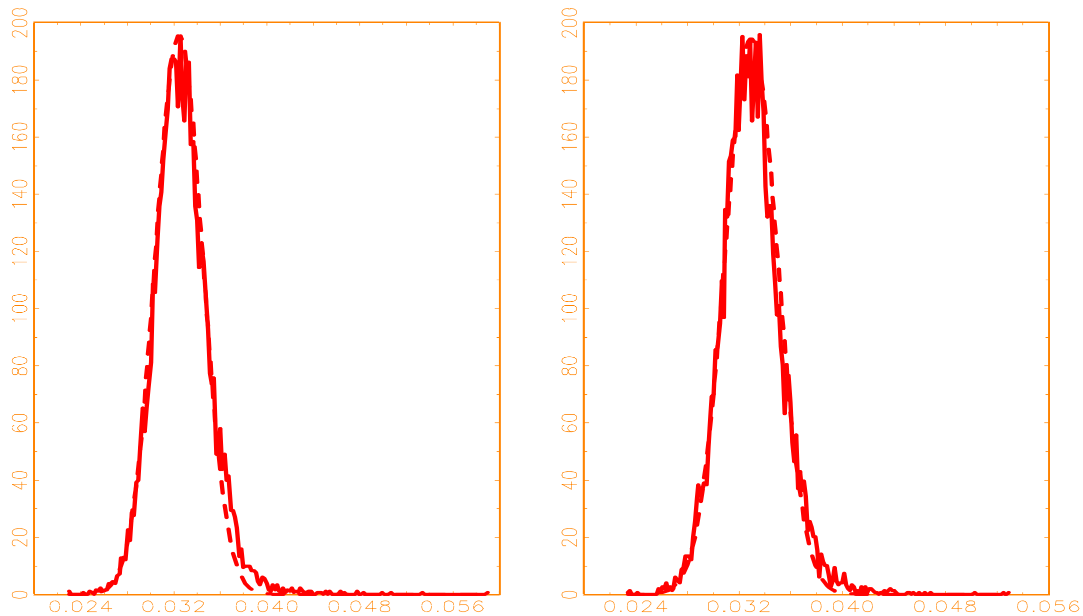

Figure 3: Distributions of the estimates of the second moment for interest rate levels equal to 6% and 7%. The dashed line is the normal distribution dictated from the asymptotic theory while the continuous line is the real distribution of the estimates. We employ a second order Gaussian kernel. The value of the bandwidth parameter is set equal to .015.

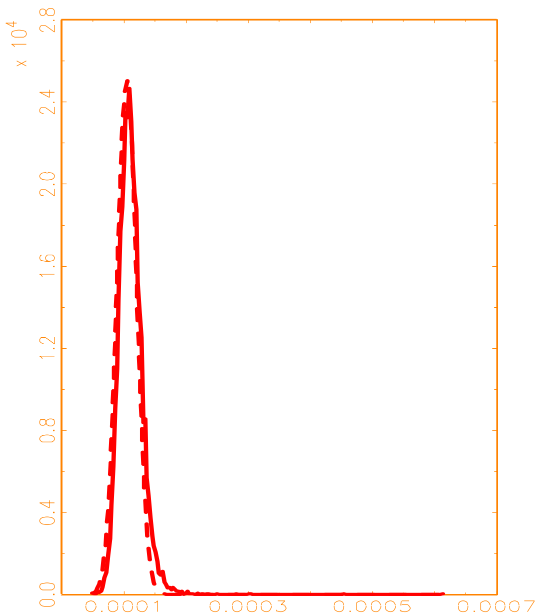

Figure 4: Distributions of the estimates of the fourth moment for interest rate levels equal to 6% and 7%. The dashed line is the normal distribution dictated from the asymptotic theory while the continuous line is the real distribution of the estimates. We employ a second order Gaussian kernel. The value of the bandwidth parameter is set equal to .015.

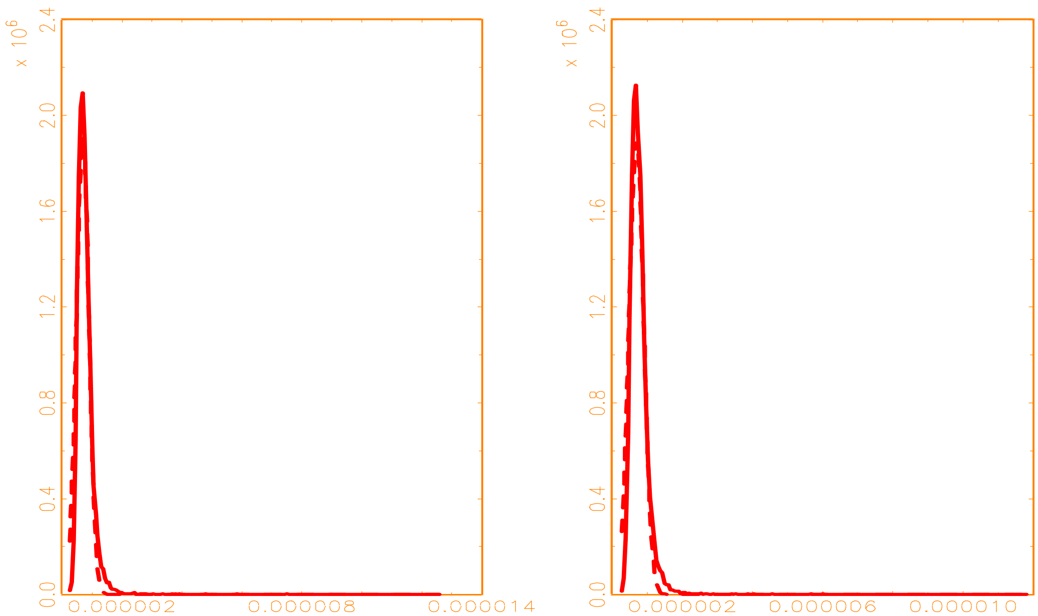

Figure 5: Distributions of the estimates of the sixth moment for interest rate levels equal to 6% and 7%. The dashed line is the normal distribution dictated from the asymptotic theory while the continuous line is the real distribution of the estimates. We employ a second order Gaussian kernel. The value of the bandwidth parameter is set equal to .015.