Threshold estimation of jump-diffusion models and interest rate modeling*

Cecilia Mancini[§] and Roberto Renò[*]

September 15, 2006

Abstract

We introduce nonparametric estimators of the coefficients of a univariate jump-diffusion process when observations are recorded discretely. We allow the drift, diffusion and intensity function to be level dependent. We also show that the estimator of the diffusion coefficient is consistent even if the jump process has infinite activity. Our results rely on the fact that it is possible to disentangle the discontinuous part of the state variable through those squared increments between observations exceeding a suitable *threshold*. We use our estimators to reexamine the estimation of models for the short interest rate. With respect to alternative estimates, the proposed threshold estimators provide very different results from those in the literature.

* We wish to acknowlege Michael Johannes and Angelo Melino for useful comments, as well as partecipants at the CIREQ conference on realized volatility, held in Montreal in April, 2006. All errors and omissions are our own.

§ Dipartimento di Matematica per le Decisioni, via C. Lombroso 6/17, Università di Firenze, email: cecilia.mancini@dmd.unifi.it

* Dipartimento di Economia Politica, Piazza S. Francesco 7, Università di Siena, email: reno@unisi.it

1 Introduction

In this paper, we focus on nonparametric estimation of univariate models with stochastic volatility and jumps. We describe the evolution of a state variable, e.g. an interest rate or a logarithmic asset price, in the form of a semimartingale X which is constructed from a drift component, a continuous diffusion with stochastic volatility, to model continuous changes of the state variable, and by a discontinuous part to model its abrupt changes. The model is:

$$dX_t = \mu(X_t)dt + \sigma(X_t)dW_t + dJ_t, \qquad t \in [0, T],$$

where W_t is a standard Brownian motion, and J_t can either be a Lévy process, with finite or infinite activity (with finite variation), or a finite activity doubly stochastic compound Poisson process with level-dependent intensity $\lambda(X_{t-})$. Suitable assumptions on the functions $\mu(\cdot), \sigma(\cdot), \lambda(\cdot)$ are detailed in the text.

The main contribution of this paper is to introduce a fully non-parametric estimator of the functions $\mu(x), \sigma(x)$, as well as of $\lambda(x)$ when J is a doubly stochastic Poisson process. The estimators are based on data sampling in the form of n observations of the state variable X_t in a time span $[0, T]$. We also dispose of an estimate of times and sizes of the jumps of J_t. In particular, the proposed estimator of the diffusion coefficient is proved to be consistent even in presence of jumps with infinite activity and finite variation, such as the Variance Gamma process (Madan, 2001), the CGMY model (Carr et al., 2002) or the class of α-stable processes provided that the Blumenthal-Gatoor index is less than one, see Cont and Tankov (2004).

In the recent literature, many intriguing methodologies have been proposed to separate the variations in the state variable X_t due to the diffusive part from those due to jumps with a feasible econometric technique. In parametric models, parameterized functional forms are imposed for the drift and diffusion functions and for the jump component. In this framework Jiang and Oomen (2004) develop estimators based on a weighted sum of squared increments in an affine asset pricing model where the jump part has finite activity. Ait-Sahalia (2004) shows that it is possible to disentangle jumps from the continuous variations using a maximum likelihood approach. Aït-Sahalia and Jacod (2006) construct a threshold based estimator of the volatility when each source of randomness is a stable Lévy process.

In a nonparametric framework, Barndorff-Nielsen and Shephard (2006, 2004); Woerner (2005); Mancini (2004a); Jacod (2006) estimate the *integrated* volatility. Barndorff-Nielsen and Shephard (2004) show that, when the volatility is independent of the Brownian motion, the *power variation* of the state variable is a consistent estimator of the integral of the corresponding power of the volatility, even in presence of a finite activity jump process. Woerner (2005); Barndorff-Nielsen et al. (2006); Jacod (2006) develop extensions to the case of infinite activity jump component. Barndorff-Nielsen and Shephard (2006) develop the original theory of the *bi-power variation*, which allows the estimation of the integrated volatility in presence of finite activity jumps. The theory of bi-power variation has inspired new directions for research on financial data, see e.g. Andersen et al. (2006); Huang and Tauchen (2005); Tauchen and Zhou (2006) among others. It also allows to test for the presence of jumps. An alternative test has been devised by Jiang and Oomen (2006).

The primary aim of this paper is to estimate the *local* volatility as a function of the level of the state variable. Nonparametric estimation of the diffusion coefficient $\sigma(\cdot)$, and/or the drift coefficient $\mu(\cdot)$ has been studied, in absence of the jump component, by Florens-Zmirou (1993); Jiang and Knight (1997); Stanton (1997); Jiang (1998); Bandi and Phillips (2003), see also the review of Fan (2005).

In presence of jumps, the only nonparametric estimator of the local volatility we are aware of has been proposed by Bandi and Nguyen (2003) and Johannes (2004). Their model contains a finite activity jump part, not necessarily of Lévy type since it has stochastic jump intensity. Their theory is based on nonparametric estimation of the moments of the state variable. Loosely speaking, the presence of jumps is revealed by excess kurtosis. The estimator proposed here is basically different. We build on the work of Mancini (2004a,b), who shows that when the interval between two observations shrinks, since the diffusive part tends to zero at a known rate, namely the modulus of continuity of the Brownian motion, it is possible to tell when there are jumps in the interval. In the case of finite jump activity, this allows to identify asymptotically both jump times and sizes. By removing the estimated jump contribution to the dynamics of the state variable, we get an estimate of its continuous path, on which we can implement nonparametric techniques.

The importance of such an estimate stems mainly from the fact that stochastic volatility models with jumps are used in a variety of financial applications. For interest rate modeling, Das (2002); Piazzesi (2005) show that the role of jumps is relevant in incorporating newly released information in interest rate levels. The statistical and economic role of jumps in interest rate modeling is further discussed in Johannes (2004). Bond pricing for jump-diffusions is discussed in Eberlein and Raible (1999). Jumps are also very important for derivative pricing, since option writers and buyers are aware of the possibility of sudden changes of the underlying, so that they demand a higher risk premium which affects the term structure of implied volatility, see Bakshi et al. (1997); Bates (2000); Eraker et al. (2003); Andersen et al. (2002); Pan (2002). Also pure-jump processes have received a lot of attention, see Madan (2001); Carr et al. (2002). An important problem for risk management and hedging strategies is to identify the contribution given to X separately by each component, see e.g. Andersen et al. (2005).

For this reason, after assessing the consistency of the proposed estimators in finite samples using Monte Carlo simulations, we use them in the framework of interest rates modeling. We first show that the estimation of the variance is reliable in realistic simulations of the short term interest rates, and that it provides superior performance with respect to the classical estimator, which is designed for continuous stochastic processes.

In order to estimate a model for the short rate, we first have to choose a proxy for it. We show that the 7-day Eurodollar deposit rate should not be used as a proxy, since it contains an inherent jump process triggered by liquidity reasons. When we estimate and exclude the jump component, we find almost the same volatility on the 7-day rate and on the three-month T-bill rate. We then compare our results with those in Johannes (2004).

This paper is organized as follows. Section 2 reviews the results on nonparametric estimation in the literature. The results when J has finite jump activity are presented in Section 3. In Section 4 we show that our proposed estimator of the diffusion coefficient is consistent even when J has infinite jump activity with finite variation. In Section 5 we test the proposed estimators on simulated experiments. In Section 6 we the threshold estimators to estimate jump-diffusion models of the short rate, and we compare the obtained results with those in the recent literature. Section 7 concludes.

2 Model setup

In this Section, we set up the model and the assumptions and we recall some preliminary results. We model the evolution of an (observable) state variable by a stochastic process X_t in the time interval $[0, T]$. In financial applications, X_t can be thought of as the short interest rate, a foreign exchange rate, or the logarithm of an asset price or of a stock index.

We work in a filtered probability space (Ω, $(\mathcal{F}_t)_{t\in[0,T]}$, $\mathcal{F}$, P), satisfying the usual conditions (Protter, 1990), where W is a standard Brownian motion and J is a pure jump Lévy process (Cont and Tankov, 2004) or a doubly stochastic Poisson process with finite activity (Brémaud, 1981). We then assume that $(X_t)_{t\in[0,T]}$ is a real process such that $X_0 \in \mathbb{R}$ and

$$dX_t = \mu(X_t)dt + \sigma(X_t)dW_t + dJ_t,\ t \in]0, T]. \tag{2.1}$$

We require the following assumption throughout all the paper.

Assumption 2.1 *$\mu_t \doteq \mu(X_t)$, $\sigma_t \doteq \sigma(X_t)$ are progressively measurable processes with cadlag paths such to guarantee that the SDE (2.1) has a unique strong solution which is adapted and right continuous with left limits on $[0, T]$, see Jacod (1979); Ikeda and Watanabe (1981); Protter (1990).*

When J is a Lèvy process, it can be decomposed as the sum of the jumps larger than one and the sum of the compensated jumps smaller than one. Following this decomposition, we write $J \equiv J_1 + \tilde{J}_2$, where

$$J_{1s} \doteq \int_0^s \int_{|x|>1} x m(dt, dx) \text{ and } \tilde{J}_{2s} \doteq \int_0^s \int_{|x|\le 1} x[m(dt, dx) - \nu(dx)dt],$$

m being the jumps random measure of X, and ν being the Lévy measure of J. The process J_1 of the jumps with size larger than one is a finite activity compound Poisson process and can be also written $J_{1s} = \sum_{\ell=1}^{N_s^1} \gamma_\ell^1$, see Cont and Tankov (2004). For simplicity we write N in place of N^1, and γ in place of γ^1.

When J is a doubly stochastic Poisson process, we have $J_s = \sum_{\ell=1}^{N_s} \gamma_\ell$, where γ_ℓ are not any longer bound to be larger than one.

Typically, we do not observe X_t continuously, but in form of $n+1$ discrete time observations $\{X_0, X_{t_1}, ..., X_{t_{n-1}}, X_{t_n}\}$, with a constant interval $t_i - t_{i-1} = \delta$ between observations, and $t_n = T$. Extension to the case of deterministic unevenly spaced date is straightforward, as in Mancini (2004a).

For a given semimartingale Z and a given cadlag process H, we use the following notations:

- $\Delta_i Z = Z_{t_i} - Z_{t_{i-1}}$, the increment of Z between t_{i-1} and t_i
- $\Delta Z_t = Z_t - Z_{t-}$ the size of the jump, if any, at time t
- $[Z]_t$ denotes the quadratic variation of Z. $[Z]_t^c$ denotes the quadratic variation of the continuos martingale part of Z.

- We denote by $(\tau_j)_{j\in\mathbb{N}}$ the jump instants of J_1 and by $\tau^{(i)}$ the instant of the first jump in $]t_{i-1}, t_i]$, if $\Delta_i N \geq 1$
- $H.W$ is the process given by the stochastic integral $\left(\int_0^t H_s dW_s\right)_{t\in[0,T]}$
- $\tilde{m}$ is the compensated measure of the Lévy measure m: $\tilde{m}(dt, dx) = m(dt, dx) - \nu(dx)dt$.

We borrow from classical kernel estimation theory, so we introduce the bandwidth.

Definition 2.2 *A bandwidth parameter is a sequence of real numbers h such that as $n \to \infty$ we have $h \to 0$ and $nh \to \infty$.*

An example of bandwidth parameter which is very popular in applications (Scott, 1992) is the following:

$$h = h_s \hat{\sigma} n^{-\frac{1}{5}} \tag{2.2}$$

where h_s is a real constant to be tuned, and $\hat{\sigma}$ is the sample standard deviation.

When $J \equiv 0$ we denote our process by Y. In the case of no jumps, Florens-Zmirou (1993) proves the following proposition.

Proposition 2.3 *(Florens-Zmirou, 1993) Define*

$$S^n(x) \doteq \frac{n\sum_{i=1}^n I_{\{|Y_{t_i}-x|<h\}}(\Delta_i Y)^2}{T\sum_{i=1}^n I_{\{|Y_{t_i}-x|<h\}}}. \tag{2.3}$$

Assume that:

- *$\mu(x)$ is bounded and has two continuous bounded derivatives;*
- *$\sigma(x)$ is uniformly bounded, bounded away from zero and has three continuous bounded derivatives;*
- *the bandwidth parameter satisfies $nh^4 \to 0$ as $n \to \infty$*

Then $S^n(x)$ converges to $\sigma^2(x)$ in probability for all x visited by Y.

This theorem is proved with the following strategy. Recall that the local time of Y is given by the following almost sure limit (Revuz and Yor, 2001):

$$L_t(x) = \lim_{\varepsilon\to 0}\frac{1}{2\varepsilon}\int_0^t I_{]x-\varepsilon,x+\varepsilon[}(Y_\tau)d[Y]_\tau, \tag{2.4}$$

Then, in Florens-Zmirou (1993) it is shown that

$$\frac{1}{2h}\sum_{i=1}^n I_{\{|Y_{t_i}-x|<h\}}\frac{T}{n} \tag{2.5}$$

is a consistent approximation of $L_T(x)/\sigma^2(x)$ and

$$\frac{1}{2h}\sum_{i=1}^n I_{\{|Y_{t_i}-x|<h\}}(\Delta_i Y)^2 \tag{2.6}$$

is a consistent approximation of $L_T(x)$.

Using the indicator function is not necessary. Jiang and Knight (1997); Bandi and Phillips (2003) provide a result analogous to proposition 2.3 with a continuous kernel replacing the indicator function.

Definition 2.4 *A kernel K is a non-negative real function such that $\int_{-\infty}^{+\infty} K(s)ds = 1$.*

An example of smooth kernel is the Gaussian density. The indicator function used by Florens-Zmirou (1993), namely $K(u) = I_{\{|u|<1\}}$, is non smooth. The choice of the kernel function is usually found to be irrelevant in financial applications, while much more important is the choice of the bandwidth parameter. Jiang and Knight (1997) show that

$$S_K^n(x) = \frac{n\sum_{i=0}^{n-1} K\left(\frac{Y_{t_i}-x}{h}\right)(\Delta_i Y)^2}{T\sum_{i=1}^{n} K\left(\frac{Y_{t_i}-x}{h}\right)} \tag{2.7}$$

is a consistent estimator of $\sigma^2(x)$.

Under the further assumption that $nh^3 \to 0$ both Florens-Zmirou (1993) and Jiang and Knight (1997) prove also the asymptotic Normality for $S^n(x)$ and $S_K^n(x)$ and provide the asymptotic standard errors, which are a function of the local time, see also the recent extension of Bandi and Phillips (2003).

3 Threshold estimation: finite jump activity

In this section, we concentrate on nonparametric estimation of model (2.1) in the case of finite activity of the jump component. We first concentrate on volatility estimation with fixed T and letting $n \to \infty$. Then, in the case of level dependent drift and jump intensity, we devise estimators of the corresponding functions letting both $n, T \to \infty$. Here and throughout the paper, all proofs are postponed to Appendix A.

3.1 Estimation of the diffusion function

We now focus on defining a nonparametric estimator of $\sigma^2(\cdot)$ when J is a finite activity process. Our model is now

$$X_t = Y_t + J_{1,t}, \tag{3.1}$$

where $Y_t = \int_0^t \mu(X_u)du + \int_0^t \sigma(X_u)dW_u$ is the diffusion part of X and $J_{1,t} = \sum_{k=1}^{N_t} \gamma_k$, with $J_{1,0^-} = 0$.

It is important to remark that, in order to estimate σ, which drives the infinitesimal (local) behavior of Y, it is enough to keep the time horizon T fixed, so that, as the time lag δ tends to zero, we have *infill* asymptotics. This fact is well known in the literature of continuous univariate diffusions, see e.g. Bandi and Phillips (2003), and we show that it holds even if we introduce jumps.

A fundamental tool for our aim is to disentangle, from the discrete observations of X, the contributions given by the jumps and those given by the diffusion part. To this purpose, we borrow some results from Mancini (2004a). Remind that $\delta = T/n$; we have:

Theorem 3.1 *(Mancini, 2004a) If J is finite activity with $P\{\gamma_k = 0\} = 0$ for all k, and if $\vartheta(u)$ is a real deterministic function such that*

$$\lim_{\delta \to 0} \vartheta(\delta) = 0 \quad \text{and} \quad \lim_{\delta \to 0} \frac{\delta \log \frac{1}{\delta}}{\vartheta(\delta)} = 0$$

then for P*-almost all ω, $\exists \bar{\delta}(\omega)$ such that $\forall \delta < \bar{\delta}(\omega)$ we have*

$$\forall i = 1, ..., n, \quad I_{\{\Delta_i N = 0\}}(\omega) = I_{\{(\Delta_i X)^2 \le \vartheta(\delta)\}}(\omega). \tag{3.2}$$

•

The above Theorem may be rephrased as: almost surely, for sufficiently small δ, if the squared increments $(\Delta_i X)^2$ in the interval $[t, t+\delta]$ are smaller than a *threshold* $\vartheta(\delta)$, which vanishes slower than the modulus of continuity of the Brownian motion (Karatzas and Shreve, 1988), then there are no jumps in that interval. On the contrary, jumps occurred only in those intervals in which $(\Delta_i X)^2$ is above the threshold.

As a consequence, following Mancini (2004b), we get an estimate of the whole jump process J_1 using:

$$\hat{J}_{1,t} = \sum_{\{i: t_i \le t\}} \hat{\gamma}_{\tau^{(i)}}, \tag{3.3}$$

where

$$\hat{\gamma}_{\tau^{(i)}} \doteq \Delta_i X I_{\{(\Delta_i X)^2 > \vartheta(\delta)\}}. \tag{3.4}$$

This is proved in the following:

Proposition 3.2 *(Mancini, 2004b) Assume that J is a compound Poisson process such that $P\{\gamma_\ell = 0\} = 0$ for each ℓ; moreover let $\mu(X_s), \sigma(X_s)$ be cadlag, T fixed and $\delta \to 0$.*
If $\vartheta(\delta)$ is such that $\frac{\delta \ln \frac{1}{\delta}}{\vartheta(\delta)} \to_\delta 0$, then, for every $\varepsilon > 0$,

$$P\left(\bigcup_{i=1}^{n} \{ n^k |\hat{\gamma}_{\tau^{(i)}} - \gamma_{\tau^{(i)}} I_{\{\Delta_i N \ge 1\}}| > \varepsilon \} \right) \to_P 0 \quad \forall k \in [0, \frac{1}{2}[.$$

•

Another consequence of Theorem 3.1 is that it allows to get an estimate $\hat{Y} = X - \hat{J}_1$ of the continuous part Y of (2.1), which can be used, to get the level-dependent volatility $\sigma^2(x)$.

Theorem 3.3 *Let X be a jump-diffusion process as in (3.1). Let the following conditions hold*

1. *$\sigma(x)$ is continuous and bounded, $\sigma^2(x) > 0$ for all x; $\mu(x)$ is bounded;*

2. *$P\{\gamma_k = 0\} = 0$ for all k;*

3. *as $\delta \to 0$ both the threshold function $\vartheta(\delta)$ and $\frac{\delta \ln \frac{1}{\delta}}{\vartheta(\delta)}$ tend to zero;*

4. *as $n \to \infty$, $\exists \beta > 1 : nh^\beta \to \infty$; $\dfrac{\sqrt{\delta \ln \frac{1}{\delta}}}{h^2} \to 0$;*

5. *K and K' are bounded.*

Then for all x visited by X

$$\hat{\sigma}_n^2(x) = \frac{\sum_{i=1}^n K\left(\frac{X_{t_{i-1}}-x}{h}\right)(\Delta_i X)^2 I_{\{(\Delta_i X)^2 \le \vartheta(\delta)\}}}{\sum_{i=1}^n K\left(\frac{X_{t_{i-1}}-x}{h}\right)\delta} \to_P \sigma^2(x). \tag{3.5}$$

Remark 3.4 *Theorem 3.3 holds even if the threshold $\vartheta(\delta)$ varies also with time according to:*

$$\vartheta_t(\delta) = c_t \tilde{\vartheta}(\delta)$$

where $\tilde{\vartheta}$ satisfies the threshold conditions, and c_t is a bounded deterministic function which is also bounded away from zero, after replacing $I_{\{(\Delta_i X)^2 \le \vartheta(\delta)\}}$ with $I_{\{(\Delta_i X)^2 \le \vartheta_{t_i}(\delta)\}}$.

Remark 3.4 will be extremely useful in financial applications, as will be shown later in the paper.

3.2 Estimating the drift and the intensity function

When J_{1t} is a doubly stochastic Poisson process, and $\lambda_t \equiv \lambda(X_{t-})$, it is possible to estimate the drift and the jump intensity functions by letting $T \to \infty$ and $T/n \to 0$. For the estimate of $\mu(x)$ and $\lambda(x)$ we rely on results in Bandi and Nguyen (2003). The estimator for $\lambda(x)$ is devised using the fact that:

$$\frac{\sum_{i=1}^n K\left(\dfrac{X_{t_{i-1}}-x}{h}\right)\Delta_i N}{\sum_{i=1}^n K\left(\dfrac{X_{t_{i-1}}-x}{h}\right)}$$

is the Nadaraya-Watson estimator of $E\left[\Delta N | X = x\right] \sim \lambda(x)\delta$.

Theorem 3.5 *Let the following conditions hold:*

1. *$\mu(x), \sigma(x), \lambda(x)$ are bounded; $\mu(x)$ and $\sigma(x)$ are twice continuously differentiable and satisfy local Lipschitz and growth conditions; $\lambda(x) \ge 0$, $\sigma^2(x) > 0$; μ and λ are integrable on $\mathbb{R}$.*

2. *X is Harris recurrent;*

3. *$\forall \varepsilon > 0 \quad P\{|\gamma_k| < \varepsilon\} \le c\varepsilon$ and the jump sizes $\{\gamma_k\}_k$ are independent of N;*

4. *the bandwidth parameter h is chosen as $h = \delta^\gamma$, with $\gamma \in]0, 1/2[$ and such that $\exists \psi > 0$: $n\delta^{3/2-\gamma(1+\psi)}\sqrt{\ln \frac{1}{\delta}} \to 0$;*

5. *$\vartheta(\delta) = \delta^\eta, \eta \in]0,1[$, with $n\delta^{1+\eta/2} \to 0$;*

6. *the kernel K is symmetric around zero, bounded, differentiable with integrable first derivative; K is square-integrable.*

7. *$\delta \to 0, T \to \infty$ and $h \to 0$, and $\frac{\sqrt{\delta \ln \frac{1}{\delta}}}{h} L_X(T,x) \overset{a.s.}{\to} 0$, $hL_X(T,x) \overset{a.s.}{\to} \infty$ for all visited x.*

Define:

$$\hat{\mu}_n(x) = \frac{\sum_{i=1}^n K\left(\frac{X_{t_{i-1}} - x}{h}\right) \Delta_i X \cdot I_{\{(\Delta_i X)^2 \le \vartheta(\frac{T}{n})\}}}{\sum_{i=1}^n K\left(\frac{X_{t_{i-1}} - x}{h}\right) \delta}. \tag{3.6}$$

Then for each x visited by X we have

$$\hat{\mu}_n(x) \to_P \mu(x).$$

Moreover, if K' is bounded and $c_{i,n}$ is a double array of constants with $i = 1, \dots, n$ such that $\sup_i |1 - c_{i,n}| \to 0$ as $n \to \infty$, define:

$$\hat{\lambda}_n(x) = \frac{\sum_{i=1}^n K\left(\frac{X_{t_{i-1}} - x}{h}\right) c_{i,n} I_{\{(\Delta_i X)^2 \ge \vartheta(\frac{T}{n})\}}}{\sum_{i=1}^n K\left(\frac{X_{t_{i-1}} - x}{h}\right) \delta}. \tag{3.7}$$

Then, for each x visited by X, we have:

$$\hat{\lambda}_n(x) \to_P \lambda(x).$$

Remark 3.6 *The assumption $\forall \varepsilon > 0 \quad P\{|\gamma_k| < \varepsilon\} \le c\varepsilon$ is fulfilled for Gaussian and exponential jump sizes. Moreover, it is satisfied when for instance each $|\gamma_i|$ is bounded away form zero or the laws $\gamma_i(P)$ have a density on $\mathbb{R}$ which is continuous in a neighborhood of zero.*

The multiplying factors $c_{i,n}$ can play an important role after making assumptions on the distribution of the jump sizes. Since we estimate the intensity using only variations which were larger than the threshold, we discard all jumps whose size is less than the threshold. If we assume the jump sizes to be Normally distributed with mean 0 and variance σ_J^2, we can attenuate considerably this problem setting:

$$c_{i,n} = \frac{1}{2\mathcal{N}(-\sqrt{\vartheta}/\sigma_J)} \tag{3.8}$$

where $\mathcal{N}(x)$ is the cumulative Normal distribution function. Given the time series of estimated jumps obtained via (3.4), we can estimate σ_J using, for example, a simple method of moments estimator, using the fact that if X is a Normal random variable with mean 0 and variance σ_J^2, the second moment $m_2(c)$ of X conditional to $|X| \ge c$ is given by:

$$m_2(c) = \sigma_J^2 + \frac{\sigma_J c \exp\left(\frac{-c^2}{2\sigma_J^2}\right)}{\mathcal{N}(-c\sigma_J)\sqrt{2\pi}} \tag{3.9}$$

Matching $m_2(\sqrt{\vartheta})$ in (3.9) to the observed variance of jump sizes provides an estimator of σ_J^2. Results of Section 6 indicate that the correction (3.8) delivers an unbiased estimator of the jump intensity.

4 Threshold estimation: infinite jump activity

In this section we assume that J is a Lévy process, so that

$$X = Y + J_1 + \tilde{J}_2, \tag{4.1}$$

with $\tilde{J}_{2s} \doteq \int_0^s \int_{|x|\leq 1} x[m(dt,dx) - \nu(dx)dt]$ and $J_{1s} = \int_0^s \int_{|x|>1} xm(dt,dx) = \sum_{k=1}^{N_t} \gamma_k$.
$\tilde{J}_2$ is an *infinite activity* jump process: every trajectory of $\tilde{J}_2$ jumps infinitely many times on every finite time interval, since $\nu(\mathbb{R} - \{0\}) = +\infty$.

Any Lévy process is such that

$$\int_{|x|\leq 1} x^2 \nu(dx) < +\infty. \tag{4.2}$$

However, if the exponent in (4.2) is less than 2, the corresponding integral can be infinite. To measure how frenetic is the jump activity, we can use the *Blumenthal-Gatoor index* α which is defined as:

$$\alpha \doteq \inf \left\{ \delta \geq 0 : \int_{|x|\leq 1} |x|^\delta \nu(dx) < +\infty \right\}.$$

The smaller α, the milder the activity. By definition, we have $\alpha \in [0,2]$. For instance, a finite activity Poisson compound process has $\alpha = 0$. An α-stable process has Blumenthal-Gatoor index equal to α. This index is very important for our purposes. We will consider Lévy processes whose Lévy measure has a density $f(x)$ which behaves like:

$$\frac{F(|x|)}{|x|^{1+\alpha}}$$

around the origin, where F tends to a finite limit as $x \to 0$ and α is the Blumenthal-Gatoor index of J. As a consequence, for $\varepsilon \to 0$

$$\int_{|x|\leq\varepsilon} |x|^k \nu(dx) \sim \int_{|x|\leq\varepsilon} |x|^{k-1-\alpha} dx$$

for every $k \geq 0$.

Assumption 4.1 *We will assume in the following that ν has a density $f(x)$ and F has a finite limit for $x \to 0$.*

Assumption 4.1 is fulfilled by the most commonly used infinite jump activity models (e.g. NIG, VG, CGMY). In what follows, we show that the estimator (3.5) still works, with some refinements, even if the jump activity is infinite, but with finite jump variation. To our knowledge, this is the first nonparametric kernel estimator which solves the problem of identifying the diffusion coefficient $\sigma(x)$ in presence of infinite activity jumps.

Theorem 4.2 *Let X be as in (4.1) and T fixed. Assume 4.1 and assume that:*

1. *$\sigma(x)$ is continuous, bounded with σ^2 bounded away from zero, for all x; $\mu(x)$ is bounded;*

2. *J has a.s. paths with finite variation (a.s. $\sum_{s\leq T} |\Delta J_s| < \infty$);*

3. *$\vartheta(\delta) = \delta^\eta$, with $\eta \in]0,1[$ such that $\frac{\sqrt{\delta |\ln \delta|}}{h^2} \to 0$; $\frac{\delta^\eta}{h} \to 0$;*

4. *K and K' are bounded.*

Then, for all x visited by X, if $\hat{\sigma}_n^2(x)$ is defined as in (3.5), we have:

$$\hat{\sigma}_n^2(x) \xrightarrow{P} \sigma^2(x)$$

Remark 4.3 *We need to assume that J has finite variation in order to ensure that $L_T^X(x)$ is cadlag in x (Protter 1990, theorem 56, ch. 4). The Variance Gamma model and all Lévy models with Blumenthal-Gatoor index $\alpha < 1$ fulfill this assumption.*

Remark 4.4 *The Gaussian kernel $K(u) = \frac{e^{-u^2/2}}{\sqrt{2\pi}}$ satisfies the assumptions of Theorems 3.3 and 4.2. For all $\eta \in]0,1[$, there exists $\gamma \in]0, \frac{1}{4}[$ such that $\vartheta(\delta) = \delta^\eta$ and $h = \delta^\gamma$ fulfill the assumptions of Theorems, 3.3, 3.5 and 4.2.*

5 Estimating the diffusion coefficient on simulated data

When implementing the threshold estimator considered so far, a number of problems arise. These can be summarized as follows.

1. The estimator can be used only on discretely sampled data, for which the limit $\delta \to 0$ does not hold. Letting δ be smaller and smaller, for example with high-frequency financial data, does not solve completely this problem, since it needs careful modeling of high-frequency data peculiarities (seasonalities, bid-ask bounce, microstructure effect). Thus asymptotic properties should be compared with *finite δ* properties.

2. When dealing with discretely sampled data, if a jump occurs but the size of the jump is less than the threshold, the jump is not detected by the threshold estimator. This effect spuriously leads to overestimate the volatility and underestimate the intensity. Conversely, a movement of diffusive nature larger than the threshold, e.g. a very large movement due to unexpectedly high volatility, could be misleadingly detected as a jump. This spuriously leads to underestimate the volatility and overestimate the intensity.

3. When a jump is detected, the indicator function cancels both the jump and the diffusive part. This effect tends to spuriously underestimate the volatility.

The considerations above lead us to assess the finite δ properties of the proposed estimator on Monte Carlo simulations of jump-diffusion models. In our simulated experiment, we focus on the estimation of the diffusion coefficient, since this is the most challenging task. Since in the final part of this paper we focus on interest rate modeling, in order to get realistic simulations, we adopt the following continuous-time model for the short interest rate r_t:

$$\begin{aligned}
dr_t &= \kappa_1(b_t - r_t)dt - \bar{\kappa}\lambda r_t dt + \gamma\sqrt{r_t}dW_{1,t} + (e^{Z_t} - 1)r_t dN_t \\
db_t &= \kappa_3(\theta - b_t)dt + \eta_2\sqrt{b_t}dW_{2,t}, \\
Z_t &\simeq \mathcal{N}(\mu_J, \sigma_J^2), \\
\bar{\kappa} &= \mathbb{E}[e^{Z_t} - 1] = e^{\left(\mu_J + \frac{1}{2}\sigma_J^2\right)},
\end{aligned} \tag{5.1}$$

where $W_{1,2}$ are independent Brownian motions and λ is the constant intensity of the Poisson process N_t. This is a modified version of the $SV_1J - SD$ model estimated by Andersen et al. (2004) on a time series of 3-month Treasury Bills annualized rates. We use their estimated parameters, that is $\hat{\kappa}_1 = 1.4312, \hat{\theta} = 5.74\%, \hat{\kappa}_3 = 0.3319, \hat{\eta}_2 = 0.0560, \mu_J = 0, \hat{\sigma}_J = 0.0266, \hat{\lambda} = 5.2980$ and the average value of the estimated volatility process, $\gamma = e^{(-0.5 \cdot 7.0461)}$. The starting values r_0 and b_0 are sampled from their unconditional distribution.

Note that the coefficients μ and σ in model (5.1) do not strictly fulfill the assumptions of Theorem 3.3, under which our estimator converges to the volatility $\sigma^2(x)$, since μ_t depends on both r_t and b_t and $\sigma(r_t) = \gamma\sqrt{r_t}$ does not satisfy the differentiability condition. However the estimated parameters guarantee that each path of r_t keeps bounded and bounded away from zero on $[0, T]$. Moreover, the assumption of boundedness of $\sigma(r)$ is not relevant in finite samples: we can always extend the deterministic function $\sigma(r)$ in the regions where there are no observations of r_t in a way such that the modified function fulfills the assumptions.

When implementing the estimator, we have to set a threshold ϑ_t to identify jumps. In what follows, we omit the dependence of ϑ_t from δ since on actual data, as well as simulated, the interval δ between adjacent observations is fixed. A variation is identified as a jump if $(\Delta_i X)^2 \geq \vartheta_t$. Setting ϑ_t as a constant would lead to the following problem. It is well known that volatility of financial prices is persistent, so that there are periods of high volatility and periods of low volatility. If we use the same threshold for the two periods, we clearly increase the probability of detecting a jump when large movements of the short rate are due to high volatility, or not detecting a jump because the threshold is too high compared to the volatility of the period.

To circumvent this difficulty, we use a time-varying threshold which is computed using an auxiliary model, exploiting Remark 3.4. The auxiliary model we propose is a GARCH$(1, 1)$:

$$\begin{aligned} r_t - r_{t-1} &= \bar{r} + \sqrt{h_t} \cdot \varepsilon_t \\ h_t &= \omega + \alpha(r_{t-1} - r_{t-2})^2 + \beta h_{t-1} \end{aligned} \tag{5.2}$$

where ε_t are standardized IID innovations and $\bar{r}, \omega, \alpha, \beta$ are constants. We then use the estimated values[1] $\hat{\omega}, \hat{\alpha}, \hat{\beta}$ to filter the conditional variance h_t. [2]

Using a GARCH$(1, 1)$ model for filtering has a twofold motivation. First, this is a very popular model, it is very simple to estimate, and it is well known to fit financial data quite well. Second, even if the GARCH$(1, 1)$ is a misspecified model (for example, in our simulation experiments the actual model is given by the diffusion (5.1)), the results in Nelson (1992); Nelson and Foster (1994) guarantee that it provides optimal volatility forecasts when the time interval δ between two observations shrinks to zero. Thus it is the "optimal" model, in an asymptotic sense, that is when $\delta \to 0$. In small and finite samples, GARCH$(1, 1)$ can be regarded as a suitable approximation of the data when the time interval between observations is sufficiently small. [3]

[1]In principle, the presence of jumps may distort the GARCH estimates. However, the procedure can be iterated reestimating the GARCH model after removing the jumps detected in the first step.

[2]We estimate these values on a very long simulation of model (5.1). Estimates and standard errors are: $\hat{\omega} = 0.518(0.384) \cdot 10^{-10}, \hat{\alpha} = 0.00234(0.00046), \hat{\beta} = 0.99735(0.00054)$. To initialize h_t, we use the unconditional variance of r_t.

[3]The GARCH(1,1) we propose is a one-sided filter. Since the technique is implemented ex-post, using two-sided filters may improve threshold estimation. In the applications of this paper, the advantage of using more complicated filters has been found to be negligible.

Finally, we set the threshold equal to a multiple of the filtered conditional variance h_t:

$$\vartheta_t = c \cdot h_t. \tag{5.3}$$

The parameter c has to be calibrated. Motivated by extensive simulation experiments, we set $c = 3^2 = 9$, that is we detect as jumps the variations which are three conditional standard deviations away from zero. Under the assumption of no jumps and GARCH variance, setting $c = 9$ is equivalent to erroneously detecting as jumps roughly 0.5% of the variations.

We test the threshold estimator on the simulations of model (5.1), discretized according to a second-order Euler scheme with $\delta = \Delta t = 1/252$, to get daily observations. The length of every simulated time series is set to 5,000 observations, and we use 5,000 replications. On each simulation, we implement the classical estimator of Florens-Zmirou (1993); Jiang and Knight (1997), given by (2.7) and the proposed threshold estimator (3.5). When implementing nonparametric estimation, we use a Gaussian kernel, and the bandwidth (2.2) with $h_s = 3$.[4]

Figure 5.1 shows the result on the simulations of model (5.1). While the classical estimator is biased in estimating the correct variance, the threshold estimator correctly estimates the variance with a slight bias for large and small interest rates. Monte Carlo simulations allow us to estimate the efficiency in detecting jumps, see Appendix B. On average, 35.3% of the jumps are detected, while the noise is 0.40%, that is 0.40% of the diffusive variations are erroneously detected as jumps.

The bias for the classical estimator is not actually very severe. However, in interest rates, jump sizes are quite small (in our simulations, $\hat{\sigma}_J = 0.026$ implies typical movements of 5% of the current interest rate level), even if jumps are quite frequent (around five per year in our simulations). If we just double the jump size standard deviation $\hat{\sigma}_J$, the advantage in using the threshold estimator increases, as shown in Figure 5.2. While the threshold estimator is still unbiased in small samples, the classical estimator is largely biased and displays huge confidence bands. With larger jumps, the percentage of detected jumps raises to 57.5%, while the percentage of detected jumps which were diffusive variations becomes 0.25%.

Finally, with Monte Carlo simulations we can assess the bias introduced by the proposed threshold estimator in the absence of jumps. As discussed before, in this case some jumps are still detected when the larger variations occur. When simulating diffusions with $\lambda = 0$, the threshold estimator erroneously detects as jumps 0.46% of the observations, as expected. However, this has a negligible impact on the estimate, as shown in Figure 5.3, even if the discarded variations are the larger in the sample, that is they are those who contribute more to the total variance.

Concluding, using the threshold estimator cancels the bias which affects the classical estimator in presence of jumps. The importance in using a threshold estimator increases with increasing jump size and frequency, since the bias of the classical estimator increases very rapidly. Finally, if there are no jumps, the difference between the classical estimator and the threshold estimator is negligible. These results seem to suggest that, for financial data, using the threshold estimator is a superior alternative.

[4] In the literature on nonparametric estimation, it is common to use an oversmoothed h_s instead of the optimal value $h_s = 1.06$ used for estimation of Gaussian densities. Varying h_s in the range $[2, 4]$ does not change the results noticeably.

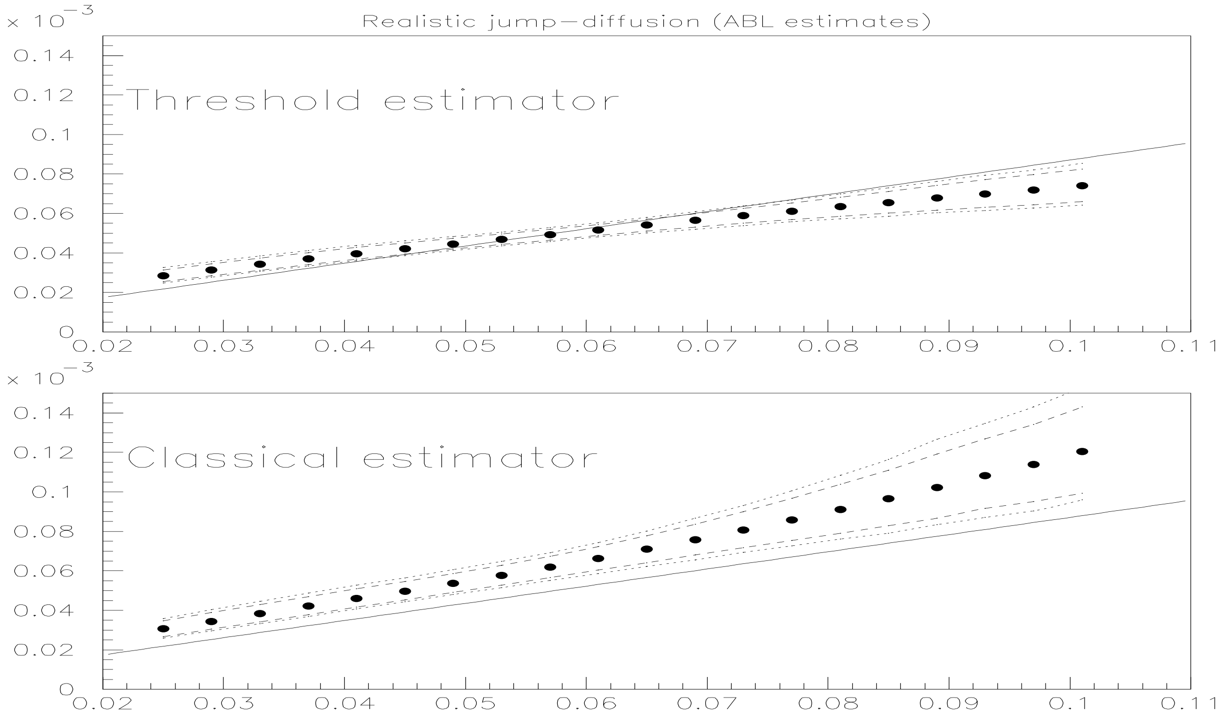

Figure 5.1: Top: Average estimate (bullets) of $\sigma^2(r)$ with the threshold estimator (3.5) on simulations of model (5.1). Bottom: Average estimate (bullets) of $\sigma^2(r)$ with the classical estimator (2.7). In both figures, the solid line is the generated variance function, which is linear. Dashed lines are $(10\%, 90\%)$ and dotted lines are $(5\%, 95\%)$ Monte Carlo confidence bands. We use $5,000$ replications of $5,000$ observations.

6 Estimation of short rate models

In this Section, we implement the proposed estimator on interest rate time series to reexamine nonparametric estimation of jump-diffusion models of the short rate. For sake of comparison, we compare the results obtained with the threshold estimator both to those in the literature on nonparametric estimation of diffusive models, and especially with the Bandi and Nguyen (2003) estimator.

6.1 Proxying the short rate

Our aim is to estimate a univariate jump-diffusion model for the short rate. The short rate is inherently unobservable, and it has to be proxied using interest rates of short term zero coupon bonds, see Chapman et al. (1999) for a tight discussion. This problem is similar to that of approximating the derivative of a function, which is not observed continuously, by using finite increments.

The literature has focused on two time series: the 7-day Eurodollar deposit time series, e.g. in Ait-Sahalia (1996);

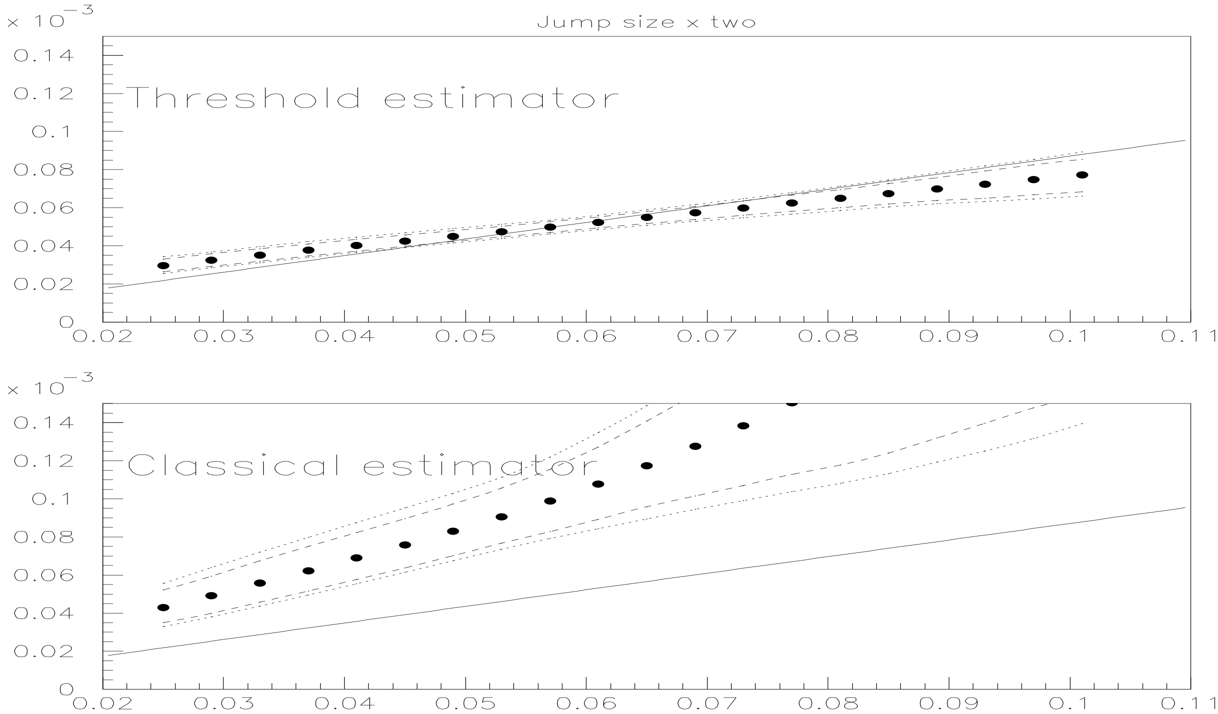

Figure 5.2: Top: Average estimate (bullets) of $\sigma^2(r)$ with the threshold estimator (3.5) on simulations of model (5.1) with σ_J doubled. Bottom: Average estimate (bullets) of $\sigma^2(r)$ with the classical estimator (2.7). In both figures, the solid line is the generated variance function, which is linear, and the dashed lines are $(10\%, 90\%)$ and dotted lines are $(5\%, 95\%)$ Monte Carlo confidence bands. We use $5,000$ replications of $5,000$ observations.

Bandi (2002); Hong and Li (2005), and the 3-month Treasury Bill time series, e.g. in Stanton (1997); Jiang (1998). Both time series yield an approximation of the short rate, so they should look quite similar. However, as it will be clear from the analysis performed in this Section, there are peculiarities which affect interest rate instruments of different maturity, even if the maturities are very close.

In our first application, we consider the time series of daily observations in the period starting in June 1973 and ending in February, 1995, for a total of 5,505 observations, for both the 7-day Eurodollar time series and the 3-month T-Bill. The time series of the 7-day Eurodollar deposit coincides with that employed by Ait-Sahalia (1996), Bandi (2002) and, more recently, by Hong and Li (2005) for their nonparametric estimates. The 3-month T-bill time series is a subset of that used by Stanton (1997).

Figure 6.1 displays the two time series. Visually, they look pretty similar. The 7-day series is slightly higher in levels. The average difference between the 7-day and the 3-month time series is 1.08%, with a standard deviation of 1.04%. Thus, there is a slight term structure effect. However, the main difference between the two is that, while the 3-month time series looks more like a continuous diffusive process, the 7-day time series displays frequent spikes, with a periodic fashion. Indeed it is well known, see e.g. Duffee (1996); Hamilton (1996), that interest rate instru-

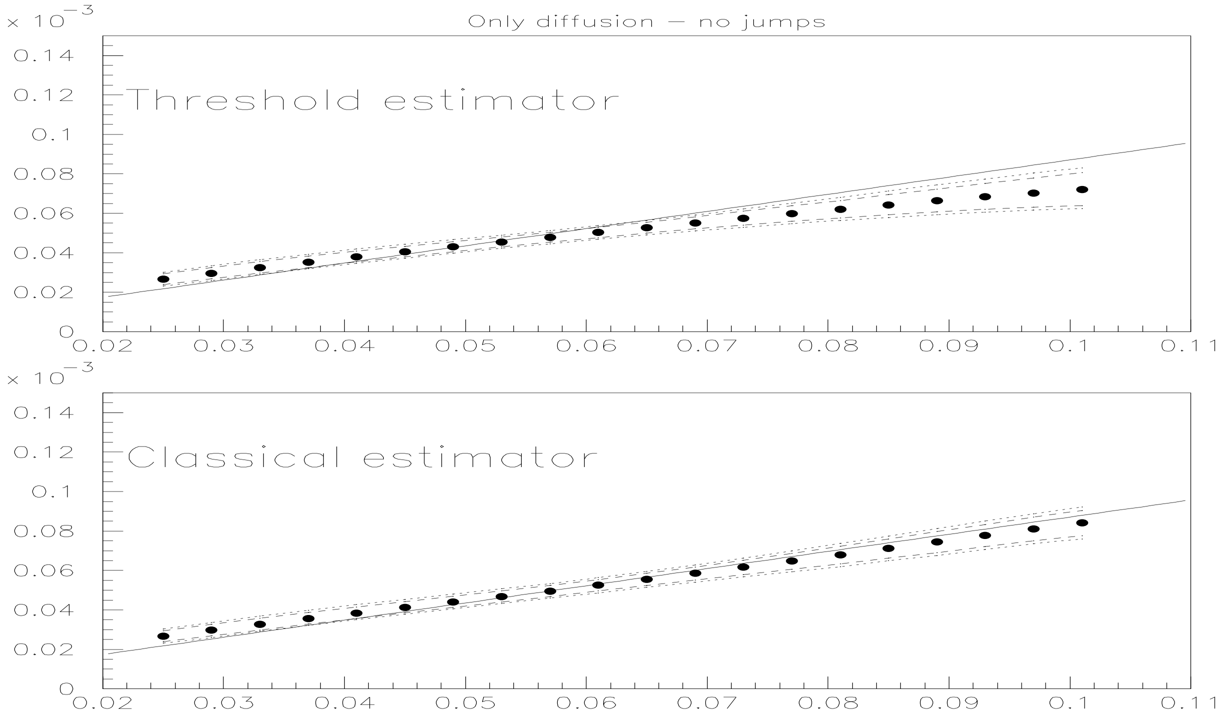

Figure 5.3: Top: Average estimate (bullets) of $\sigma^2(r)$ with the threshold estimator (3.5) on simulations of model (5.1) with $\lambda = 0$ (no jumps). Bottom: Average estimate (bullets) of $\sigma^2(r)$ with the classical estimator (2.7) (bullets). In both figures, the solid line is the generated variance function, which is linear, and the dashed lines are $(10\%, 90\%)$ and dotted lines are $(5\%, 95\%)$ Monte Carlo confidence bands. We use $5,000$ replications of $5,000$ observations.

ments below three months display idiosyncratic features which are mostly due to calendar and liquidity effects, e.g. reserve management, and spikes are induced by liquidity shortages which nevertheless do not affect higher maturity instruments.

Thus, if we implement the classical diffusion estimator (2.7) and the drift estimator (3.6) with the threshold set at an infinite level, that is the same estimators used by Stanton (1997), on the two time series, we get the estimates displayed in Figure 6.2 and Figure 6.3. Confidence bands are obtained via Monte Carlo simulation as explained in Appendix B. For the drift function, the two estimates are quite consistent, even if the 7-day time series is estimated to be more mean-reverting than the 3-month time series. For the diffusion coefficient, big differences are instead observed. The two estimates are completely different, since the spikes in the 7-day time series raise enormously the observed variance of the short rate. This is rather annoying, since the two time series are supposed to proxy for the same economic variable, the short rate.

It is important to remark that the estimate of the diffusion and drift coefficients on the 7-day time series are almost identical to the estimate proposed by Bandi (2002) for the drift and the diffusion function of a continuous-time model,

Figure 6.1: Top: the 7-day Eurodollar deposit daily time series, from June 1973 to February 1995, for a total of 5505 observations. Bottom: the 3-month T-bill daily time series, same time span.

which are overlapped[5] in Figures 6.2 and 6.3, on the very same time series[6].

It is clear that the problem is that in both the time series there is a jump component which distorts the variance estimate. However, the jump component is much more relevant for the 7-day time series. Thus, if we properly implement a jump-diffusion model, once jumps are detected and deleted, the diffusion and the drift coefficient should look similar on the two time series. That is, the major difference between the two time series is the jump component. Loosely speaking, the two time series represent nearly the same variable, but the 7-day rate is "with jumps", and the 3-month rate is "without jumps". Thus, they are a suitable battlefield for the threshold estimator proposed in this paper.

The threshold estimator is implemented in the following way. First, a GARCH$(1,1)$ model is estimated[7] on the 3-month time series, and it is used as an auxiliary model to filter the conditional variance. We then implement the

[5]The estimates obtained in Bandi (2002) and, later in the paper, by Johannes (2004) have been obtained by scanning the images in their papers and using a digitalization software. Thus, they should be regarded as indicative, even if the error of the digitalization procedure is negligible.

[6]Simulation results show that the Bandi and Phillips (2003) estimator, used in Bandi (2002), and the Florens-Zmirou (1993) estimator, used in Stanton (1997), yield nearly identical results in estimating the diffusion function, see e.g. Renò et al. (2004).

[7]Regarding model (5.2), the following estimates are used: $\hat{\omega} = 0.75 \cdot 10^{-8}, \hat{\alpha} = 0.091379, \hat{\beta} = 0.906878$. h_t is initialized with its estimated unconditional variance.

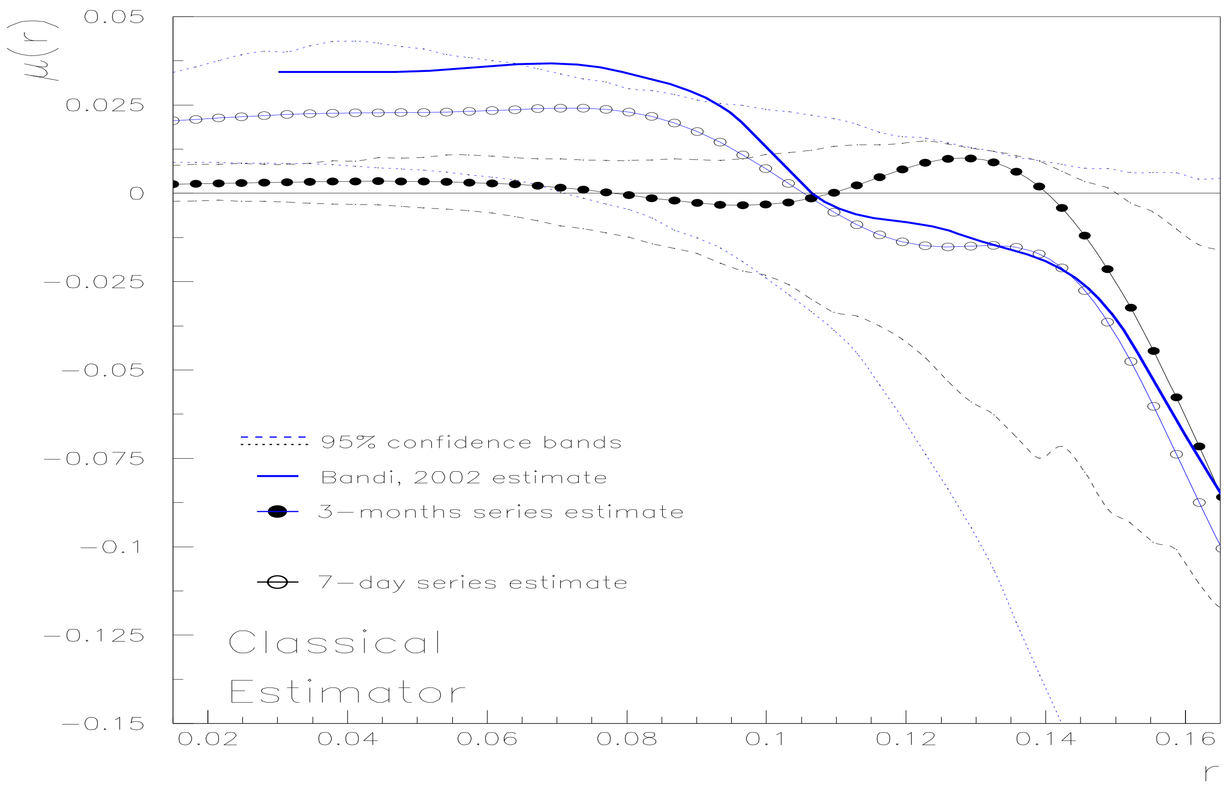

Figure 6.2: Estimates of the drift function $\mu(r)$ on the two time series, obtained with the estimator proposed in Stanton (1997). The dashed lines are the $(10\%, 90\%)$ confidence bands for the estimate on the 3-month time series, computed via Monte Carlo simulation. The dotted lines are confidence bands for the estimate on the 7-day time series. The solid line is the estimate obtained in Bandi (2002).

estimator (3.5) with the following threshold:

$$\vartheta_t = 9 \cdot h_t \tag{6.1}$$

where $(h_t)_{t=1,\dots,T}$ is the filtered variance from GARCH(1,1) estimation, see the discussion of simulated experiments. This procedure cuts observations whose variations are three conditional standard deviations away from zero. By using a conditionally varying threshold, we are more sensible in detecting jumps when the variance is high, that is when a large movement could be more likely due to the diffusive component instead of a jump. We use the same estimates of the GARCH model for the two time series, while the threshold is different since the time series of squared differences are different. The time series of detected jumps is shown in Figure 6.4. As expected, we find much more jumps (607 jumps, 11.0% of observations) on the 7-day time series than on the 3-month time series (85 jumps, 1.54% of observations), and jump sizes are much larger on the 7-day time series.

We then estimate the intensity function $\lambda(r)$ for the two time series using the estimator (3.7) with the correction (3.8). Here and throughout the rest of the paper, the threshold estimator is implemented with a Gaussian kernel and with the bandwidth (2.2) with $h_s = 3$ when estimating the drift and diffusion function, and with $h_s = 5$ when estimating

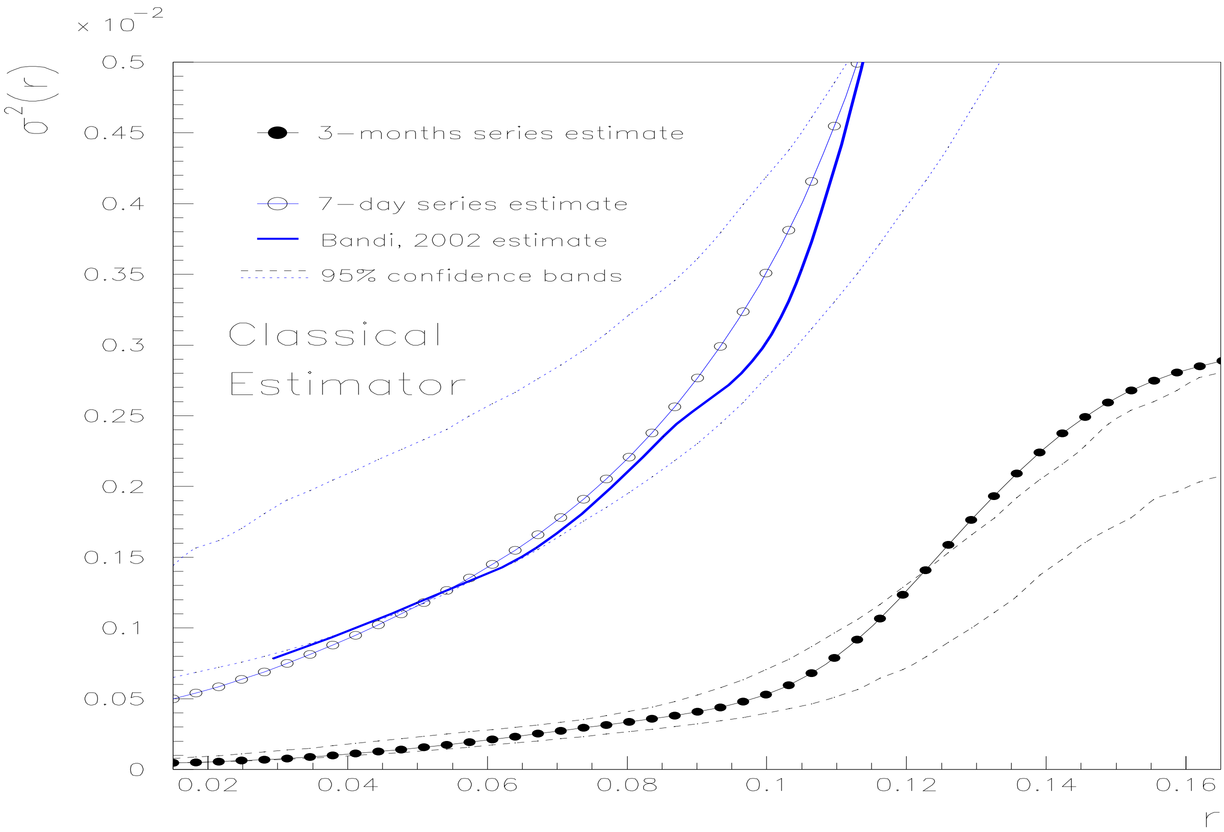

Figure 6.3: Estimates of the diffusion function $\sigma^2(r)$ on the two time series, obtained with the Florens-Zmirou (1993) estimator (2.7). The dashed lines are the $(10\%, 90\%)$ confidence bands for the estimate on the 3-month time series, computed via Monte Carlo simulation. The dotted lines are confidence bands for the estimate on the 7-day time series. The solid line is the estimate obtained in Bandi (2002).

the intensity function. An higher smoothing parameter in the latter case is needed, since the number of jumps is not very large.

Figure 6.5 shows the result. The estimates are very different on the two samples. On the 7-day time series, we find a typical intensity of 35 jumps per year, and an intensity function which is not monotone. The estimate on the 7-day time series is close to the estimated intensity in Das (2002), who estimates a jump-diffusion model on U.S. federal fund rates, and finds an intensity of $13 - 17$ jumps per year. Confidence bands indicate a severe downward bias[8] in estimation, see Appendix B.

For the 3-month time series, we get a typical intensity of 5 jumps per year, and the intensity is increasing with the level of interest rate. It is in very good agreement with the estimate of Andersen et al. (2004) of around 5 jumps per year. Confidence bands indicate that the proposed intensity estimator is unbiased.

[8]When confidence bands are below the estimate, it means that we get a lower estimate than the generated one, thus the estimate is downward biased.

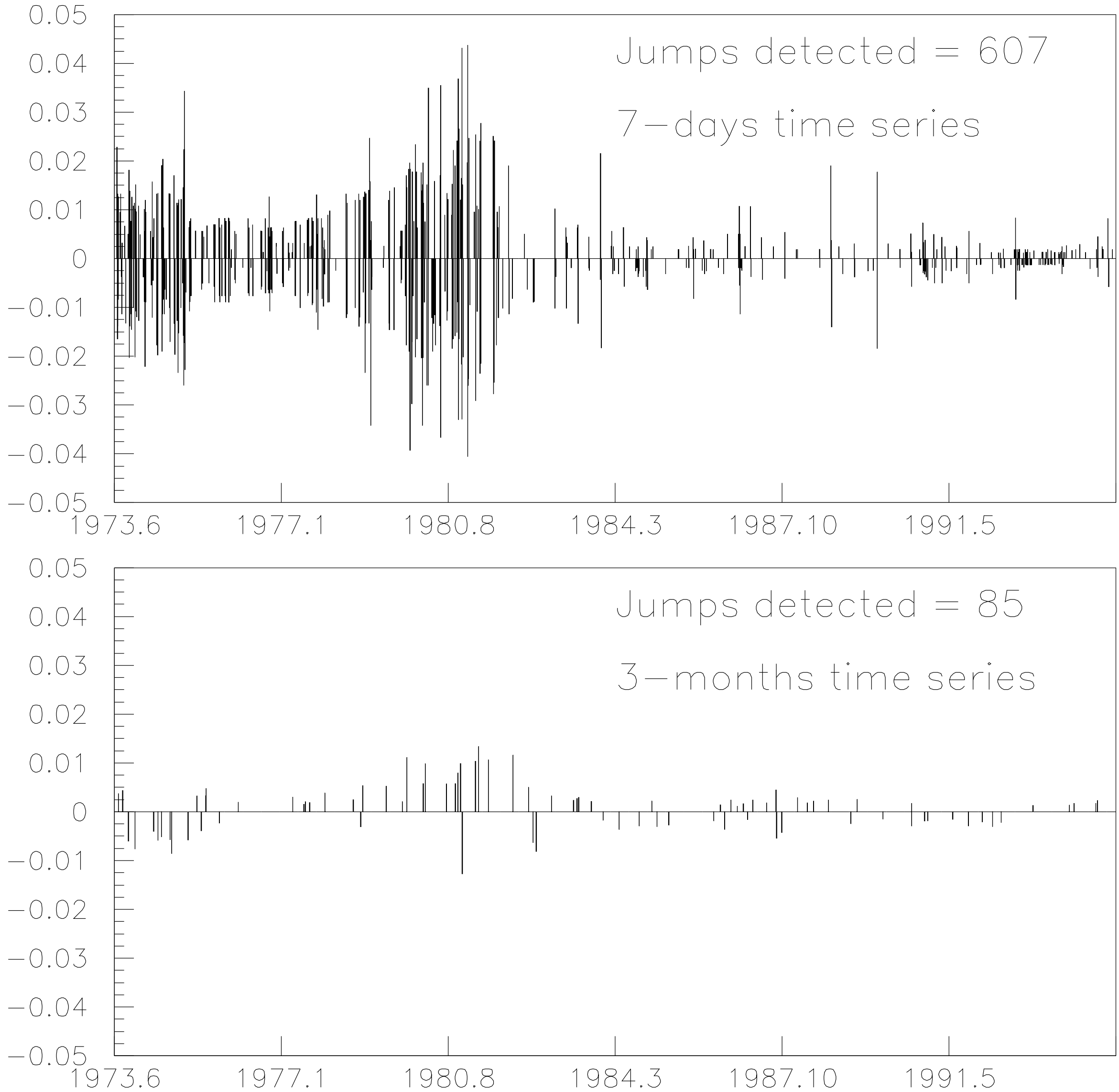

Figure 6.4: Top: The time series of detected jumps on the 7-day time series. Bottom: The time series of detected jumps on the 3-month time series.

Clearly, we are disentangling the idiosyncratic noise of the 7-day Eurodollar deposit rate, which is driven by liquidity reasons, and is displayed in the form of discontinuous variations, from the continuous variations of the short rate. The detected jumps on the 3-month time series have instead been shown to be mostly due to macroeconomic announcements (Johannes, 2004; Piazzesi, 2005).

We now turn to the estimate of the drift and diffusion coefficient. Estimation results are displayed in Figure 6.6 for

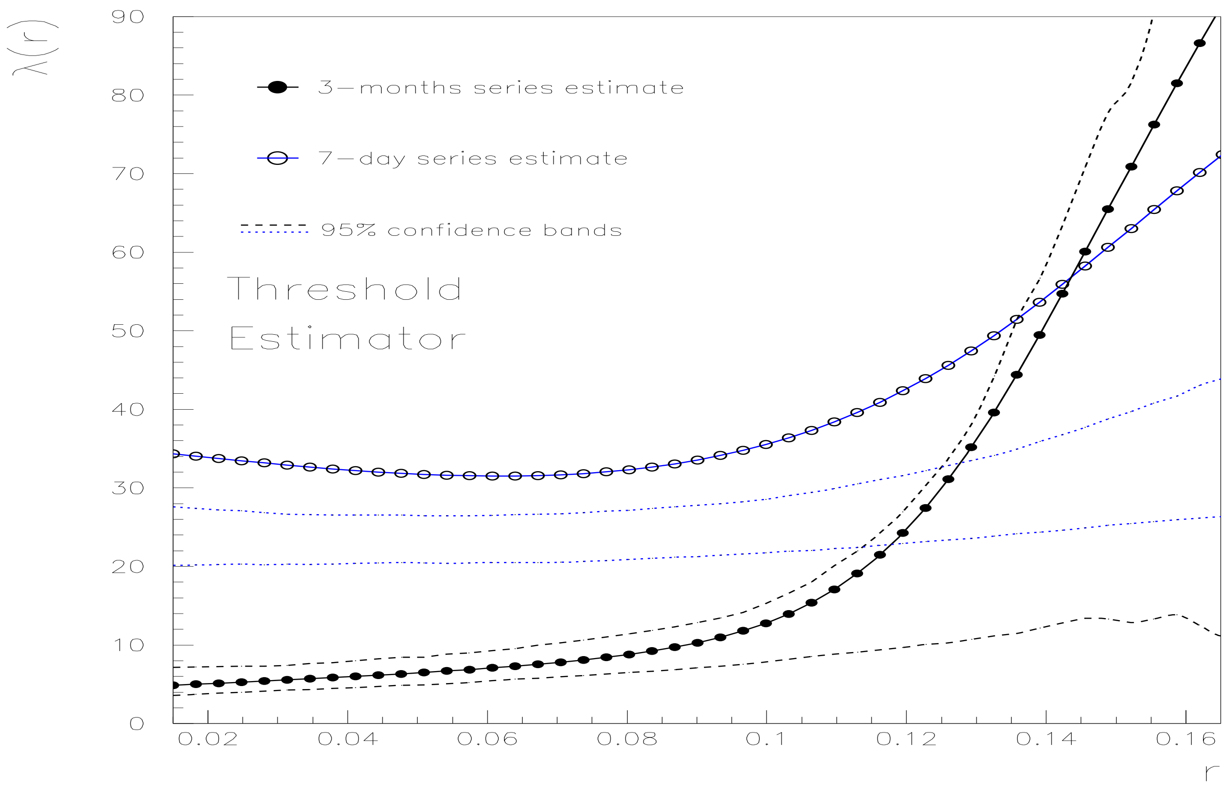

Figure 6.5: Estimates of the jump intensity function $\lambda(r)$ on the two time series, obtained with the threshold estimator (3.7) proposed in this paper. The dashed lines are the $(10\%, 90\%)$ confidence bands for the estimate on the 3-month time series, computed via Monte Carlo simulation. The dotted lines are confidence bands for the estimate on the 7-day time series.

the diffusion coefficient and in Figure 6.7 for the drift coefficient. Both figures are exactly on the same scale of Figure 6.2 and Figure 6.3 respectively, to allow the reader to directly compare the two.

The results are compelling. After filtering for the jump components, the two estimates of the diffusion function do actually look the same, as it should be, considering the above discussion. For the 3-month time series, the difference between the estimate with the threshold estimator and with the classical estimator is not very large, with the classical estimator providing a larger estimate. For the 7-day time series the situation is completely different; with the same threshold, we detect a large number of jumps, and this makes a large difference in estimating the variance function. The same considerations apply to the estimate of the drift coefficient.

The confidence bands obtained in Figure 6.6 on the 7-day time series suggest that the estimate is upward biased[9]. This problem is originated by the severe bias in estimating the intensity function on the 7-day time series, see Figure 6.5. Since the jump size is Normally distributed in Monte Carlo experiments, most of the jumps are below the threshold,

[9]That is, when we estimate the diffusion function on a model with the drift, diffusion and intensity estimated on the 7-day time series, we get a larger estimate for the diffusion coefficient than the generated one.

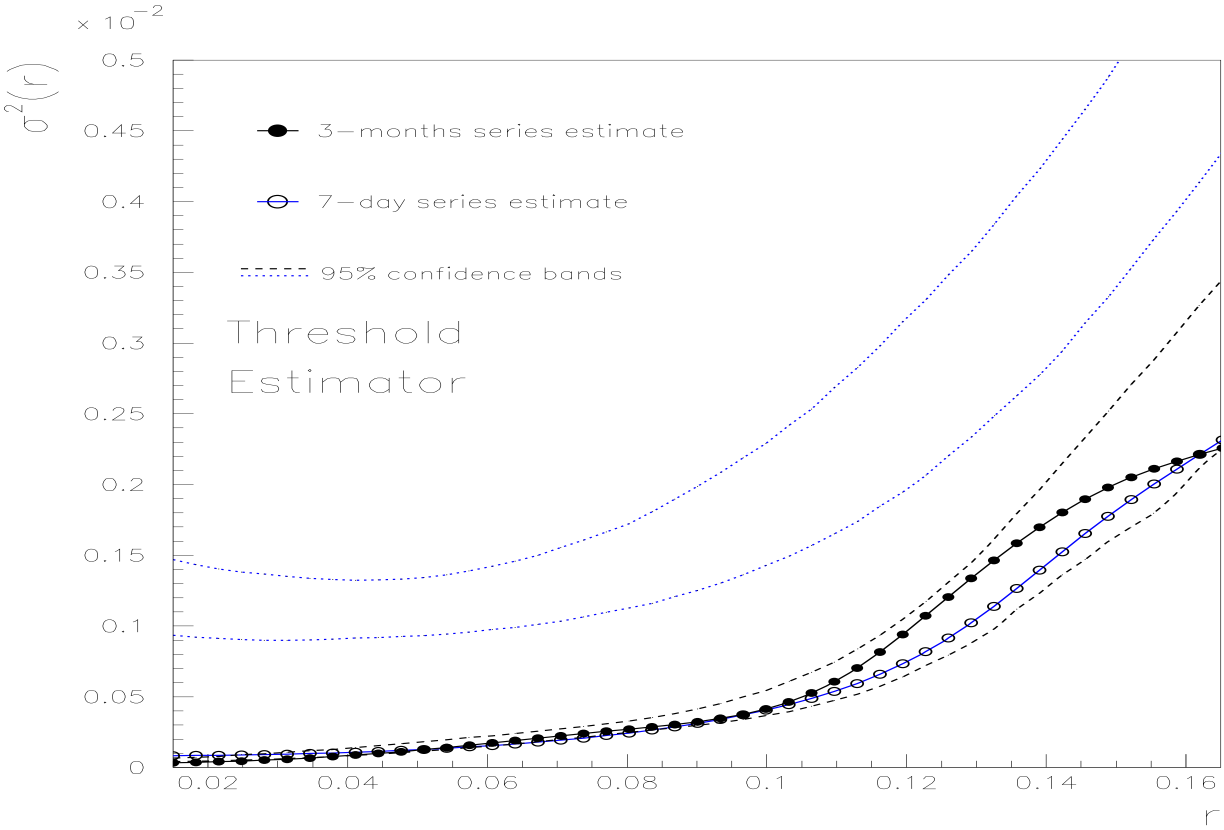

Figure 6.6: Estimates of the diffusion function $\sigma^2(r)$ on the two time series, obtained with the estimator (3.5) proposed in this paper. The dashed lines are the $(10\%, 90\%)$ confidence bands for the estimate on the 3-month time series, computed via Monte Carlo simulation. The dotted lines are confidence bands for the estimate on the 7-day time series.

and the intensity is underestimated. All the jumps which are not detected contribute spuriously to the estimated variance, which turns out to be larger.

Generally speaking, in small and finite samples the unbiasedness of the estimator always depends on the specific data generating process. Thus, the small sample analysis, performed e.g. via Monte Carlo simulations, is mandatory to assess the reliability of the results.

Our simulations of the estimated model also allow to provide an estimate of the efficiency in detecting jumps, see Appendix B. On the simulated paths of the model estimated on 3-month time series, we get an efficiency of 39.47% in detecting jumps, and a noise of 0.17%. On the 7-day time series the efficiency is lower: 26.70% and also the noise is lower: 0.02%.

Concluding, if the purpose is estimating continuous time models, the 7-day time series should not be used, while it is better to use the 3-month time series. However, when carefully implementing a threshold estimator, the estimates of the continuous part of the two time series are consistent.

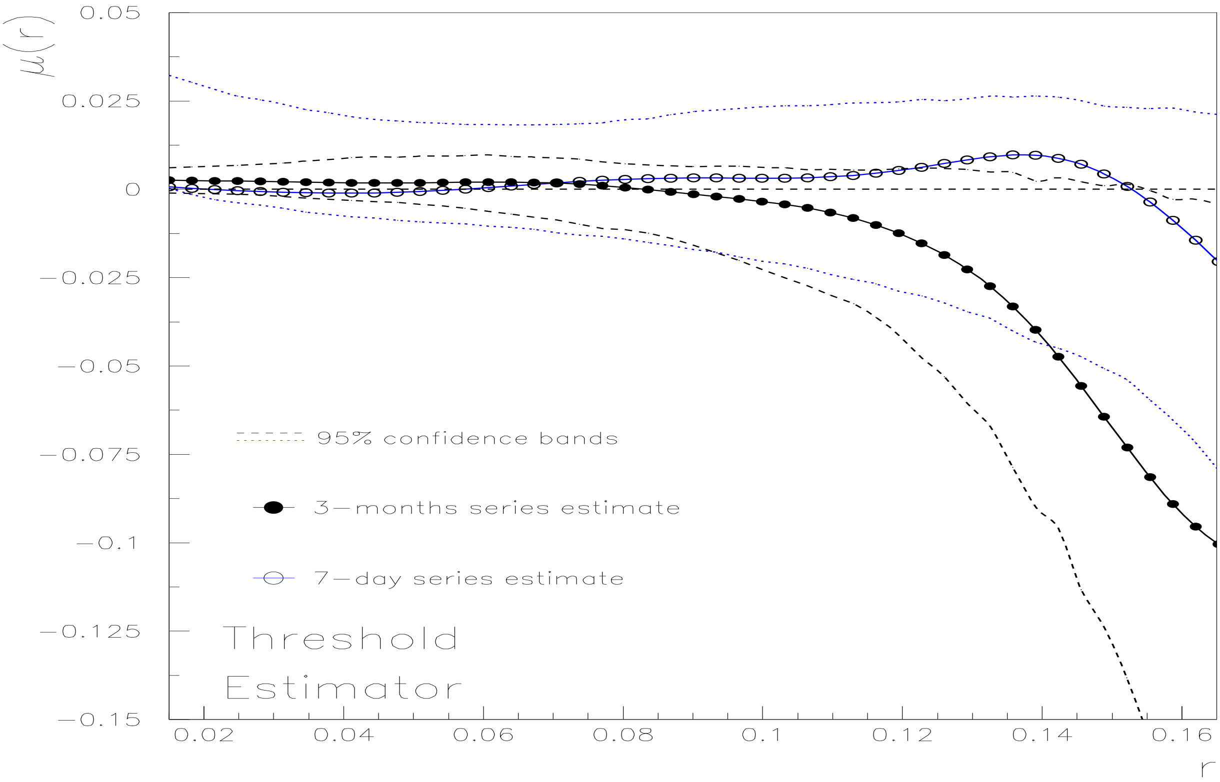

Figure 6.7: Estimates of the drift function $\mu(r)$ on the two time series, obtained with the estimator (3.6) proposed in this paper. The dashed lines are the $(10\%, 90\%)$ confidence bands for the estimate on the 3-month time series, computed via Monte Carlo simulation. The dotted lines are confidence bands for the estimate on the 7-day time series.

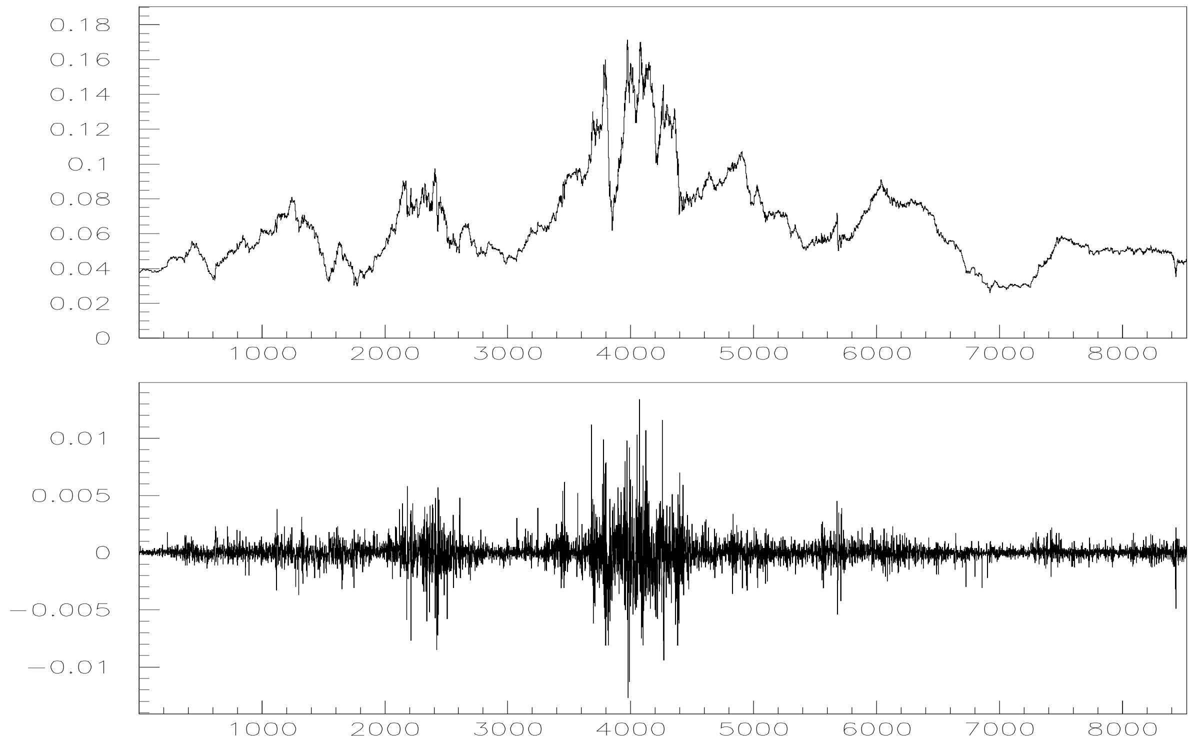

Figure 6.8: Top: The time series used in Johannes (2004). Bottom: the time series of first differences.

6.2 A comparison with the Bandi-Nguyen-Johannes estimator

In this Section, we directly compare the results obtained with the threshold estimator with those obtained with the estimator proposed by Bandi and Nguyen (2003) and Johannes (2004). To this purpose, we use the very same dataset used in Johannes (2004), that is the time series of daily 3-month Treasury-bills annualized rate from January, 1965 to February, 1999, for a total of 8,522 observations. The time series of interest rate levels and first differences are plotted in Figure 6.8. Observations range from the lower value of 2.61% to the larger value of 17.14%. On this data set, we use the threshold estimator to estimate the squared diffusion function $\sigma^2(r)$, the drift function $\mu(r)$ and the intensity function $\lambda(r)$.

The estimate of the squared diffusion function is plotted in Figure 6.9 and directly compared to that obtained in Johannes (2004).[10] We get a large difference between the two methodologies. The estimate obtained with the threshold estimator is less oscillating and considerably and significantly higher, for all interest rate levels. Substantial differences are also observed on the intensity function in the opposite direction. Both methods find that the intensity of jumps increases with the level of interest rates, but Johannes (2004) estimates an intensity ranging from 15 to 30 jumps

[10]In Johannes (2004), the drift and diffusion functions are obtained for the logarithmic differences of the short rate. To compare his results to ours, which are obtained on short rate differences, we have to trasform his estimates according to Ito's Lemma.

per year when the short rate is less than 10%, while, with the threshold estimator, we find in the same interest rate range an intensity from 5 to 10 jumps per year. For short rates larger than 10% the intensity becomes much larger for both estimators. While it is difficult, a priori, to disentangle which of the two estimates is the correct one, simple considerations suggest, in our opinion, that the estimates of the intensity by Johannes (2004) is too high for interest rate data. First, the intensity of 15-30 jumps per year looks quite counterfactual, since it implies around one jump per week. Second, parametric estimates on similar data are much closer to the estimate obtained with the threshold estimator. For example, Andersen et al. (2004) estimate around 5 jumps per year. Summarizing, the threshold estimator seems to suggest more variance coming from continuous movements, and less jump intensity.

For the drift, we find that the estimate obtained with the threshold estimator is very similar to that obtained with the Bandi-Nguyen-Johannes estimator, see Figure 6.10. Looking at confidence bands, we observe significant mean-reversion only for rates less than 3% and rates larger than 14%. This is in substantial agreement with the result in Jones (2003), who reports substantial random walk behavior of the short rate in the central part of the distribution, and evidence of mean reversion for extreme values of the short rate. The presence of mean reversion in interest rate data is quite controversial, see e.g. Pritsker (1998); Chapman and Pearson (2000). Our result confirm that it is very difficult to detect significant mean reversion in interest rates movements, if not for extreme levels.

Finally, it is remarkable to observe that the threshold estimator for the intensity function is unbiased for the jump-diffusion considered, as tested with Monte Carlo confidence bands. On Monte Carlo simulations of the estimated model, we find an efficiency in detecting jumps of 34.79% and a noise of 0.17%.

Concluding, the threshold estimator and the Bandi-Nguyen-Johannes estimator provide consistent estimates for the drift term, while the two methodologies provide a very different estimate of the diffusion and intensity function, with the threshold estimator being closer to parametric estimates of the jump intensity.

7 Conclusions

In this paper, we introduce threshold estimators for univariate jump-diffusion models in which the drift coefficient, the diffusion coefficient and the jump coefficient are level dependent.

The proposed estimator is based on the fact that, when the time interval δ between two observations tends to zero, jumps can be detected using a threshold which tends to zero slower than the modulus of continuity of the Brownian motion. We use this intuition to devise nonparametric estimators of the jump-diffusion process, extending the preceeding literature on nonparametric estimation of continuous diffusions and of diffusions plus finite activity jump processes. Moreover, we show that the consistency of the diffusion coefficient still hold if the activity of jumps is infinite.

Monte Carlo simulations show that, in finite samples, the threshold estimator for the diffusion coefficient of realistic interest rate processes is unbiased and performs fairly better than the classical estimator.

Data analysis helps us in better understanding the peculiarities of short rate dynamics. First, we show that the difference between the 3-month rate and the 7-day rate is due to the jump component which affects the second and

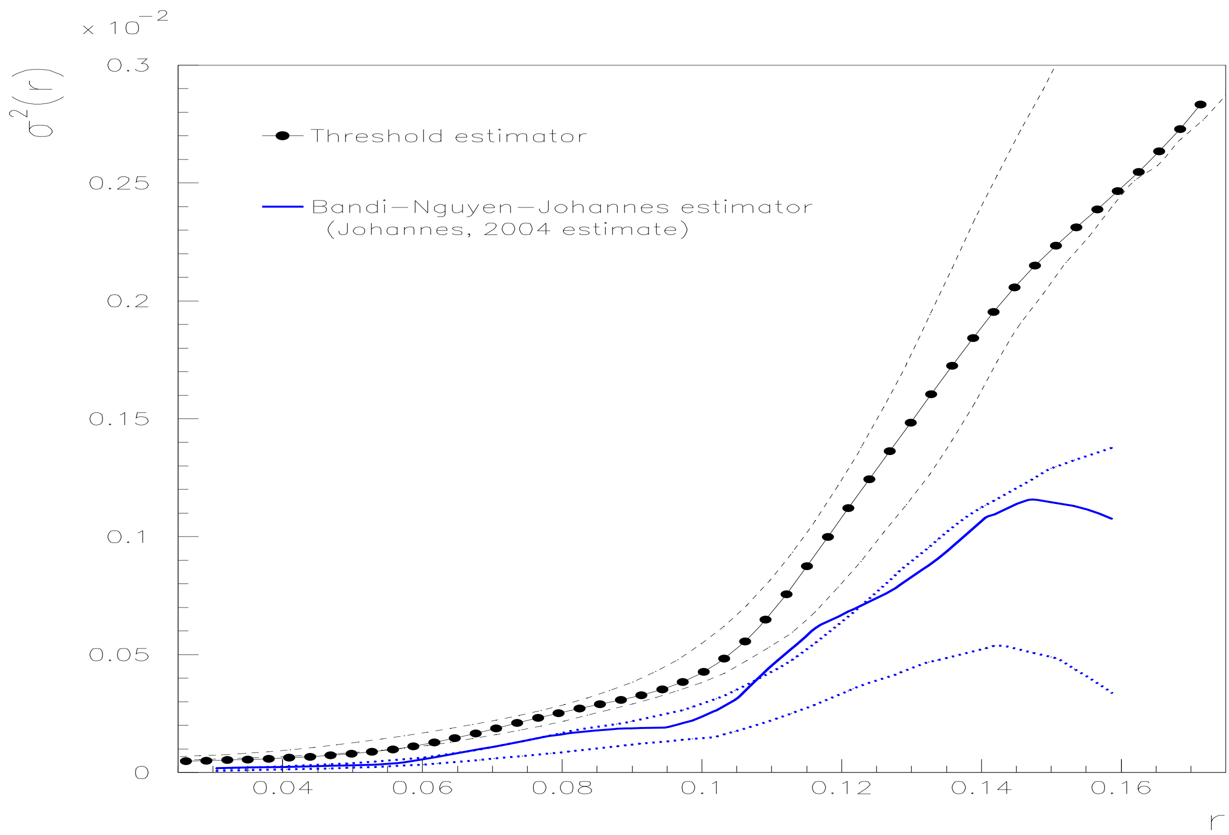

Figure 6.9: Estimates of the diffusion function $\sigma^2(r)$ with the threshold estimator (3.5), together with the estimate on the same time series published by Johannes (2004). Dashed lines are the $(10\%, 90\%)$ Monte Carlo confidence bands for the threshold estimator. Dotted lined are the $(10\%, 90\%)$ Monte Carlo confidence bands published by Johannes (2004).

that is due to liquidity reasons. Our estimates imply that the 7-day time series, and generally instruments with maturity below three months, should not be used when estimating continuous diffusions. Once the jump component is detected and discarded, the two rates display the same diffusive pattern, as it should be since both are used as proxies of the short rate.

We also use data analysis to compare the performance of the threshold nonparametric estimator to that of the estimator of Bandi and Nguyen (2003) and Johannes (2004). The threshold estimator has an important theoretical advantage, namely it is capable to detect both jump sizes and times. Moreover, we also show that the threshold estimator provides quite different estimates with respect to the alternative, especially regarding the different contributions to the total variance of the short interest rate. We find that more variance has to be ascribed to the diffusive part than to the jump part, with respect to the findings in Johannes (2004): We find less jump intensity and a significantly larger diffusion function.

We think that our results can be useful not only in the framework of interest rate modeling, but more generally for

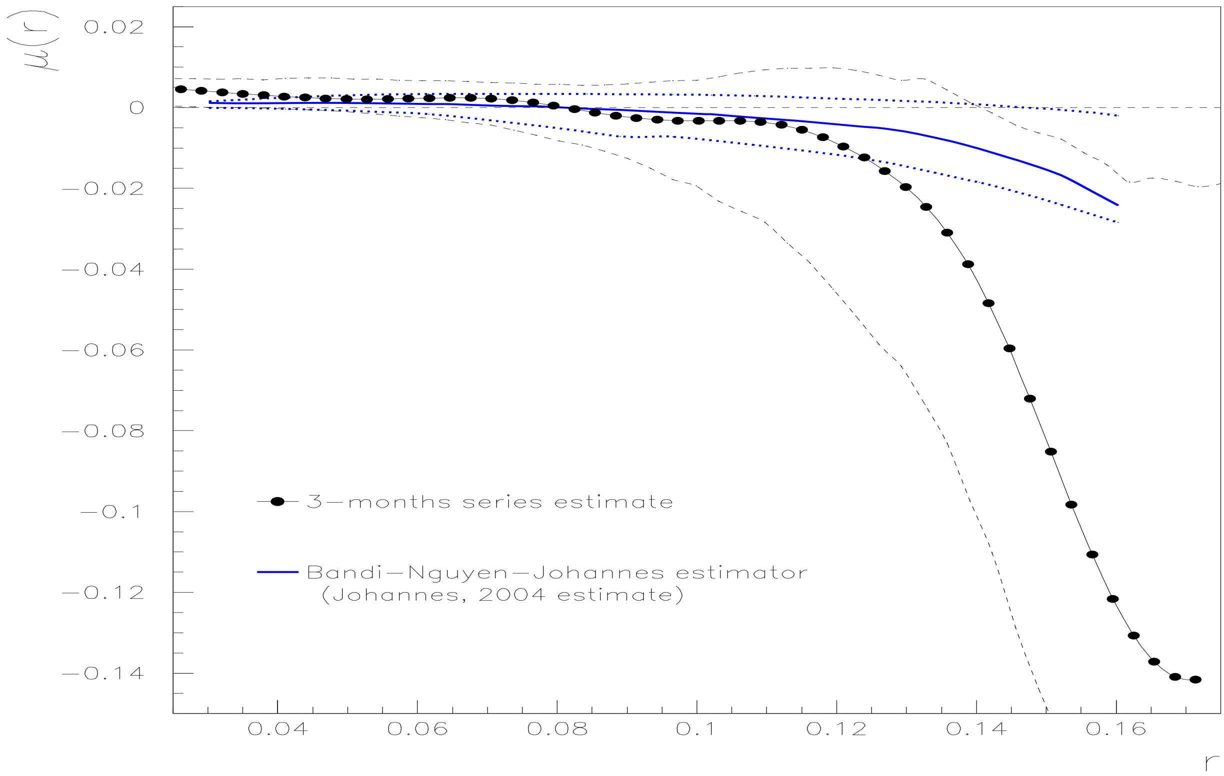

Figure 6.10: Estimates of the drift function $\mu(r)$ with the threshold estimator (3.6), together with the estimate on the same time series published by Johannes (2004). Dashed lines are the $(10\%, 90\%)$ Monte Carlo confidence bands for the threshold estimator. Dotted lined are the $(10\%, 90\%)$ Monte Carlo confidence bands published by Johannes (2004).

asset and derivatives pricing. Further research on this topic is under development.

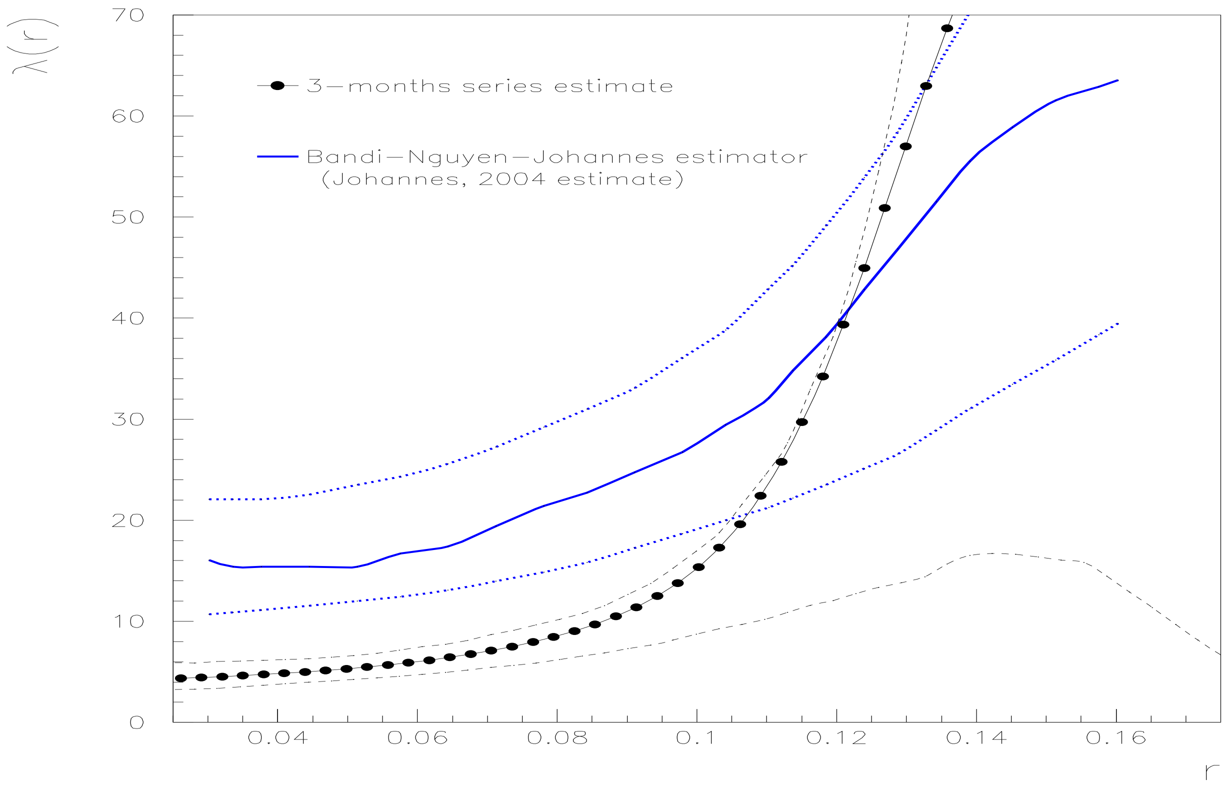

Figure 6.11: Estimates of the intensity function $\lambda(r)$ with the threshold estimator (3.7), together with the estimate on the same time series published by Johannes (2004). Dashed lines are the $(10\%, 90\%)$ Monte Carlo confidence bands for the threshold estimator. Dotted lined are the $(10\%, 90\%)$ Monte Carlo confidence bands published by Johannes (2004).

References

Ait-Sahalia, Y. (1996). Nonparametric pricing of interest rate derivative securities. *Econometrica 64*(3), 527–560.

Ait-Sahalia, Y. (2004). Disentangling diffusion from jumps. *Journal of Financial Economics 74*, 487–528.

Aït-Sahalia, Y. and J. Jacod (2006). Discretely sampled Lévy processes: estimation. *Annals of Statistics*. Forthcoming.

Andersen, T., L. Benzoni, and J. Lund (2002). Towards an empirical foundation for continous-time equity return models. *Journal of Finance 57*, 1239–1284.

Andersen, T., L. Benzoni, and J. Lund (2004). Stochastic volatility, mean drift and jumps in the short-term interest rate. Working Paper.

Andersen, T., T. Bollerslev, and F. Diebold (2005). Roughing it up: Including jump components in the measurement, modeling and forecasting of return volatility. Manuscript.

Andersen, T., T. Bollerslev, and D. Dobrev (2006). No-arbitrage semi-martingale restrictions for continuous-time volatility models subject to leverage effect and jumps: Theory and testable distributional implications. *Journal of Econometrics*. Forthcoming.

Bakshi, G., C. Cao, and Z. Chen (1997). Empirical performance of alternative option pricing models. *Journal of Finance 52*, 2003–2049.

Bandi, F. (2002). Short-term interest rate dynamics: a spatial approach. *Journal of Financial Economics 65*, 73–110.

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

Bandi, F. and P. Phillips (2003). Fully nonparametric estimation of scalar diffusion models. *Econometrica 71*(1), 241–283.

Barndorff-Nielsen, O. E. and N. Shephard (2004). Power and bipower variation with stochastic volatility and jumps. *Journal of Financial Econometrics 2*, 1–48.

Barndorff-Nielsen, O. E. and N. Shephard (2006). Econometrics of testing for jumps in financial economics using bipower variation. *Journal of Financial Econometrics 4*, 1–30.

Barndorff-Nielsen, O. E., N. Shephard, and M. Winkel (2006). Limit theorems for multipower variation in the presence of jumps. *Stochastic processes and applications 116*, 796–806.

Bates, D. (2000). Post-'87 crash fears in the S&P 500 futures option market. *Journal of Econometrics 94*, 181–238.

Brémaud, P. (1981). *Point processes and queues*. New York: Springer Verlag.

Carr, P., H. Geman, D. Madan, and M. Yor (2002). The fine structure of asset returns: an empirical investigation. *Journal of Business 75*(2), 305–332.

Chapman, D., J. Long, and N. Pearson (1999). Using proxies for the short rate: When are three months like an instant? *Review of Financial Studies 12*(4), 763–806.

Chapman, D. and N. Pearson (2000). Is the short rate drift actually nonlinear? *Journal of Finance 55*(1), 355–388.

Cont, R. and P. Tankov (2004). *Financial Modelling with Jump Processes*. Chapman & Hall - CRC.

Das, S. (2002). The surprise element: Jumps in interest rates. *Journal of Econometrics 106*, 27–65.

Duffee, G. (1996). Idiosyncratic variation of Treasury Bill yield spread. *Journal of Finance 51*, 527–552.

Eberlein, E. and S. Raible (1999). Term structure models driven by general Lévy processes. *Mathematical Finance 9*, 31–53.

Eraker, B., M. Johannes, and N. Polson (2003). The impact of jumps in equity index volatility and returns. *Journal of Finance 58*, 1269–1300.

Fan, J. (2005). A selective overview of nonparametric methods in finance. *Statistical Science 20*(4), 317–337.

Florens-Zmirou, D. (1993). On estimating the diffusion coefficient from discrete observations. *Journal of Applied Probability 30*, 790–804.

Hamilton, J. (1996). The daily market for federal funds. *Journal of Political Economy 104*, 26–56.

Hong, Y. and H. Li (2005). Nonparametric specification testing for continuous-time models with applications to term structure of interest rates. *Review of Financial Studies 18*(1), 37–84.

Huang, X. and G. Tauchen (2005). The relative contribution of jumps to total price variance. *Journal of Financial Econometrics 3*(4), 456–499.

Ikeda, N. and S. Watanabe (1981). *Stochastic differential equations and diffusion processes*. North Holland.

Jacod, J. (1979). *Calcul Stochastique et problème de martingales*, Volume 714 of *Lecture Notes in Mathematics*. Springer.

Jacod, J. (2006). Asymptotic properties of realized power variations and associated functionals of semimartingales. Working paper.

Jiang, G. (1998). Nonparametric modeling of U.S. interest rate term structure dynamics and implications on the prices of derivative securities. *Journal of Financial and Quantitative Analysis 33*(4), 465–497.

Jiang, G. and J. Knight (1997). A nonparametric approach to the estimation of diffusion processes, with an application to a short-term interest rate model. *Econometric Theory 13*, 615–645.

Jiang, G. and R. Oomen (2004). Estimating latent variables and jump diffusion models using high frequency data. Working Paper.

Jiang, G. and R. Oomen (2006). A new test for jumps in asset prices. Working Paper.

Johannes, M. (2004). The statistical and economic role of jumps in continuous-time interest rate models. *Journal of Finance 59*, 227–260.

Jones, C. (2003). Nonlinear mean reversion in the short-term interest rate. *Review of Financial Studies 16*, 765–791.

Karatzas, I. and E. Shreve (1988). *Brownian Motion and Stochastic Calculus*. Springer-Verlag.

Madan, D. (2001). Purely discontinuous asset price processes. In J. Cvitanic, E. Jouini, and M. Musiela (Eds.), *Advances in Mathematical Finance*. Cambrdige University Press.

Mancini, C. (2004a). Estimating the integrated volatility in stochastic volatility models with Lévy type jumps. Working Paper, Università di Firenze.

Mancini, C. (2004b). Estimation of the parameters of jump of a general Poisson-diffusion model. *Scandinavian Actuarial Journal 1*, 42–52.

Nelson, D. (1992). Filtering and forecasting with misspecified ARCH models I: Getting the right variance with the wrong model. *Journal of Econometrics 52*, 61–90.

Nelson, D. and D. Foster (1994). Asymptotic filtering theory for univariate ARCH models. *Econometrica 62*(1), 1–41.

Pan, J. (2002). The jump-risk premia implicit in options: Evidence from an integrated time series study. *Journal of Financial Economics 63*, 3–50.

Piazzesi, M. (2005). Bond yields and the federal reserve. *Journal of Political Economy 113*(2), 311–344.

Pritsker, M. (1998). Nonparametric density estimation and tests of continuous time interest rate models. *Review of Financial Studies 11*(3), 449–487.

Protter, P. (1990). *Stochastic Integration and Differential Equations*. Springer-Verlag.

Renò, R., A. Roma, and S. Schaefer (2004). A comparison of alternative nonparametric estimators of the diffusion coefficient. Quaderni del Dipartimento di Economia Politica, Università di Siena.

Revuz, D. and M. Yor (2001). *Continuous Martingales and Brownian Motion*. Springer.

Sato, K. (2001). Basic results on Lévy processes. In O. Barndorff-Nielsen, T. Mikosch, and S. I. Resnick (Eds.), *Lévy Processes, Theory and Applications*, pp. 3–37. Birkhäuser.

Scott, D. (1992). *Multivariate Density Estimation*. John Wiley & Sons.

Stanton, R. (1997). A nonparametric model of term structure dynamics and the market price of interest rate risk. *Journal of Finance 52*, 1973–2002.

Tauchen, G. and H. Zhou (2006). Identifying realized jumps on financial markets. Working Paper.

Woerner, J. (2005). Estimation of integrated volatility in stochastic volatility models. *Applied Stochastic Models in Business and Industry 21*, 27–44.

A Mathematical appendix

We recall that $\delta = \frac{T}{n}$, $t_i = i\delta$, $X = Y + J$. We define

$$K_{i-1} \doteq K\left(\frac{X_{t_{i-1}} - x}{h}\right) \text{ and } K_s \doteq \left(\frac{X_s - x}{h}\right),$$

$$\Delta_i \hat{Y} = \Delta_i X I_{\{(\Delta_i X)^2 \leq \vartheta(\delta)\}} \quad \Delta_i \hat{J} = \Delta_i X I_{\{(\Delta_i X)^2 > \vartheta(\delta)\}},$$

by an upper bar we indicate the supremum of the absolute value, e.g.

$$\bar{\sigma} = \sup_x |\sigma(x)|.$$

Lemma A.1 *(Mancini, 2004a) Take $T = 1$. If μ_s and σ_s have cadlag paths, we have*

$$\max_{i=1..n} \frac{|\int_{t_{i-1}}^{t_i} \mu_u du + \Delta_i \sigma.W|}{\sqrt{\delta \ln \frac{1}{\delta}}} \leq (\sqrt{2}\tilde{\bar{\sigma}} + \sqrt{2} + \tilde{\bar{\mu}}) \doteq \Lambda, \tag{A.1}$$

and

$$\max_{i=1..n} \sup_{s \in](i-1)\delta, i\delta]} \frac{|Y_s - Y_{t_{i-1}}|}{\sqrt{\delta \ln \frac{1}{\delta}}} \leq \Lambda \tag{A.2}$$

where $\tilde{\bar{\sigma}}(\omega) = \sup_{s\in[0,T]} |\sigma_s(\omega)|$, $\tilde{\bar{\mu}}(\omega) = \sup_{s\in[0,T]} |\mu_s(\omega)|$ and Λ are a.s. finite. •

We denote by $L_T(x)$ the local time of X at time T and point x. We define the local time of the jump-diffusion process as in Protter (1990), p. 165. We will use the characterization of Corollary 3, p. 178. We define the following modification of local time:

$$L_T^\star(x) = L_T(x) \int_{\mathbb{R}_+} K(u)du + L_T(x^-) \int_{\mathbb{R}_-} K(u)$$

Proposition A.2 *For fixed $T > 0$, if*

1. *$\sigma(x)$ is continuos, and $\sigma^2(x) > 0$ for all x;*
2. *J has finite activity;*
3. $\dfrac{\sqrt{\delta \ln \frac{1}{\delta}}}{h^2} \to 0$;
4. *K and K' are bounded*

then we have

$$\hat{L}_T^\star(x) = \frac{1}{h} \sum_{i=1}^n K\left(\frac{X_{t_{i-1}} - x}{h}\right) \delta \to \frac{L_T^\star(x)}{\sigma^2(x)},$$

as δ and $h \to 0$.

Proof of Proposition A.2. *Consistency of $\hat{L}^\star_T(x)$, fixed T, finite jump activity.*

We set, without loss of generality, $T = 1$, so that $\frac{T}{n} = \frac{1}{n} = \delta$. Recall that $t_i = i\delta$. Set $L \doteq L_1$, L being the local time of X at time T. We follow the same lines as in Bandi and Phillips (2003) with the necessary modifications for our framework. We have

$$\hat{L}^\star(x) = \frac{1}{h}\sum_{i=1}^{n} K_{i-1}\delta = \frac{1}{h}\left(\sum_{i=1}^{n} K_{i-1}\delta - \int_0^T K_s ds\right) + \frac{1}{h}\int_0^T K_s ds. \tag{A.3}$$

Note that the integrals $\int_0^T K_s ds$ and $\int_0^T K_{s-} ds$ coincide.

We show that the first term is negligible and the second one tends to $\frac{L^\star(x)}{\sigma^2(x)}$. For each n define

$$I_{0,n} = \{i \in \{1..n\} : \Delta_i N = 0\} \text{ and } I_{1,n} = \{i \in \{1..n\} : \Delta_i N \neq 0\}.$$

The first term of (A.3) is dominated by

$$\frac{1}{h}\left|\sum_{i=1}^{n} K_{i-1}\delta - \int_0^T K_s ds\right|$$

$$\leq \frac{1}{h}\sum_{i\in I_{0,n}}\int_{t_{i-1}}^{t_i} |K_s - K_{i-1}|ds + \frac{1}{h}\sum_{i\in I_{1,n}}\int_{t_{i-1}}^{t_i} |K_{i-1} - K_s|ds. \tag{A.4}$$

We have that $\frac{1}{h}\sum_{i\in I_{1,n}}\int_{t_{i-1}}^{t_i} |K_{i-1} - K_s|ds$ is dominated by $N_T\frac{2\bar{K}\delta}{h} \to 0$, given assumption 2 of Proposition A.2. The first term in (A.4) is dominated by

$$\frac{1}{h}\sum_{i\in I_{0,n}}\int_{t_{i-1}}^{t_i}\left|K'\left(\frac{\tilde{X}_{is} - x}{h}\right)\right|\frac{|X_s - X_{t_{i-1}}|}{h}ds = O_{a.s.}\left(\frac{\sqrt{\delta \ln\frac{1}{\delta}}}{h^2}\right)\bar{K}'T \to 0, \tag{A.5}$$

for suitable values $\tilde{X}_{is}$ in $]X_{t_{i-1}}, X_s[$, $i \in I_{0,n}$, where X is continuous.

The second term of (A.3) coincides with

$$\frac{1}{h}\int_0^T K\left(\frac{X_{s-} - x}{h}\right)\frac{d[X]^c_s}{\sigma^2(X_{s-})}$$

and by the occupation time formula (Protter, 1990) the latter term coincides with

$$\frac{1}{h}\int_{\mathbb{R}} K\left(\frac{a-x}{h}\right)\frac{1}{\sigma^2(a)}\,L(a)da =$$

$$\int_{\mathbb{R}_+} K(u)\frac{1}{\sigma^2(uh+x)}\,L(uh+x)du + \int_{\mathbb{R}_-} K(u)\frac{1}{\sigma^2(uh+x)}\,L(uh+x)du$$

which a.s. tends to $\frac{L^\star(x)}{\sigma^2(x)}$, and therefore our assertion is proved. •

Proof of Theorem 3.3. *Consistency of $\hat{\sigma}^2(x)$, fixed T, finite jump activity.*

Set $T = 1$, so that $\delta\frac{T}{n} = \frac{1}{n}$, and $L \doteq L_1$, the local time of X at time $T = 1$. Set $\hat{Y} = X - \hat{J}_1$. By (3.2), estimator (3.5) a.s. coincides with

$$\frac{\sum_{i=1}^n K_{i-1}(\Delta_i Y)^2 I_{\{\Delta_i N=0\}}}{\sum_{i=1}^n K_{i-1}\delta} =$$

$$= \frac{\sum_{i=1}^n K_{i-1}(\Delta_i Y)^2}{\sum_{i=1}^n K_{i-1}\delta} - \frac{\sum_{i=1}^n K_{i-1}(\Delta_i Y)^2 I_{\{\Delta_i N\neq 0\}}}{\sum_{i=1}^n K_{i-1}\delta}. \tag{A.6}$$

The second term is negligible, since

$$\left|\frac{\sum_{i=1}^n K_{i-1}(\Delta_i Y)^2 I_{\{\Delta_i N\neq 0\}}}{\sum_{i=1}^n K_{i-1}\delta}\right| \leq \Lambda^2 \bar{K}\delta \ln\frac{1}{\delta}\frac{N_T}{\sum_{i=1}^n K_{i-1}\delta},$$

and, denoting by $\psi = 1-\frac{1}{\beta}$, by Proposition A.2,

$$\text{Plim }\left(\delta\ln\frac{1}{\delta}\right)\frac{1}{\sum_{i=1}^n K_{i-1}\delta} = \text{Plim }\ln(n)\frac{\sigma^2(x)}{nhL^\star(x)} = O\left(\frac{n^\psi}{nh}\frac{\ln n}{n^\psi}\right),$$

which tends to zero.

Now we deal with the first term of (A.6), which coincides with

$$\frac{\frac{1}{h}\left(\sum_{i=1}^n K_{i-1}(\Delta_i Y)^2 - \int_0^T K_{s-}d[Y]_s\right) + \frac{1}{h}\int_0^T K_{s-}d[Y]_s}{\frac{1}{h}\sum_{i=1}^n K_{i-1}\delta}. \tag{A.7}$$

For the integral in the numerator of (A.7) we have:

$$\begin{aligned}\frac{1}{h}\int_0^T K\left(\frac{X_{s-}-x}{h}\right)d[Y]_s ds &= \frac{1}{h}\int_0^T K\left(\frac{X_{s-}-x}{h}\right)\sigma^2(X_{s-})ds = \\ &= \frac{1}{h}\int_0^T K_{s-}d[X]_s^c = \frac{1}{h}\int_{\mathbb{R}} K\left(\frac{a-x}{h}\right)L(a)da = \\ = \int_{\mathbb{R}_+} K(u)\,L(uh+x)du &+ \int_{\mathbb{R}_-} K(u)\,L(uh+x)du \to L^\star(x)\end{aligned} \tag{A.8}$$

as $h\to 0$. Using proposition A.2 we then have:

$$\frac{\frac{1}{h}\int_0^T K_{s-}d[Y]_s}{\frac{1}{h}\sum_{i=1}^n K_{i-1}\delta} \to \sigma^2(x).$$

We then show that the remainder of the numerator of (A.7) tends to zero. Note that, as a consequence of Ito's formula, for all $i = 1..n$ we have

$$(\Delta_i Y)^2 = [Y]_{t_i} - [Y]_{t_{i-1}} + 2\int_{t_{i-1}}^{t_i}(Y_u - Y_{t_{i-1}})dY_u, \tag{A.9}$$

and note that

$$\frac{1}{h}\sum_{i\in I_{1,n}}\left(K_{i-1}(\Delta_i Y)^2 - \int_{t_{i-1}}^{t_i} K_{s-}\sigma^2(X_{s-})ds\right) \leq N_T\bar{K}\frac{\Lambda^2\delta\ln\frac{1}{\delta}+\bar{\sigma}^2\delta}{h} \to 0,$$

so that the Plim of the first term of (A.7) coincides with the Plim of

$$\begin{aligned}&\frac{1}{h}\sum_{i\in I_{0,n}}\left(K_{i-1}(\Delta_i Y)^2 - \int_{t_{i-1}}^{t_i} K_{s-}\sigma^2(X_{s-})ds\right) = \\ &\frac{1}{h}\sum_{i\in I_{0,n}} K_{i-1}\left(\int_{t_{i-1}}^{t_i}\sigma^2(X_{s-})ds + 2\int_{t_{i-1}}^{t_i}(Y_u - Y_{t_{i-1}})dY_u - \int_{t_{i-1}}^{t_i} K_{s-}\sigma^2(X_{s-})ds\right) = \\ &\frac{1}{h}\sum_{i\in I_{0,n}}\int_{t_{i-1}}^{t_i}(K_{i-1}-K_{s-})\sigma^2(X_{s-})ds + \frac{2}{h}\sum_{i\in I_{0,n}}\int_{t_{i-1}}^{t_i} K_{i-1}(Y_{u-}-Y_{t_{i-1}})\mu(X_{u-})du + \\ &+\frac{2}{h}\sum_{i\in I_{0,n}}\int_{t_{i-1}}^{t_i} K_{i-1}(Y_{u-}-Y_{t_{i-1}})\sigma(X_{u-})dW_u.\end{aligned} \tag{A.10}$$

Since σ is bounded, using the same technique of equation (A.5) we get that the first term of (A.10) tends to zero. By the boundedness of μ and K, the second term is $O\left(\frac{\sqrt{\delta \ln \frac{1}{\delta}}}{h}\right) \to 0$. Finally, the last term of (A.10) has the same Plim as U_1, where

$$U_r = \frac{2}{h} \sum_{i=1}^{[nr]} \int_{t_{i-1}}^{t_i} K_{i-1}(Y_u - Y_{t_{i-1}})\sigma(X_u) dW_u.$$

U_1 tends to zero in probability, since U_r is a martingale with quadratic variation at $T = 1$ equal to

$$[U]_1 = \frac{4}{h^2} \sum_{i=1}^{n} \int_{t_{i-1}}^{t_i} K_{i-1}^2 (Y_u - Y_{t_{i-1}})^2 \sigma^2(X_u) du = O\left(\frac{\delta \ln \frac{1}{\delta}}{h^2}\right) \to 0.$$

•

In what follows, we use the following classical result, see (Revuz and Yor, 2001), p. 153, exercise 3.6:

Proposition A.3 *Let H to be a real process, which is progressively measurable, and s.t.*

$$\int_0^T H_s^2 ds \le M^2 T \quad a.s.,$$

then

$$P\{\sup_{t\in[0,T]} \left|\int_0^t H_s dW_s\right| \ge c\} \le 2e^{-\frac{c^2}{2M^2T}}.$$

Corollary A.4 *(Mancini, 2004b) If σ is bounded,*

$$P\left\{\sup_{t\in[t_{i-1},t_i]} \left|\int_{t_{i-1}}^t \sigma_s dW_s\right| \ge c\right\} \le 2e^{-\frac{c^2}{2\bar\sigma^2\delta}},$$

and thus, $\forall i$

$$P\left\{\left|\int_{t_{i-1}}^{t_i} \sigma_s dW_s\right| \ge c\right\} \le 2e^{-\frac{c^2}{2\bar\sigma^2\delta}}.$$

•

Lemma A.5 *If λ is bounded, then $P\{\Delta_i N \ge 2\} = O(\delta^2)$.*

Proof. We have

$$P\{\Delta_i N \ge 2\} = E[E[I_{\{\Delta_i N \ge 2\}}|X]] =$$

$$E\left[\left(e^{\int_{t_{i-1}}^{t_i} \lambda(X_s)ds} - \int_{t_{i-1}}^{t_i} \lambda(X_s)ds - 1\right) e^{-\int_{t_{i-1}}^{t_i} \lambda(X_s)ds}\right]$$

which has the same limit, as $\delta \to 0$, of

$$E\left[\frac{\left(\int_{t_{i-1}}^{t_i} \lambda(X_s)ds\right)^2}{2}\right] \le \bar\lambda \frac{\delta^2}{2}.$$

•

Proof of Theorem 3.5. *Consistency of $\hat{\mu}(x)$, $T \to \infty$, finite jump activity.*

Note that since $T \to \infty$ as $\delta \to 0$, in an infinite horizon framework it is no more guranteed that (3.2) holds.
Our proof borrows heavily from the proof of Theorem 2 in Bandi and Nguyen (2003). However, it differs in the fact that our estimator eliminates the jumps in the terms $\Delta_i X$.

First step: we start by showing that

$$\text{Plim } \hat{\mu}_n(x) = \text{Plim } \frac{\sum_{i=1}^n K_{i-1}\Delta_i Y}{\sum_{i=1}^n K_{i-1}\delta},$$

which is an implication of:

$$\text{Plim } \frac{1}{h^{1+\psi}} \sum_{i=1}^n K_{i-1}(\Delta_i \hat{Y} - \Delta_i Y) = 0. \tag{A.11}$$

The left hand side in (A.11) is equal to

$$\text{Plim } \frac{1}{h^{1+\psi}} \sum_{i=1}^n K_{i-1}(\Delta_i J - \Delta_i \hat{J})$$

$$= \frac{1}{h^{1+\psi}} \sum_{i=1}^n K_{i-1} \left[\sum_{\ell=1}^{\Delta_i N} \gamma_\ell - \Delta_i X I_{\{(\Delta_i X)^2 > \vartheta(\delta)\}} \right] \left[I_{\{\Delta_i N = 0\}} + I_{\{\Delta_i N = 1\}} + I_{\{\Delta_i N \geq 2\}} \right]. \tag{A.12}$$

We now show that each term tends to zero in probability. Let us begin with the term multiplying $I_{\{\Delta_i N \geq 2\}}$:

$$P\left\{ \frac{1}{h^{1+\psi}} \sum_{i=1}^n K_{i-1} \left[\sum_{\ell=1}^{\Delta_i N} \gamma_\ell - \Delta_i X I_{\{(\Delta_i X)^2 > \vartheta(\delta)\}} \right] I_{\{\Delta_i N \geq 2\}} \neq 0 \right\} \leq P\{\cup_{i=1}^n \{\Delta_i N \geq 2\}\}$$

which, by Lemma A.5, is $O(n\delta^2) \to 0$. Let us now deal with the term multiplying $I_{\{\Delta_i N = 0\}}$:

$$\left| \frac{1}{h^{1+\psi}} \sum_{i=1}^n K_{i-1} \left[\sum_{\ell=1}^{\Delta_i N} \gamma_\ell - \left(\int_{t_{i-1}}^{t_i} \mu(X_s) ds + \Delta_i(\sigma.W) + \sum_{\ell=1}^{\Delta_i N} \gamma_\ell \right) I_{\{(\Delta_i X)^2 > \vartheta(\delta)\}} \right] I_{\{\Delta_i N = 0\}} \right|$$

$$\leq \frac{1}{h^{1+\psi}} \sum_{i=1}^n |K_{i-1}| \left| \int_{t_{i-1}}^{t_i} \mu(X_s) ds + \Delta_i(\sigma.W) \right| I_{\{|\int_{t_{i-1}}^{t_i} \mu(X_s)ds + \Delta_i(\sigma.W)| > \sqrt{\vartheta(\delta)}, \Delta_i N = 0\}},$$

however $|\int_{t_{i-1}}^{t_i} \mu(X_s)ds + \Delta_i(\sigma.W)| \leq |\int_{t_{i-1}}^{t_i} \mu(X_s)ds| + |\Delta_i(\sigma.W)|$, thus last term is dominated by

$$\frac{1}{h^{1+\psi}} \sum_{i=1}^n |K_{i-1}| \left| \int_{t_{i-1}}^{t_i} \mu(X_s) ds + \Delta_i(\sigma.W) \right| I_{\{|\int_{t_{i-1}}^{t_i} \mu(X_s)ds| > \frac{\sqrt{\vartheta(\delta)}}{2}, \Delta_i N = 0\}} +$$

$$+ \frac{1}{h^{1+\psi}} \sum_{i=1}^n |K_{i-1}| \left| \int_{t_{i-1}}^{t_i} \mu(X_s) ds + \Delta_i(\sigma.W) \right| I_{\{|\Delta_i(\sigma.W)| > \frac{\sqrt{\vartheta(\delta)}}{2}, \Delta_i N = 0\}}.$$

Since μ is bounded we have $|\int_{t_{i-1}}^{t_i} \mu(X_s)ds| \leq \bar{\mu}\delta$ for each i, and thus a.s., for small δ, $I_{\left\{|\int_{t_{i-1}}^{t_i} \mu(X_s)ds| > \frac{\sqrt{\vartheta(\delta)}}{2}\right\}}(\omega) = 0$ and the first term above is zero. For the second term

$$P\left\{ \frac{1}{h^{1+\psi}} \sum_{i=1}^n |K_{i-1}| \left| \int_{t_{i-1}}^{t_i} \mu(X_s) ds + \Delta_i(\sigma.W) \right| I_{\{|\Delta_i(\sigma.W)| > \frac{\sqrt{\vartheta(\delta)}}{2}, \Delta_i N = 0\}} \neq 0 \right\} \leq$$

$$\leq P\left\{ \cup_i \left\{ |\Delta_i(\sigma.W)| > \frac{\sqrt{\vartheta(\delta)}}{2} \right\} \right\}$$

which is dominated, by Corollary A.4, for each T, by $2ne^{-\frac{\vartheta(\delta)}{4}\frac{1}{2\sigma^2\delta}}$ which tends to zero, since

$$ne^{-\frac{\vartheta(\delta)}{\delta}} = ne^{-\frac{\vartheta(\delta)}{\delta\ln\frac{1}{\delta}}\ln\frac{1}{\delta}} < ne^{2\ln\delta} = n\delta^2.$$

Finally we deal with the term multiplying $I_{\{\Delta_i N=1\}}$:

$$\frac{1}{h^{1+\psi}}\sum_{i=1}^{n}K_{i-1}\left[\sum_{\ell=1}^{\Delta_i N}\gamma_\ell - \left(\int_{t_{i-1}}^{t_i}\mu(X_s)ds + \Delta_i(\sigma.W) + \sum_{\ell=1}^{\Delta_i N}\gamma_\ell\right) I_{\{(\Delta_i X)^2>\vartheta(\delta)\}}\right] I_{\{\Delta_i N=1\}} =$$

$$\frac{1}{h^{1+\psi}}\sum_{i=1}^{n}K_{i-1}\gamma_{\tau^{(i)}}I_{\{|\Delta_i X|\le\sqrt{\vartheta(\delta)},\Delta_i N=1\}}+ \tag{A.13}$$

$$-\frac{1}{h^{1+\psi}}\sum_{i=1}^{n}K_{i-1}\left(\int_{t_{i-1}}^{t_i}\mu(X_s)ds + \Delta_i(\sigma.W)\right) I_{\{|\Delta_i X|>\sqrt{\vartheta(\delta)},\Delta_i N=1\}}.$$

The probability that the first term is different from zero is dominated by

$$\sum_i P\{|\Delta_i X|\le\sqrt{\vartheta(\delta)},\Delta_i N=1\} =$$

$$\sum_i P\left\{|\Delta_i X|\le\sqrt{\vartheta(\delta)},\Delta_i N=1,|\Delta_i(\sigma.W)|\le\frac{\sqrt{\vartheta(\delta)}}{2}\right\}+$$

$$+\sum_i P\left\{|\Delta_i X|\le\sqrt{\vartheta(\delta)},\Delta_i N=1,|\Delta_i(\sigma.W)|>\frac{\sqrt{\vartheta(\delta)}}{2}\right\}.$$

The second of last terms is dominated by

$$\sum_i P\left\{|\Delta_i(\sigma.W)|>\frac{\sqrt{\vartheta(\delta)}}{2}\right\}$$

which tends to zero as before. For the first one

$$\sum_i P\left\{|\Delta_i X|\le\sqrt{\vartheta(\delta)},\Delta_i N=1,|\Delta_i(\sigma.W)|\le\frac{\sqrt{\vartheta(\delta)}}{2}\right\}\le$$

$$\sum_i P\left\{|\gamma_{\tau^{(i)}}|\le\frac{3}{2}\sqrt{\vartheta(\delta)},\Delta_i N=1\right\}, \tag{A.14}$$

since $|\Delta_i X|\ge|\Delta_i J| - |\int_{t_{i-1}}^{t_i}\mu(X_s)ds| - |\Delta_i(\sigma.W)|$, so that on $\left\{|\Delta_i X|\le\sqrt{\vartheta(\delta)},\Delta_i N=1,|\Delta_i(\sigma.W)|\le\frac{\sqrt{\vartheta(\delta)}}{2}\right\}$ we have $|\Delta_i J|\le\bar{\mu}\delta+\frac{3}{2}\sqrt{\vartheta(\delta)} = O\left(\frac{3}{2}\sqrt{\vartheta(\delta)}\right)$; moreover on $\{\Delta_i N=1\}$ we have $\Delta_i J=\gamma_{\tau^{(i)}}$. Now

$$\sum_i P\left\{|\gamma_{\tau^{(i)}}|\le\frac{3}{2}\sqrt{\vartheta(\delta)},\Delta_i N=1\right\} = \sum_i P\left\{|\gamma_{\tau^{(i)}}|\le\frac{3}{2}\sqrt{\vartheta(\delta)}\right\}P\{\Delta_i N=1\} \tag{A.15}$$

by the independence of each γ_k from N. Now $P\{\Delta_i N=1\} = E[E[I_{\{\Delta_i N=1\}}|X]]\le E[\int_{t_{i-1}}^{t_i}\lambda(X_u)du]\le\bar{\lambda}\delta$, and by assumption we have $P\left\{|\gamma_{\tau^{(i)}}|\le\frac{3}{2}\sqrt{\vartheta(\delta)}\right\}\le c\sqrt{\vartheta(\delta)}$. Therefore (A.15) is dominated by $\sum_i c\sqrt{\vartheta(\delta)}\,\bar{\lambda}\delta = O(n\delta^{1+\eta/2})\to 0$.

Finally we show that the second term of (A.13) tends to zero in probability:

$$\left|\frac{1}{h^{1+\psi}}\sum_{i=1}^{n}K_{i-1}\int_{t_{i-1}}^{t_i}\mu(X_s)ds I_{\{|\Delta_i X|>\sqrt{\vartheta(\delta)},\Delta_i N=1\}}\right|$$

$$\leq \frac{1}{h^{1+\psi}} \sum_{i=1}^{n} |K_{i-1}| \bar{\mu} \delta I_{\{\Delta_i N = 1\}} \leq \bar{K} \bar{\mu} \delta \frac{N_T}{T} \frac{T}{h^{1+\psi}}.$$

We have that $E\left[\frac{N_T}{T}\right] \leq \bar{\lambda}$ so the term above is $O_{a.s.}(\delta^{2-\gamma(1+\psi)} n) \to 0$ since $\delta^{3/2-\gamma(1+\psi)} n \to 0$ by assumption. Moreover, for all T:

$$\left| \frac{1}{h^{1+\psi}} \sum_{i=1}^{n} K_{i-1} \Delta_i(\sigma.W) I_{\{|\Delta_i X| > \sqrt{\vartheta(\delta)}, \Delta_i N = 1\}} \right| \leq$$

$$\frac{\bar{K}}{h^{1+\psi}} \sqrt{\delta \ln \frac{1}{\delta}} \bar{\sigma} \frac{N_T}{T} T = O_{a.s.}\left(n \delta^{3/2-\gamma(1+\psi)} \sqrt{\ln \frac{1}{\delta}} \right) \to 0.$$

<u>*Second step*</u>: now

$$\frac{\sum_{i=1}^{n} K_{i-1} \Delta_i Y}{\sum_{i=1}^{n} K_{i-1} \delta} = \alpha_{n,T}(x) + \beta_{n,T}(x), \tag{A.16}$$

where $\alpha_{n,T}(x) = \frac{\sum_{i=1}^{n} K_{i-1} \int_{t_{i-1}}^{t_i} \mu(X_{s-}) ds}{\sum_{i=1}^{n} K_{i-1} \delta}$ and $\beta_{n,T}(x) = \frac{\sum_{i=1}^{n} K_{i-1} \int_{t_{i-1}}^{t_i} \sigma(X_{s-}) dW_s}{\sum_{i=1}^{n} K_{i-1} \delta}$ coincide with the $\alpha_{n,T}(x)$ and $\beta_{n,T}(x)$ defined by equation (104) in Bandi and Nguyen (2003), where it is shown that a.s.

$$\alpha_{n,T}(x) \to \mu(x) \text{ and } \beta_{n,T}(x) \to 0$$

for each x visited by the path $\{X_s\}_{s \in \mathbb{R}_+}$, as n and T jointly diverge.

Consistency of $\hat{\lambda}(x)$, $T \to \infty$, finite jump activity.

$$\hat{\lambda}_n(x) = \frac{\sum_{i=1}^{n} K_{i-1} \Delta_i \hat{N}}{\sum_{i=1}^{n} K_{i-1} \delta} - \frac{\sum_{i=1}^{n} K_{i-1} \Delta_i \hat{N} (1 - c_{in})}{\sum_{i=1}^{n} K_{i-1} \delta}.$$

The second term is dominated by $\sup_i |1 - c_{in}| \frac{\sum_{i=1}^{n} K_{i-1} \Delta_i \hat{N}}{\sum_{i=1}^{n} K_{i-1} \delta}$. Since $\sup_i |1 - c_{in}| \to 0$, it is sufficient to show that $\frac{\sum_{i=1}^{n} K_{i-1} \Delta_i \hat{N}}{\sum_{i=1}^{n} K_{i-1} \delta}$ converges to $\lambda(x)$. We write the last ratio as

$$\frac{\frac{1}{h} \sum_{i=1}^{n} K_{i-1} \Delta_i N + \frac{1}{h} \sum_{i=1}^{n} K_{i-1} (\Delta_i \hat{N} - \Delta_i N)}{\frac{1}{h} \sum_{i=1}^{n} K_{i-1} \delta}. \tag{A.17}$$

As in equation (A.12), $\frac{1}{h} \sum_{i=1}^{n} K_{i-1} (\Delta_i \hat{N} - \Delta_i N) = o_P(h^{\psi})$, so that the limit in probability of (A.17) coincides with

$$\text{Plim} \frac{\frac{1}{h} \sum_{i=1}^{n} K_{i-1} \Delta_i N}{\frac{1}{h} \sum_{i=1}^{n} K_{i-1} \delta} =$$

$$\text{Plim} \frac{\sum_{i=1}^{n} K_{i-1} (\Delta_i N - \int_{t_{i-1}}^{t_i} \lambda(X_{s-}) ds)}{\sum_{i=1}^{n} K_{i-1} \delta} + \frac{\frac{1}{h} \sum_{i=1}^{n} K_{i-1} \int_{t_{i-1}}^{t_i} \lambda(X_{s-}) ds}{\frac{1}{h} \sum_{i=1}^{n} K_{i-1} \delta} \tag{A.18}$$

The second term tends to $\lambda(x)$ in probability, as for equation (106) in Bandi and Nguyen (2003), with $\lambda(X_{s-})$ in place of $\mu(X_{s-})$. The first term of (A.18) is equal to $\gamma_{n,T}$ in equation (104) in Bandi and Nguyen (2003), with $c(X_{s^-}, y) \equiv 1$, see also equation (8) in Bandi and Nguyen (2003), and it tends a.s. to 0 as shown in Bandi and Nguyen (2003), equation (126). •

To show Theorem 4.2 we need the following lemmas.

Lemma A.6 *Under the assumptions $\vartheta(\delta) \to 0$, $\frac{\delta \ln \frac{1}{\delta}}{\vartheta(\delta)} \to 0$, for any fixed $c > 0$, we have a.s. for δ small enough*

$$\forall i = 1..n \quad I_{\{|\Delta_i Y| > c\sqrt{\vartheta(\delta)}\}}(\omega) = 0.$$

Proof. $\{|\Delta_i Y| > c\sqrt{\vartheta(\delta)}\} = \left\{ \frac{|\Delta_i Y|}{\sqrt{\delta \ln \frac{1}{\delta}}} > c \frac{\sqrt{\vartheta(\delta)}}{\sqrt{\delta \ln \frac{1}{\delta}}} \right\}$ and $\frac{\sqrt{\vartheta(\delta)}}{\sqrt{\delta \ln \frac{1}{\delta}}} \to \infty$ by assumption, while a.s. $\frac{|\Delta_i Y|}{\sqrt{\delta \ln \frac{1}{\delta}}} \leq \Lambda < \infty$. •

Lemma A.7 *Under the assumptions 4.1, $\vartheta(\delta) = \delta^\eta$, $\eta \in]0,1[$, we have a.s. for δ small enough*

$$\forall i = 1..n \quad I_{\{(\Delta_i \tilde{J}_2)^2 > \vartheta(\delta)/4\}}(\omega) = 0.$$

Proof. On $\{(\Delta_i \tilde{J}_2)^2 > \vartheta(\delta)/4\}$ for small δ we have (Mancini, 2004a):

$$\Delta_i \tilde{J}_2 = \int_{t_{i-1}}^{t_i} \int_{\sqrt{\vartheta(\delta)}/2 < |x| \leq 1} x m(dt, dx) = \Delta_i \tilde{J}_{2m}^{>} + \Delta_i \tilde{J}_{2c},$$

where

$$\Delta_i \tilde{J}_{2m}^{>} \doteq \int_{t_{i-1}}^{t_i} \int_{\sqrt{\vartheta(\delta)}/2 < |x| \leq 1} x \tilde{m}(dt, dx), \quad \Delta_i \tilde{J}_{2c} \doteq \int_{t_{i-1}}^{t_i} \int_{\sqrt{\vartheta(\delta)}/2 < |x| \leq 1} x \nu(dx) dt,$$

so

$$\{(\Delta_i \tilde{J}_2)^2 > \vartheta(\delta)/4\} \subset \{2(\Delta_i \tilde{J}_{2m}^{>})^2 + 2(\Delta_i \tilde{J}_{2c})^2 > \vartheta(\delta)/4\} \subset$$

$$\left\{ (\Delta_i \tilde{J}_{2m}^{>})^2 > \frac{\vartheta(\delta)}{16} \right\} \cup \left\{ (\Delta_i \tilde{J}_{2c})^2 > \frac{\vartheta(\delta)}{16} \right\}. \tag{A.19}$$

Now, $\{(\Delta_i \tilde{J}_{2c})^2 > \frac{\vartheta(\delta)}{16}\} = \{\delta \vartheta(\delta)^{-\frac{\alpha}{2}} > c\}$, which is empty for sufficiently small δ, for a positive constant c, uniformly in i, since $\frac{\delta}{\vartheta(\delta)^{\frac{\alpha}{2}}} = \left(\frac{\delta}{\vartheta(\delta)}\right)^{\frac{\alpha}{2}} \delta^{1-\frac{\alpha}{2}} \to 0$.
Moreover $\forall i$ also each $\{(\Delta_i \tilde{J}_{2m}^{>})^2 > \frac{\vartheta(\delta)}{16}\}$ is empty for small δ. In fact in probability it coincides with $\{(\tilde{J}_{2m,\delta}^{>})^2 > \frac{\vartheta(\delta)}{16}\}$ which is a subset of

$$\left\{ \sup_{s \leq \delta} (\tilde{J}_{2m,s}^{>})^2 > \frac{\vartheta(\delta)}{16} \right\} = \left\{ \sup_{s \leq \delta} \frac{|J_{2m,s}^{>}|}{\delta^{1/2-\beta}} \, \delta^{1/2-\beta} > \sqrt{\vartheta(\delta)}/4 \right\}, \tag{A.20}$$

where we take $\beta > 0$ such that $1 - \eta - 2\beta > 0$. Since by the Doob inequality $\sup_{s \leq \delta} \frac{|J_{2m,s}^{>}|}{\delta^{1/2-\beta}} \xrightarrow{P} 0$, passing to a subsequence we have the a.s. convergence to zero, and in particular $\sup_{s \leq \delta} \frac{|J_{2m,s}^{>}|}{\delta^{1/2-\beta}} < 1$ a.s. for small δ, so (A.20) is subset of

$$\{\delta^{1/2-\beta} > \sqrt{\vartheta(\delta)}/4\} = \{\delta^{1/2-\beta} \vartheta(\delta)^{-1/2} > 1/4\}$$

which is empty for small δ, by the choice we have done of β and the assumption that $\vartheta(\delta) = \delta^\eta$ with $\eta \in]0,1[$. •

Lemma A.8 *Under the assumptions 4.1, $\vartheta(\delta) = \delta^\eta$, $\eta \in]0,1[$, given a double array of real numbers a_{ni} with $i = 1, \ldots, n$, if $\sup_{i=1,\ldots,n} |a_{ni}| \to 0$ then*

$$Plim \sum_i |a_{ni}| I_{\{(\Delta_i X)^2 \leq \vartheta(\delta)\}} = Plim \sum_i |a_{ni}| I_{\{(\Delta_i \tilde{J}_2)^2 \leq 4\vartheta(\delta), \Delta_i N = 0\}}.$$

Proof. On $\{(\Delta_i X)^2 \le \vartheta(\delta)\}$ we have $|\Delta_i J| - |\Delta_i Y| \le |\Delta_i X| \le \sqrt{\vartheta(\delta)}$ and thus, for small δ, $|\Delta_i J| \le 2\sqrt{\vartheta(\delta)}$, so that

$$\text{Plim} \sum_i |a_{ni}| I_{\{(\Delta_i X)^2 \le \vartheta(\delta)\}} \le \text{Plim} \sum_i |a_{ni}| I_{\{(\Delta_i J)^2 \le 4\vartheta(\delta)\}}.$$

However Plim $\sum_i |a_{ni}| I_{\{(\Delta_i J)^2 \le 4\vartheta(\delta), \Delta_i N \neq 0\}} \le \sup_i |a_{ni}| N_T \to 0$, and thus

$$\text{Plim} \sum_i |a_{ni}| I_{\{(\Delta_i X)^2 \le \vartheta(\delta)\}} \le \text{Plim} \sum_i |a_{ni}| I_{\{(\Delta_i J)^2 \le 4\vartheta(\delta), \Delta_i N = 0\}} =$$

$$\text{Plim} \sum_i |a_{ni}| I_{\{(\Delta_i \tilde{J}_2)^2 \le 4\vartheta(\delta), \Delta_i N = 0\}}.$$

On the other hand

$$\text{Plim} \sum_i |a_{ni}| (I_{\{(\Delta_i J)^2 \le 4\vartheta(\delta), \Delta_i N = 0\}} - I_{\{(\Delta_i X)^2 \le \vartheta(\delta)\}}) \xrightarrow{P} 0,$$

since

$$\{(\Delta_i J)^2 \le 4\vartheta(\delta), \Delta_i N = 0\} - \{(\Delta_i X)^2 \le \vartheta(\delta)\} =$$

$$\{(\Delta_i J)^2 \le 4\vartheta(\delta), \Delta_i N = 0, (\Delta_i X)^2 > \vartheta(\delta)\} \subset$$

$$\{|\Delta_i J| \le 2\sqrt{\vartheta(\delta)}, \Delta_i N = 0, |\Delta_i Y| + |\Delta_i \tilde{J}_2| > \sqrt{\vartheta(\delta)}\} \subset$$

$$\{|\Delta_i J| \le 2\sqrt{\vartheta(\delta)}, |\Delta_i Y| > \sqrt{\vartheta(\delta)}/2\} \cup \{|\Delta_i J| \le 2\sqrt{\vartheta(\delta)}, \Delta_i N = 0, |\Delta_i \tilde{J}_2| > \sqrt{\vartheta(\delta)}/2\}$$

and by lemma A.6, a.s. for small δ, $I_{\{|\Delta_i Y| > \sqrt{\vartheta(\delta)}/2\}}(\omega) = 0, \forall i$, so

$$\sum_i |a_{ni}| I_{\{|\Delta_i J| \le 2\sqrt{\vartheta(\delta)}, |\Delta_i Y| > \sqrt{\vartheta(\delta)}/2\}} = 0.$$

Finally, by lemma A.7, a.s. for small δ also $I_{\{|\Delta_i \tilde{J}_2| > \sqrt{\vartheta(\delta)}/2\}}(\omega) = 0, \ \forall i$, so that

$$\sum_i |a_{ni}| I_{\{|\Delta_i J| \le 2\sqrt{\vartheta(\delta)}, \Delta_i N = 0, |\Delta_i \tilde{J}_2| > \sqrt{\vartheta(\delta)}/2\}} = 0.$$

•

Proof of Theorem 4.2.

Consistency of $\hat{\sigma}^2(x)$, fixed T, infinite jump activity. Recall that T is fixed. Set $T = 1$. Let α be the Blumenthal-Gatoor index of J.

First of all, on $\{(\Delta_i \tilde{J}_2)^2 \le 4\vartheta(\delta), \Delta_i N = 0\}$ we have

$$\frac{1}{h}(\Delta_i X)^2 K_{i-1} \le \frac{1}{h}(\Delta_i \tilde{J}_2 + \Delta_i Y)^2 \bar{K} = O\left(\frac{\vartheta(\delta)}{h}\right) \to 0,$$

and on $\{(\Delta_i X)^2 \le \vartheta(\delta)\}$ we have

$$\frac{1}{h}(\Delta_i X)^2 = O\left(\frac{\vartheta(\delta)}{h}\right) \to 0,$$

uniformly on i, so that by lemma A.8 the limit in probability of $\hat{\sigma}_n^2(x)$ coincides with

$$\text{Plim} \frac{\frac{1}{h}\sum_{i=1}^n K_{i-1}(\Delta_i X)^2 I_{\{(\Delta_i \tilde{J}_2)^2 \le 4\vartheta(\delta), \Delta_i N = 0\}}}{\frac{1}{h}\sum_{i=1}^n K_{i-1}\delta} \tag{A.21}$$

$$= \frac{\frac{1}{h}\sum_{i=1}^n K_{i-1}(\Delta_i Y)^2 + \frac{1}{h}\sum_i K_{i-1}\left[(\Delta_i \tilde{J}_2 + \Delta_i Y)^2 I_{\{|\Delta_i \tilde{J}_2| \le 2\sqrt{\vartheta(\delta)}, \Delta_i N = 0\}} - (\Delta_i Y)^2\right]}{\frac{1}{h}\sum_{i=1}^n K_{i-1}\delta}.$$

We now show that last term is equal to

$$\text{Plim}\ \frac{\frac{1}{h}\sum_{i=1}^{n} K_{i-1}(\Delta_i Y)^2}{\frac{1}{h}\sum_{i=1}^{n} K_{i-1}\delta}. \tag{A.22}$$

It is sufficient to prove that Plim $\frac{\sum_{i=1}^n K_{i-1}\left[(\Delta_i \tilde{J}_2+\Delta_i Y)^2 I_{\{|\Delta_i \tilde{J}_2|\le 2\sqrt{\vartheta(\delta)},\Delta_i N=0\}}-(\Delta_i Y)^2\right]}{h} = 0$. To this purpose, we have to prove the following equalities:

$$\text{Plim}\ \frac{\sum_{i=1}^n K_{i-1}[(\Delta_i \tilde{J}_2+\Delta_i Y)^2 I_{\{|\Delta_i \tilde{J}_2|\le 2\sqrt{\vartheta(\delta)},\Delta_i N=0\}}-(\Delta_i Y)^2]}{h} = \tag{A.23}$$

$$\text{Plim}\ \frac{\sum_{i=1}^n K_{i-1}[(\Delta_i \tilde{J}_2+\Delta_i Y)^2 I_{\{|\Delta_i \tilde{J}_2|\le 2\sqrt{\vartheta(\delta)}\}}-(\Delta_i Y)^2]}{h} = \tag{A.24}$$

$$\text{Plim}\ \frac{\sum_{i=1}^n K_{i-1}\left[(\Delta_i Y)^2\left(\frac{\Delta_i \tilde{J}_2 I_{\{|\Delta_i \tilde{J}_2|\le 2\sqrt{\vartheta(\delta)}\}}}{\Delta_i Y}+1\right)^2 I_{\{|\Delta_i \tilde{J}_2|\le 2\sqrt{\vartheta(\delta)}\}}-(\Delta_i Y)^2\right]}{h} = \tag{A.25}$$

$$\text{Plim}\ \frac{\sum_{i=1}^n K_{i-1}[(\Delta_i Y)^2 I_{\{|\Delta_i \tilde{J}_2|\le 2\sqrt{\vartheta(\delta)}\}}-(\Delta_i Y)^2]}{h} = \tag{A.26}$$

$$\text{Plim}\ -\frac{\sum_{i=1}^n K_{i-1}(\Delta_i Y)^2 I_{\{|\Delta_i \tilde{J}_2|> 2\sqrt{\vartheta(\delta)}\}}}{h} = 0 \tag{A.27}$$

Let us prove each step. We get (A.24) showing that

$$\text{Plim}\ \frac{\sum_{i=1}^k(\Delta_i \tilde{J}_2+\Delta_i Y)^2 I_{\{|\Delta_i \tilde{J}_2|\le 2\sqrt{\vartheta(\delta)},\Delta_i N\neq 0\}}}{h} = 0.$$

In fact the above term is $O\left(N_1\frac{(\sqrt{\vartheta(\delta)}+\sqrt{\delta\ln\frac{1}{\delta}})^2}{h}\right)=O\left(\frac{\vartheta(\delta)}{h}\right)\to 0$ by assumption.

(A.25) is straightforward.

Now, in order to get (A.26), denote by:

$$\Delta_i \tilde{J}_{2m} \doteq \int_{t_{i-1}}^{t_i}\int_{|x|\le\sqrt{\vartheta(\delta)}/2} x\tilde{m}(dt,dx)$$

and recall that

$$\Delta_i \tilde{J}_{2c} \doteq \int_{t_{i-1}}^{t_i}\int_{\sqrt{\vartheta(\delta)}/2<|x|\le 1} x\nu(dx)dt,$$

so that (Mancini, 2004a)

$$(\Delta_i \tilde{J}_2)^2 I_{\{|\Delta_i \tilde{J}_2|\le 2\sqrt{\vartheta(\delta)}\}} = \Delta_i \tilde{J}_{2m}+\Delta_i \tilde{J}_{2c}.$$

We note that $(\sigma.W)_s = B_{IV_s}$ is a time changed Brownian motion, where $IV_s=\int_0^T\sigma^2(X_u)du$, so that

$$|\Delta_i\sigma.W| = |B_{IV_{t_i}}-B_{IV_{t_{i-1}}}|$$

which dominates $\ge(1-\varepsilon)\sqrt{2\Delta_i IV|\ln\Delta_i IV|}\ge(1-\varepsilon)\sqrt{\underline{\sigma}^2\delta|\ln\underline{\sigma}^2\delta|}$ by the monotonicity of $x\ln|\ln x|$, where $\underline{\sigma}^2 = \inf_x|\sigma^2(x)|$. Moreover, uniformly in i, for small δ, we have the same lower bound for $\Delta_i Y = \int_{t_{i-1}}^{t_i}\mu(X_s)ds + \int_{t_{i-1}}^{t_i}\sigma(X_s)dW_s = \left(\frac{\int_{t_{i-1}}^{t_i}\mu(X_s)ds}{\int_{t_{i-1}}^{t_i}\sigma(X_s)dW_s}+1\right)\int_{t_{i-1}}^{t_i}\sigma(X_s)dW_s$. Therefore, for all $\varepsilon>0$, for small δ,

$$\left|\frac{\Delta_i \tilde{J}_{2m}}{\Delta_i Y}\right|^2 \le \frac{|\Delta_i \tilde{J}_{2m}|^2}{\underline{\sigma}^2(1-\varepsilon)^2\delta|\ln\underline{\sigma}^2\delta|}$$

and

$$E\left[\frac{|\Delta_i \tilde{J}_{2m}|^2}{\delta|\ln\delta|}\right] \leq \frac{\delta\int_{|x|\leq\sqrt{\vartheta(\delta)}} x^2\nu(dx)}{(1-\varepsilon)^2\delta|\ln\delta|} = O\left(\frac{\vartheta(\delta)^{1-\alpha/2}}{|\ln\delta|}\right) \to 0. \tag{A.28}$$

On the other hand

$$\left|\frac{\Delta_i \tilde{J}_{2c}}{\Delta_i Y}\right|^2 \leq \frac{\delta^2\left(\int_{|x|\in]2\sqrt{\vartheta(\delta)},1]}|x|\nu(dx)\right)^2}{\underline{\sigma}^2(1-\varepsilon)^2\delta|\ln\underline{\sigma}^2\delta|} \to 0. \tag{A.29}$$

Now, the equality of (A.26) with the left hand side of (A.27) is straightforward. The conclusion in (A.27) follows from lemma A.7, and (A.22) is proved.

Now, from (A.22) on, we use a reasoning similar to that in the proof of Theorem 3.3. Namely (A.7) and (A.8) hold because of the occupation time formula (Protter, 1990), as soon as L has a cadlag version. By (A.9), for each $i = 1..n$ we can write

$$\frac{1}{h}\sum_{i=1}^{n}\left(K_{i-1}(\Delta_i Y)^2 - \int_{t_{i-1}}^{t_i} K_{s-}\sigma^2(X_{s-})ds\right)$$

$$= \frac{1}{h}\sum_i\int_{t_{i-1}}^{t_i}(K_{i-1}-K_{s-})\sigma^2(X_{s-})ds + \frac{2}{h}\sum_i\int_{t_{i-1}}^{t_i}K_{i-1}(Y_{u-}-Y_{t_{i-1}})\mu(X_{u-})du \tag{A.30}$$

$$+\frac{2}{h}\sum_i\int_{t_{i-1}}^{t_i}K_{i-1}(Y_{u-}-Y_{t_{i-1}})\sigma(X_{u-})dW_u, \tag{A.31}$$

From Sato (2001), we have that a.s.:

$$\limsup_{\delta\to 0}\frac{|J_\delta|}{\sqrt{2\delta\ln|\ln\delta|}} = 0,$$

thus $\max_{i=1..n}\sup_{s\in Ii}|X_{s-}-X_{i\delta}| = O_P(\sqrt{\delta\ln|\ln\delta|})$, which allows to show that each term of (A.30) and (A.31) tends to zero in probability as in Theorem 3.3. •

B Monte Carlo confidence bands

In this Section, we briefly explain how confidence bands are computed in finite samples. First of all, we assume that the data generating process is a doubly stochastic Poisson process with level dependent intensity, and that jump sizes are Normally distributed. Clearly, reliability of confidence bands hinges on this assumption.

We use the simulations to get confidence bands. Once obtained the estimates $\hat{\mu}, \hat{\sigma}, \hat{\lambda}$ on a given sample, we simulate the paths of the corresponding univariate jump-diffusion model. We use $1,000$ replications. Paths have the same length of the sample on which estimation took place, and the starting value of the short rate is chosen uniformly randomly between the minimum and the maximum observations. On each sample, we compute the nonparametric estimates and record the minimum and the maximum value obtained. We use exactly the same parameters used in estimation, as well as the same GARCH(1,1) model. For the jump component, we use $dJ_t = \exp(Z_t - 1)\, r_t dN_t$, where N is a Poisson process and $Z_t \simeq \mathcal{N}(0, \sigma_J^2)$ is a Normal distribution. The standard deviation of jumps, σ_J, is estimated on the time series of detected jumps. We can also assess the efficiency and noise in detecting jumps (signal and background efficiency). Since on generated paths we know exactly the location of the jumps, we can compute N_1 : the number of detected jumps which correspond to an actually simulated jump; N_2 : the number of detected jumps which do no correspond to an actually simulated jump; $\hat{N}_J = N_1 + N_2$ gives the total number of detected jumps, while N_J is the number of actually generated jumps. If $N+1$ is the number of simulated data, we define the efficiency as the ratio N_1/N_J and the noise as the ratio $N_2/(N - N_J)$.