QUADERNI

Università degli Studi di Siena
DIPARTIMENTO DI ECONOMIA POLITICA

ROBERTO RENO'

Nonparametric Estimation of the Diffusion Coefficient via Fourier Analysis, with Aplication to Short Rate Modeling

n. 440 - Novembre 2004

Nonparametric estimation of the diffusion coefficient via Fourier analysis, with application to short rate modeling*

Roberto Renò
Dipartimento di Economia Politica
Università di Siena
e-mail: `reno@unisi.it`

October 27, 2004

Abstract

In this paper a new fully nonparametric estimator of the diffusion coefficient is introduced, based on Fourier analysis of the observed trajectory. The proposed estimator is proved to be consistent and asymptotically normally distributed. After testing the estimator on Monte Carlo simulations, we use it to estimate an univariate model of the short rate with available interest rate data. Data analysis helps shedding new light on the functional form of the diffusion coefficient.

Jel Classification: C14, C6, E43

*I wish to acknowledge Mavira Mancino, Maurizio Pratelli, Antonio Roma and Rosario Rizza for reviewing the manuscript, and especially Marzia De Donno for pointing out an error in an older version. All errors and omissions are my own responsibility.

1 Introduction

The last decade witnessed a growing literature in the field of the diffusion coefficient estimation. The main motivation underlying this strand of research is that the diffusion coefficient, which is called volatility, plays a fundamental role in practically every financial application. We concentrate on univariate models of the kind:

$$dX_t = \mu(X_t)dt + \sigma(X_t)dW_t \tag{1.1}$$

where W_t is a standard real Brownian motion and the real functions $\mu(x)$ and $\sigma(x)$ are such that a unique solution X_t of the stochastic differential equation (1.1) exists. X_t can be any variable; however, in this paper we concentrate on short rate modeling, that is $X_t = r_t$. Our specific problem is then to estimate the diffusion term $\sigma(r)$ when we observe a discrete realization of the process r, namely n observation $\hat{r}_{t_1}, \ldots, \hat{r}_{t_n}$ in the interval $[0, T]$.

The methods for measuring volatility can be coarsely divided into parametric and nonparametric. The parametric approach consists in specifying the function $\sigma(r) = \sigma(r; \vec{\theta})$, with $\vec{\theta}$ being a vector of real parameters. As a popular example, a large parametric class has been explored by Chan et al. (1992), who study the following model:

$$dr_t = \beta(\alpha - r_t)dt + \sigma r_t^{\gamma} dW_t,$$

where $\beta, \alpha, \sigma, \gamma$ are real numbers. This specification nests many popular one-factor models, like the constant variance model of Vasicek (1977), for $\gamma = 0$ or the square-root diffusion of Cox et al. (1985), for $\gamma = 0.5$. The methodology is to estimate $\vec{\theta}$ through point estimation. For interest rates diffusions, this can be done via maximum likelihood (Duffie et al., 2002) or GMM, direct (Chan et al., 1992) or simulated (Gallant and Tauchen, 1996; Dai and Singleton, 2000). The advantage of parametric models is that closed form solutions exist for bond and derivative pricing. On the other hand, the advantage of nonparametric specification is clearly its flexibility. For example, Jiang (1998) shows that the nonparametric specification provides more accurate prices for bonds and derivatives. However, for bond and derivative pricing one has to resort to Monte Carlo simulations since closed form solutions are out of reach.

One important example of nonparametric estimator for the diffusion coefficient is that proposed by Ait-Sahalia (1996a). The Ait-Sahalia estimator is based on the fact that, if the spot rate r_t is driven by equation (1.1), then we have:

$$\sigma^2(r) = \frac{2\int_{-\infty}^{r} \mu(y)\pi(y)dy}{\pi(r)} \tag{1.2}$$

where $\pi(r)$ is the unconditional distribution of r. Equation (1.2) allows, given two out of the three functions $\mu(r), \sigma(r), \pi(r)$, to obtain the third after integration or derivation. This estimator is not fully nonparametric, since to get an estimate of the variance a specification of the drift term is needed. Ait-Sahalia (1996a) suggests to specify the drift $\mu(r)$ as an affine function of r, then to estimate the conditional variance $\sigma(r)$. Given the drift $\mu(r)$, the estimator (1.2) still depends on the unconditional distribution $\pi(r)$. However, we can obtain an estimate of $\pi(r)$ with a nonparametric technique (Scott, 1992) and replace $\pi(r)$ in (1.2) with its estimate. Suppose our observations are equally spaced, and denote them by $\hat{r}_i, i = 1, \ldots, n$. Then the nonparametric estimator of the density is given by:

$$\hat{\pi}(r) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{r - \hat{r}_i}{h}\right) \tag{1.3}$$

where $K(\cdot)$ is the kernel function and h a bandwidth parameter which depends on n. One popular way of estimating densities through formula (1.3) is the histogram, where the kernel function is an indicator function of a compact real interval centered around zero.

Ait-Sahalia (1996a) estimation results on interest rate data show that a departure from classical univariate models (e.g. CIR and Vasicek) is observed. In a related paper, Ait-Sahalia (1996b) rejected almost all the most popular one-factor model used for modeling interest rates. However this rejection is very controversial. The problem is that estimating (1.3) in finite samples can be problematic if the mean reversion is low. For example, Pritsker (1998) shows that the rejection of the most popular parametric models comes from severe underestimation of confidence interval for testing the null in small samples.

The most remarkable example of a fully nonparametric estimator of the diffusion coefficient can be found in Florens-Zmirou (1993). She introduces an estimator which is conceptually different from that used in Ait-Sahalia (1996a), since it does not need any assumption on the drift. Florens-Zmirou (1993) shows that, given discretely sampled data, the diffusion coefficients in (1.1) may be estimated by:

$$\hat{\sigma}^2(r) = \frac{n \sum_{i=1}^{n-1} (\hat{r}_{i+1} - \hat{r}_i)^2 K\left(\frac{r - \hat{r}_i}{h}\right)}{T \sum_{i=1}^{n} K\left(\frac{r - \hat{r}_i}{h}\right)}. \tag{1.4}$$

The variance estimator (1.4) looks more appealing since there is the same kernel in the numerator and in the denominator, so biases in finite samples coming from nonparametric estimation of the density could cancel out. The estimator (1.4) has been used by Jiang and Knight (1997) on Canadian interest rates, and by Stanton (1997) on U.S. interest rates. In both those papers, the Authors conclude in favor of a departure from standard models, and they suggest a strong mean reversion for values of the spot rate r less than 3% and larger than 15%, see Chapman and Pearson (2000) for a discussion on these results.

An estimator similar to that of Florens-Zmirou (1993) has been proposed by Bandi and Phillips (2003), and studied in Bandi (2002). The estimator proposed in Bandi and Phillips (2003) is the following:

$$\hat{\sigma}^2(r) = \frac{n \sum_{i=1}^{n} K\left(\frac{r - \hat{r}_i}{h}\right) \left(\frac{1}{m_i} \sum_{j=0}^{m_i} \left[\hat{r}_{t_{i,j}+1} - \hat{r}_{t_{i,j}}\right]^2\right)}{T \sum_{i=1}^{n} K\left(\frac{r - \hat{r}_i}{h}\right)} \tag{1.5}$$

where $t_{i,j}$ is a subset of indexes such that

$$t_{i,0} = \inf \{t \geq 0 : |\hat{r}_t - \hat{r}_i| \leq \varepsilon_s\},$$

and

$$t_{i,j+1} = \inf \{t \geq t_{i,j} + \Delta t : |\hat{r}_t - \hat{r}_i| \leq \varepsilon_s\},$$

m_i is the number of times that $|\hat{r}_t - \hat{r}_i| \leq \varepsilon_s$, ε_s is a parameter to be selected and Δt is the time step between adjacent observations. We refer the reader to the cited paper for details. It is important to remark that Bandi and Phillips do not require the process (1.1) to be stationary, but only the weaker condition to be recurrent. This condition can be important theoretically, since Bandi (2002) and many others, e.g. Ball and Torous (1996), show that there is no strong support to the assumption of stationarity of interest rate data. Watching carefully expressions (1.4) and (1.5), we can see that the difference between Florens-Zmirou and Bandi-Phillips estimators is that, while the Florens-Zmirou estimator weights the observation r_t with the quadratic variation at time t, the Bandi-Phillips estimator weights the observation r_t with the average quadratic variation at all observations which are "close" to r_t.

Other estimators have been proposed, in a more general framework, by Jacod (1999). Hoffmann (1999) proposes a wavelet estimator which is consistent in $\mathcal{L}^p$. Both these authors study convergence rate properties for their estimator, showing that the Florens-Zmirou (1993) estimator is not optimal. However, asymptotic distribution can only be assessed for the estimator (1.4).

In this paper we introduce a new fully nonparametric estimator of the diffusion coefficient of an univariate stochastic differential equation. The estimator is fully nonparametric in the sense that we do not impose any restriction on the functional form of the drift term. Moreover, it is developed under mild regularity conditions for the stochastic differential equation (1.1). As for the estimator in Bandi and Phillips (2003), the stationarity assumption is not strictly required, being substituted by the milder assumption of recurrency. The estimator is proved to be consistent in the $\mathcal{L}^2$ sense, and asymptotically normally distributed. In order to assess the asymptotic properties, we borrow from the limit theory for semimartingales, and in particular of the convergence of a semimartingale to a process with independent increments. The asymptotic distribution turns out to be identical to that of Florens-Zmirou (1993). However, the estimator is basically different, and this allows us to re-examine the estimation of the diffusion coefficient with the available interest rate data.

This paper is structured as follows. Section 2 shows how to estimate volatility from discrete observation in an interval $[0, T]$. Section 3 presents the estimator, and shows its consistency and asymptotic normality, using limit theory for semimartingales. In Section 4 we implement the estimator for measuring the variance of the diffusion of short interest rates, and compare our results to those in the literature, and in particulare with the results of Stanton (1997). Finally, Section 5 concludes.

2 Volatility estimation

We start with a result stated by Malliavin and Mancino (2002) which will be very useful in the following. We work in the filtered probability space $(\Omega, \mathcal{F}_t, \mathcal{P})$ satisfying the usual conditions

(Protter, 1990), and define X_t as the solution of the following process:

$$\begin{cases} dX_t = \mu(X_t)dt + \sigma(X_t)dW_t \\ X_0 = x_0 \end{cases} \tag{2.1}$$

where $\sigma(\cdot), \mu(\cdot)$ are bounded, differentiable functions such that a unique solution of the stochastic differential equation (2.1) exists in the interval $[0,T]$ and W_t is a standard Brownian motion. We will write $X_t(\omega)$ to explicit the dependence of X from $t \in [0,T]$ and $\omega \in \Omega$.

In this case X_t is a semimartingale, and its quadratic variation (Jacod and Shiryaev, 1987) is given by:

$$[X,X]_t = \int_0^t \sigma^2(X_s)ds \tag{2.2}$$

Without loss of generality, we consider the solution X_t in the interval $[0,2\pi]$, and define the Fourier coefficients of dX and σ^2 as follows:

$$\begin{aligned} a_0(dX) &= \frac{1}{2\pi}\int_0^{2\pi} dX_t & a_0(\sigma^2) &= \frac{1}{2\pi}\int_0^{2\pi} \sigma^2(X_t)dt \\ a_k(dX) &= \frac{1}{\pi}\int_0^{2\pi} \cos(kt)dX_t & a_k(\sigma^2) &= \frac{1}{\pi}\int_0^{2\pi} \cos(kt)\sigma^2(X_t)dt \\ b_k(dX) &= \frac{1}{\pi}\int_0^{2\pi} \sin(kt)dX_t & b_k(\sigma^2) &= \frac{1}{\pi}\int_0^{2\pi} \sin(kt)\sigma^2(X_t)dt \end{aligned} \tag{2.3}$$

It is worth noting that these integrals can be defined on each trajectory, for almost every trajectory in Ω, see Föllmer (1979).

There are many ways to reconstruct $\sigma^2(X_t)$ given its Fourier coefficients. One way is the Fourier-Fejer formula:

$$\sigma^2(X_t) = \lim_{M\to\infty} \sum_{k=0}^{M} \left(1 - \frac{k}{M}\right) \left[a_k(\sigma^2)\cos(kt) + b_k(\sigma^2)\sin(kt)\right] \tag{2.4}$$

Convergence of Fourier sums is in $\mathcal{L}^2([0,2\pi])$ norm. We now state the main result:

Theorem 2.1 *Consider a process X_t satisfying (2.1), and define the Fourier coefficients of dX and σ^2 as in (2.3). Given an integer $n_0 > 0$, we have almost surely:*

$$a_0(\sigma^2) = \lim_{N\to\infty} \frac{\pi}{N+1-n_0} \sum_{k=n_0}^{N} a_k^2(dX) = \lim_{N\to\infty} \frac{\pi}{N+1-n_0} \sum_{k=n_0}^{N} b_k^2(dX) \tag{2.5}$$

$$a_q(\sigma^2) = \lim_{N\to\infty} \frac{2\pi}{N+1-n_0} \sum_{k=n_0}^{N} a_k(dX)a_{k+q}(dX) = \lim_{N\to\infty} \frac{2\pi}{N+1-n_0} \sum_{k=n_0}^{N} b_k(dX)b_{k+q}(dX) \tag{2.6}$$

$$b_q(\sigma^2) = \lim_{N\to\infty} \frac{2\pi}{N+1-n_0} \sum_{k=n_0}^{N} a_k(dX)b_{k+q}(dX) = \lim_{N\to\infty} \frac{2\pi}{N+1-n_0} \sum_{k=n_0}^{N} b_k(dX)a_{k+q}(dX) \tag{2.7}$$

Proof. See Appendix A. □

Corollary 2.2 *The Fourier coefficients of $\sigma^2(X_t)$ can be computed almost surely as:*

$$a_0(\sigma^2) = \lim_{N\to\infty} \frac{\pi}{N+1-n_0} \sum_{k=n_0}^{N} \frac{1}{2}\left(a_k^2(dX) + b_k^2(dX)\right) \tag{2.8}$$

$$a_q(\sigma^2) = \lim_{N\to\infty} \frac{\pi}{N+1-n_0} \sum_{k=n_0}^{N} \left(a_k(dX)a_{k+q}(dX) + b_k(dX)b_{k+q}(dX)\right) \tag{2.9}$$

$$b_q(\sigma^2) = \lim_{N\to\infty} \frac{\pi}{N+1-n_0} \sum_{k=n_0}^{N} \left(a_k(dX)b_{k+q}(dX) + b_k(dX)a_{k+q}(dX)\right) \tag{2.10}$$

From Theorem 2.1 we get immediately an estimator of the integrated volatility. Indeed:

$$\int_0^{2\pi} \sigma^2(X_s)ds = 2\pi a_0(\sigma^2) \tag{2.11}$$

where $a_0(\sigma^2)$ is given by formula (2.5). The following Theorem provides asymptotic confidence intervals for the Fourier coefficients of volatility, in the case of constant σ:

Theorem 2.3 *Assume volatility is a constant, $\sigma(\cdot) = \sigma \in \mathbb{R}$. As $N \to \infty$, we have:*

$$\sqrt{N+1-n_0}\left(a_0(\sigma^2) - \sigma^2\right) \to \mathcal{N}\left(0, 2\sigma^4\right) \tag{2.12}$$

$$\sqrt{N+1-n_0}\; a_q(\sigma^2) \to \mathcal{N}\left(0, \sigma^4\right) \tag{2.13}$$

$$\sqrt{N+1-n_0}\; b_q(\sigma^2) \to \mathcal{N}\left(0, \sigma^4\right) \tag{2.14}$$

where the above limit is in distribution.

Proof. We have already shown in the proof of Theorem 2.1 that $a_k(dX), b_k(dX)$ can be replaced by $a_k(dv), b_k(dv)$ where $dv = dX - \mu(t)dt$, since all the contribution of the coefficients of μ vanish a.s. as $N \to \infty$. We start from the fact that $a_k(dv), b_k(dv)$ are Gaussian random variables with zero mean. Let $\sigma(\cdot) = \sigma$. We then have:

$$\mathbb{E}[a_k^2(dv)] = \mathbb{E}[b_k^2(dv)] = \frac{1}{\pi}\sigma^2 \tag{2.15}$$

and

$$\mathbb{E}[a_k^4(dv)] = \mathbb{E}[b_k^4(dv)] = \frac{3}{\pi^2}\sigma^4 \tag{2.16}$$

Moreover, from the orthogonality of the trigonometric base, if $k \neq h, \mathbb{E}[a_k(dv)a_h(dv)] =$ $= \mathbb{E}[b_h(dv)b_k(dv)] = 0$ and, for every $k, h, \mathbb{E}[a_k(dv)b_k(dv)] = 0$. Thus $a_k(dv), b_h(dv)$ are all independent, thus if $k \neq h$, $a_k^2(dv) + b_k^2(dv)$ is independent of $a_h^2(dv) + b_h^2(dv)$. Then standard central limit theorem yields the result. We get the result for $a_q(\sigma^2), b_q(\sigma^2)$ with the same reasoning, since $\mathbb{E}[a_k(dv)a_{k+q}(dv)] = \mathbb{E}[a_k(dv)b_{k+q}(dv)] = 0$ and $\mathbb{E}[(a_k(dv)a_{k+q}(dv))^2] = \sigma^4/\pi^2$. □

It is sometimes convenient to rewrite equation (2.4) as:

$$\sigma^2(X_t) = \lim_{M\to\infty} \sum_{k=-M}^{M} \left(1 - \frac{k}{M}\right) A_k(\sigma^2) e^{ikt}, \tag{2.17}$$

where

$$A_k(\sigma^2) = \begin{cases} \frac{1}{2}(a_k(\sigma^2) - ib_k(\sigma^2)), & k \geq 1 \\ \frac{1}{2}a_0(\sigma^2), & k = 0 \\ \frac{1}{2}(a_{|k|}(\sigma^2) + ib_{|k|}(\sigma^2)) & k \leq -1 \end{cases} \tag{2.18}$$

For the implementation of the estimator, we follow Barucci and Renò (2002a). Since we observe the process X_t only at discrete times $t_1, \ldots, t_n$, we set $X_t = X_{t_i}$ in the interval $t_i \leq t < t_{i+1}$. Using interpolation techniques different from this we get a bias in the volatility measurement (Barucci and Renò, 2002b). Then the Fourier coefficients of the price can be computed as:

$$a_k(dX) = \frac{1}{\pi}\int_0^{2\pi} \cos(kt)dX_t = \frac{X_{2\pi} - X_0}{\pi} - \frac{k}{\pi}\int_0^{2\pi} \sin(kt)X_t dt, \tag{2.19}$$

then using:

$$\frac{k}{\pi}\int_{t_i}^{t_{i+1}} \sin(kt)X_t\, dt = X_{t_i}\frac{k}{\pi}\int_{t_i}^{t_{i+1}} \sin(kt)dt = X_{t_i}\frac{1}{\pi}[\,\cos(kt_i) - \cos(kt_{i+1})\,]. \tag{2.20}$$

Before computing (2.19), we add a linear trend such that we get $X_{2\pi} = X_0$, which does not affect the volatility estimate. Then we stop the expansions (2.8-2.10) at a properly selected frequency N. For equally spaced data, the maximum N which prevents aliasing effects is $N = \frac{n}{2}$, see Priestley (1979). Finally, we have to select the maximum M in (2.17). M should be a function of N such that $M(N) \to \infty$ when $N \to \infty$. In the applications of this paper, we will use evenly spaced data only, then we select $N = \frac{n}{2}$, $M = \frac{N}{2} = \frac{n}{4}$, and $n_0 = 1$.

3 Nonparametric estimation of the diffusion coefficient

We consider the SDE (2.1) in the interval $[0, T]$. We assume the following:

Assumption 3.1 *Given the SDE (2.1), we have:*

i) $x_0 \in \mathcal{L}^2(\Omega)$ *is independent of* W_t, $t \in [0, T]$ *and measurable with respect to* $\mathcal{F}_0$.

ii) $\mu(x)$ *and* $\sigma(x)$, *defined on a compact interval I, are once continuously differentiable.*

iii) There exists a constant K *such that* $0 < \sigma(x) \leq K$ *and* $|\mu(x)| \leq K$.

iv) (Feller condition for non-explosion). Given:

$$S(\alpha) = \int_0^\alpha e^{\int_0^y -\frac{2\mu(x)}{\sigma^2(x)}dx} dy, \tag{3.1}$$

$$V(\alpha) = \int_0^\alpha S'(y) \int_0^y \frac{2}{S'(x)\sigma^2(x)} dxdy, \tag{3.2}$$

then $V(\alpha)$ diverges at the boundaries of I.

Assumption 3.1 insures existence and uniqueness of a strong solution. Asking for the Feller condition allows to deal with models which, as noted by Ait-Sahalia (1996a), do not satisfy global Lipschitz and growth condition (e.g. CIR). Moreover, Feller condition is necessary and sufficient for recurrence in I, see the discussion in Bandi and Phillips (2003). Alternatively, one can ask for global Lipschitz and growth conditions on μ and σ (Karatzas and Shreve, 1988).

Asking for a bounded volatility (and drift) is harmless from an econometric point of view, since we always observe a finite, thus bounded, set of observations. For example, in the CIR model the variance is unbounded and proportional to $\sqrt{X_t}$. However, since estimation is on a finite sample, the observations $\hat{X}_i$ are bounded and $\sqrt{X_t}$ is indistinguishable from $\min(\sqrt{X_t}, \sqrt{\max \hat{X}_i})$.

Moreover we will consider a kernel function for nonparametric estimation with the following properties:

Assumption 3.2 *We define a kernel $K(\cdot)$ a bounded function in $\mathcal{L}^2(\mathbb{R})$ which is continuously differentiable, positive, with $\int_{\mathbb{R}} K(s)ds = 1$ and such that $\lim_{s\to\pm\infty} K(s) = 0$ faster than any inverse polynomial.*

A typical choice is the Gaussian kernel:

$$K(s) = \frac{1}{\sqrt{2\pi}} e^{-\frac{s^2}{2}} \tag{3.3}$$

We will moreover consider a sequence h_n such that:

Assumption 3.3 *$(h_n)_{n\in\mathbb{N}}$ is a real sequence such that, as $n \to \infty$, we have $h_n \to 0$ and $nh_n \to \infty$.*

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

$$h_n = h_s \hat{\sigma} n^{-\frac{1}{5}} \tag{3.4}$$

where h_s is a real constant to be tuned, and $\hat{\sigma}$ is the sample standard deviation. We will assume 3.1,3.2,3.3 holding throughout all the paper.

Consider the solution process X_t with $t \in [0, T]$. We now consider the fact that the process X_t is usually recorded at equally spaced times. When subdividing the interval in n steps of equal length, we use, as shorthand notation, $X_i = X_{iT/n}$, that is $X_0 = x_0$, $X_n = X_T$. Moreover, we set $t_i = iT/n$. Assumption 3.1 implies that X_t is a continuous semimartingale, thus we can define its local time (Revuz and Yor, 1998) as:

$$L_t(x) = \lim_{\varepsilon \to 0} \frac{1}{2\varepsilon} \int_0^t I_{]x-\varepsilon, x+\varepsilon[}(X_\tau) d\tau \tag{3.5}$$

We can estimate the local time of a diffusion via the following approximation:

$$L_t^n(x) = \frac{1}{nh_n} \sum_{i=1}^{[nt]} K\left(\frac{X_i - x}{h_n}\right) \tag{3.6}$$

where $[x]$ is the integer part of x. We have indeed:

Proposition 3.4 *If, as $n \to \infty$ we have $nh_n^4 \to 0$, then $L_t^n(x) \to L_t(x)$ in the $\mathcal{L}^2$ sense. The convergence is almost sure if $\frac{\log n}{nh_n^2} \to 0$.*

Proof. These are Propositions 1 and 2 in Florens-Zmirou (1993). □

The estimator of the diffusion coefficient proposed in Florens-Zmirou (1993); Stanton (1997); Jiang and Knight (1997) is based on the following quantity:

$$V_t^n(x) = \frac{1}{Th_n} \sum_{i=0}^{n-1} K\left(\frac{X_i - x}{h_n}\right) (X_{i+1} - X_i)^2 \tag{3.7}$$

We have indeed,

Proposition 3.5 *If $nh_n^4 \to 0$ as $n \to \infty$, then $V_t^n(x)$ converges to $\sigma^2(x) L_t(x)$ in the $\mathcal{L}^2$ sense.*

Proof. This is Proposition 3 in Florens-Zmirou (1993). □

Thus dividing $V_t^n(x)$ by $L_t^n(x)$ we get a consistent estimator of $\sigma^2(x)$. In this paper, relying on the result of Section 2, we want to substitute the quantity (3.7) with the following:

$$U_t^n(x) = \frac{1}{Tnh_n} \sum_{i=0}^{n-1} K\left(\frac{X_i - x}{h_n}\right) \hat{\sigma}^2(t_i) \tag{3.8}$$

where $\hat{\sigma}^2(t_i)$ is computed via (2.4) on the observed trajectory of X_t.

We then define the estimator:

$$S^n(x) = \frac{\sum_{i=1}^{n} K\left(\frac{X_i - x}{h_n}\right) \hat{\sigma}^2(t_i)}{T \sum_{i=1}^{n} K\left(\frac{X_i - x}{h_n}\right)} \tag{3.9}$$

We now prove that $S^n(x)$ is a consistent estimator of $\sigma^2(x)$.

Theorem 3.6 *If $nh_n^4 \to 0$ as $n \to \infty$, then $S^n(x)$ is a consistent estimator of $\sigma^2(x)$ in the $\mathcal{L}^2$ sense.*

Proof. Suppose $\mu(x) = 0$ in (2.1). For every $\omega \in \Omega$, consider the solution $X_t(\omega)$ of equation (2.1), and define the process Y_t^ω defined by:

$$dY_t^\omega = \sigma(X_t(\omega))dW_t' \tag{3.10}$$

where W_t' is a standard Brownian in an auxiliary probability space $(\Omega', (\mathcal{F}_t')_{0\le t\le T}, P')$, and $\sigma(X_t(\omega))$ is the realization of $\sigma(X_t)$. Assumption (3.1) guarantees the existence of the solution of (3.10). Then we can construct random variables in Ω by taking expectations in Ω'. We denote by E' the expected value in Ω' and by $\mathbb{E}$ the expected value in Ω.

Using equation (2.17), we get for $0 \le t < s \le T$ and almost surely in Ω:

$$\begin{aligned} \frac{1}{s-t}E'[(Y_t^\omega - Y_s^\omega)^2] &= \frac{1}{s-t}\int_t^s \sigma^2(X_u(\omega))du = \\ &= \frac{1}{s-t}\int_t^s \sum_{q=-\infty}^{+\infty} A_q(\sigma^2)e^{iqu}du = \\ &= \sum_{q=-\infty}^{+\infty} A_q(\sigma^2)e^{iqt}\frac{e^{iq(s-t)}-1}{iq(s-t)} = \\ &= \sigma^2(X_t(\omega)) + \sum_{q=-\infty}^{+\infty} A_q(\sigma^2)e^{iqt}\left(\frac{e^{iq(s-t)}-1}{iq(s-t)} - 1\right) = \\ &= \sigma^2(X_t(\omega)) + F(s,t) \end{aligned} \tag{3.11}$$

since the Fourier series can be integrated term by term. Moreover, since the integral of a Fourier series converges uniformly, we have almost surely in Ω:

$$\lim_{s\to t} F(s,t) = 0 \tag{3.12}$$

Now it is simple to prove that, almost surely:

$$\mathbb{E}\left[E'[(Y_t^\omega - Y_s^\omega)^2]|\mathcal{F}_s\right] = \mathbb{E}[(X_t - X_s)^2|\mathcal{F}_s], \tag{3.13}$$

where X is the solution of (2.1), since both are equal to $\int_s^t \mathbb{E}[\sigma^2(X_u)|\mathcal{F}_s]du$. Moreover, using $\mathbb{E}[(X_t - X_s)^4|\mathcal{F}_s] = 3\mathbb{E}^2[\int_s^t \sigma^2(X_u)du|\mathcal{F}_s]$ and Cauchy-Schwartz inequality, we get almost surely:

$$\mathbb{E}\left[E'^2[(Y_t^\omega - Y_s^\omega)^2]|\mathcal{F}_s\right] \le \mathbb{E}[(X_t - X_s)^4|\mathcal{F}_s], \tag{3.14}$$

Now, let us denote the $\mathcal{L}^2(\Omega)$ norm of X by $\|X\|^2 = \mathbb{E}[X^2]$. We then use almost sure identity (3.11) and get:

$$\begin{aligned} \left\|U_t^n(x) - \sigma^2(x)L_t^n(x)\right\| = \Bigg\| &\frac{1}{Th_n}\sum_{i=1}^{[nt]-1} K\left(\frac{X_i - x}{h_n}\right)E'\left[\left(Y_{i+1}^\omega - Y_i^\omega\right)^2\right] + \\ &-\sigma^2(x)L_t^n(x) - \frac{1}{Tnh_n}\sum_{i=1}^{[nt]-1} K\left(\frac{X_i - x}{h_n}\right)F(t_i, t_{i+1})\Bigg\| \end{aligned} \tag{3.15}$$

When expanding the square in the norm (3.15), we can use the fact that $\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X|\mathcal{F}_{t_i}]]$, then use (3.13) and (3.14), to get:

$$\begin{aligned} &\left\| U^n(x) - \sigma^2(x) L^n(x) \right\| \leq \\ &\leq \left\| V^n(x) - \sigma^2(x) L^n(x) - \frac{1}{Tnh_n} \sum_{i=1}^{[nt]-1} K\left(\frac{X_i - x}{h_n}\right) F(t_i, t_{i+1}) \right\| \leq \\ &\leq \left\| V^n(x) - \sigma^2(x) L_t^n(x) \right\| + \left\| \frac{1}{Tnh_n} \sum_{i=1}^{[nt]-1} K\left(\frac{X_i - x}{h_n}\right) F(t_i, t_{i+1}) \right\| . \end{aligned} \tag{3.16}$$

Both terms converge to zero: the first, because of Proposition 3.5; the second given Proposition 3.4 and since $F(t_i, t_{i+1}) = o\left(\frac{1}{n}\right)$.

If $\mu(x) \neq 0$, then (3.13) becomes:

$$\mathbb{E}\left[E'[(Y_t^\omega - Y_s^\omega)^2]\right] = \mathbb{E}[(X_t - X_s)^2] - \mathbb{E}\left[\left(\int_s^t \mu(X_u) du\right)^2\right], \tag{3.17}$$

and the second term of the r.h.s vanishes as $s \to t$ i.e. $n \to \infty$. □

We want now to assess the asymptotic normality of $S^n(x)$. We first state two lemmas.

Lemma 3.7 *If $nh_n^3 \to 0$ when $n \to \infty$, then,*

$$\sum_{i=1}^{[nt]} \mathbb{E}\left[\frac{n}{h_n} \left(K\left(\frac{X_i - x}{h_n}\right) [(X_{i+1} - X_i)^2 - \sigma^2(x)/n] \right)^2 \middle| \mathcal{F}_{i-1} \right] \to \sigma^4(x) L_t(x) \tag{3.18}$$

where the above convergence is in probability.

Proof. This is Lemma 2(*b*) in Florens-Zmirou (1993). □

Lemma 3.8 *Let $g(x) : \mathbb{R} \to \mathbb{R}$ be a continuously differentiable bounded function. Let $nh_n^4 \to 0$ when $n \to \infty$. Consider:*

$$G_t(x) = \frac{1}{\sqrt{nh_n}} \sum_{i=1}^{[nt]} K\left(\frac{X_i - x}{h_n}\right) [g(X_i) - g(x)] \tag{3.19}$$

then, as $n \to \infty$, $G_t(x) \to 0$ in the $\mathcal{L}^1$ sense, and thus in probability.

Proof. We have:

$$\mathbb{E}[|G_t(x)|] \leq \mathbb{E}\left[\frac{1}{\sqrt{nh_n}} \sum_{i=1}^{[nt]} K\left(\frac{X_i - x}{h_n}\right) |g(X_i) - g(x)| \right] \tag{3.20}$$

We now divide the sum in terms such that $|X_i - x| \leq \frac{1}{n}$ and their complementary. Then:

$$
\begin{aligned}
\mathbb{E}[|G(x)|] &\leq \mathbb{E}\left[\frac{1}{\sqrt{nh_n}} \sup_{|X_i - x| \leq \frac{1}{n}} |g(X_i) - g(x)| \sum_{|X_i - x| \leq \frac{1}{n}} K\left(\frac{X_i - x}{h_n}\right) + \right. \\
&\left. + \frac{1}{\sqrt{nh_n}} \sum_{|X_i - x| > \frac{1}{n}} K\left(\frac{X_i - x}{h_n}\right) |g(X_i) - g(x)| \right] \leq \\
&\leq \mathbb{E}\left[\sqrt{nh_n} \sup_{|X_i - x| \leq \frac{1}{n}} |g(X_i) - g(x)| \sum_{i=1}^{[nt]} \frac{1}{nh_n} K\left(\frac{X_i - x}{h_n}\right) + \right. \\
&\left. + \frac{1}{\sqrt{nh_n}} \sup_{|X_i - x| > \frac{1}{n}} |g(X_i) - g(x)| \sum_{|X_i - x| > \frac{1}{n}} K\left(\frac{X_i - x}{h_n}\right) \right]
\end{aligned}
\tag{3.21}
$$

Now, using Taylor's rule we get that $\sup_{|X_i - x| \leq \frac{1}{n}} |g(X_i) - g(x)| = o(\frac{1}{n})$. Then, using Proposition 3.4, we have that the first term goes to zero as $n \to \infty$. The second term goes to zero given the boundedness of g and the fact that the kernel goes to zero when its argument goes to infinity faster than inverse polynomials. □

We finally state the main result of this paper. The idea is still to substitute $n(X_{i+1} - X_i)^2$ with $\hat{\sigma}^2(t_i)$ in (3.18), with the remainder vanishing in probability.

Theorem 3.9 *If* $nh_n^3 \to 0$ *then*

$$
\sqrt{nh_n}\left(\frac{S_n(x)}{\sigma^2(x)} - 1\right) \xrightarrow{\mathcal{L}} \frac{1}{\sqrt{L_T(x)}} \mathcal{N}(0, 1), \tag{3.22}
$$

where the above convergence is in law, and $\mathcal{N}(0, 1)$ *is a standard Normal variable.*

Proof. Consider the discrete filtration $\mathcal{F}_i = \mathcal{F}_{t_i}$, $i = 0, \ldots, n$. Define the following:

$$
\Theta_{i+1} = \sqrt{\frac{1}{nh_n}} K\left(\frac{X_i - x}{h_n}\right) \left[\sigma^2(X_i) - \sigma^2(x)\right] \tag{3.23}
$$

Since $\sigma^2(\cdot)$ is bounded, Θ_i is bounded, and it is adapted to $\mathcal{F}_{i-1}$ (it is actually a constant with respect to $\mathcal{F}_{i-1}$). We want now to verify the conditions of Theorem B.6. In the limit $n \to \infty$, we have the following:

i) $\sum_{i=0}^{[nt]-1} \mathbb{E}[\Theta_{i+1} | \mathcal{F}_i] = \sum_{i=0}^{[nt]-1} \Theta_{i+1}$ which tends to 0 in probability, given Lemma 3.8.

ii) We have to prove that $\sum_{i=0}^{[nt]-1} \mathbb{E}[\Theta_{i+1}^2|\mathcal{F}_i] \to 0$. We can use formula (3.11) and use the same reasoning leading to the proof of Theorem 3.6 getting:

$$\sum_{i=0}^{[nt]-1} \mathbb{E}\left[\Theta_{i+1}^2|\mathcal{F}_i\right] = \\ \sum_{i=0}^{[nt]-1} \mathbb{E}\left[\frac{1}{Tnh_n}\left\{K\left(\frac{X_i - x}{h_n}\right)\left[nE'\left[(Y_{i+1}^\omega - Y_i^\omega)^2\right] - F(t_i, t_{i+1}) - \sigma^2(x)\right]\right\}^2 \middle| \mathcal{F}_i\right] \le \\ \le \sum_{i=0}^{[nt]-1} \mathbb{E}\left[\frac{1}{Tnh_n}\left\{K\left(\frac{X_i - x}{h_n}\right)\left[n(X_{i+1} - X_i)^2 - F(t_i, t_{i+1}) - \sigma^2(x)\right]\right\}^2 \middle| \mathcal{F}_i\right] \tag{3.24}$$

From this inequality and Lemma 3.7 we get the result.

iii) We have to prove conditional Lindeberg condition. We have:

$$\mathbb{E}\left[\left|\mathbb{E}\left[\Theta_i^2|\mathcal{F}_{i-1}\right]\right|\right] = \sum_{|\Theta_i|>\varepsilon} \Theta_i^2. \tag{3.25}$$

Now, the sum (3.25) is bounded by $\sigma^4(x)L_t(x)$; moreover, we can rewrite $|\Theta_i| > \varepsilon$ as:

$$K\left(\frac{X_i - x}{h_n}\right)\left|\sigma^2(X_i) - \sigma^2(x)\right| > \varepsilon\sqrt{nh_n} \tag{3.26}$$

The l.h.s of equation (3.26) is bounded, thus as $n \to \infty$ we have $\varepsilon\sqrt{nh_n} \to \infty$ and the sum (3.25) vanishes in probability.

Thus, we fulfill the assumptions of Theorem B.6, then if we define $Y_t^n(x)$ as:

$$Y_t^n(x) = \sum_{i=0}^{[nt]-1} \Theta_{i+1}(x) \tag{3.27}$$

than we have that $Y_t^n(x)$ converges in law to the continuous martingale M_t with quadratic variation $[M, M]_t = \sigma^4(x)L_t(x)$. We then set $M_t = B(\sigma^4(x)L_t(x))$, where $B(t)$ is a standard Brownian motion. Now consider:

$$Z_t^n(x) = \sum_{i=0}^{[nt]-1} \left(W_{t_{i+1}} - W_{t_i}\right) \tag{3.28}$$

where W_t is the standard Brownian motion in (2.1). It is clear that $Z_t^n(x)$ converges in law to the standard Brownian motion W_t. Moreover we have:

$$\sum_{i=0}^{[nt]-1} \mathbb{E}\left[\Theta_{i+1}\left(W_{t_{i+1}} - W_{t_i}\right)|\mathcal{F}_i\right] = 0 \tag{3.29}$$

By equation (3.29), we get that M_t and W_t are orthogonal. We can also write $B(t) = M_{T(t)}$, where $T(t) = \inf_s\left(\frac{s}{\sigma^4(x)L_s(x)}\right)$. Then, by Knight's Theorem C.1 we get that $B(t)$ and W_t are independent

Brownian motions. Then $B(t)$ and $L_t(x)$ are independent, since the filtration generated by X_t is included in the filtration generated by W_t. We then have that $Y_t^n(x) \to \sqrt{L_t(x)}\sigma^2(x)\mathcal{N}(0,1)$, where $\mathcal{N}(0,1)$ is a standard normal random variable independent of $L(x)$. Since $L_t^n(x)$ converges in probability to $L_t(x)$, we have the desired result. □

The above result can be easily generalized to a multivariate framework, and, more importantly, if the observations X_i are not equally spaced, but if they are such that, as $n \to \infty$, then we have $sup|t_i - t_{i-1}| \to 0$.

We tested the estimator on simulations of the Vasicek (1977) model:

$$dr_t = k(\alpha - r_t)dt + \sigma dW_t, \tag{3.30}$$

which displays mean reversion and constant variance. We simulate the model with parameters resembling actual interest rates distribution. For example, the annualized variance of 3-months T-bill is around 3%, so we keep this value throughout all our simulations. From previous literature, it is well known that two parameters play a crucial role: the choice of the bandwidth parameter h_s in (3.4), and the mean reversion parameter k. The choice of the bandwidth parameter has been long debated in the literature. While consistency of nonparametric estimators is independent of h_s, convergence rates and small sample properties depend crucially on it. Scott (1992) suggests the choice of $h_s = 1.06$, as it is optimal with respect to the mean integrated standard error criterion. However, typical choices in the literature were larger than this value. For example, Stanton (1997) uses $h_s = 4$, while Ait-Sahalia (1996a) uses $h_s = 5$.

The role of mean-reversion is even more debated. The strength of mean reversion is measured by the parameter k. Estimates of k in the literature range around 0.1, which is a very low value. Thus many studies argue that in the available interest rate data there is no mean reversion at all, if not for extreme values (Ait-Sahalia, 1996a; Bandi, 2002). Chapman and Pearson (2000) show, via Monte Carlo evidence, that nonparametric methods could be biased toward finding non-linearities in the drift even if the drift is linear. Jones (2003) concludes in a similar way, and he argues that mean reversion in available interest rate data is so weak, if present, that its detection is very difficult with any statistical method.

We then select a grid of (h_s, k) ranging across the values of interest. We select $h_s = 1.06, 3, 5$ and $k = 0.05, 0.5, 5.0$. In all the replications of the model (3.30), we draw the starting value from the marginal distribution, and we use sample of $n = 8,000$ observations, for comparison with the data sets in the next Section. For every value of h_s and k, we use $5,000$ replications.

Figure 1 shows the average measurements on Monte Carlo replications, together with estimated confidence intervals, obtained with the estimator (3.9). Simulations show that the Fourier estimator is unbiased in small samples for the selected parameters. As expected, confidence intervals are broader for smaller mean reversion and smaller bandwidth parameter, and broader for large and small interest rates, which are less frequent.

4 Data Analysis

In this Section, we turn to the analysis of interest rate data. Our aim is to estimate the diffusion coefficient of the univariate model:

$$dr_t = \mu(r_t)dt + \sigma(r_t)dW_t \tag{4.1}$$

when discretely observing the short rate r_t in a time interval $[0, T]$. Since the spot rate is inherently unobservable, we use proxies for it, typically the three-months rate as it is common in the literature, see Duffee (1996); Chapman et al. (1999). Alternatively, one can regard the model (4.1) as a model for the three-months rate itself.

We first test the methodology on the same data set in Jiang (1998), that is the daily time series of the annualized yields of the three-months U.S. Treasury Bill, from January 1962 to January 1996, for a total of $8,503$ observations. The minimum and the maximum of the yield in this time span are 2.61 % and 17.14 %, thus the estimates of the diffusion coefficient outside of this interval are an artifact of the nonparametric estimation procedure. Figure 2 shows the estimation results. The estimate obtained with the Fourier estimator (3.9), using as bandwith parameter (3.4) with $h_s = 4$, is the solid line. Confidence intervals are computed according to (3.22), using estimated local times via equation (3.6). The Fourier estimator is implemented with the maximal $N = n/2$ and with $M = n/4$. The Fourier estimator confirms the departure from standard parametric models, such as the Vasicek variance $\sigma^2(r) = k$, or the CIR variance $\sigma^2(r) = kr$. In order to better clarify this point, we consider the parametric model of Chan et al. (1992) which nests many popular one-factor models including CIR and Vasicek:

$$dr_t = \mu(\alpha - r_t)dt + \sigma r_t^{\gamma} dW_t. \tag{4.2}$$

This model has been estimated in Jiang (1998) on the same data set using indirect inference. Parameter estimates are $\alpha = 0.079(0.044)$, $\mu = 0.093(0.100)$, $\gamma = 1.474(0.008)$, $\sigma = 0.794(0.019)$, where standard errors are in brackets. Figure 2 shows the function σr^{γ} for comparison with the nonparametric estimate. While the shape of the two estimates is increasing in both case, and the two estimates are compatible for r around 8%, we get a significantly higher estimate at low interest rates, and a significantly lower estimate at high interest rates when using the nonparametric method. It is clear that we are exploiting the flexibility provided by the nonparametric methodology.

We then compare the estimates with those obtained with other nonparametric estimators. To this purpose, we use the same data set used in Stanton (1997), that is the daily time series of the annualized yields of the three-months U.S. Treasury Bill, from January 1965 to July 1995, for a total of $n = 7,975$ observations (minimum 2.61 %, maximum 17.15 %). Thus, we can directly compare our results with those obtained in Stanton (1997). From this perspective, we use the same value $h_s = 4$ used by Stanton. Figure 3 shows that the two estimates are quite different. With respect to the estimate obtained in Stanton (1997), the Fourier estimate coincides only in the central part of the distribution, i.e. $r \simeq 11\%$, while it is higher for smaller values of r and considerably smaller for larger values of r. For larger values of r confidence intervals are wider, since the local time is small for the paucity of observations in that zone. For further comparison purposes, we also estimate the variance with the nonparametric estimator proposed in Bandi and Phillips (2003), with the same bandwidth parameter $h_s = 4.0$ and $\varepsilon_s = 1.5\%$ in (1.5). The result obtained with the estimator proposed in Bandi and Phillips (2003) is almost identical to that obtained with the Stanton estimator, confirming that the empirical performance of the two estimators is nearly the same. We do not report the estimate obtained with the Ait-Sahalia method, since it is very different from those obtained here, and it is very unstable at the level of mean reversion displayed by three-months interest rates, see Renò et al. (2004).

Finally, we estimate the diffusion coefficient on the full data set at our disposal, that is the daily yields on the three-months Treasury Bill from from 4 February 1960 to 11 December 2003, for a total of 10,944 observation (minimum 0.79% on 19 June 2003, maximum 17.14% on 11 December 1980) and the daily yields on the ten-years Treasury Note from 2 February 1962 to 11 December 2003, for a total of 10,447 observations (minimum 3.10% on 13 June 2003, maximum 15.51% on 4 September 1981). Figure 4 shows the results which are in line with the previous findings. We also find that the volatility of the longer maturity contract is less than the shorter one, as it is well known. We leave the reader the judgment on the opportunity for using such a long data set when estimating the model, given the heterogeneity of the economic conditions which drove the interest rate evolution: it is quite clear that the answer to this question depends on the specific application.

5 Conclusions

In this paper a new nonparametric estimator for the diffusion coefficient based on discrete observations is introduced. This estimator is based on a result on volatility estimation derived in Malliavin and Mancino (2002). The estimator is proved to be asymptotically consistent and normally distributed, and asymptotic confidence intervals are provided.

This nonparametric estimators, as well others, can be used in a variety of applications. We used it to compute the diffusion coefficient for daily time series of short interest rates. Our results are in line with those in the literature, but with some peculiarity. We show that our nonparametric estimates is quite different from standard parametric specifications. Moreover, the estimate with the Fourier estimator provides larger variances for interest rates smaller than 9% and smaller variances for interest rates larger than 12% than the variances obtained on the same data set by Stanton (1997). We conclude that the estimator proposed here can be a very useful tool for the issue of correctly specifying the diffusion term of stochastic models for interest rates.

References

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

Ait-Sahalia, Y. (1996b). Testing continuous time models of the spot interest rate. *Review of Financial Studies 9*(2), 385–426.

Ball, C. and W. Torous (1996). Unit roots and the estimation of interest rate dynamics. *Journal of Empirical Finance 3*, 215–238.

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

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

Barucci, E. and R. Renò (2002a). On measuring volatility and the GARCH forecasting performance. *Journal of International Financial Markets, Institutions and Money 12*, 183–200.

Barucci, E. and R. Renò (2002b). On measuring volatility of diffusion processes with high frequency data. *Economics Letters 74*, 371–378.

Chan, K., A. Karolyi, F. Longstaff, and A. Sanders (1992). An empirical comparison of alternative models of the short-term interest rate. *Journal of Finance 47*(3), 1209–1227.

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.

Cox, J., J. Ingersoll, and S. Ross (1985). A theory of the term structure of interest rates. *Econometrica 53*, 385–406.

Dai, Q. and K. Singleton (2000). Specification analysis of affine term-structure models. *Journal of Finance 55*(5), 1943–1978.

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

Duffie, D., J. Pan, and K. Singleton (2002). Transform analysis and asset pricing for affine junp-diffusions. *Econometrica 68*, 1343–1376.

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

Föllmer, H. (1979). Calcul d'Ito sans probabilites. *Séminaire de Probabilités XV.*

Gallant, R. and G. Tauchen (1996). Which moments to match? *Econometric Theory*, 657–681.

Hoffmann, M. (1999). L_p estimation of the diffusion coefficient. *Bernoulli 5*(3), 447–481.

Jacod, J. (1999). Non-parametric kernel estimation of the coefficient of a diffusion. *Scandinavian Journal of Statistics 27*, 83–96.

Jacod, J. and A. N. Shiryaev (1987). *Limit Theorems for Stochastic Processes.* Springer.

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. J. (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.

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.

Knight, F. B. (1971). A reduction of continuous square integrable martingale to brownian motion. In *Lecture Notes in Mathematics*, Volume 190, pp. 19–31. Springer-Verlag.

Malliavin, P. and M. Mancino (2002). Fourier series method for measurement of multivariate volatilities. *Finance & Stochastics 6*(1), 49–61.

Priestley, M. (1979). *Spectral Time Series Analysis.* Wiley.

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. Working Paper.

Revuz, D. and M. Yor (1998). *Continuous Martingales and Brownian Motion.* Springer-Verlag.

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.

Vasicek, O. (1977). An equilibrium characterization of the term structure. *Journal of Financial Economics 5*, 177–188.

A Proof of Theorem 2.1

Theorem 2.1 is a slightly extended version of the main theorem in Malliavin and Mancino (2002). The proof presented here is similar.

Proof of Theorem 2.1. Fix $\omega \in \Omega$ such that the stochastic integral can be defined on ω, which can be done almost surely in Ω (Föllmer, 1979), and define:

$$\hat{\sigma}(t) = \sigma(X_t(\omega)), \quad \hat{\mu}(t) = \mu(X_t(\omega)) \tag{A.1}$$

Then define:

$$\hat{X}_t = \int_0^{2\pi} \hat{\sigma}(t)dW_t + \int_0^{2\pi} \hat{\mu}(t)dt \tag{A.2}$$

with $\hat{X}_0 = x_0$. Consider the mapping from $\varphi : \Omega \to \Omega'$ which associates to ω the trajectory $W_t(\omega)$ in the Wiener space Ω'. It is clear that this mapping preserves probabilities. Moreover we have:

$$X_t(\omega) = \hat{X}_t(\varphi(\omega)), \quad a_k(dX_t(\omega)) = a_k(d\hat{X}_t(\varphi(\omega))), \quad b_k(dX_t(\omega)) = b_k(d\hat{X}_t(\varphi(\omega))) \tag{A.3}$$

Now we prove almost sure convergence in Ω'. Then, suppose that convergence does not hold for a set $A \subset \Omega$ such that $\mathcal{P}(A) > 0$. This would imply that convergence does not hold in the set $\varphi(A) \subset \Omega'$ such that $\mathcal{P}(\phi(A)) > 0$, but we proved almost sure convergence in Ω' then convergence is almost sure in Ω as well.

Denote by $\mathbb{E}$ the expected value in Ω' and consider first the case $\mu(\cdot) = 0$.

We choose $k, h \in \mathbb{N}$ such that $k > h \geq 1$. We have:

$$\begin{aligned}
\mathbb{E}\left[a_k(d\hat{X})a_h(d\hat{X})\right] &= \mathbb{E}\left[\frac{1}{\pi}\int_0^{2\pi} \cos(kt)d\hat{X}_t \cdot \frac{1}{\pi}\int_0^{2\pi} \cos(ht)d\hat{X}_t\right] = \\
&= \mathbb{E}\left[\frac{1}{\pi^2}\int_0^{2\pi} \cos(kt)\hat{\sigma}(t)dW(t) \cdot \int_0^{2\pi} \cos(ht)\hat{\sigma}(t)dW(t)\right] = \\
&= \frac{1}{\pi^2}\int_0^{2\pi} \hat{\sigma}^2(t)\cos(kt)\cos(ht)dt.
\end{aligned} \tag{A.4}$$

by the contraction formula.

Using the following identity:

$$2\cos(kt)\cos(ht) = \cos[(k-h)t] + \cos[(k+h)t] \tag{A.5}$$

we get:

$$\mathbb{E}\left[a_k(d\hat{X})a_h(d\hat{X})\right] = \frac{1}{2\pi}\left[a_{k-h}(\hat{\sigma}^2) + a_{k+h}(\hat{\sigma}^2)\right] \tag{A.6}$$

Moreover we have:

$$\left\|\hat{\sigma}^2\right\|^2_{\mathcal{L}^2} = \sum_{k=0}^{+\infty}\left(a_k^2(\hat{\sigma}^2) + b_k^2(\hat{\sigma}^2)\right) \tag{A.7}$$

Now fix an integer $n_0 > 0$ and define, for $q \in \mathbb{N}$:

$$U_N^q = \frac{1}{N+1-n_0}\sum_{k=n_0}^{N} a_k(d\hat{X})a_{k+q}(d\hat{X}) \tag{A.8}$$

Using (A.6) after taking expectations we get:

$$\mathbb{E}[U_N^q] = \frac{1}{N+1-n_0}\frac{1}{2\pi}\sum_{k=n_0}^{N}\left(a_q(\hat{\sigma}^2) + a_{2k+q}(\hat{\sigma}^2)\right) = \frac{1}{2\pi}a_q(\hat{\sigma}^2) + R_N. \tag{A.9}$$

Where

$$|R_N| = \frac{1}{N+1-n_0}\frac{1}{2\pi}\left|\sum_{k=n_0}^{N} a_{2k+q}(\hat{\sigma}^2)\right| \leq \frac{1}{\sqrt{N+1-n_0}}\left\|\hat{\sigma}^2\right\|_{\mathcal{L}^2} \tag{A.10}$$

by Schwartz inequality, thus

$$a_q(\hat{\sigma}^2) = \lim_{N\to\infty} \mathbb{E}[U_N^q]. \tag{A.11}$$

We want now to proof that $a_q(\hat{\sigma}^2) = \lim_{N\to\infty} U_N^q$. To do so we compute:

$$\mathbb{E}^2[U_N^q] = \frac{1}{(N+1-n_0)^2}\sum_{n_0\leq k_1,k_2\leq N} \mathbb{E}\left[a_{k_1}(d\hat{X})a_{k_1+q}(d\hat{X})\right]\mathbb{E}\left[a_{k_2}(d\hat{X})a_{k_2+q}(d\hat{X})\right] \tag{A.12}$$

Using the fact that $a_k(d\hat{X})$ is a Gaussian random variable with mean 0, we use a well known formula for the product of four Gaussian random variables to compute:

$$\begin{aligned}\mathbb{E}[(U_N^q)^2] &= \frac{1}{(N+1-n_0)^2}\sum_{n_0\leq k_1,k_2\leq N} \mathbb{E}\left[a_{k_1}(d\hat{X})a_{k_1+q}(d\hat{X})a_{k_2}(d\hat{X})a_{k_2+q}(d\hat{X})\right] = \\ &= \frac{1}{(N+1-n_0)^2}\sum_{n_0\leq k_1,k_2\leq N} \Big(\mathbb{E}\left[a_{k_1}(d\hat{X})a_{k_1+q}(d\hat{X})\right]\mathbb{E}\left[a_{k_2}(d\hat{X})a_{k_2+q}(d\hat{X})\right] + \\ &\qquad +\mathbb{E}\left[a_{k_1}(d\hat{X})a_{k_2}(d\hat{X})\right]\mathbb{E}\left[a_{k_1+q}(d\hat{X})a_{k_2+q}(d\hat{X})\right] \\ &\qquad + \mathbb{E}\left[a_{k_1}(d\hat{X})a_{k_2+q}(d\hat{X})\right]\mathbb{E}\left[a_{k_1+q}(d\hat{X})a_{k_2}(d\hat{X})\right]\Big)\end{aligned} \tag{A.13}$$

We now use equation (A.6) to get:

$$\begin{aligned}\mathbb{E}[(U_N^q - \mathbb{E}[U_N^q])^2] &= \frac{1}{4\pi^2(N+1-n_0)^2}\cdot \\ &\cdot \sum_{n_0\leq k_1,k_2\leq N}\left[\left(a_{k_1+k_2}(\hat{\sigma}^2) + a_{|k_1-k_2|}(\hat{\sigma}^2)\right)\left(a_{k_1+k_2+2q}(\hat{\sigma}^2) + a_{|k_1-k_2|}(\hat{\sigma}^2)\right) + \right. \\ &\left. + \left(a_{k_1+k_2+q}(\hat{\sigma}^2) + a_{|k_1-k_2-q|}(\hat{\sigma}^2)\right)\left(a_{k_1+k_2+q}(\hat{\sigma}^2) + a_{|k_1-k_2+q|}(\hat{\sigma}^2)\right)\right]\end{aligned} \tag{A.14}$$

Finally we use Cauchy-Schwartz:

$$\begin{aligned}&\mathbb{E}[(U_N^q - \mathbb{E}[U_N^q])^2] \leq \\ &\leq \frac{1}{4\pi^2(N+1-n_0)^2}\cdot \\ &\cdot\left[\left(\sum_{n_0\leq k_1,k_2\leq N}\left(a_{k_1+k_2}(\hat{\sigma}^2) + a_{|k_1-k_2|}(\hat{\sigma}^2)\right)^2 \cdot \sum_{n_0\leq k_1,k_2\leq N}\left(a_{k_1+k_2+2q}(\hat{\sigma}^2) + a_{|k_1-k_2|}(\hat{\sigma}^2)\right)^2\right)^{\frac{1}{2}} + \right. \\ &\left. + \left(\sum_{n_0\leq k_1,k_2\leq N}\left(a_{k_1+k_2+q}(\hat{\sigma}^2) + a_{|k_1-k_2-q|}(\hat{\sigma}^2)\right)^2 \cdot \sum_{n_0\leq k_1,k_2\leq N}\left(a_{k_1+k_2+q}(\hat{\sigma}^2) + a_{|k_1-k_2+q|}(\hat{\sigma}^2)\right)^2\right)^{\frac{1}{2}}\right] \leq \\ &\leq \frac{2}{\pi^2(N+1-n_0)}\left\|\hat{\sigma}^2\right\|_{\mathcal{L}^2}^2\end{aligned} \tag{A.15}$$

The above inequality proves (2.6), since $\mathcal{L}^2$ convergence to a constant implies almost sure convergence. If we now repeat the calculation (A.4) replacing a_k, a_h with a_k, b_h, we have:

$$\mathbb{E}\left[a_k(d\hat{X})b_h(d\hat{X})\right] = \int_0^{2\pi} \hat{\sigma}^2(t)\cos(kt)\sin(ht)dt. \tag{A.16}$$

We now use the identity:

$$2\cos(kt)\sin(ht) = \sin[|k-h|t] + \sin[(k+h)t] \tag{A.17}$$

and we get:

$$\mathbb{E}\left[a_k(d\hat{X})b_h(d\hat{X})\right] = \frac{1}{2\pi}\left[b_{k-h}(\hat{\sigma}^2) + b_{k+h}(\hat{\sigma}^2)\right] \tag{A.18}$$

We then get formula 2.7 by computing the expected value of:

$$V_N^q = \frac{1}{N+1-n_0}\sum_{k=n_0}^{N} a_k(d\hat{X})b_{k+q}(d\hat{X}),\ W_N^q = \frac{1}{N+1-n_0}\sum_{k=n_0}^{N} b_k(d\hat{X})a_{k+q}(d\hat{X}) \tag{A.19}$$

The second part of formula (2.6) comes in the same way from the identity:

$$2\sin(kt)\sin(ht) = \cos[|k-h|t] - \cos[(k+h)t] \tag{A.20}$$

Formula (2.5) comes in the same way from:

$$\mathbb{E}[a_k^2(d\hat{X})] = \mathbb{E}[b_k^2(d\hat{X})] = \frac{1}{2\pi}\left[2a_0(\hat{\sigma}^2) - a_{2k}(\hat{\sigma}^2)\right] \tag{A.21}$$

If $\mu(\cdot) \neq 0$, then in all previous computation we replace $d\hat{X}$ with dv defined by $dv = d\hat{X} - \hat{\mu}dt$. Now, all the extra terms depending on μ vanish asymptotically since:

$$\int_0^{2\pi} \hat{\mu}^2(t)dt = \sum_{k=0}^{+\infty}\left(a_k^2(\hat{\mu}) + b_k^2(\hat{\mu})\right). \tag{A.22}$$

□

B Limit Theorems

Definition B.1 *A process with independent increments (PII) in a filtered probability space is a cadlag adapted $\mathbb{R}$-valued process X such that $X_0 = 0$ and that $\forall\ 0 \leq s \leq t$ the variable $X_t - X_s$ is independent of $\mathcal{F}_s$.*

Definition B.2 *A truncation function $h(x)$ is a bounded Borel real function with compact support which behaves like x near the origin.*

For every semimartingale X, we define its characteristics (B, C, ν) as follows. Let h be a truncation function. We define $X(h) = X - \sum_{s \leq t}[\Delta X_s - h(\Delta X_s)]$. $X(h)$ is a special semimartingale and we can write its canonical decomposition:

$$X(h) = X_0 + M(h) + B(h) \tag{B.1}$$

where $M(h)$ is a local martingale and $B(h)$ a predictable process of finite variation.

Definition B.3 *The characteristics of X is the triplet (B, C, ν) defined by:*

i) $B = B(h)$ in (B.1)

ii) $C = [X^c, X^c]$ i.e. the quadratic variation of the continuous martingale part of X

iii) ν is the compensator of the random measure associated with the jumps of X.

We then have that B is a predictable process of finite variation, C is a continuous process of finite variation and ν is a predictable measure on $\mathbb{R}^+ \times \mathbb{R}$. Extension to the multivariate case is straightforward. If X is a PII with $X_0 = 0$ and without fixed times of discontinuity, then Levy-Kinthchine formula holds:

$$\mathbb{E}[e^{iuX_t}] = \exp\left(iuB_t - \frac{u^2}{2}C_t + \int_{\mathbb{R}^+} \left(e^{iux} - 1 - iuh(x)\right) \nu_t(dx)\right). \tag{B.2}$$

Then next theorem provides the characteristics of the processes of the following kind:

$$Y_t = \sum_{i=1}^{[nt]} \Theta_i \tag{B.3}$$

where $[x]$ is the integer part of x. In definition (B.3), we assume that there is a discrete filtration $(\mathcal{F}_i)_{i \in \mathbb{N}}$, and we need Θ_i to be adapted to $\mathcal{F}_{i-1}$. We can think of Θ_i as the difference process $X_i - X_{i-1}$ of an adapted discrete process X_i.

Theorem B.4 *Let h be any truncation function. Then the characteristics of Y_t defined as in (B.3) is:*

$$\begin{cases} B_t = \sum_{i=1}^{[nt]} \mathbb{E}[h(\Theta_i)|\mathcal{F}_{i-1}] \\ C_t = 0 \\ g * \nu = \int\int_{[0,T]\times\Omega} g d\nu = \sum_{i=1}^{[nt]} \mathbb{E}[g(\Theta_i) I_{\{\Theta_i \neq 0\}}|\mathcal{F}_{i-1}] \end{cases} \tag{B.4}$$

Proof. This is Theorem II.3.11 in Jacod and Shiryaev (1987). □

If $h^2 * \nu < \infty$, $\forall t \in [0,T]$, we can define the following:

$$\tilde{C}_t = C_t + h^2 * \nu_t - \sum_{s \le t} (\Delta B_s)^2 \tag{B.5}$$

We then have the following convergence theorem:

Theorem B.5 *Fix a truncation function h. Assume that X^n is a sequence of processes, and X is a PII semimartingale without fixed time of discontinuity. Denote by (B^n, C^n, ν^n) the characteristics of X^n and by (B, C, ν) the characteristics of X. Define $\tilde{C}$ by equation (B.5). Moreover assume the following:*

i) $\sup_{s \le t} |B^n_s - B_s| \to 0$ *in probability*, $\forall t \in [0,T]$

ii) $\tilde{C}^n_t \to \tilde{C}_t$ *in probability* , $\forall t \in [0,T]$

iii) $g * \nu^n_t \to g * \nu_t$ *in probability*, $\forall t \in [0,T], g \in \mathcal{C}_1(\mathbb{R})$

Then $X_n \to X$ in distribution.

Proof. This is Theorem VIII.2.17 in Jacod and Shiryaev (1987). □

We then show the following Theorem, which will be useful in our analysis:

Theorem B.6 *Consider the process Y^n_t defined in (B.3) on $(\Omega, (\mathcal{F}_i)_{i\in\mathbb{N}}, P)$ with Θ_i bounded and adapted to $\mathcal{F}_{i-1}$, and assume the following:*

i) $\sum_{i=1}^{[nt]} \mathbb{E}[\Theta_i | \mathcal{F}_{i-1}] \to 0$ *in probability*

ii) $\sum_{i=1}^{[nt]} \mathbb{E}[\Theta_i^2 | \mathcal{F}_{i-1}] \to V_t$ *in probability*

iii) $\forall\, \varepsilon > 0$, $\sum_{i=1}^{[nt]} \mathbb{E}[\Theta_i^2 I_{\{|\Theta_i| > \varepsilon\}} | \mathcal{F}_{i-1}] \to 0$ *in probability (conditional Lindeberg condition)*

Then Y_t converges in distribution to the continuous martingale M_t with quadratic variation $[M,M]_t = V_t$.

Proof. We have to prove conditions $i) - iv)$ of theorem B.5. We compute the characteristics (B^n, C^n, ν^n) of Y^n by theorem B.4, with $h(x) = x \wedge \sup \Theta_i$ and $\tilde{C}$ by (B.5). The characteristics of M_t is $(0, V_t, 0)$.

i) Is implied in *iv*)

ii) We have $B'^n = B^n$. Since Θ_i is bounded, this follows directly.

iii) Comes directly from the definition of C'^n and the fact that $\sum_{s\leq t}(\Delta B_s)^2 = \sum_{i=1}^{[nt]} \mathbb{E}[\Theta_i^2|\mathcal{F}_{i-1}] \to 0$ from 1.

iv) If $g \in \mathcal{C}_1$ there exist real numbers k, K such that $|g(x)| \leq Kx^2 I_{\{|x|>k\}}$, thus the conditional Lindeberg condition implies $g * \nu^n \to 0$.

□

C Knight's Theorem

The following Theorem has been proved in Knight (1971). It allows to transform a vector of orthogonal square-integrable continuous martingales into a vector of independent Brownian motions via a suitable time change.

Theorem C.1 *Let $M_1, \ldots, M_n$ be orthogonal square-integrable martingales, and consider the time changes:*

$$T_i(t) = \begin{cases} \inf_s [B_i, B_i]_s > t & if\ this\ is\ finite \\ +\infty & otherwise \end{cases} \tag{C.1}$$

Then the transformed variables:

$$X_i(t) = \begin{cases} B_i(T_i(t)) & if\ T_i(t) < \infty \\ B_i(\infty) + W_i(t - [B_i, B_i]_\infty) & otherwise \end{cases} \tag{C.2}$$

where $W_1, \ldots W_n$ is an n-dimensional Brownian motion independent of X_i, are an n-dimensional Brownian motion relative to their generated filtration.

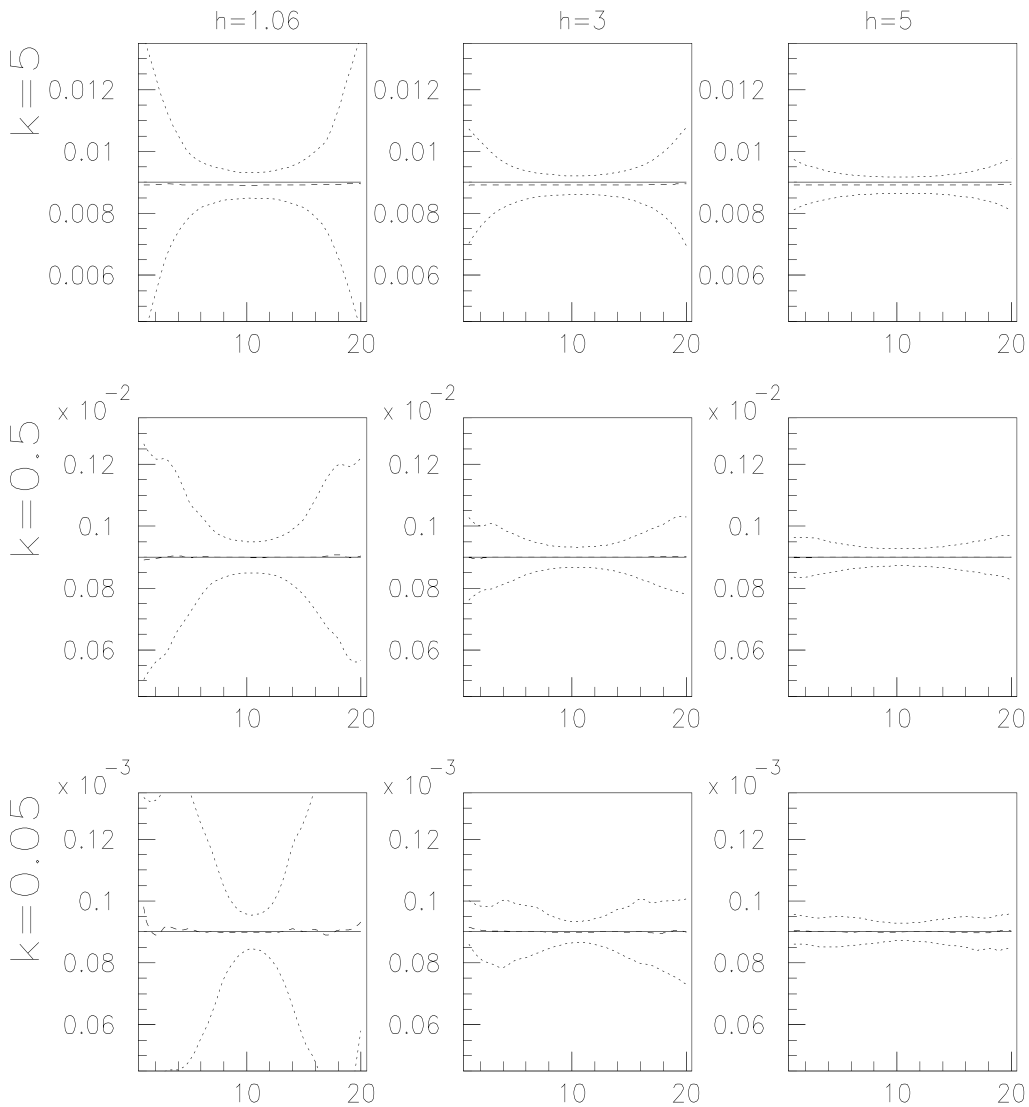

Figure 1: Variance $\hat{\sigma}^2(r)$ obtained with the Fourier estimator (3.9) on $5,000$ replications of $8,000$ observations of the Vasicek model (3.30), with different values of h_s and k, and $\alpha = 10.5\%$, $\sqrt{\sigma^2/2k} = 3\%$. Solid line: the generated variance σ^2. Dashed line: average estimates. Dotted lines: 5% and 95% confidence intervals.

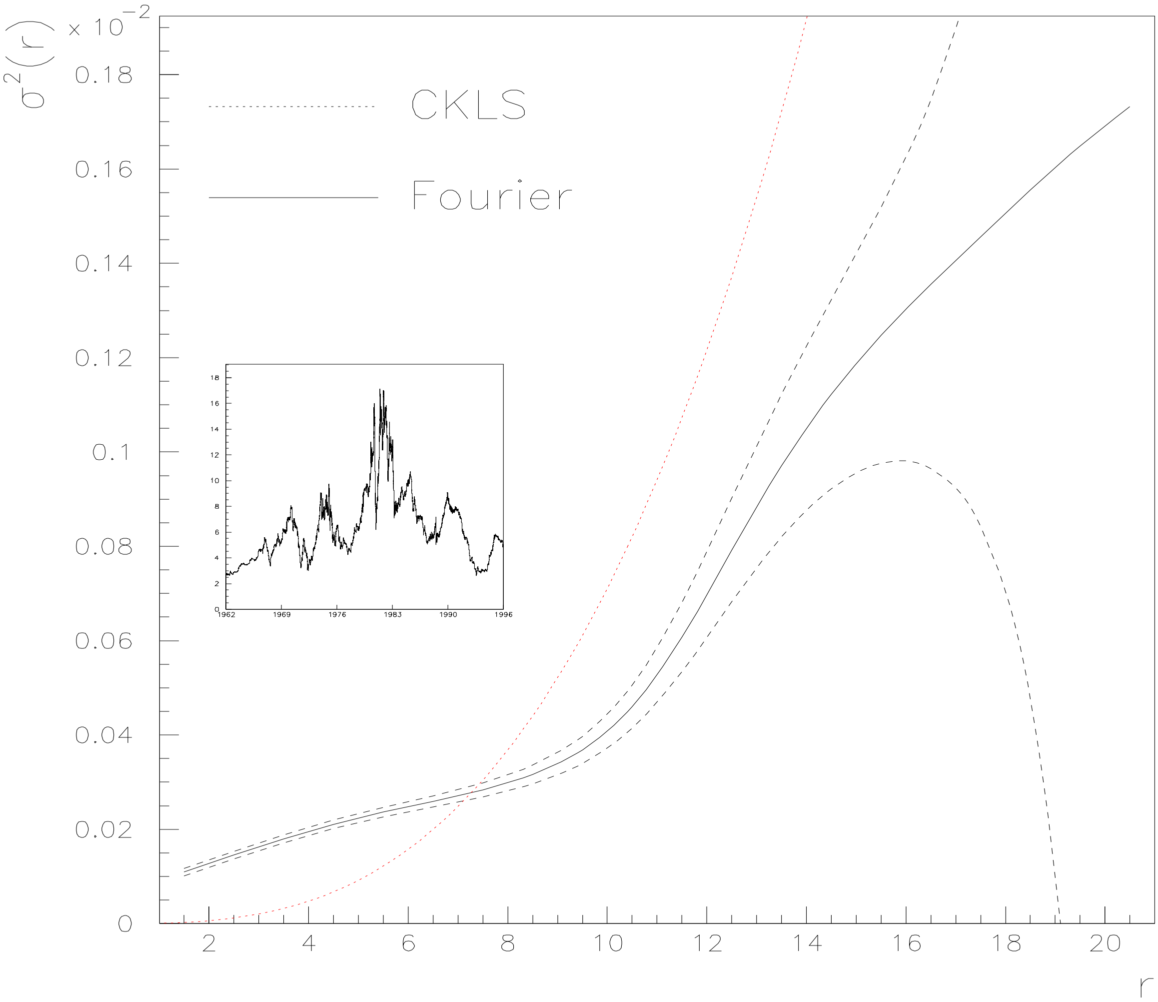

Figure 2: Estimate of the diffusion coefficient $\sigma^2(r)$ on Jiang (1998) data set. Solid line: Fourier estimate. Dashed lines: 5% and 95% confidence intervals. Dotted line: estimate obtained with the parametric model (4.2). In the inset, the time series of the yields on three-months T-bill under study, from January 1966 to January 1996.

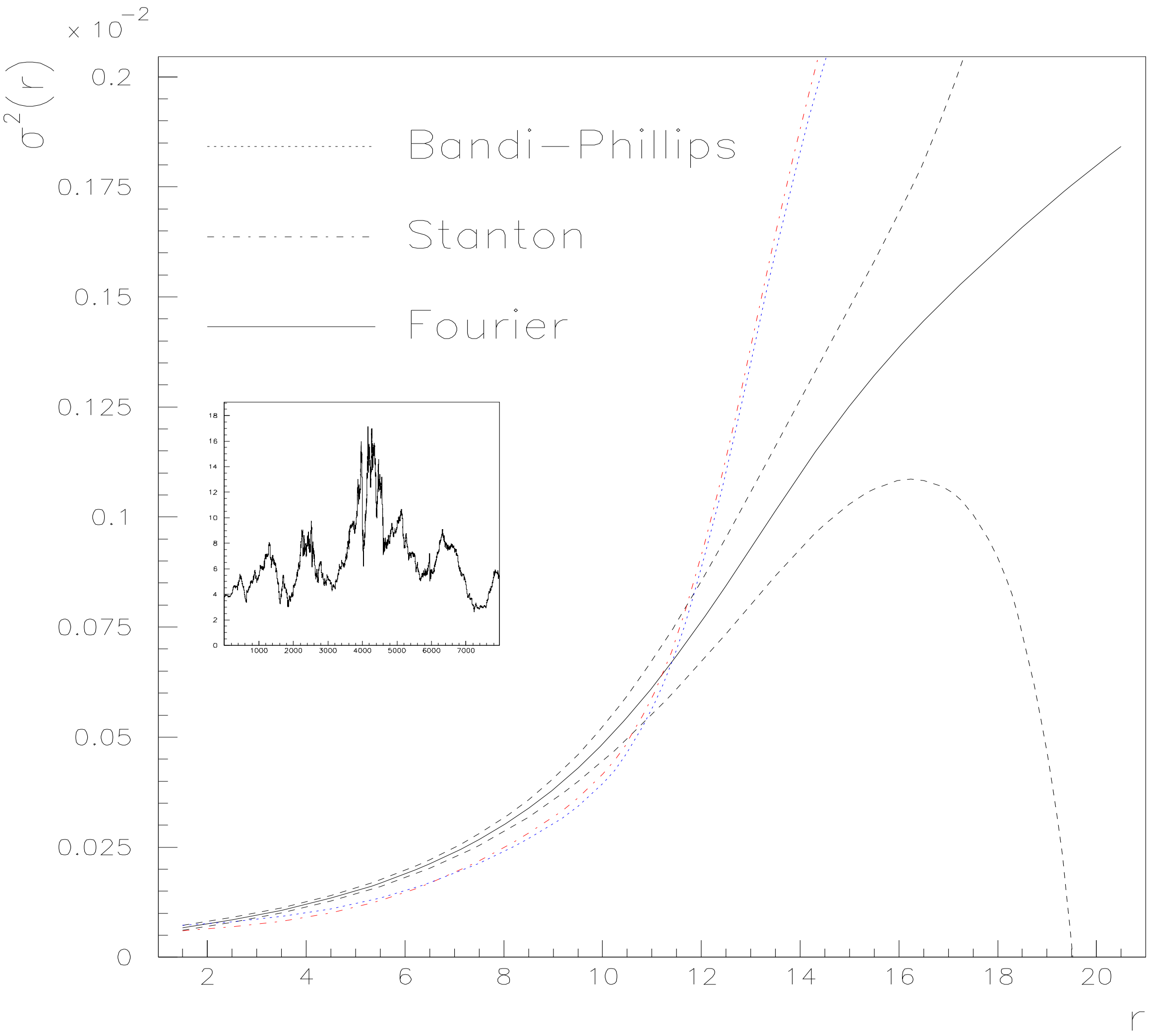

Figure 3: Estimate of the diffusion coefficient $\sigma^2(r)$ on Stanton data set. Solid line: Fourier estimate. Dashed lines: 5% and 95% confidence intervals. Dashed-dotted line: Stanton estimate. Dotted line: estimate obtained with the estimator (1.5) of Bandi and Phillips. In the inset, the time series of the yields on three-months T-bill under study, from January 1965 to July 1995.

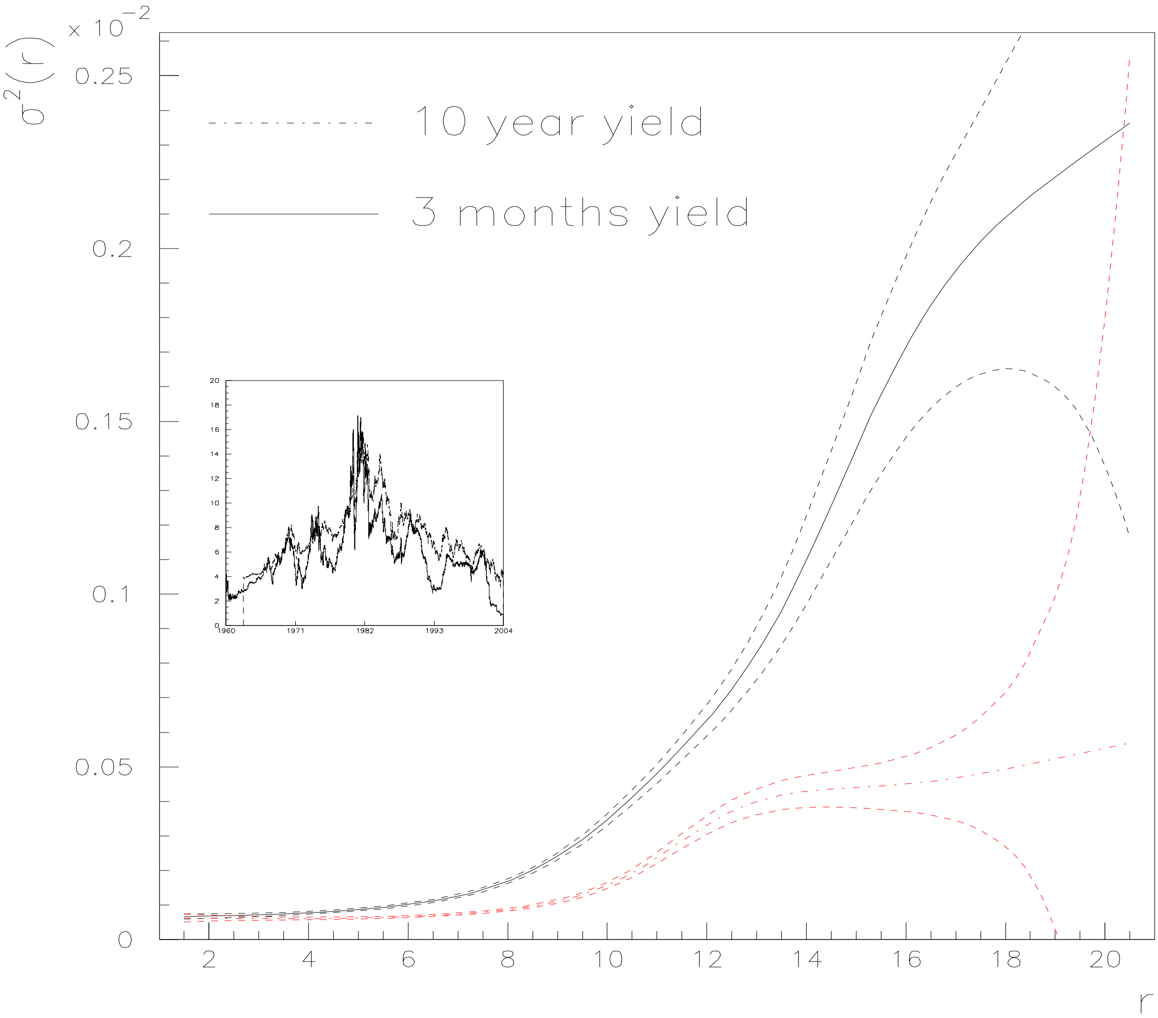

Figure 4: Estimate of the diffusion coefficient $\sigma^2(r)$ on the full 1960-2003 data set. Solid line: diffusion coefficient of the yield of the 3-months Treasury Bill. Dashed-dotted line: diffusion coefficient of the yield of the 10-year Treasury Note (from 1962). Dashed lines: 5% and 95% confidence intervals. In the inset, the time series under study.