QUADERNI

Università degli Studi di Siena
DIPARTIMENTO DI ECONOMIA POLITICA

ROBERTO RENO'
ANTONIO ROMA
STEPHEN SCHAEFER

A Comparison of Alternative Nonparametric Estimators of the Short Rate Diffusion Coefficient

n. 445 - Dicembre 2004

Abstract - In this paper we discuss the estimation of the diffusion coefficient of one-factor models for the short rate via non-parametric methods. We test the estimators proposed by Ait Sahalia (1996a), Stanton (1997) and Bandi and Phillips (2003) on Monte Carlo simulation of the Vasicek and CIR model and show that all estimators, especially that proposed by Ait-Sahalia (1996a), are problematic for values of the mean reversion coefficient typically displayed by interest rate data. Moreover all estimators depend crucially on the choice of the bandwith parameter. Data analysis shows that the estimators lead to different estimates on the data set analyzed by Ait-Sahalia (1996a) and Stanton (1997); moreover we show that the two data set are inherently different.

JEL Classification: C14, E43

Roberto Renò, Dipartimento di Economia Politica, Università degli Studi di Siena

Antonio Roma, Dipartimento di Economia Politica, Università degli Studi di Siena

Stephen Schaefer, London Business School, Londra

1 Introduction

In one-factor models of the term structure the "factor" is invariably taken as the short rate:

$$dr_t = \mu(r_t)dt + \sigma(r_t)dZ. \quad (1)$$

A number of different models for the short rate have been proposed and tested (Chan et al., 1992), which parameterize the drift and diffusion coefficients $\mu(r_t)$, $\sigma(r_t)$ through the specification

$$dr_t = (a + br_t)dt + \sigma r_t^\gamma dz \quad (2)$$

where, for example, $\gamma = 0$ corresponds to the Vasicek (1977) Gaussian model and $\gamma = 1/2$ to the Cox et al. (1985) (CIR) square root model. A number of estimators for this latter process are known (Bianchi et al., 1994).

Recently a number of different non-parametric estimators of the diffusion coefficient have been proposed. Ait-Sahalia (1996a,b) has proposed a non-parametric estimator of the function $\sigma(r)$, based on the a-priori parametric specification of the function $\mu(r)$. Stanton (1997) has proposed an approximate non-parametric estimator for both the functions $\mu(r)$ and $\sigma(r)$. Bandi and Phillips (2003) proposed an estimator very similar to that of Stanton. These estimators are based on the non-parametric estimation of the conditional density of the short rate. Stanton showed that, as the sampling frequency increases, his estimator converges, at a speed depending on the order of the discrete approximation used, to the infinitesimal drift and diffusion functions.

The advantage of the non-parametric specification is clearly its flexibility. When applied to actual interest rate data, these non-parametric estimators produce functions $\mu(r)$ and $\sigma(r)$ that appear non linear in r, and depart from benchmark parametric models such as the CIR square root model. This is evident from Ait-Sahalia (1996a) and Stanton (1997) applications. Chapman and Pearson (2000) used Monte Carlo simulation to investigate the properties of the Stanton estimator for the drift and diffusion function, as well as a drift estimator developed using Ait Sahalia's approach. Adopting the CIR model as the null hypothesis, they conclude that the non-linearity of the drift function $\mu(r)$ estimated through the non-parametric model may not be indicative of a truly non-linear underlying model. They conclude however that the Stanton (1997) diffusion estimator is capable of identifying with some precision the square root form of the non-linearity in the volatility parameter. No analysis is carried out of the Ait Sahalia volatility estimator.

In this paper we examine the Ait Sahalia non-parametric diffusion estimator. We show that, for realistic parameters values, the estimated volatility will be a non-linear function of the interest rate even when the actual volatility is a constant. In contrast, the Stanton (1997) estimator appears to be reasonably accurate. Furthermore we show that the AS estimator is biased in a manner that accounts for much of the non-linear behavior of the estimated variance in the constant variance case. Finally, we show that using actual (positive) interest rate data, the approach may result in negative estimates of variance.

2 The Ait-Sahalia estimator

2.1 Theory

The Ait-Sahalia estimator (AS) is based on the fact that, if a variable $r(t)$ follows a diffusion described by:

$$dr(t) = \mu(r)dt + \sigma(r)dW(t) \tag{3}$$

in which $W(t)$ is a standard Brownian motion and $\mu(r)$ and $\sigma(r)$ are such that a unique solution of the stochastic differential equation (3) exists, then:

$$\sigma^2(r) = \frac{2\int_{-\infty}^{r} \mu(x)\pi(x)dx}{\pi(r)} \tag{4}$$

where $\pi(r)$ is the unconditional distribution of r under the diffusion (3). Given two of the three functions $\mu(r), \sigma(r), \pi(r)$, equation (4) allows the third to be obtained via integration or differentiation.

Ait-Sahalia (1996a) suggests specifying the drift $\mu(r)$ as an affine function of r and then estimating the conditional variance $\sigma^2(r)$. Ait-Sahalia (1996a) also suggests replacing $\pi(r)$ in (4) with an estimate derived using a non-parametric approach (Scott, 1992). Suppose we have T observations of r, denoted by $\hat{r}_i, i = 1, \ldots, T$. Then the non-parametric estimator of $\pi(r)$ is given by:

$$\hat{\pi}(r) = \frac{1}{Th_s}\sum_{i=1}^{T} K\left(\frac{r - \hat{r}_i}{h_s}\right) \tag{5}$$

where K is a kernel function that depends on h_s. A typical choice of the kernel, suggested by Ait-Sahalia (1996a), is the Gaussian kernel,

$$K(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}. \tag{6}$$

For the Gaussian kernel the optimal smoothing parameter is given by

$$h_s = h\hat{\sigma}T^{-\frac{1}{5}}, \tag{7}$$

where $\hat{\sigma}^2 = Var[\hat{r}_i]$ and $h = 1.06$ (Scott, 1992, p.131)[1]. It is widely recognized that the choice of the kernel function is much less important than the choice of an appropriate smoothing parameter (Scott (1992) p.133). Although nominally a "non-parametric" approach, the bandwidth h_s (i.e. the scalar h) is in fact a parameter that has to be appropriately selected.

Under these assumptions, the Ait-Sahalia estimator is given by:

$$\hat{\sigma}^2(r) = \frac{2\sum_{i=1}^{T}\int_{-\infty}^{r}\mu(s)K\left(\frac{s-\hat{r}_i}{h_s}\right)ds}{\sum_{i=1}^{T}K\left(\frac{r-\hat{r}_i}{h_s}\right)} \tag{9}$$

with $K(x)$ given by (6) and h_s by (7).

2.2 Nonparametric estimate of the density of persistent data

While much is known about the asymptotic behavior of non-parametric estimators, their finite sample properties are largely unknown and their robustness, especially with respect to the choice of the bandwidth parameter, h, is suspect. For the AS estimator, we show that estimating the unconditional density of data, the denominator in (4), is particularly critical.

To study the small-sample properties of the AS and other estimators we use a Monte Carlo approach and generate sample paths for the interest rate that follow an Ornstein-Uhlenbeck ('Vasicek') process as in the Vasicek (1977) model:

$$dr(t) = k(\alpha - r(t))dt + \sigma dW(t). \tag{10}$$

For this model the diffusion coefficient is, of course, constant and equal to σ^2. The range of possible interest rates is $(-\infty, \infty)$ and the unconditional density function, $\pi(r)$, is a Normal distribution with mean α and variance $\sigma^2/2k$. Although the Vasicek process might be viewed

[1]The optimal smoothing parameter is chosen with reference to the Mean Integrated Square Error (MISE) of the density estimator, $\hat{f}$, compared to a reference density, f, given by:

$$MISE = \int_t \left(\hat{f}(t) - f(t)\right)^2 dt. \tag{8}$$

Different reference densities imply a different optimal smoothing parameter. See Scott (1992) p. 142 for the conversion ratio of the smoothing parameter under different reference densities.

as an unrealistic model of the short-term rate – because, for example, it admits negative values – it is quite adequate to illustrate our first point that, when the degree of mean reversion is (realistically) low, reliable estimating the unconditional density is very difficult with persistent data.

We replicate 1,000 paths of the Vasicek diffusion model (10)for a range of values for k, the mean reversion coefficient, and N, the number of observations. In each case we set $\alpha = 8.3\%$ and $\sigma = 0.1$. For each sample path we estimate the mean and the standard deviation of $r(t)$ and, as $N \to +\infty$, these should converge to α and $\sigma/\sqrt{2k}$ respectively, i.e. the mean and standard deviation of the unconditional distribution. Table 1 reports the average values of the mean and standard deviation of $r(t)$ for the replications, along with confidence bands and shows that the rate of convergence of the sample estimates can be very slow when mean reversion is low. In particular, the standard deviation of the unconditional distribution is significantly underestimated in small samples. For example with $k = 0.05$ and $\sigma = 0.01$ the standard deviation of the unconditional distribution is 3.16%. But with $N = 10,000$ observations, the estimated standard deviation has a mean of 1.79% and a 95th percentile of 2.781%. The underestimation is even larger at $N = 5,000$, and still appreciable at $N = 50,000$, even with $k = 0.5$. On the other hand, with high mean reversion ($k = 5$) the estimates are more reliable. With $N = 10,000$ observations, for example, and $k = 5$, the 5^{th} and 95^{th} percentiles of the estimate of the standard deviation are 0.289 and 0.339 when the true value is 0.312. Moreover, with a small sample size, even the mean is poorly estimated. In the case of the mean, unlike the standard deviation, the estimate is unbiased but the dispersion of the estimates can be very large. Obviously, since densities are constrained to be integrated to one, if the unconditional standard deviation is systematically underestimated, the estimated nonparametric density will have a peak at the estimated mean (which will be unlikely to be the true mean) and will have thinner tails than the true unconditional distribution. The problem of estimating kernels with persistent data has been also studied by Pritsker (1998), who finds out that "inference based on the asymptotic distribution of kernel density estimators can be very misleading".

These results show how estimates of the unconditional density of $r(t)$ vary with the length of the data series and the degree of mean reversion. Where, in the range we have considered, do actual data fit? First, even with daily data, $5,000$ observations represent around 20 years of data, $10,000$ observations 40 years and $50,000$ observations an entirely unrealistic 200 years. The sample size in Stanton paper is $N = 7,795$, and to our knowledge is the largest ever used. AS has $N = 5,505$ observations. Thus AS's sample size is approximately equal to our "small" sample and Stanton's lies between our "small" and "medium" sized samples. Both samples, however, are “large” in an economic sense, however, in that both rely on model stability parameters for a period of 20-30 years. This is particularly worrisome in the case of the mean parameter, α, and many authors from Brennan and Schwartz onwards have advocated a two (or more) factor process for the short rate in which the mean itself changes stochastically (Brennan and Schwartz, 1980; Dai and Singleton, 2000) This misspecification is a further potential problem for the non-parametric estimator.

The second parameter that strongly affects the reliability of small sample estimates of the unconditional density of $r(t)$ is the degree of mean reversion and Table 2 reports some estimates of the mean reversion parameter, k, from four studies in the literature. Three of these report estimates in the range $0.1-0.2$; the exception is the study by Ait-Sahalia (1996a) who reports a value of 0.978. AS uses a daily series of rates on 7-day Eurodollar futures. These data have some features, discussed below, that cast doubt on the reliability of this high estimate of k

In summary, our conclusion from this part of the analysis is that the likely value of k is in the region of 0.1-0.2 and that, in this case, even with sample sizes of 5,000 and 10,000, reliable estimation of the unconditional density is almost impossible.

2.3 Performance of the estimator

In this section we carry out a Monte-Carlo analysis of the AS estimator (12) when the data is generated by the Vasicek model. Initially we assume that the drift is known, that is, in computing the estimator of $\sigma^2(r)$ we use the values of k and α used to generate the simulated samples. An advantage of using a constant conditional variance is that in this case, with a Gaussian kernel, we can easily find an analytic expression for the estimator of $\sigma^2(r)$ in (9). Indeed, in the case of the Vasicek model (10), it is straightforward to show that the Ait Sahalia estimator has the form:[2]

$$\hat{\sigma}^2(r) = 2kh_s^2 + \frac{h_s\sqrt{2\pi}\sum_{t=1}^{T} k(\alpha - \hat{r}_t)\left[1 + Erf\left(\frac{r-\hat{r}_t}{\sqrt{2}h_s}\right)\right]}{\sum_{t=1}^{T} e^{-\frac{(r-\hat{r}_t)^2}{2h_s^2}}}. \tag{12}$$

From equation (7), as the number of observations increases ($T \to \infty$) and if the second moment of the unconditional distribution is finite, then $h_s \to 0$ and the first term on the right

[2]The function $Erf(x)$ is defined as:

$$Erf(x) = \frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}dt \tag{11}$$

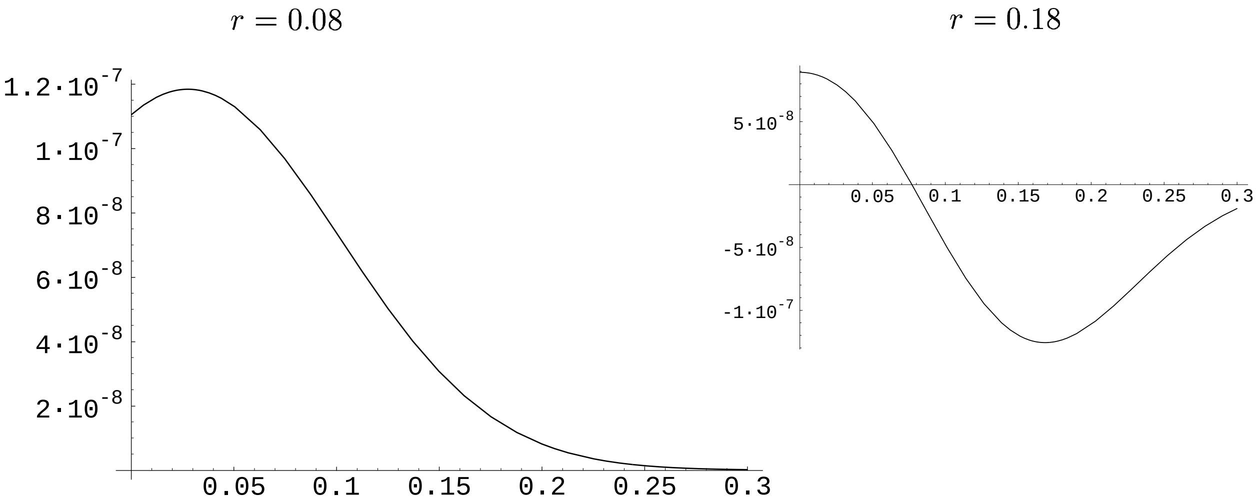

Figure 1: Numerator with Ait-Sahalia parameters conditioned on $r = 8\%$ (left) and $r = 18\%$ (right).

hand side of (12), a constant, goes asymptotically to zero. The asymptotic estimator is then given by:

$$\sigma^2(r) = \lim_{T\to\infty} \frac{h_s\sqrt{2\pi}\sum_{t=1}^{T} k(\alpha - \hat{r}_t)\left[1 + Erf\left(\frac{r-\hat{r}_t}{\sqrt{2}h_s}\right)\right]}{\sum_{t=1}^{T} e^{-\frac{(r-\hat{r}_t)^2}{2h_s^2}}} \tag{13}$$

However, in any finite sample, the conditional variance estimator (12) not only displays sampling error but, in the case of the Vasicek model when the numerator may become negative, is not even guaranteed always to be positive. The two panels of Figure 1 show the behavior of the numerator as a function of $r(t)$ for $h = 4$, conditioning on $r = 0.08$ and $r = 0.18$ respectively, using the FGLS parameter values estimated by Ait-Sahalia (1996a). As the estimated conditional variance will be approximately proportional to the expected value of the numerator under the unconditional distribution of the interest rate $r(t)$, it is clear that the estimated conditional variance may become negative under Ait-Sahalia's method.

Figures 2,3,4 show, respectively, the true density, the numerator in (4) and the true variance $\sigma^2(r)$ together with their estimates computed from a single simulated series of size N. We show the results for values of N between 5,000, i.e., approximately twenty years of daily data, and 5,000,000, i.e., approximately 20,000 years of daily data. By using a single simulation, we study the convergence of the estimator as $N \to \infty$.

In the previous subsection we showed that with low mean reversion the variance of the unconditional distribution is systematically underestimated and mean poorly estimated. These deficiencies are clearly illustrated in Figure 2. For low and average mean reversion we obtain reliable estimates of the density only for $N = 50,000$ or larger. Errors in estimating the marginal density can have a dramatic effect on the estimate of the numerator, as illustrated in Figure 3, again, especially with low or average mean reversion. This, in turn, affects the variance estimate which is very poor for small sample sizes (Figure 4).

The particular case for $k = 0.05$, $N = 5,000$, at the bottom-left corner of Figures 2,3,4 is worth noting. Here, in the center of the distribution, the density is overestimated (Figure xx) and the numerator underestimated. These two effects combine to produce substantial underestimation of the variance. Asymptotically, the estimates are quite good but, for low mean reversion ($k = 0.05$), only for values of N larger than around $500,000$. In contrast, for high mean reversion, which unfortunately is not a feature of the data, we obtain reasonable estimates of the density, numerator and variance even in small samples.

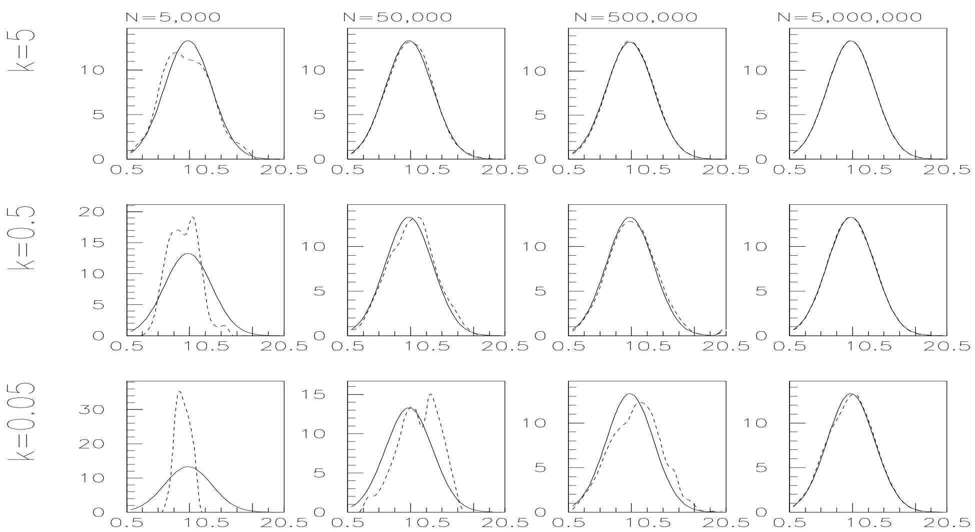

Figure 2: Non-parametric estimates of the density (dashed line) on a single simulation of the Vasicek model (10) of length N, for different values of N, as displayed, with $\alpha = 8.3\%$, $\sigma = 0.1$, $h = 1.06$ and different levels of mean reversion:$k = 5$ (top row), $k = 0.5$ (central row) and $k = 0.05$ (bottom row). For comparison, the true density is shown (solid line).

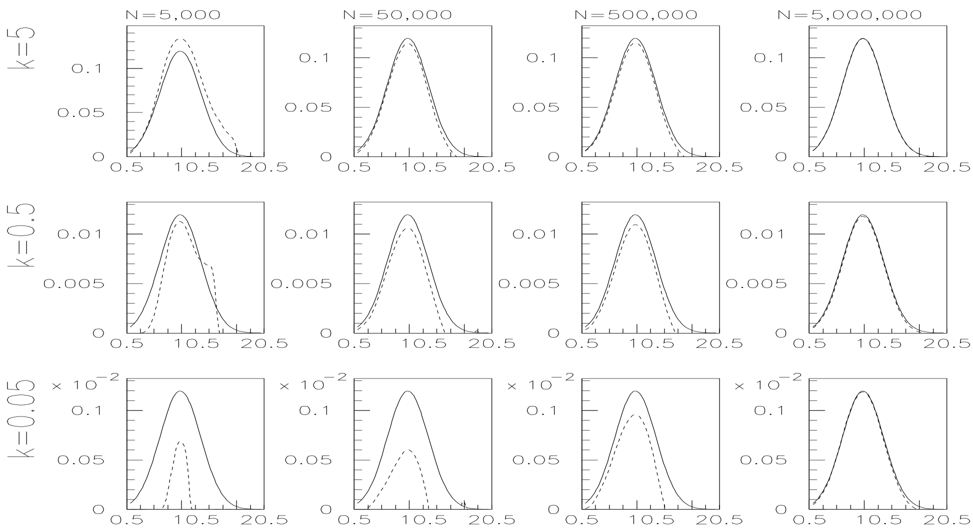

Figure 3: Non-parametric estimates of the numerator in the Ait-Sahalia estimator (dashed line), on a single simulation of the Vasicek model (10) of length N, for different values of N, as displayed, with $\alpha = 8.3\%$, $\sigma = 0.1$, $h = 1.06$ and different levels of mean reversion:$k = 5$ (top row), $k = 0.5$ (central row) and $k = 0.05$ (bottom row). For comparison, the true numerator is shown (solid line).

2.4 Small sample properties

We now turn to compute the small sample properties of the estimator. The charts in Figure 5 were computed using 1,000 replications of samples paths with $N = 6,000$. Here and throughout all the paper, we draw the starting point for each sample path from the (true) unconditional distribution. The charts in the first column of Figure 5 show the mean value of $\hat{\sigma}$, together with the 5% and 95% confidence limits, for values of k of 5.0 (first row), 0.5 (second row) and 0.05 (third row). In the first column, the estimator is computed using the *true* values of k and α (as the theory requires). For the charts in the second column we use the true value of k and an estimated value of κ. For the third column, both estimated values of both parameters are used.

Some of the results are striking. For relatively $k = 5.0$ the performance of the estimator, even with N as "low" as $6,000$, is quite good and, interesting, the performance actually improves when estimated values of k and α are used in place of the true values. The improvement is particularly marked when the true value of α is replaced by it's estimate and the reason for this is that, by using the estimated value the mean of the unconditional distribution and the sample mean coincide and the numerator and denominator of the estimator become more "consistent". For $k = 0.5$ and 0.05, however, the performance of the estimator is quite poor even when estimated values of k and α are used. In particular, for large values of r the mean is increasing in r even though the true value of r in this case is, of course, constant.

2.5 The choice of the bandwith parameter h

A critical issue in any nonparametric study is the choice of the bandwith parameter h and the literature provides no widely accepted recipe for this choice. In this section, we study the dependence of the Ait-Sahalia estimator on h.

Figures 6 and 7 show the results of simulations of the Vasicek models with the same parameters as in Figure 5, but with $h = 3$ and $h = 5$ respectively. There are two main reasons for considering a higher value of h. First, it will lead to a smoother estimate of the density in the tails of the distribution where fewer observations are available. Second, we ameliorate the problem of underestimating $\hat{\sigma}$ when h_s is selected as in equation (7); since we know that the

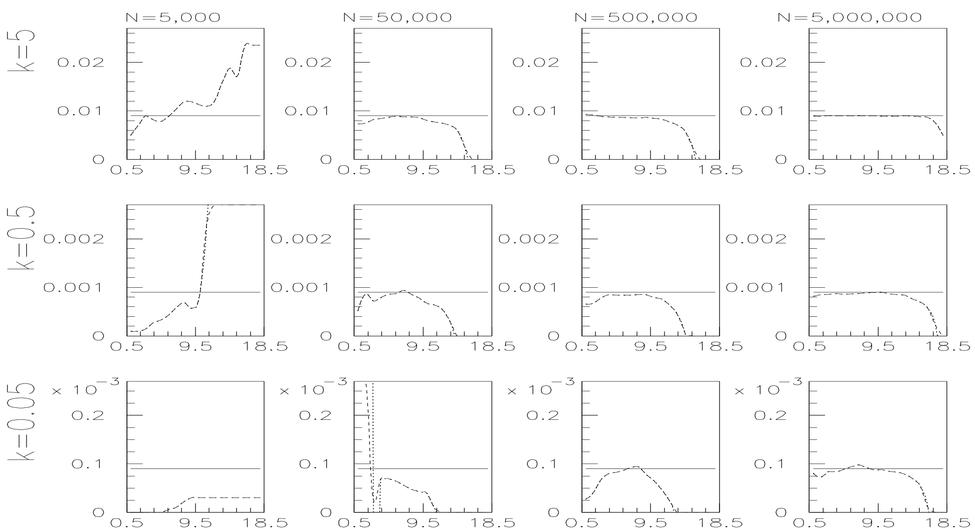

Figure 4: Ait-Sahalia estimator of the variance (dashed line), on a single simulation of the Vasicek model (10) of length N, for different values of N, as displayed, with $\alpha = 8.3\%$, $\sigma = 0.1$, $h = 1.06$ and different levels of mean reversion:$k = 5$ (top row), $k = 0.5$ (central row) and $k = 0.05$ (bottom row). For comparison, the true variance is shown (solid line).

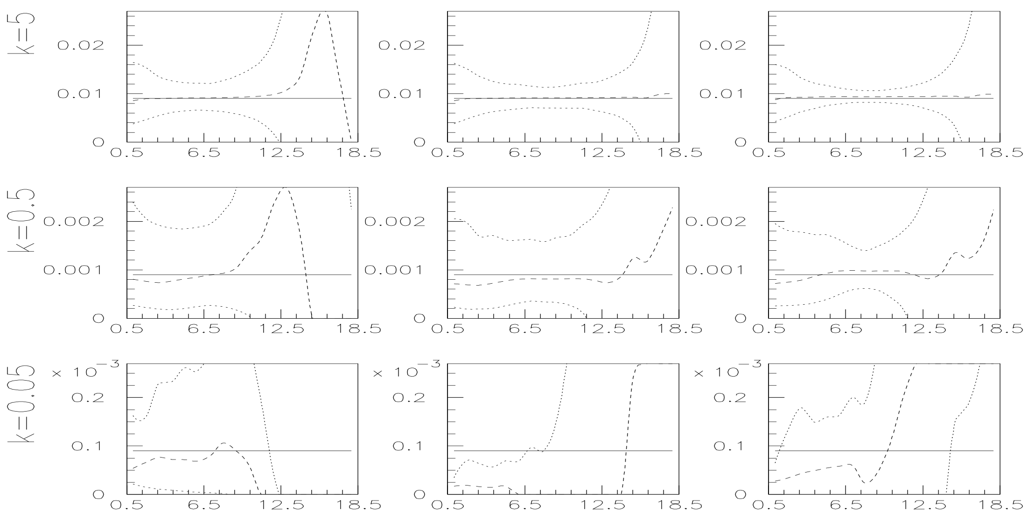

Figure 5: Average Ait-Sahalia estimator computed on 1,000 simulation of the Vasicek model (10), with $\alpha = 8.3\%$, $\sigma = \sqrt{0.03k}$, $\mathbf{h = 1.06}$, sample size $N = 6,000$ and different levels of mean reversion: $k = 5$ (top row), $k = 0.5$ (central row) and $k = 0.05$ (bottom row). In the first column, we use the true values of α and k when estimating $\sigma^2(r)$; in the second column, we use the true value of k and estimate α from data; in the third column, we estimate both k and α from data.The average estimate is the dashed line, while the true value is the solid line. The dotted line show 95% and 5% confidence limits.

estimate $\hat{\sigma}$ will be smaller than the actual value, we heuristically increase h.

Figures 6 and 7 show that the estimator is highly sensitive on the value on h in small samples ($N = 6,000$). In particular, for high and moderate mean reversion the higher values of h lead to an upward bias. Note that in the expression (12) h_s is a scale factor for the estimator of the variance in small sample. As expected, by increasing h we reduce the explosive behavior for large values of r, since the estimate of the density in the tail is smoother. If we repeat the experiment with larger sample size, the estimator converges slowly to the generated variance.

2.6 Performance when short rate follows the CIR Process

It is possible that our results so far are specific to the Ornstein-Uhlenbeck case when the variance of the short rate is not a function of its level. In this section we therefore repeat the

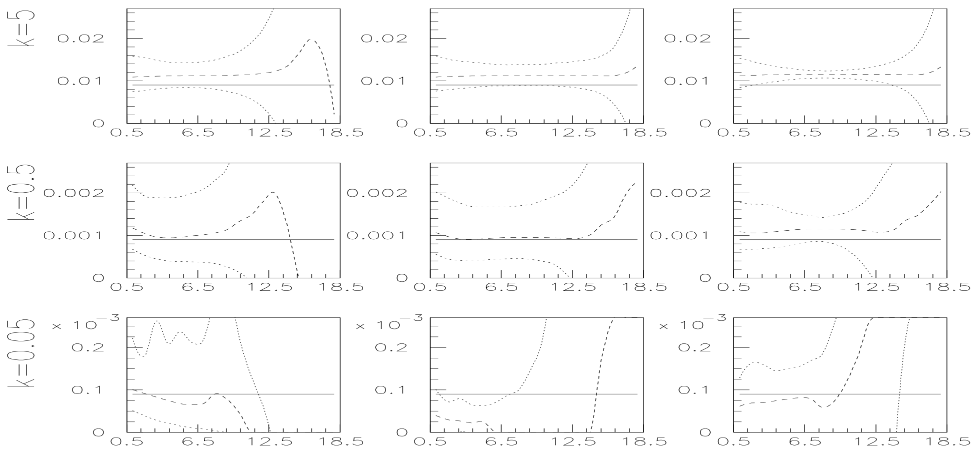

Figure 6: As in Figure 5, but with $h = 3$.

analysis just described but under the assumption that the true model describing the evolution of the short rate is the Cox et al. (1985) model (CIR):

$$dr(t) = k(\alpha - r(t))dt + \sigma\sqrt{r(t)}dW(t). \tag{14}$$

As before we compute 1,000 sample paths for the short rate, each of length 6,000, and, in Figure 8, show the mean value of $\hat{\sigma}$, together with the 5% and 95% confidence limits. The parameter values are given in the caption. The results in the three columns are for values of the bandwidth parameter H of 1.06, 3.0 and 5.0.

Once again, the results are not encouraging. For $h = 1.06$ the estimator works reasonably only when there is only a low level of persistence in the data, i.e. for high k. The results for $k = 0.5$ and 0.05 are, as in the previous case, poor.

In summary, the small sample performance of the Ait-Sahalia estimator turns is poor except in the case when the short rate displays a very high degree of mean reversion. In this case it is possible to obtain a reasonable estimate of the marginal density. However, in the empirically relevant case when the degree of mean reversion is low, the estimator does not produce reliable estimates of the variance and can easily suggest the presence of a non-linear relation between the variance of interest rates and their level when, in fact, no such relation exists.

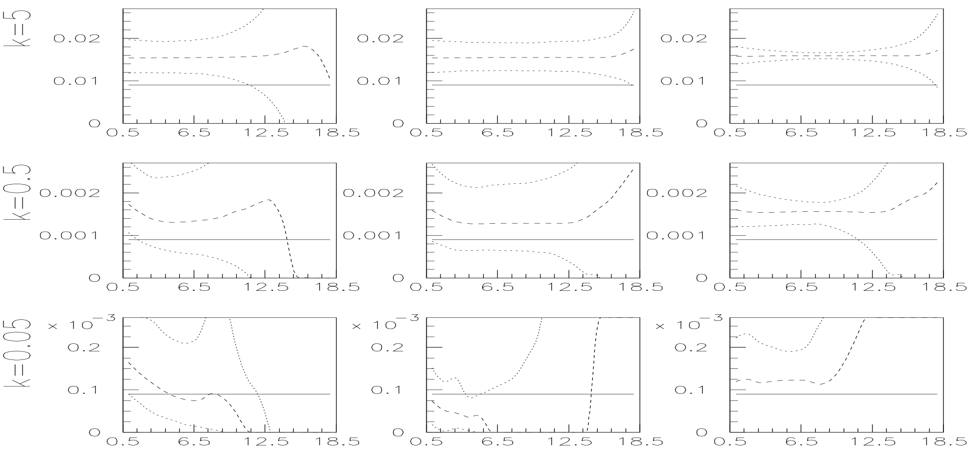

Figure 7: As in Figure 5, but with $h = 5$.

3 Analysis of the Stanton Estimator

The estimator proposed by Stanton (1997) differs from the Ait-Sahalia estimator and, in particular, does not require that the drift is known. Stanton (1997) shows that, given discretely sampled data, the drift and diffusion coefficients in (3) may be represented as:

$$\mu(r_t) = \frac{1}{\Delta} E_t[r_{t+\Delta} - r_t] + O(\Delta) \tag{15}$$

$$\sigma(r_t) = \sqrt{\frac{1}{\Delta} E_t[(r_{t+\Delta} - r_t)^2]} + O(\Delta) \tag{16}$$

Disregarding higher order terms, the drift and diffusion coefficients for a given level of r may be approximated, as the conditional expectations of (15) and (16)These expected values may be computed using the non-parametric estimate of the conditional density, which is obtained via the Kernel estimator, as previously described:

$$\hat{\mu}(r) = E[r_{t+1} - r_t | r_t = r] = \frac{\sum_{t=1}^{T-1} (\hat{r}_{t+1} - \hat{r}_t) K\left(\frac{r - \hat{r}_t}{h_s}\right)}{\sum_{t=1}^{T} K\left(\frac{r - \hat{r}_t}{h_s}\right)} \tag{17}$$

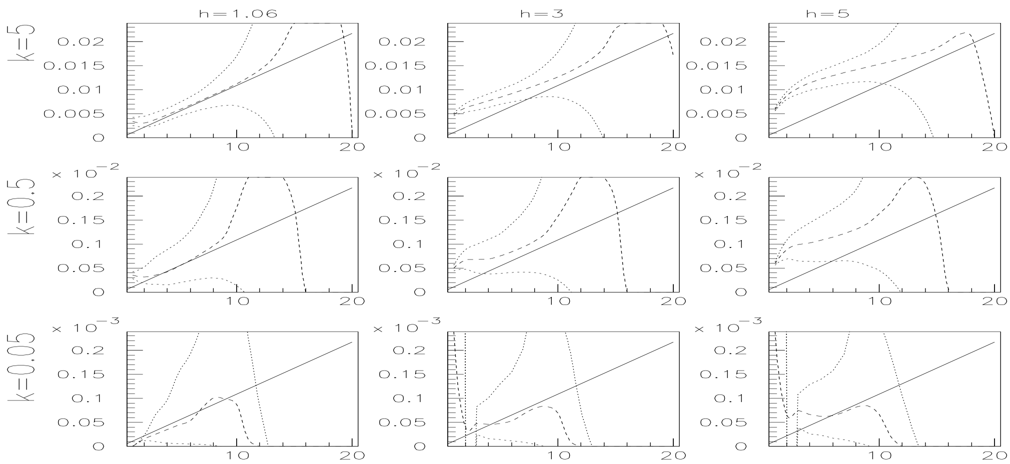

Figure 8: Average Ait-Sahalia estimator computed on 1,000 simulation of the CIR model (14), with $\alpha = 8.3\%$, $\sigma = \sqrt{0.005k/\alpha}$, sample size $N = 6,000$ and different levels of mean reversion: $k = 5$ (top row), $k = 0.5$ (central row) and $k = 0.05$ (bottom row), and bandwith parameter: $h = 1.06$ (first column), $h = 3$ (second column) and $h = 5$ (third column). The average estimate is the dashed line, while the true value is the solid line. The dotted line show 95% and 5% confidence limits.

$$\hat{\sigma}^2(r) = E[(r_{t+1} - r_t)^2_t | r_t = r] = \frac{\sum_{t=1}^{T-1} (\hat{r}_{t+1} - \hat{r}_t)^2 K\left(\frac{r - \hat{r}_t}{h_s}\right)}{\sum_{t=1}^{T} K\left(\frac{r - \hat{r}_t}{h_s}\right)} \tag{18}$$

The variance estimator suggested by Stanton (18) has the attractive feature that the same non-parametric estimator appears in both the numerator and the denominator and so errors arising from this source may offset at least to some extent. Even hough both the numerator and the denominator (which is the same as that for Figure 2) are not well approximated in small samples, the bias is, in this case, in the same direction and so cancels out in the ratio. As a consequence the variance is well estimated even in small samples.

Our results confirm that this is indeed the case. Figure 9 shows the estimate the numerator in (18) and of the estimated variance from a single simulation of the Vasicek model (10) with the same parameters as used previously. As before, the single path simulation is used to test the properties of the estimator as the sample size goes to infinity.

In the Ait-Sahalia estimator the Kernel estimate of the marginal density also appears but

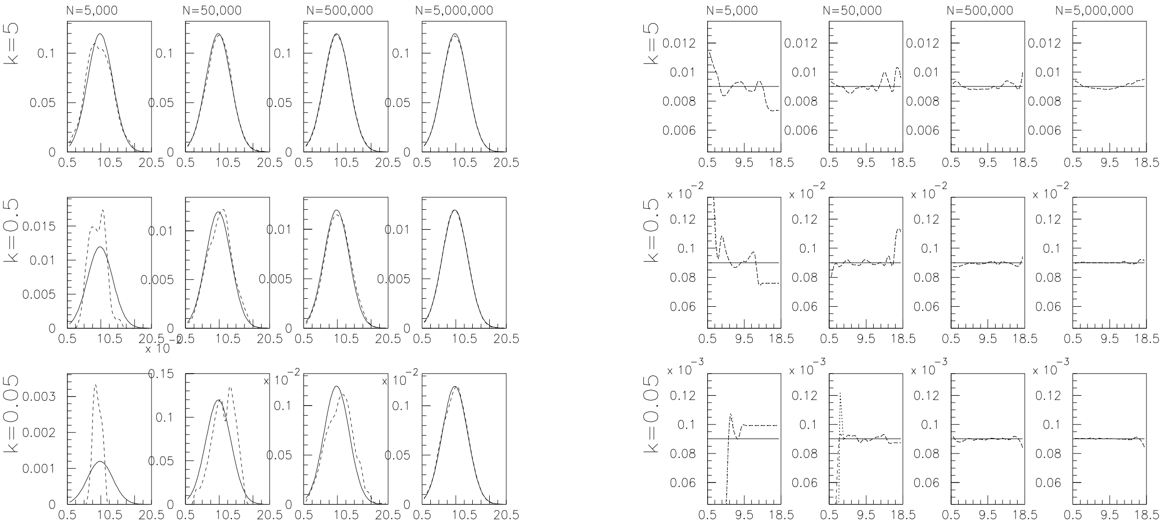

Figure 9: Non-parametric estimates of the numerator (left) and variance (right) for the Stanton estimator (dashed line), on a single simulation of the Vasicek model (10) of length N, for different values of N, as displayed, with $\alpha = 8.3\%$, $\sigma = 0.1$, $h = 1.06$ and different levels of mean reversion:$k = 5$ (top row), $k = 0.5$ (central row) and $k = 0.05$ (bottom row). For comparison, the true numerator is shown (solid line).

only in the denominator and so, in this case, there is no scope for errors to offset as in the Stanton estimator.

Even in the case of the Stanton estimator, however, h is still a crucial parameter that needs to be fine-tuned. Figure 10 shows that, in the case of the Vasicek model, the performance of the estimator is improved substantially by increasing h: this causes the width of the confidence intervals to shrink without introducing a bias. Things are slightly different when the Stanton method is used to estimate variance for data generated by the CIR model (14). The results are shown in figure 11 which shows that, while the Stanton estimator perform much better than the Ait-Sahalia estimator, for larger values of h (3.0 and 5.0) small biases arise for both low and high values of r. [3]

[3]Stanton uses $h = 4$

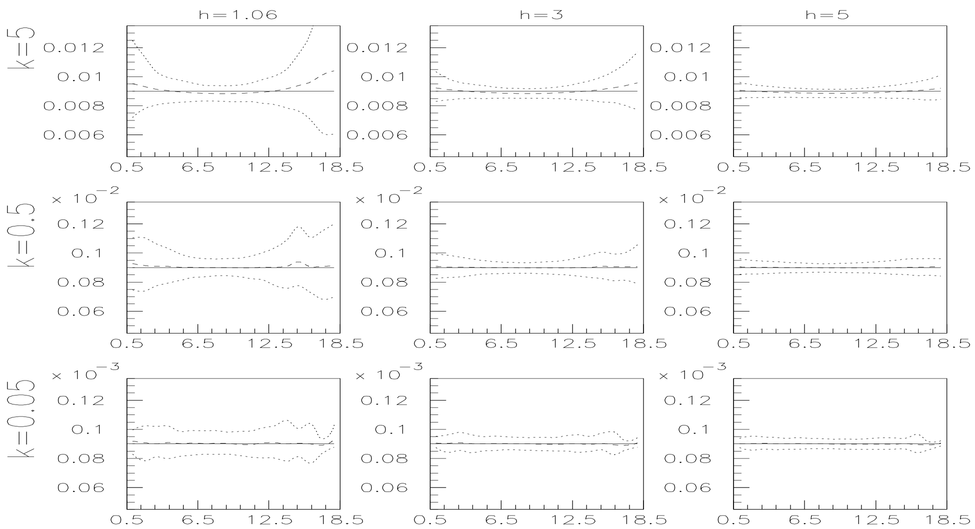

Figure 10: Average Stanton estimator computed on 1,000 simulation of the Vasicek model (10), with $\alpha = 8.3\%$, $\sigma = \sqrt{0.03k}$, sample size $N = 6,000$ and different levels of mean reversion: $k = 5$ (top row), $k = 0.5$ (central row) and $k = 0.05$ (bottom row), and bandwith parameter: $h = 1.06$ (first column), $h = 3$ (second column) and $h = 5$ (third column). The average estimate is the dashed line, while the true value is the solid line. The dotted line show 95% and 5% confidence limits.

4 Analysis of the Bandi-Phillips estimator

The estimator proposed in Bandi and Phillips (2003) is the following:

$$\hat{\sigma}^2(r) = \frac{\sum_{i=1}^{T} K\left(\frac{r-\hat{r}_i}{h_s}\right)\left(\frac{1}{m_i}\sum_{j=0}^{m_i}\left[\hat{r}_{t_{i,j}+1} - \hat{r}_{t_{i,j}}\right]^2\right)}{\sum_{i=1}^{T} K\left(\frac{r-\hat{r}_i}{h_s}\right)} \tag{19}$$

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

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

and

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

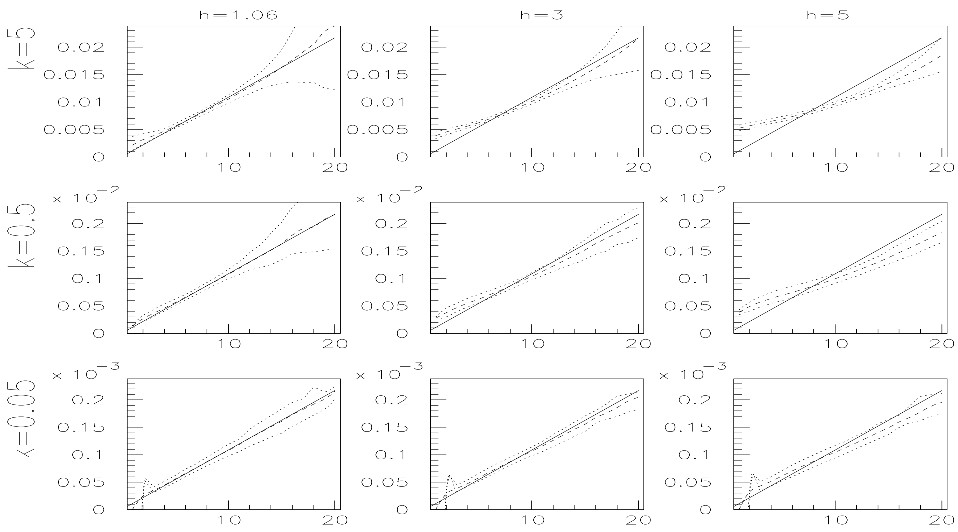

Figure 11: Average Stanton estimator computed on 1,000 simulation of the CIR model (14), with $\alpha = 8.3\%$, $\sigma = \sqrt{0.005k/\alpha}$, sample size $N = 6,000$ and different levels of mean reversion: $k = 5$ (top row), $k = 0.5$ (central row) and $k = 0.05$ (bottom row), and bandwith parameter: $h = 1.06$ (first column), $h = 3$ (second column) and $h = 5$ (third column). The average estimate is the dashed line, while the true value is the solid line. The dotted line show 95% and 5% confidence limits.

m_i is the number of times that $|\hat{r}_t - \hat{r}_i| \leq \varepsilon_s$ and ε_s is a parameter to be selected[4]. Looking at expressions (18),(19), we can see that the difference between Stanton and Bandi-Phillips estimators is that, while the Stanton estimator weights the observation r_t by the quadratic variation at time t, the Bandi-Phillips estimator weights the observation r_t with the average quadratic variation of all observation which are "close" to r.

As for the previous estimators, we implement the Bandi-Phillips method on simulated paths of the Vasicek and CIR model. We use $\varepsilon_s = 1.5\%$ as suggested in Bandi (2002). Figures 12 and 13 show the results for the Vasicek and CIR models respectively. The performance of the Bandi-Phillips and Stanton estimators are almost identical with the errors from the former just slightly lower.

[4]See Bandi and Phillips (2003) for details.

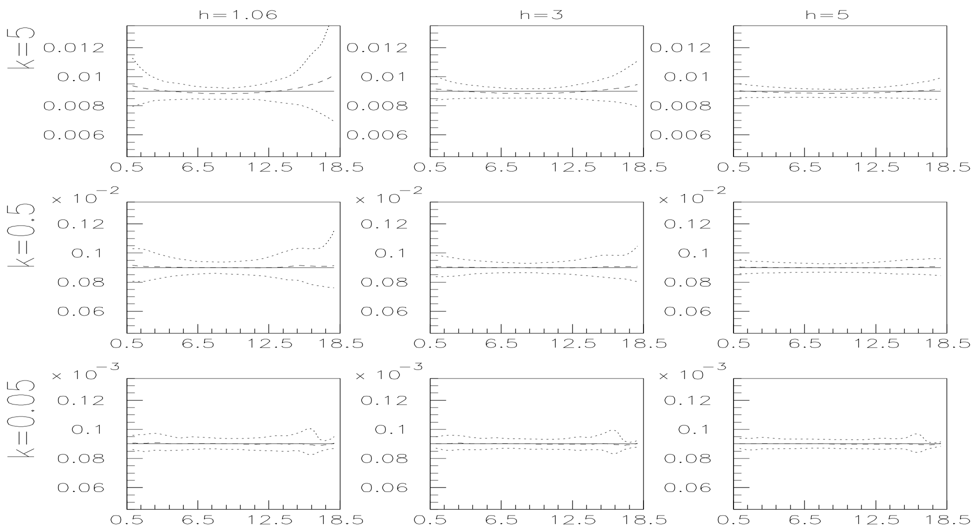

Figure 12: Average Bandi-Phillips estimator computed on 1,000 simulation of the Vasicek model (10), with $\alpha = 8.3\%$, $\sigma = \sqrt{0.03k}$, sample size $N = 6,000$ and different levels of mean reversion: $k = 5$ (top row), $k = 0.5$ (central row) and $k = 0.05$ (bottom row), and bandwith parameter: $h = 1.06$ (first column), $h = 3$ (second column) and $h = 5$ (third column). The average estimate is the dashed line, while the true value is the solid line. The dotted line show 95% and 5% confidence limits.

5 Data analysis

In this Section, we turn to an analysis of the data used in Ait-Sahalia (1996a) and Stanton (1997).

The data used in Ait-Sahalia (1996a) consist of the seven-day Eurodollar deposit rate from June 1, 1973 to February 25, 1995, a total of 5505 daily observations. Stanton (1997)'s data span an even longer period, from January 1965 to July 1995, and consist of daily yields on the three-month U.S. Treasury Bill, a total of 7975 observations. The time series of these two datasets are shown in Figure 14 and, even from visual inspection, it is clear that the two data set are very different. The Ait-Sahalia data appears much more volatile since it has many "spikes" which are typical of very short-term interest-rates. Indeed, Duffee (1996) suggests that rates on instruments with less than three months to maturity should not indeed be used for these purposes since they have too much idiosyncratic variation.

To implement the Ait-Sahalia estimator, the drift function has to be assumed. We follow

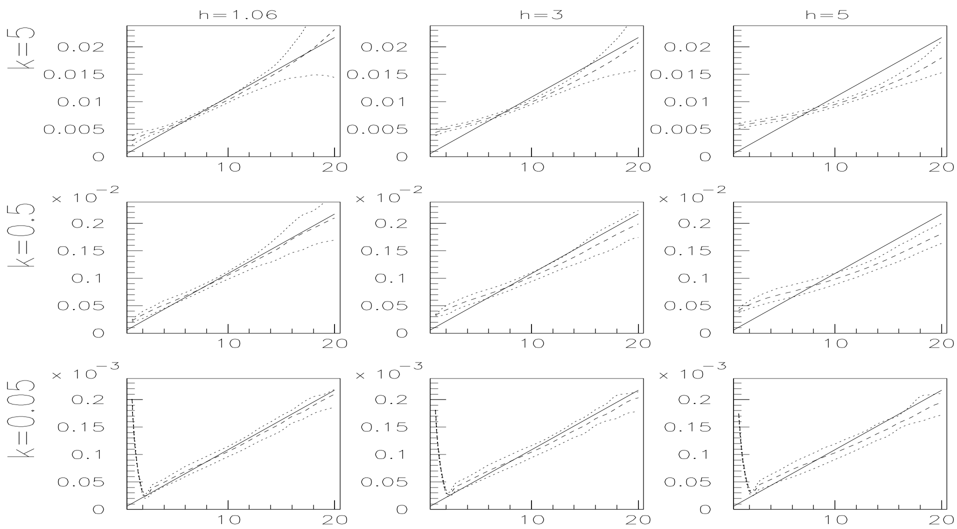

Figure 13: Average Bandi-Phillips estimator computed on 1,000 simulation of the CIR model (14), with $\alpha = 8.3\%$, $\sigma = \sqrt{0.005k/\alpha}$, sample size $N = 6,000$ and different levels of mean reversion: $k = 5$ (top row), $k = 0.5$ (central row) and $k = 0.05$ (bottom row), and bandwith parameter: $h = 1.06$ (first column), $h = 3$ (second column) and $h = 5$ (third column). The average estimate is the dashed line, while the true value is the solid line. The dotted line show 95% and 5% confidence limits.

Ait-Sahalia (1996a) and choose a linear specification $\mu(x) = k(\alpha - x)$. The parameters k, α must be estimated from the data and may be obtained in a number of ways. In Table 3 we report the parameter estimates obtained from Ait-Sahalia data using a number of alternative econometric methods. In the Table, OLS denotes Ordinary Least Squares regression, GLS denotes the use of a Cochrane-Orcutt correction for autocorrelation of the residuals and FGLS* denotes the method used by Ait-Sahalia. This is a two-stage procedure where the non-parametric variance function is used to weight observations. This dataset displays strong mean reversion.

We have estimated the conditional variance function on the Ait-Sahalia dataset using the Stanton and Bandi-Phillips methods and also using the Ait-Sahalia method with a linear drift specification. The two panels on the left of Figure 15 show the estimated variance function under the three methods and with two different bandwidths, $h = 1.06$ (the Gaussian optimal bandwith) and $h = 4$ (that used by Stanton). While the Stanton and Bandi-Phillips estimators give fairly similar results, the Ait-Sahalia estimator gives a larger variance for low interest rates

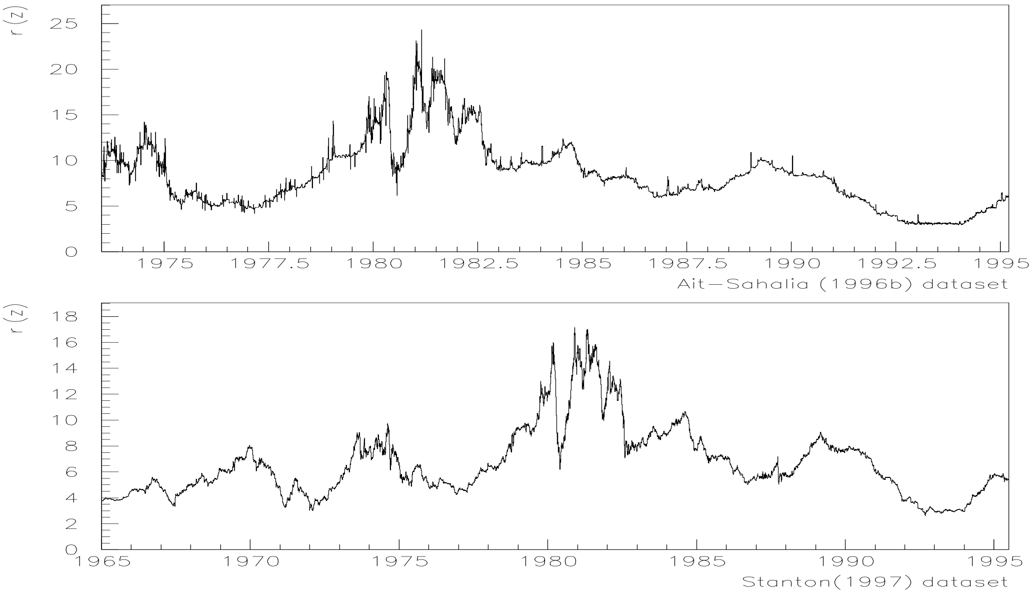

Figure 14: The time series of data used in Ait-Sahalia (1996a) (top) and Stanton (1997) (bottom).

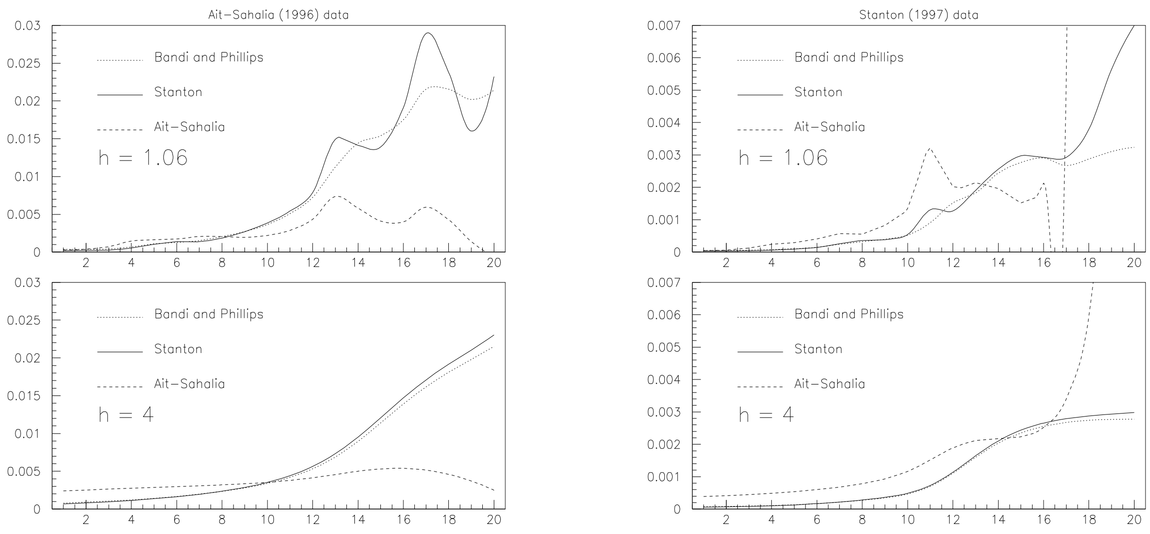

Figure 15: Estimates on data

and a smaller variance for high interest rates than the other estimators. All three estimators give smoother estimates for larger values of h.

We have also attempted to apply all three methods with the dataset used in Stanton (1997). However, when the mean reversion parameters, which are needed to apply the Ait-Sahalia method, were estimated on this dataset, it became apparent that the data are close to non-stationary. The OLS autoregressive coefficient was 0.998998 which translates into a k of $-log(0.998996) * 255 = 0.25615$. The absence of the spikes that are present in the 7-day rate used by Ait-Sahalia is almost certainly responsible for the lower estimated intensity of the mean reversion in this case. The estimated parameters were are reported in Table 4.

The two panels on the right of Figure 15 show the estimates for the Stanton (1997) dataset. With the low level of estimated mean reversion, it was impossible, using the Ait-Sahalia method, to obtain meaningful conditional variance estimates with $h = 1.06$ and for rates larger than 15%. This is consistent with the results of our Monte Carlo simulations which show instability in the Ait-Sahalia estimator when mean reversion is low. With larger values of h, all three estimators appear to be more stable. It is worth noting that, even though they span more or less the same period, for some values of r, the two data sets give rise to estimates of variance that differ by an order of magnitude. It is possible that this simply reflects the difference in the maturity of the two rates but, in our view, it is more likely to result from the presence of large spikes in the Ait-Sahalia data and their relative absence in the Stanton data.

It is clear that the conditional variance estimated by, on one hand, the Stanton and Bandi-Phillips methods and, on the other, the Ait-Sahalia method, differ substantially when the number of observations are available for the estimation is low and, in particular, for high values of the interest rate (say, above 12%). We also not that, as pointed out above, using the Ait-Sahalia technique, the estimated conditional variance may be negative. This occurs with the Stanton dataset and a bandwidth parameter of $h = 1.06$ when the interest rate on which the estimate is conditioned is large.

6 Conclusions

Our conclusions regarding the reliability of non-parametric estimators of the diffusion coefficient for the short rate are not encouraging. With the number of observations that is likely to be available, each of the three methods we have investigated – Ait-Sahalia, Stanton and Bandi-Phillips – may produce apparent non-linearities in the relation between conditional variance and the level of rates that are spurious. Applying the methods to data in which the conditional variance is in fact constant, we find that all three methods may, for a single time series of 5,000 observations (i.e., five years of daily data), produce a pattern of conditional volatility that appears strongly non-linear. We find the Ait-Sahalia method to be particularly prone to this problem but the problem also arises with the other two methods.

We also test the methods when the short rate is generated by the CIR process and find that, when the data are are persistent, i.e., exhibit low levels of mean reversion, both the Stanton and Bandi-Phillips methods are biased in the tails. This problem arises because, when the data are persistent, it is effectively impossible to estimate the density non-parametrically with sample sizes that are at all realistic. We also find that, for all three "non-parametric" methods, the choice of bandwidth remains an important parameter that needs to be determined.

Finally, although they have a substantial overlap, the datasets used by Ait-Sahalia and Stanton have entirely different properties and give rise to quite different estimates of the conditional variance of 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.

Andersen, T. and J. Lund (1997). Estimating continuous-time stochastic volatility models of the short-term interest rate. *Journal of Econometrics 77*, 343–377.

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.

Bianchi, C., R. Cesari, and L. Panattoni (1994). Alternative estimators of the Cox, Ingersoll and Ross model of the term structure of interest rates: A Monte Carlo comparison. Bank of Italy Discussion Themes.

Brennan, M. J. and E. S. Schwartz (1980). Analysing convertible bonds. *Journal of Financial and Quantitative Analysis 15*, 907–929.

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. 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). An intertemporal general equilibrium model of asset prices. *Econometrica 53*, 363–384.

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.

Durham, G. (2003). Likelihood-based specification analysis of continuous-time models of the short-term interest rate. *Journal of Financial Economics 70*(2), 463–487.

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

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.

Table 1: We simulate sample paths of N observations of the Vasicek model (10), with $\alpha = 8.3\%$, $\sigma = 0.1$, $r(0) = 8.4387\%$ for different values of k $(5, 0.5, 0.05)$ and N (5000, 50000 , 500000). The Table shows the average estimated mean and standard deviation on the simulated paths. Superscripts and subscripts show 95% and 5% confidence limits respectively. The true value of standard deviation is displayed in brackets, the true value of mean is 8.3%.

N=5,000		
	mean (%)	standard deviation (%)
$k = 5.0$	$8.34^{9.02}_{7.68}$	$3.12^{3.47}_{2.76}$ (3.16)
$k = 0.5$	$8.73^{15.1}_{2.44}$	$8.61^{11.45}_{6.09}$ (10.0)
$k = 0.05$	$10.29^{38.53}_{-18.98}$	$14.0^{22.74}_{7.87}$ (31.6)
N=10,000		
	mean (%)	standard deviation (%)
$k = 5.0$	$8.32^{8.84}_{7.83}$	$3.14^{3.39}_{2.89}$ (3.16)
$k = 0.5$	$8.53^{13.2}_{3.56}$	$9.24^{11.52}_{7.20}$ (10.0)
$k = 0.05$	$10.29^{38.53}_{-18.98}$	$17.9^{27.81}_{10.97}$ (31.6)
N=50,000		
	mean (%)	standard deviation (%)
$k = 5.0$	$8.31^{8.53}_{8.07}$	$3.15^{3.27}_{3.04}$ (3.16)
$k = 0.5$	$8.37^{10.54}_{6.09}$	$9.81^{10.98}_{8.62}$ (10.0)
$k = 0.05$	$9.02^{29.14}_{-12.18}$	$27.1^{36.57}_{19.31}$ (31.6)

Table 2: Estimate of the mean reversion parameter k in the literature.

Paper	Estimate	Model	Method
Ait-Sahalia (1996a)	0.978	$dr = (\alpha + kr)dt + \sigma(r)dW(t)$	FGLS
Chan et al. (1992)	0.1779	$dr = (\alpha + kr)dt + \sigma r^{\gamma} dW(t)$	GMM
Andersen and Lund (1997)	0.173	stochastic volatility	EMM
Durham (2003)	0.1875	$dr = (\alpha + kr)dt + \sqrt{\beta_1 + \beta_2 r} dW(t)$	Max. Likelihood
"	0.1049	$dr = (\alpha + kr)dt + \beta_1 r^{\beta_2} dW(t)$	Max. Likelihood
"	0.1056	$dr = (\alpha + kr)dt + \sqrt{\beta_1 + \beta_2 r + \beta_3 r^{\beta_4}} dW(t)$	Max. Likelihood

Table 3: Ait-Sahalia (1996b) dataset

	OLS	GLS	FGLS*
α	0.083082	0.082536	0.084387
k	1.6088	0.94014	0.97788

Table 4: Stanton (1997) dataset

	OLS	GLS
α	0.068423	0.068009
k	0.25615	0.32935