Resistant Nonparametric Analysis of the Short Term Rate

Rosario Dell'Aquila and Elvezio Ronchetti

No 2004.06
Resistant Nonparametric Analysis of the Short Term Rate

Rosario Dell’Aquila* and Elvezio Ronchetti†

July, 2004

Abstract

Aït-Sahalia (1996), Stanton (1997) and Jiang (1998) apply nonparametric and semi-parametric estimators to the short term interest rate and find strong nonlinearities in the drift function. In this paper we apply resistant techniques to the estimation of the drift and diffusion function. We show how the influential observations resulting from a resistant estimation of drift and diffusion may be used as a diagnostic tool to understand whether the estimated drift and diffusion function are broadly consistent with the assumed diffusion process. In an empirical exercise using Stanton’s (1997) and Aït-Sahalia’s (1996) data we find a clustering of influential observations in the pre 1982 period, in particular in the 1972-1974 and 1979-1982 period, suggesting that a regime change may be the dominant feature in the data rather than nonlinearities in the drift. As an additional result, we show that the bias reported in Chapman and Pearson (2000) is exaggerated because of the extrapolation of interest rate values outside the range of the simulated series.

JEL Classification: C1, C5, E4

*Quantitative Investment Research, Zürcher Kantonalbank, Switzerland and Università della Svizzera Italiana, Switzerland. e-mail: rosario.dellaquila@lu.unisi.ch

†Dept. of Econometrics, University of Geneva, Biv. Pont d’Arve 40, CH-1211 Geneva, Switzerland and Università della Svizzera Italiana, Switzerland. e-mail: elvezio.ronchetti@metri.unige.ch
1 Introduction

Aït-Sahalia (1996), Stanton (1997) and Jiang (1998) propose nonparametric and semi-parametric estimators to model the short term interest rate. Their main conclusion is, that there is a strong nonlinearity in the drift term and that virtually every diffusion process with a linear drift has to be rejected. While the estimated diffusion is similar to that estimated by Chan, Karolyi, Longstaff and Sanders (1992), there is evidence of substantial nonlinearity in the drift. This is close to zero for low and medium interest rates, but mean reversion increases sharply at higher interest rates.

This finding has been criticized along several directions. Chapman and Pearson (2000) and Abhyankar and Basu (2001) find that, due to the finite sample, the estimator in Stanton (1997) induces nonlinear estimates of the drift function even when the underlying process has a linear drift. Pritsker (1997) finds that the bandwidth that minimizes the mean integrated squared error relative to the true density is very sensitive to the autocorrelation in the data and is much larger than the optimal choice in the case of i.i.d. data.

In this paper, we re-analyze some of these facts. In particular as a first result we present a version of drift and diffusion function estimator proposed by Stanton (1997) which is resistant to outliers. By using a resistant estimation technique we expect to model the drift and diffusion function based on the majority of the data and in addition to detect outliers or deviating structures by means of the weights implied by the resistant estimation. These weights implied by the resistant estimation may therefore be seen as a simple diagnostic tool to check whether there are significant deviating structures.

When applying the resistant estimation technique to the datasets used in Stanton (1997) and Aït-Sahalia (1996) we find that the drift function estimated with the classical and the resistant kernel estimator are rather similar, while the diffusion function gets slightly flatter. Moreover, we find, that the influential observations implied by the resistant estimation of the drift and diffusion function seem to cluster in the pre-1982 period and in particular in the 1973-1974
and 1979-1982 period, the first Opec crisis and the period where the Federal Reserve changed the way it implemented its monetary policy respectively. This clustering of influential observations indicates that the even estimating non-parametrically the drift and diffusion function, we cannot fully describe some features of the data. In particular it shows the presence of periods of different volatility regimes in the data.

Therefore our results complement the results in Ang and Bekaert (2001), who find that processes with state-dependent regime shifts may replicate the shapes in the drift and diffusion function found in Aït-Sahalia (1996) and Stanton (1997). In another related analysis in a full parametric setting Dell’Aquila, Ronchetti and Trojani (2003) find evidence of a possibly misspecified process specification when analyzing the Chan, Karolyi, Longstaff and Sanders (1992) and Ahn and Gao (1999) models. In these cases the influential observations cluster around the 1979-1982 period, when the Federal Reserve changed the way it implemented its monetary policy.

These results have implications also for practitioners. Indeed, bond and option pricing models have been developed to cope with possibly nonlinear shapes of the drift function. (c.f. Ahn and Gao (1999) and Takamizava Shoji (2001)).

As a last result, in the Appendix, we review the Chapman and Pearson (2000) critique. In particular we show that the strong nonlinearity in the drift which appears in their simulations is mainly due to extrapolations outside of the support of the simulated series, which leads to a misperception of the magnitude of the truncation bias. When repeating the same exercise aggregating only over the support of the data, the strong spurious nonlinearity for the $h_{iid}$ bandwidth choice is clearly smaller and the oversmoothing of the cross-validation and Stanton bandwidth becomes evident.

The remainder of the paper is organized as follows. In Section 2 we discuss classical and resistant nonparametric estimators for drift and diffusion. Section 3 presents and discusses the data and the results. Section 4 concludes the paper.
2 Nonparametric Models of the Interest Rate Dynamics and their Estimation

2.1 Nonparametric Estimation of Drift and Diffusion

We consider a Markov process for the short-term interest rate \( \{ x_t; t \geq 0 \} \) given by

\[
dx_t = \mu(x_t)dt + \sigma(x_t)dB_t, \tag{1}\]

where \( \{ B_t; t \geq 0 \} \) is a standard Brownian motion, \( \mu: \mathbb{R} \to \mathbb{R} \) is the drift function and \( \sigma: \mathbb{R} \to \mathbb{R}^+ \) is the diffusion function. In this paper we refer to \( x_t \) as the instantaneous or short interest rate.

Under suitable restrictions on \( \mu \) and \( \sigma \) Stanton (1997) derives the following first order approximations for \( \mu \) and \( \sigma^2 \):

\[
\mu(x_t) = \frac{1}{\Delta} E[x_{t+\Delta} - x_t | x_t] + o(\Delta) \tag{2}
\]

and

\[
\sigma^2(x_t) = \frac{1}{\Delta} E[(x_{t+\Delta} - x_t)^2 | x_t] + o(\Delta), \tag{3}
\]

where \( \Delta \) denotes a discrete time step in a sequence of observations of the process \( x_t \).

The drift and diffusion function can be estimated nonparametrically using the familiar Nadaraya-Watson kernel regression estimator. In particular we can estimate separately the drift function at every gridpoint \( z_j \) by

\[
\hat{\mu}(z_j) = \frac{1}{\Delta} \frac{\sum_{t=1}^{T-1} K \left( \frac{z_j - x_{t+\Delta}^n}{h} \right) (x_{t+1}^{\Delta} - x_t^{\Delta})}{\sum_{t=1}^{T-1} K \left( \frac{z_j - x_t^{\Delta}}{h} \right)}
\]

and the diffusion function by

\[\hat{\sigma}(z_j) \]

\[= \frac{1}{\Delta} \frac{\sum_{t=1}^{T-1} K \left( \frac{z_j - x_{t+\Delta}^n}{h} \right) (x_{t+1}^{\Delta} - x_t^{\Delta})}{\sum_{t=1}^{T-1} K \left( \frac{z_j - x_t^{\Delta}}{h} \right)}
\]

\[= \frac{1}{\Delta} \frac{\sum_{t=1}^{T-1} K \left( \frac{z_j - x_{t+\Delta}^n}{h} \right) (x_{t+1}^{\Delta} - x_t^{\Delta})}{\sum_{t=1}^{T-1} K \left( \frac{z_j - x_t^{\Delta}}{h} \right)}
\]

\[= \frac{1}{\Delta} \frac{\sum_{t=1}^{T-1} K \left( \frac{z_j - x_{t+\Delta}^n}{h} \right) (x_{t+1}^{\Delta} - x_t^{\Delta})}{\sum_{t=1}^{T-1} K \left( \frac{z_j - x_t^{\Delta}}{h} \right)}
\]

\[= \frac{1}{\Delta} \frac{\sum_{t=1}^{T-1} K \left( \frac{z_j - x_{t+\Delta}^n}{h} \right) (x_{t+1}^{\Delta} - x_t^{\Delta})}{\sum_{t=1}^{T-1} K \left( \frac{z_j - x_t^{\Delta}}{h} \right)}
\]

\[= \frac{1}{\Delta} \frac{\sum_{t=1}^{T-1} K \left( \frac{z_j - x_{t+\Delta}^n}{h} \right) (x_{t+1}^{\Delta} - x_t^{\Delta})}{\sum_{t=1}^{T-1} K \left( \frac{z_j - x_t^{\Delta}}{h} \right)}
\]

\[= \frac{1}{\Delta} \frac{\sum_{t=1}^{T-1} K \left( \frac{z_j - x_{t+\Delta}^n}{h} \right) (x_{t+1}^{\Delta} - x_t^{\Delta})}{\sum_{t=1}^{T-1} K \left( \frac{z_j - x_t^{\Delta}}{h} \right)}
\]

\[= \frac{1}{\Delta} \frac{\sum_{t=1}^{T-1} K \left( \frac{z_j - x_{t+\Delta}^n}{h} \right) (x_{t+1}^{\Delta} - x_t^{\Delta})}{\sum_{t=1}^{T-1} K \left( \frac{z_j - x_t^{\Delta}}{h} \right)}
\]

\[= \frac{1}{\Delta} \frac{\sum_{t=1}^{T-1} K \left( \frac{z_j - x_{t+\Delta}^n}{h} \right) (x_{t+1}^{\Delta} - x_t^{\Delta})}{\sum_{t=1}^{T-1} K \left( \frac{z_j - x_t^{\Delta}}{h} \right)}
\]

\[= \frac{1}{\Delta} \frac{\sum_{t=1}^{T-1} K \left( \frac{z_j - x_{t+\Delta}^n}{h} \right) (x_{t+1}^{\Delta} - x_t^{\Delta})}{\sum_{t=1}^{T-1} K \left( \frac{z_j - x_t^{\Delta}}{h} \right)}
\]

\[= \frac{1}{\Delta} \frac{\sum_{t=1}^{T-1} K \left( \frac{z_j - x_{t+\Delta}^n}{h} \right) (x_{t+1}^{\Delta} - x_t^{\Delta})}{\sum_{t=1}^{T-1} K \left( \frac{z_j - x_t^{\Delta}}{h} \right)}
\]

\[= \frac{1}{\Delta} \frac{\sum_{t=1}^{T-1} K \left( \frac{z_j - x_{t+\Delta}^n}{h} \right) (x_{t+1}^{\Delta} - x_t^{\Delta})}{\sum_{t=1}^{T-1} K \left( \frac{z_j - x_t^{\Delta}}{h} \right)}
\]

\[= \frac{1}{\Delta} \frac{\sum_{t=1}^{T-1} K \left( \frac{z_j - x_{t+\Delta}^n}{h} \right) (x_{t+1}^{\Delta} - x_t^{\Delta})}{\sum_{t=1}^{T-1} K \left( \frac{z_j - x_t^{\Delta}}{h} \right)}
\]

\[= \frac{1}{\Delta} \frac{\sum_{t=1}^{T-1} K \left( \frac{z_j - x_{t+\Delta}^n}{h} \right) (x_{t+1}^{\Delta} - x_t^{\Delta})}{\sum_{t=1}^{T-1} K \left( \frac{z_j - x_t^{\Delta}}{h} \right)}
\]

\[= \frac{1}{\Delta} \frac{\sum_{t=1}^{T-1} K \left( \frac{z_j - x_{t+\Delta}^n}{h} \right) (x_{t+1}^{\Delta} - x_t^{\Delta})}{\sum_{t=1}^{T-1} K \left( \frac{z_j - x_t^{\Delta}}{h} \right)}
\]

\[= \frac{1}{\Delta} \frac{\sum_{t=1}^{T-1} K \left( \frac{z_j - x_{t+\Delta}^n}{h} \right) (x_{t+1}^{\Delta} - x_t^{\Delta})}{\sum_{t=1}^{T-1} K \left( \frac{z_j - x_t^{\Delta}}{h} \right)}
\]

\[= \frac{1}{\Delta} \frac{\sum_{t=1}^{T-1} K \left( \frac{z_j - x_{t+\Delta}^n}{h} \right) (x_{t+1}^{\Delta} - x_t^{\Delta})}{\sum_{t=1}^{T-1} K \left( \frac{z_j - x_t^{\Delta}}{h} \right)}
\]
\[ \hat{\sigma}(z_j) = \sqrt{ \frac{1}{\Delta} \sum_{t=1}^{T-1} K \left( \frac{z_j - x_{t+1}^T}{\Delta} \right) \left( x_{t+1}^T - x_{t}^T \right)^2 } \]

where \( \{x_t^T\}_{t=1}^T \) is a sample of the continuous time process \( x_t \), observed at the discrete interval \( \Delta \).

The kernel estimator is completely characterized by the choice of a particular kernel function satisfying \( \int_{-\infty}^{\infty} K(y)dy = 1 \) and by the choice of an appropriate bandwidth \( h \), which determines the width of the kernel function, and therefore the amount of smoothing at any \( z_j \). The choice of the kernel function \( K(\cdot) \) is not crucial (c.f. Härdle 1990) and as in Stanton (1997) we choose a Gaussian kernel \( K(u) = \frac{1}{\sqrt{2\pi}} \exp\left( -\frac{1}{2} u^2 \right) \).

The choice of the bandwidth is critical and non-trivial. Several bandwidth-selection procedures have been proposed in the literature. The idea behind these criteria is to minimize the mean squared prediction error, where the error is the difference at each point between the estimated function and the true function\(^2\). The bandwidth \( h \) determines the degree of smoothness of the estimated function \( \hat{m}(x_t) \). This can be seen by considering the limits for \( h \) tending to zero or to infinity, respectively. Indeed, at an observation \( y_t \), \( \hat{m}(x_t) \to y_t \) as \( h \to 0 \), while \( \hat{m}(x_t) \to \bar{y} \), as \( h \to \infty \). A common way of finding the function which minimizes the mean squared prediction error is to use cross-validation or a version of the so called least-squares cross-validation. Cross-validation selects the bandwidth which minimizes

\[ \min_{\{h_i\}} \sum_{t=1}^{T} \left[ y_t - \hat{m}_{-t,h_i}(x_t) \right]^2 w(x_t), \quad (4) \]

where \( \hat{m}_{-t,h_i}(x_t) \) is the fitted values of the kernel regression estimated at \( x_t \), when the observation at time \( t \) is not used for the estimation, and \( w(x_t) \) is a weight function which gives less weight to the observations \( x_t \) at the boundary of the sample to avoid boundary effects. Similarly least-squares cross-validation selects the bandwidth which minimizes

\(^2\)C.f. Härdle (1990) for a complete characterization of different approaches.
where \( \hat{m}_{h_i}(x_t) \) is the fitted values of the kernel regression estimated at \( x_t \) and \( \Xi(T^{-1}h_i) \) is a penalty function for small bandwidths\(^3\).

Alternatively, an optimal "plug-in" bandwidth for i.i.d. data (c.f. Silvermann (1986)) is given by

\[
h_{iid} = \hat{c} \sigma_0 T^{-1/5},
\]

where \( \sigma_0 \) is the sample standard deviation of the data and \( T \) is the sample size. This bandwidth minimizes the mean squared prediction error assuming a normal distribution for the data.

Chapman and Pearson (2000) argue that the estimator in Stanton (1997) leads to spurious nonlinearities near the end of the sample. They provide plots with estimated drift functions from a simulated CIR model with impressive nonlinearities in particular when using the \( h_{iid} \) bandwidth. In the Appendix we show that the strong nonlinearity in the drift is mainly due to extrapolating outside of the support of the simulated series which leads to a misperception of the magnitude of the truncation bias. When alternatively repeating the same exercise aggregating only over the support of the data, the strong spurious non-linearity for the \( h_{iid} \) bandwidth choice is clearly smaller and the oversmoothing of the cross-validation and bandwidth used by Stanton \( h_{Stanton} = 4 \ast h_{iid} \) becomes evident.

### 2.2 Resistant Kernel Estimation of Drift and Diffusion

We propose to estimate a resistant version of the drift function by implicitly solving a slightly modified version of (2), namely

\[
\min_{\{h_i\}} T^{-1} \sum_{t=1}^{T} [y_t - \hat{m}_{h_i}(x_t)]^2 w(x_t) \Xi(T^{-1}h_i),
\]

\(^3\)Several choices for the penalty function are available (c.f. Härdle (1990)). Chapman and Pearson (2000) for example choose the Shibata penalty function

\[
\Xi(T^{-1}h_i) = 1 + 2T^{-1}h_i^{-1}K(0).
\]
\[ E[\psi \left( \frac{(x_{t+\Delta} - x_t) - \Delta \mu(x_t)}{\sqrt{\Delta \sigma(x_t)}} \right)] | x_t] = 0, \]  

(6)

where, for instance,

\[
\psi(x) = \psi_c(x) = \begin{cases} 
-c, & x < c \\
-\psi_c, & |x| < c \\
c, & x \geq c
\end{cases}
\]

is the Huber function. Similarly to (3), we can estimate the diffusion function by solving

\[ E[\chi \left( \frac{x_{t+\Delta} - x_t}{\sqrt{\Delta \sigma(x_t)}} \right)] | x_t] = 0, \]  

(7)

where \( \chi(u) = \psi_c(u)^2 - E(\psi_c^2) \) and \( E(\psi_c^2) \) is the bias correction due to the truncation. The parameter \( c \) determines the desired degree of resistance and when \( c = \infty \), the estimator corresponds to the classical estimator proposed in Stanton (1997). The statistical properties of similar resistant estimators have been analyzed by Härdle and Tsybakov (1988). Note that resistant versions of higher order approximation given in Stanton (1997) can be derived in a similar way. We can now estimate a resistant version of the drift function \( b \), by a resistant kernel \( M \)-smoother, that is by implicitly finding for each gridpoint \( z_j \) the value \( \hat{m}(z_j) \) which solves

\[
\sum_{t=1}^{T-1} K \left( \frac{z_j - x_t^\Delta}{h} \right) \psi \left( \frac{y_t - \hat{m}(z_j)}{s(z_j)} \right) = 0,
\]  

(8)

where in simplified notation we write \( y_t = x_t^\Delta - x_{t-1}^\Delta \), \( m(z_j) = \Delta \mu(z_j) \) and \( s(z_j) = \sqrt{\Delta \sigma(z_j)} \). By linearizing Huber’s function (c.f. Appendix 5.1) we obtain the following iterative algorithm

\[
\hat{m}^{(l+1)}(z_j) = \hat{m}^{(l)}(z_j) + \frac{1}{T-1} \sum_{t=1}^{T-1} K \left( \frac{z_j - x_t^\Delta}{h} \right) \psi'(r_t^j) s(z_j),
\]  

(9)
where \( r^t_j \) is the residual given by
\[
r^t_j = \frac{y^t - \hat{m}^{(t)}(z_j)}{s(z_j)}.
\]

In a similar way, the resistant diffusion function is defined by
\[
\sum_{t=1}^{T-1} K \left( \frac{z_j - x^\Delta_t}{h} \right) \chi^t \left( \frac{y^t}{s(z_j)} \right) = 0, \tag{10}
\]
and can be computed (c.f. the Appendix 5.1) by the following iterative algorithm
\[
\tilde{s}(z_j)^{(t+1)} = \tilde{s}(z_j)^{(t)} \cdot \left( 1 + \frac{\sum_{t=1}^{T-1} K \left( \frac{z_j - x^\Delta_t}{h} \right) \chi^t(r^t_j)}{\sum_{t=1}^{T-1} K \left( \frac{z_j - x^\Delta_t}{h} \right) \chi^t(r^t_j)} \right),
\]
where the residual is given by \( r^t_j = \frac{y^t - \hat{m}(z_j)}{s(z_j)} \).

In practice, the diffusion function has to be estimated first. Typically it is estimated by choosing a slight amount of robustness (or even by using the classical diffusion estimator).

A crude way to measure the outlyingness of the observations in the sample is to interpret the resistant version as a weighted version of the classical kernel estimator. In particular we can write (8) for each gridpoint as
\[
\sum_{t=1}^{T-1} K \left( \frac{z_j - x^\Delta_t}{h} \right) \psi(r^t_j) = 0.
\]
Therefore for each gridpoint we can measure the influence of an observation on the estimated curve by
\[
\omega(x^\Delta_t) = \frac{\sum_{j=1}^n K \left( \frac{z_j - x^\Delta_t}{h} \right) \psi(r^t_j)}{\sum_{j=1}^n K \left( \frac{z_j - x^\Delta_t}{h} \right)} \tag{11}
\]
where \( r^t_j = \frac{y^t - \hat{m}(z_j)}{s(z_j)} \) and it is easy to verify, that the weights are between 0 and 1. Notice, that the sum is now taken over the gridpoints and not over the observations. Alternatively, we can measure the outlyingness directly by using the deviations from the fitted resistant curve, that is \( \omega(x^\Delta_t) = \frac{\psi(x^\Delta_t - \hat{m}(x^\Delta_t))}{\psi(x^\Delta_t - \hat{m}(x^\Delta_t))} \).
Similarly, we can measure the influence of the observations on the estimated diffusion function by
\[
\omega(x_t^\Delta) = \frac{\psi^2 \left( \frac{y_t}{s(x_t^\Delta)} \right)}{\left( \frac{y_t}{s(x_t^\Delta)} \right)^2}.
\] (12)

In addition to the robust estimation of the diffusion function, a resistant estimation of the bandwidth is required. A straightforward way would be to replace (4) by
\[
\min_{\{h_t\}} \sum_{t=1}^{T} \rho(y_t - \hat{m}_{-t,h_t}(x_t))
\]
where for instance \( \rho_c(x) = x^2/2 \) if \(|x| \leq c\) and \( \rho_c(x) = c|x| - c^2/2 \) otherwise and \( \hat{m}_{-t,h_t}(x_t) \) is the estimated drift function at \( x_t \) with bandwidth \( h_t \), when leaving out the \( t \)-th observation. Notice that \( \frac{d\rho_c}{dx} = \psi_c(x) \). A similar approach has been taken for example by Leung, Mariott and Wu (1993) and Cantoni and Ronchetti (2001). Given the difficulties of selecting a proper bandwidth in our context as reported by Pritsker (1997) and Chapman and Pearson (2000) we do not estimate a resistant bandwidth in this case, but we prefer to perform our analyses for a broad range of bandwidths.

3 Data and Empirical Results

3.1 Data

In the empirical analysis we use the same datasets as in Stanton (1997) and Aït-Sahalia (1996). The Aït-Sahalia dataset consists of 7-day Eurodollar deposit spot rate, bid-ask midpoint, 5505 observation from 1st June 1973 to 25th February 1995. The Stanton data are daily values of the secondary market yields on three Treasury Bills between January 1965 and July 1995 converted from discounts to annualized interest rates. The series in Stanton has been updated to April 2000.
3.2 Empirical Results

We first estimate the drift and the diffusion of the short rate with the estimator (2) proposed in Stanton (1997) using the same dataset as in Stanton (1997). For the reasons mentioned in the last Section, a broad range of bandwidths as in Chapman and Pearson (1997), that is $h_{iid}$, $h_{Stanton}$ and $h_{LSCV}$ were chosen. $h_{Stanton}$ is the bandwidth used in Stanton (1997) and corresponds to $h_{Stanton} = 4 \times h_{iid}$.

In the first graph of Figure 1, the drift function estimated using the Stanton (1997) dataset and estimator is plotted. The dashed, dotted and solid line correspond to the estimation of the drift function using $h_{LSCV}$, $h_{iid}$, $h_{Stanton}$ respectively. As it is well-known from the literature, we notice, that there is only very slight mean reversion for low and medium values of the interest rate $x_t$. The mean reversion gets very high for interest rate values beyond about 14%. This corresponds to the results obtained in Stanton (1997) and Aït-Sahalia (1996).

The graph in the first row of Figure 2 shows the estimated diffusion function. As it is well-known, for datasets containing pre 1982 interest rate data, the estimated function does not look like any of the diffusion functions of models like CIR or Vasicek, but rather like the diffusion function $\dot{\sigma}(x_t) = x_t^{1.5}$ estimated by Chan, Karolyi, Longstaïf and Sanders (1992).

The corresponding estimations with the resistant drift and diffusion estimator are presented in the second row of Figure 1 and Figure 2, respectively. We impose a slight amount of resistance by choosing $c = 2.0^5$. We first notice,

---

4 The penalty function was chosen as in Shibata (1981).
5 Stanton (1997) actually does not use $h_{iid}$, but uses $h_{Stanton} = 4 \times h_{iid}$ as a bandwidth following a heuristic choice; see Chapman and Pearson (2000) footnote 4 p. 360
6 Indeed when looking at the scatter plot $x_t$ versus $\Delta x_t^2$ for the Stanton (1997) and the Aït-Sahalia (1996) dataset (Figure 5), we notice, that few of the 7975 observations in the Stanton (1997) dataset are outlying with respect to the bulk of the data and determine the explosive shape of the diffusion function for high values of the interest rates. Therefore we prefer to
that the shape of the drift function estimated with the resistant kernel estimator looks rather similar to the drift function estimated with the classical kernel estimator, while the diffusion function is now slighly flatter.

The most important information can be gained by looking at the weights resulting from the resistant drift and diffusion estimation. In the first graph of Figure 3 the weights resulting from the resistant drift estimation of the Stanton (1997) and in the second graph the weights implied by the resistant estimation of the diffusion function are displayed. In both cases, the same bandwidth as in Stanton (1997) was chosen. Other bandwidth choices give a very similar structure for the weights. A weight of 1 corresponds to no downweighting. The weights implied by the resistant estimation of the drift and diffusion function are rather similar. In particular we notice, that there is a clustering of influential observations in the 1972-1974 and 1979-1982 period and that the majority of the weights cluster in the period before 1982. 389 out of 498 outlying observations are in the pre-1982 period. Notice, that the 1974 period corresponds to the first Opec crisis, while the 1979-1982 corresponds to a change in the implementation of the monetary policy by the Federal Reserve. When performing the same analysis with the Aït-Sahalia dataset we find a more extreme situation, see Figure 4. 283 out of 308 outlying observations are in the pre-1982 period.

In addition, when looking more closely to the outlying observations for the Aït-Sahalia dataset in the post 1982 period, we immediately see, that the outlying observation are located approximately about 7 days before the end of the year. Since we are analyzing the 7 day US euro-rate, these influential observations are a well-known liquidity effect at the end of the year. It is questionable whether this observations should be modelled, but in any case the weights provide valuable information to the analyst about the data.

use an estimator with a slight amount of resistance, since when looking at the occurrence of these observations, we notice, that they all cluster in the 1979-1982 period, the time of the well-known monetary experiment of the Federal Reserve and the only period where we have the occurrence of very high interest rate data. The results regarding the influential points are similar for different choices of $c$. 11
The structure of the weights for the drift and diffusion function suggest, that there is a regime shift in the volatility. Similarly, for the drift function, the weights cluster in the same pre-1982 period, even when taking the classical diffusion function as auxiliary estimation of the diffusion in order to be as near as possible to the analysis in Stanton (1997). Notice, that the classical estimator in Stanton (1997) estimates $\mu$ and $\sigma$ separately. The weights resulting from a resistant estimation of drift and diffusion can be seen as good diagnostic tool to check whether there are substantial deviating structures which are not consistent with the assumed process specification.

Our results can also be viewed as a complement of the Ang and Bekaert (2002) analysis, who document how a regime-switching model with state dependent transition probabilities between regimes can replicate the patterns for the drift and diffusion function found by non-parametric studies.

In a related analysis in a full parametric GMM setting Dell’Aquila, Ronchetti and Trojani (2003) find evidence of a possibly misspecified process specification when analyzing the Chan, Karolyi, Longstaff and Sanders (1992) and Ahn and Gao (1999) specification. In their case the influential observations cluster around the 1979-1982 period, where the Federal Reserve changed the way it implemented its monetary policy. In our nonparametric setting, the clustering is extended to the pre-1982 period and includes 1979-1982.

These results have implications for practitioners. Indeed, bond and option pricing models have been developed to cope with possibly nonlinear shapes of the drift function. (c.f. Ahn and Gao (1999) and Takamizava Shoji (2001)). In view of the analysis above, we would question these attempts. In particular looking at the drift estimation in the post-1982 period there seems to be no need for a nonlinear drift specification.
4 Conclusions

Aït-Sahalia (1996), Stanton (1997), Jiang (1998) report strong nonlinearities in the drift function. We develop an outlier resistant version of the Stanton (1997) estimator for the drift and diffusion function. We apply the resistant estimators to the Aït-Sahalia and Stanton dataset and find a clustering of influential observations occurs in the pre-1982 period and in particular in the 1972-1974 and 1979-1982 period. This indicates that the estimated nonparametric drift and diffusion function cannot fully describe the data structure. In particular different regimes seem still to be present. Our analysis complements the results found in Ang and Bekaert (2002), who document how a regime-switching model with state dependent transition probabilities between regimes can replicate the patterns for the drift and diffusion function found by non-parametric studies.

If one is willing to use pre-1982 data for the analysis, our results indicate that further research is needed focussing on whether nonlinear model may cope with regime switches in the data.
References


5 Appendix:

5.1 Resistant Drift and Diffusion Estimators

Define the error term and the residuals by $e_t^j = \frac{y_t - m(z_j)}{s(z_j)}$ and $r_t^j = \frac{y_t - \hat{m}(z_j)}{s(z_j)}$ respectively. In the subsequent analysis we will write shortly $e_t$ for $e_t^j$ and $r_t$ for $r_t^j$. Noting that $r_t = e_t + \frac{m(z_j) - \hat{m}(z_j)}{s(z_j)}$ and expanding $\psi(r_t)$ around $\psi(e_t)$ we obtain

$$\psi(r_t) = \psi\left(\frac{y_t - \hat{m}(z_j)}{s(z_j)}\right) \approx \psi(e_t) + \psi'(e_t) \frac{m(z_j) - \hat{m}(z_j)}{s(z_j)}.$$

Hence by (8)

$$\sum_{t=1}^{T-1} K \left(\frac{z_j - x_t^j}{h}\right) \left[\psi(e_t)s(z_j) + \psi'(e_t)m(z_j)\right] \approx \sum_{t=1}^{T-1} K \left(\frac{z_j - x_t^j}{h}\right) \psi'(e_t)\hat{m}(z_j)$$

and thus

$$\hat{m}(z_j) \approx \frac{\sum_{t=1}^{T-1} K \left(\frac{z_j - x_t^j}{h}\right) [\psi(e_t)s(z_j) + \psi'(e_t)m(z_j)]}{\sum_{t=1}^{T-1} K \left(\frac{z_j - x_t^j}{h}\right) \psi'(e_t)}.$$

Hence the following iterative algorithm can be used to determine the drift function $\hat{m}(z_j)$

$$\hat{m}^{(l+1)}(z_j) \approx \hat{m}^{(l)}(z_j) + \frac{\sum_{t=1}^{T-1} K \left(\frac{z_j - x_t^j}{h}\right) [\psi(r_t)s(z_j)]}{\sum_{t=1}^{T-1} K \left(\frac{z_j - x_t^j}{h}\right) \psi'(r_t)}.$$

In a similar way, we can define the error term by $e_t = \frac{y_t}{s(z_j)}$, and the residual by $r_t = \frac{y_t - \hat{m}(z_j)}{s(z_j)}$, (where we have omitted the $j$ index for brevity) and approximate $\chi(r_t)$ by

$$\chi(r_t) = \chi\left(\frac{y_t}{s(z_j)}\right) \approx \chi(e_t) - \chi'(e_t) \frac{\epsilon_t}{s(z_j)} [s(z_j) - s(z_j)].$$

(13)
Hence, inserting (13) in (10) we obtain

\[
\sum_{t=1}^{T-1} K \left( \frac{z_j - x_t^2}{h} \right) \left[ \chi(\epsilon_t) - \chi'(\epsilon_t) \epsilon_t \left( \frac{\hat{s}(z_j)}{s(z_j)} - 1 \right) \right] \cong 0
\]

and therefore we can estimate \( \hat{s}(z_j) \) in a resistant way by means of the following iterative algorithm

\[
\hat{s}(z_j)^{(l+1)} = \hat{s}(z_j)^{(l)} \cdot \left( 1 + \frac{\sum_{t=1}^{T-1} K \left( \frac{z_j - x_t^2}{h} \right) \chi(r_t)}{\sum_{t=1}^{T-1} K \left( \frac{z_j - x_t^2}{h} \right) \chi'(r_t) r_t} \right).
\]


Chapman and Pearson (2000) (CP henceforth) study the finite-sample properties of the Stanton (1997) and Aït-Sahalia (1996) estimators of drift and diffusion function by means of simulated sample paths of a CIR diffusion process. Although the drift function is linear, both estimator show nonlinearities of the magnitude reported in Stanton (1997) and Aït Sahalia (1996). CP argue that this is due to a truncation bias arising in small samples. While it is true that there is a bias due to truncation, the large bias reported in CP especially for the \( h_{iid} \) bandwidth arises because the values for every interest rate are calculated for all interest rate values between 0.0 and 0.2 and not only for those values in the simulated sample. Indeed for the values chosen by CP the simulated samples will very often have no values near 0 nor values larger than 0.2. As a consequence, the possibly small bias becomes very large. When aggregating only using gridpoints in the domain of the interest rate series, a much smaller bias appears. In additions it becomes clear that the \( h_{LSCV} \) and \( h_{Stanton} \) oversmooth the data.

Following CP we study the finite-sample properties of their estimators by applying them to simulated sample paths of a square root diffusion process

\[
\frac{dx_t}{x_t} = \kappa(\theta - x_t)dt + \sigma \sqrt{x_t} dB_t,
\]

(14)
where \( \theta \) is the long run mean of \( x_t \), \( \kappa \) determines the speed at which the process returns to the long run mean, and \( \sigma \) is the volatility. The unconditional moments of the square-root process are

\[
E(x) = \theta
\]

\[
Var(x) = \frac{\theta \sigma^2}{2\kappa}, \tag{15}
\]

and

\[
Corr(x_{t+\Delta}, x_t) = \exp(-\kappa\Delta).
\]

CP construct sample paths from (14) for a given set of parameter values\(^7\) and apply the kernel regression estimators from Stanton (1997) to the simulated square root data. The results for the drift and diffusion function for the parameter choice \((\kappa, \theta, \sigma) = (0.21459, 0.085711, 0.07830)\) are reported in Figure 6 and 7 respectively. Each graph has the same structure: the solid line is the true drift function, the dashed line is the pointwise mean and the dotted lines are the pointwise 25th and 75th percentiles at each gridpoint across the 1000 simulations.

The left hand side of Figure 6 displays the drift estimates when the aggregation takes place over each interest rate value in the grid between 0.0 and 0.2, thus extrapolating as in CP, while the right hand side displays the drift estimates when the aggregation takes place only over those gridpoints values contained in the support of the series. All simulation are performed for a sample size of 7500, which broadly corresponds to the sample size of the Stanton dataset.

\(^7\)The parameter values \((\kappa, \theta, \sigma)\) are \((0.21459, 0.085711, 0.07830)\) which imply a monthly autocorrelation coefficient of 0.982, consistent with the upper end estimates of this parameter based on U.S. interest rate data, and \((0.85837, 0.085711, 0.15660)\), which imply a first-order (monthly) autocorrelation that is equal to that of the Eurodollar data used in Ait-Sahalia (1996). Given a \((\kappa, \theta)\), the value of \( \sigma \) results from setting equation (15) equal to the sample variance in the Ait-Sahalia data set.
Following CP the most striking spurious nonlinearity are obtained when estimating the drift using the \( h_{iid} \) bandwidth. Comparing the left and the right hand side of Figure 6 for this case, we notice that the bias reported in CP which is illustrated in the left column is very large compared to the right column. In addition we notice, that when we avoid extrapolation, \( h_{Stanton} \) and \( h_{LSCV} \) oversmooth the estimated drift function. Notice, that the truncation bias and the boundary bias are acting in different directions. This is also the reason why in CP the jackknife kernel estimator which eliminates the boundary bias leads to strong nonlinearities also when using \( h_{Stanton} \) and \( h_{LSCV} \).

Similar results arise for the diffusion function. Our conclusion is that the \( h_{iid} \) seems an acceptable bandwidth, in both cases. As for the drift function, \( h_{Stanton} \) and \( h_{LSCV} \) seem to oversmooth the data.

It seems therefore that the results in CP are not conclusive with regard to the magnitude of the truncation bias and the choice of the bandwidth. In addition, further research about the impact of regime switches in the data in the choice of the bandwidth is needed.
Figure 1: Plot of the estimated classical and resistant drift function using the Stanton (1997) dataset and three different bandwidth parameter choices, $h_{\text{LSCV}}$ (dashed line), $h_{\text{iid}}$ (dotted line) and $h_{\text{Stanton}}$ (solid line).

6 Figures
Figure 2: Plot of the estimated classical and resistant diffusion function using the Stanton (1997) dataset and three different bandwidth parameter choices, $h_{LSCV}$ (dashed line), $h_{iid}$ (dotted line) and $h_{Stanton}$ (solid line).
Figure 3: Plot of the weights resulting from the resistant estimation of the drift and diffusion function using the Stanton dataset (updated until April 2000). The graph in the first row shows the interest rate series. The second and third row shows the weights resulting from the resistant estimation of the drift and diffusion function respectively. A weight of 1 corresponds to no downweighting.
Figure 4: Plot of the weights resulting from the resistant estimation of the drift and diffusion function using the Aït-Sahalia dataset. The graph in the first row shows the interest rate series. The second and third row shows the weights resulting from the resistant estimation of the drift and diffusion function respectively. A weight of 1 corresponds to no downweighting.
Figure 5: Plot of the annualized interest rate versus the squared one-day changes in interest rates.
Figure 6: The drift function using the estimator in Stanton (1997) for the true drift defined as $\mu(x) = \kappa(\theta - x)$, where $\kappa = 0.21459, \theta = 0.085711$. In these simulations, the diffusion function is $\sigma(x) = \sigma \sqrt{x}$, where $\sigma = 0.07830$. The solid line is the true drift. The dashed line is the pointwise mean across the 1000 simulations with $T=7500$, and the dotted lines are the pointwise 25th and 75th percentiles across the 1000 simulations. The left hand side displays the estimated aggregate value when aggregating pointwise, independently of the domain of the series as in Chapman and Pearson (2000). The right hand side displays the results, when aggregation takes place depending on the domain of the series.
Figure 7: The diffusion function using the estimator in Stanton (1997) for the true diffusion function defined as $\sigma(x) = \sigma \sqrt{x}$, where $\sigma = 0.07830$. In these simulations, the drift function is $\mu(x) = \kappa(\theta - x)$ where $\kappa = 0.21459, \theta = 0.085711$. The solid line is the true diffusion. The dashed line is the pointwise mean across the 1000 simulations with $T=7500$, and the dotted lines are the pointwise 25th and 75th percentiles across the 1000 simulations. The left hand side displays the estimated aggregate value when aggregating pointwise, independently of the domain of the series as in Chapman and Pearson (2000). The right hand side displays the results, when aggregation takes place only for the domain of the series.