Empirical comparisons in short-term interest rate models using nonparametric methods

Arapis, Manuel and Gao, Jiti

The University of Western Australia, The University of Adelaide

23 September 2004

Nonparametric Kernel Estimation and Testing in Continuous–Time Financial Econometrics

Manuel Arapis and Jiti Gao[1], The University of Western Australia, WA 6009, Australia

Abstract

This study applies the nonparametric estimation procedure to the diffusion process modeling the dynamics of short-term interest rates. This approach allows us to operate in continuous-time, estimating the continuous-time model, despite the use of discrete data. Three methods are proposed. We apply these methods to two important financial data. After selecting an appropriate bandwidth for each data set, empirical comparisons indicate that the specification of the drift has a considerable impact upon the pricing of derivatives, through its effect on the diffusion function. Indeed, this impact is more substantial than that reported in the literature.

KEY WORDS: Diffusion process, drift function, kernel density estimation, stochastic volatility.

1. INTRODUCTION

The application of continuous-time mathematics to the field of finance dates back to 1900 when Louis Bachelier wrote a dissertation on the theory of speculation. Since Bachelier, the continuous-time approach to pricing assets such as derivative securities has evolved into a fundamental finance tool. The recent rapid expansion of asset pricing theory may be largely attributable to the seminal work of Merton (1973) and Black and Scholes (1973). Their work changed the way in which finance and asset valuation was viewed by practitioners, consequently laying the foundation for the theory of pricing derivative securities. Many papers have since been written on the valuation of derivatives, creating important extensions to the original model.

A time series model used extensively in finance is the continuous-time diffusion, or Itô process. In modeling the dynamics of the short-term riskless rate process $\{r_t\}$, for example, the applicable diffusion process is:

$$dr_t = \mu(r_t)dt + \sigma(r_t)dB_t \tag{1.1}$$

where $\mu(\cdot) = \mu(\cdot, \theta)$ and $\sigma(\cdot) = \sigma(\cdot, \theta)$ are the drift and volatility functions of the process respectively, and can be indexed by θ, a vector of unknown parameters, and B_t is the standard

[1] *Address for correspondence:* Professor Jiti Gao, Department of Statistics, School of Mathematics and Statistics, The University of Western Australia, Crawley WA 6009, Australia. Email: jiti@maths.uwa.edu.au

Brownian motion. The diffusion function is also referred to as the instantaneous variance. The model developed by Merton specified the drift and diffusion functions as constant. This assumption has since been relaxed by most researchers interested in refining the model in order to describe the behavior of interest rates. The prices generated by such modified models are generally believed to better reflect those observed in the market.

A vast array of models has been put forward in an attempt to explain the aberrant behavior of the short-term riskless rate, which is one of the fundamental prices determined in practice today. Some of the most popular models developed are shown below:

$$dr = \alpha(\beta - r)dt + \sigma dB, \tag{1.2}$$
$$dr = \alpha(\beta - r)dt + \sigma r^{1/2} dB, \tag{1.3}$$
$$dr = \alpha(\beta - r)dt + \sigma r dB, \tag{1.4}$$
$$dr = r\{\kappa - (\sigma^2 - \kappa\alpha)r\}dt + \sigma r^{3/2} dB, \tag{1.5}$$
$$dr = \kappa(\alpha - r)dt + \sigma r^{\rho} dB, \tag{1.6}$$
$$dr = (\alpha_{-1} r^{-1} + \alpha_0 + \alpha_1 r + \alpha_2 r^2)dt + \sigma r^{3/2} dB. \tag{1.7}$$

Model (1.2) was used by Vasicek (1977) to derive a discount bond price model. Unlike the model developed by Merton (1973) and Black and Scholes (1973), whose respective process follow Brownian motion with drift and Geometric Brownian Motion (GBM), Vasicek utilizes the Ornstein-Uhlenbeck process. This model has the feature of mean-reversion, where the process tends to be pulled to its long-run trend of β with force α. This force is proportional to the deviation of the interest rate from its mean. This model specifies the volatility of the interest rate as being constant. By definition, the volatility function generates the erratic fluctuations of the process around its trend. Cox, Ingersoll and Ross (CIR) (1985) utilize model (1.3) to model term-structure. It is the square root process. Not only does the drift have mean-reversion, but the model also implies the volatility $\sigma(\cdot)$, of the process increases at a rate proportional to $\sqrt{r}$. Thus the diffusion increases at a rate proportional to r. Model (1.4) (see Brennan and Schwartz 1980) was developed to price convertible bonds. It not only possesses the mean-reversion property, but the model also implies that the instantaneous variance $\sigma^2(\cdot)$ of the process increases at a rate proportional to r^2. Model (1.5) is the inverse of the CIR process discussed in Ahn and Gao (1999) and then Aït-Sahalia (1999). Model (1.6) is the constant elasticity of volatility model proposed in Chan, Karolyi, Longstaff and Sanders (1992). The nonlinear drift model (1.7) was proposed in Aït-Sahalia (1996a).

As well as the recent developments made in the application of continuous-time diffusion processes to the finance world, there has also been much work done in the adoption of statistical methods for the estimation of these continuous-time models. The main estimation techniques encountered in the majority of the literature (see Sundaresan 2001) include maximum likelihood (ML), generalized method of moments (GMM) and, more recently, nonparametric approaches. ML and GMM both require us to firstly parameterize the underlying model of interest. That is, we apply these methods to estimate the parameters of the diffusion process, such that they are consistent with the restrictions we have imposed on the model by the parameterizations. This is comparable to fitting a linear regression to nonlinear phenomena for reasons of convenience. It thus seems reasonable that we look for an approach which places the least restrictions on models so that we have empirical rather than analytical tractability.

Empirical researchers have recently shown a preference for nonparametric alternatives. Its only prerequisite is that accurate data is used. Such an approach is useful when approximating very general distributions, and has the additional advantage of not requiring the functional form specification of the drift and diffusion functions in our model of the short-term riskless rate (1.1). By leaving the diffusion process unspecified, the resulting functional forms specified by this method should result in a process that follows asset price data closely. This method requires a smooth density estimator of the marginal distribution $\pi(\cdot)$ and utilization of a property of (1.1), similar to that of a normal random variable whose distribution is explained entirely by its first two moments, to characterize the marginal and conditional densities of the interest rate process. The first two moments of the normal distribution are its mean and variance. For the case of the diffusion process, they are the drift and diffusion functions. Thus, the functions are formed such that they are consistent with the observed distribution of the available data.

Aït-Sahalia (1996a) was among the first to pioneer the nonparametric approach. The paper notes, as with Chan, *et al* (1992) and Ahn and Gao (1999) that one of the most important features of the process given by (1.1) in its ability to accurately model the term structure of interest rates, is the specification of the diffusion function $\sigma^2(\cdot)$. By qualifying the restriction on the drift function $\mu(\cdot)$, to the linear parametric class $\mu(r_t;\theta) = \beta(\alpha - r_t)$, which is consistent with the majority of prior research, the form of the diffusion function is left unspecified and estimated nonparametrically. Jiang and Knight (1997), however, argue that this is effectively a semiparametric approach because of the linear restriction imposed on the drift function. Jiang and Knight (1997) were able to develop an identification and estimation procedure for both the drift and diffusion functions of a general Itô diffusion process. They too have used nonparametric kernel estimators of the marginal density function based on discretely sampled data and the property of the Itô diffusion process, analogous to that used by Aït-Sahalia (1996a). In contrast to Aït-Sahalia (1996a), the drift function is left unspecified. Jiang and Knight (1997) suggest that the diffusion term can be identified first because it is of lower order than the drift. It is noted that the diffusion term is of order $\sqrt{dt}$ whereas the drift term is of order dt. These estimators as with that of Aït-Sahalia (1996a) are shown to be pointwise consistent and asymptotically normal. Other important approaches include Stanton (1997), Pritsker (1998), Chapman and Pearson (2000), Chen, Härdle and

Li (2003), Fan and Zhang (2003), Gao and King (2003, 2004), Chen and Gao (2004, 2005), Durham (2004), Hong and Li (2004), Jones (2004), and Li, Pearson and Poteshman (2004).

Unlike the work by Aït-Sahalia (1996a) and Jiang and Knight (1997), in order to avoid undersmoothing we propose an improved and simplified nonparametric approach to the estimation of both the drift and diffusion functions, and then establish the mean integrated square error (MISE) of each nonparametric estimator for the case where the bandwidth is still proportional to $T^{-1/5}$, where T is the number of the data under consideration. We then apply the proposed nonparametric approach to (i) the Federal Funds rate data, sampled monthly between January 1963 and December 1998 and (ii) the Eurodollar deposit rates, bid-ask midpoint and sampled daily from June 1, 1973 to February 25, 1995. Three nonparametric methods for estimating the drift and diffusion functions are established. For each data set, we compute these estimators. We see that bandwidth selection is both difficult and critical to the application of the nonparametric approach. After empirical comparisons we then suggest for each given set of data, the best fitting model and bandwidth which produces the most acceptable results. We see that the imposition of the linear mean-reverting drift does in fact affect the estimation of the diffusion function. Differences between the three diffusion estimators suggest the drift function may have a greater effect on pricing derivatives than what is quoted in the literature.

We also propose a novel test on both sets of data and their subsets to reinforce our conclusions regarding nonlinearity in the drift. The proposed test itself is novel methodologically in the following aspects. First, we establish that the size of the proposed test is asymptotically correct under any model in the null. Second, we also show that the test is asymptotically consistent when the null is false. Third, the implementation of the proposed test uses a range of bandwidth values instead of using an estimation optimal value based on a cross–validation selection criterion. Fourth, the implementation of the proposed test is based on a simulated P–value rather than an asymptotic critical value of the standard normal distribution

The paper is organized as follows. In Section 2, we describe the approach taken in estimating nonparametrically the unknown functions of (1.1). We develop our own three different estimators for the drift and diffusion functions. Section 3 proposes a novel test and establishes some asymptotic results. Section 4 includes the application of the proposed estimators to the two sets of data, drawing empirical comparisons with the three methods. In Section 5, remarks on the comparisons and analyses are offered, with some direction for further research. Mathematical details are relegated to the appendix.

2. NONPARAMETRIC ESTIMATION

The nonparametric approach to density estimation allows modeling of data where no priori ideas about the data exist. A given specification of the diffusion process can be accepted or rejected based on a comparison of the densities implied by the parametric model with its nonparametric density estimates (Aït-Sahalia 1996b). Then it is conceivable to work in the reverse, through density matching, so that a model may be developed by first estimating the density such that the model cannot be rejected. Given T discrete interest rate observations with sampling interval (equivalently the time between successive observations) Δ, the kernel density estimate of the marginal density is given by

$$\hat{\pi}(r) = \frac{1}{T}\sum_{t=1}^{T}\frac{1}{h}K\left(\frac{r-r_t}{h}\right) \tag{2.1}$$

where $K(\cdot)$ is the kernel function and h is the kernel bandwidth. The bandwidth is a smoothing or scaling parameter. It can be seen as a contraposing measure of the amount of information one wishes to attain. Large values of h tend to oversmooth the estimate and thus hide structure while small values of h tend to undersmooth and thus might produce excessive and confusing modes. As yet, we have not been provided with a definite form or value for h that should be employed for use in this analysis (this is a topic in its own right, needing further research). By comparing the nonparametric marginal density, drift and diffusion estimates acquired by use of some existing 'good' bandwidth values, we can suggest the most appropriate bandwidth to use for each of our two different sets of financial data. Whereas bandwidth selection is critical for optimal results, the selection of the kernel does not have a significant bearing on the overall result. We therefore utilize the normal kernel function $K(x) = \frac{1}{\sqrt{2\pi}}\exp\{-\frac{x^2}{2}\}$ throughout this paper. Our nonparametric marginal density estimate is,

$$\hat{\pi}(r) = \frac{1}{\sqrt{2\pi}hT}\sum_{t=1}^{T}\exp\left\{\frac{-(r-r_t)^2}{2h^2}\right\}. \tag{2.2}$$

We now state the first result of this paper.

Proposition 2.1. *Assume that Assumptions A.1, A.2 and A.3(i) listed in the appendix hold. Then*

$$\begin{aligned} E\left\{\int\left[\hat{\pi}(r)-\pi(r)\right]^2 dr\right\} &= \frac{1}{2\sqrt{\pi}}\cdot\frac{1}{Th}+\frac{h^4}{4}\int\left(\pi''(r)\right)^2 dr + o\left(\frac{1}{Th}\right)+o(h^{-4}) \\ &= \left[\frac{c}{2\sqrt{\pi}}+\frac{c^4}{4}\int\left(\pi''(r)\right)^2 dr\right]\cdot T^{-\frac{4}{5}} + o\left(T^{-\frac{4}{5}}\right) \end{aligned}$$

when $h = cT^{-\frac{1}{5}}$ *for some* $c > 0$.

The proof of Proposition 2.1 is relegated to the appendix.

Before proposing estimators for the drift and diffusion functions in (1.1), we consider a nonparametric time series model of the form:

$$Y_t = f(X_t) + g(X_t)e_t,\ t = 1, 2, \ldots, T \tag{2.3}$$

where $f(\cdot)$ and $g(\cdot) > 0$ are unknown functions defined on $R^1 = (-\infty, \infty)$, and $\{(X_t, Y_t)\}$ and $\{e_t\}$ are strongly stationary time series with $E[e_t|X_t] = 0$ and $E[e_t^2|X_t] = 1$. It is well known that $f(\cdot)$ can be estimated by

$$\hat{f}(x) = \frac{\sum_{t=1}^{T} Y_t \, K\left(\frac{x-X_t}{h}\right)}{\sum_{t=1}^{T} K\left(\frac{x-X_t}{h}\right)}. \tag{2.4}$$

Similarly, the conditional variance $g^2(x) = E[(Y_t - f(X_t))^2|X_t = x]$ can be estimated by

$$\hat{g}^2(x) = \frac{\sum_{t=1}^{T}(Y_t - \hat{f}(X_t))^2 K\left(\frac{x-X_t}{h}\right)}{\sum_{t=1}^{T} K\left(\frac{x-X_t}{h}\right)}. \tag{2.5}$$

These estimators of $f(\cdot)$ and $g^2(\cdot)$ are known as the Nardaraya-Watson estimators. This methodology is the basis for the derivation of the estimators $\hat{\mu}_1(\cdot)$ and $\hat{\sigma}_1^2(\cdot)$ of Method 1 below. The remaining two methods rely on relationships between the drift, diffusion and marginal density functions which alternatively describe the usual diffusion process.

From the Fokker-Planck equation (see (2.2) of Aït-Sahalia 1996a), we can obtain

$$\frac{d^2}{dr^2}(\sigma^2(r)\pi(r)) = 2\frac{d}{dr}(\mu(r)\pi(r)). \tag{2.6}$$

Integrating and rearranging (2.6) yields

$$\mu(r) = \frac{1}{2\pi(r)} \frac{d}{dr} \left[\sigma^2(r)\pi(r)\right]. \tag{2.7}$$

Or alternatively, integrating (2.6) twice yields

$$\sigma^2(r) = \frac{2}{\pi(r)} \int_0^r \mu(x)\pi(x)dx. \tag{2.8}$$

These equations allow us to estimate the drift, $\mu(\cdot)$ given a specification of the diffusion, $\sigma^2(\cdot)$ and marginal density, $\pi(\cdot)$, or the diffusion term given the drift and marginal density estimates.

2.1 *Method 1*

The drift and diffusion functions can be alternatively interpreted as

$$\mu(r_t) = \lim_{\delta \to 0} E\left[\frac{r_{t+\delta} - r_t}{\delta} | r_t\right]. \tag{2.9}$$

$$\sigma^2(r_t) = \lim_{\delta \to 0} E\left[\frac{[r_{t+\delta} - r_t]^2}{\delta} | r_t\right], \quad 0 < t < \infty. \tag{2.10}$$

Stanton (1997) refers to these estimators as first order approximations for $\mu(\cdot)$ and $\sigma^2(\cdot)$. The author constructs a family of approximations to the drift and diffusion functions, and estimates the approximations nonparametrically.

Equations (2.9) and (2.10) suggest using an approximate form of model (1.1) below:

$$r_{(t+1)\Delta} - r_{t\Delta} = \mu(r_{t\Delta})\Delta + \sigma(r_{t\Delta}) \cdot (B_{(t+1)\Delta} - B_{t\Delta}),\ t = 1, 2, \cdots, \tag{2.11}$$

where Δ is the time between successive observations. In practice, Δ is small but fixed, as most continuous-time models in finance are estimated with monthly, weekly, daily, or higher frequency observations.

Now, (2.4) and (2.9) suggest estimating $\mu(\cdot)$ by

$$\hat{\mu}_1(r) = \frac{\sum_{t=1}^{T-1} K\left(\frac{r - r_{t\Delta}}{h}\right)\left(\frac{r_{(t+1)\Delta} - r_{t\Delta}}{\Delta}\right)}{\sum_{t=1}^{T} K\left(\frac{r - r_{t\Delta}}{h}\right)}. \tag{2.12}$$

Multiplying the numerator and denominator by $\frac{1}{Th}$ gives

$$\begin{aligned}
\hat{\mu}_1(r) &= \frac{\frac{1}{\Delta Th}\sum_{t=1}^{T-1} K\left(\frac{r - r_{t\Delta}}{h}\right)\left(r_{(t+1)\Delta} - r_{t\Delta}\right)}{\frac{1}{Th}\sum_{t=1}^{T} K\left(\frac{r - r_{t\Delta}}{h}\right)} \\
&= \frac{1}{\Delta Th\hat{\pi}(r)} \sum_{t=1}^{T-1} K\left(\frac{r - r_{t\Delta}}{h}\right)\left(r_{(t+1)\Delta} - r_{t\Delta}\right) \\
&= \frac{1}{\Delta Th\hat{\pi}(r)\sqrt{2\pi}} \sum_{t=1}^{T-1} \exp\left(-\frac{(r - r_{t\Delta})^2}{2h^2}\right) \cdot \left(r_{(t+1)\Delta} - r_{t\Delta}\right), \qquad (2.13)
\end{aligned}$$

when $K(x) = \frac{1}{\sqrt{2\pi}} \exp\{-\frac{x^2}{2}\}$. Similarly, by (2.5) and (2.10) we estimate $\sigma^2(\cdot)$ by

$$\hat{\sigma}_1^2(r) = \frac{1}{\Delta Th\ \hat{\pi}(r)\sqrt{2\pi}} \sum_{t=1}^{T-1} \exp\left(-\frac{(r - r_{t\Delta})^2}{2h^2}\right) \cdot \left(r_{(t+1)\Delta} - r_{t\Delta}\right)^2. \tag{2.14}$$

Let $\hat{m}_1(r) = \hat{\mu}_1(r)\hat{\pi}(r)$, $m(r) = \mu(r)\pi(r)$, $\hat{V}_1(r) = \hat{\sigma}_1^2(r)\hat{\pi}(r)$, and $V(r) = \sigma^2(r)\pi(r)$. Since $\frac{\hat{m}_1(r)}{\pi(r)}$ and $\hat{\mu}_1(r)$ have the same asymptotic property for the MISE, we only establish the MISE for $\hat{m}_1(r)$ below. The same reason applies to explain why $\hat{V}_1(r)$ has been introduced. We now have the following propositions and their proofs are relegated to the appendix.

Proposition 2.2. *Assume that Assumptions A.1, A.2 and A.3(iii) listed in the appendix hold. Then*

$$\begin{aligned}
E\left\{\int \left[\hat{m}_1(r) - m(r)\right]^2 dr\right\} &= \frac{1}{Th} \cdot \frac{1}{2\sqrt{\pi}} \int \left(\mu^2(r) + \sigma^2(r)\Delta^{-1}\right)\pi(r)dr + \frac{h^4}{4}\int \left(m''(r)\right)^2 dr \\
&+ o\left(\frac{\Delta^2}{Th}\right) + o(h^{-4}).
\end{aligned}$$

(i) *When Δ is fixed but $h = c\,T^{-\frac{1}{5}}$ for some $c > 0$, we have*

$$E\left\{\int \left[\hat{m}_1(r) - m(r)\right]^2 dr\right\} = \frac{C_1}{Th} + C_2 h^4 = C_3 \cdot T^{-\frac{4}{5}} + o\left(T^{-\frac{4}{5}}\right). \tag{2.15}$$

(ii) *When* $\Delta = c_1\ h^2$ *and* $h = c_2\ T^{-\frac{1}{7}}$ *for some* $c_1 > 0$ *and* $c_2 > 0$, *we have*

$$E\left\{\int \left[\hat{m}_1(r) - m(r)\right]^2 dr\right\} = \frac{D_1}{Th\Delta} + D_2 h^4 = D_3 \cdot T^{-\frac{4}{7}} + o\left(T^{-\frac{4}{7}}\right). \qquad (2.16)$$

Proposition 2.3. *Assume that Assumptions A.1, A.2 and A.3(iii) listed in the appendix hold. Then*

$$\begin{aligned} E\left\{\int \left[\hat{V}_1(r) - V(r)\right]^2 dr\right\} &= \frac{1}{Th}\cdot\frac{1}{2\sqrt{\pi}}\int \left(\mu^4(r) + 3\sigma^4(r)\Delta^{-2} + 6\mu^2(r)\sigma^2(r)\Delta^{-1}\right)\pi(r)dr + o\left(\frac{1}{Th}\right) \\ &+ \int \left(\Delta\mu^2(r)\pi(r) + \frac{\Delta^2}{2}h^2\left(\mu^2(r)\pi(r)\right)'' + \frac{h^2}{2}V''(r)\right)^2 dr + o(h^4). \end{aligned}$$

(i) *When* Δ *is fixed but* $h = d\ T^{-\frac{1}{5}}$ *for some* $d > 0$, *we have*

$$E\left\{\int \left[\hat{V}_1(r) - V(r)\right]^2 dr\right\} = \frac{C_1}{Th\Delta^2} + C_2\Delta^2 + C_3 h^4 = C_1 T^{-\frac{4}{5}} + C_2\Delta^2. \qquad (2.17)$$

(ii) *When* $\Delta = d_1\ h^2$ *and* $h = d_2\ T^{-\frac{1}{7}}$ *for some* $d_1 > 0$ *and* $d_2 > 0$, *we have*

$$E\left\{\int \left[\hat{V}_1(r) - V(r)\right]^2 dr\right\} = \frac{C_1}{Th\Delta^2} + C_2\Delta^2 + C_3 h^4 = D_1 T^{-\frac{4}{7}} + o\left(T^{-\frac{4}{7}}\right). \qquad (2.18)$$

Propositions 2.2 and 2.3 show that while $\hat{m}_1(r)$ attains the optimal MISE rate of $T^{-\frac{4}{5}}$, $\hat{V}_1(r)$ normally cannot attain the optimal MISE rate. When the drift of model (1.1) vanishes (i.e. $\mu(r) \equiv 0$), however, the optimal MISE rate of $T^{-\frac{4}{5}}$ can be achieved for $\hat{V}_1(r)$.

By making use of the conditional expectation expression and the nonparametric marginal density estimator $\hat{\pi}(\cdot)$, we can forego the prior necessities of having to specify either of the otherwise unknown functions, $\mu(\cdot)$ and $\sigma^2(\cdot)$ in order to calculate the other.

2.2 *Method 2*

This method adopts a similar approach to that taken by Jiang and Knight (1997). They have estimated $\sigma^2(\cdot)$ by

$$\hat{\sigma}^2_{JK}(r) = \frac{\sum_{t=1}^{T-1} TK\left(\frac{r_{t\Delta_T} - r}{h}\right)\left[r_{(t+1)\Delta_T} - r_{t\Delta_T}\right]^2}{\sum_{t=1}^{T} NK\left(\frac{r_{t\Delta_T} - r}{h}\right)}, \qquad (2.19)$$

which is comparable to $\hat{\sigma}_1^2(\cdot)$, where T is the number of interest rates observed, N is the time length, Δ_T depends on T and $\Delta_T \to 0$ as $T \to \infty$. Jiang and Knight (1997) estimate the drift by

$$\hat{\mu}_{JK}(r) = \frac{1}{2}\left(\frac{d\hat{\sigma}^2_{JK}(r)}{dr} + \hat{\sigma}^2_{JK}(r)\frac{\sum_{t=1}^{T}\frac{1}{h}K'\left(\frac{r_{t\Delta_T} - r}{h}\right)}{\sum_{t=1}^{T} K\left(\frac{r_{t\Delta_T} - r}{h}\right)}\right), \qquad (2.20)$$

but as we shall see, this is unnecessarily complicated and can be simplified further by making use of the normal kernel and the estimators of Method 1.

Multiplying $\hat{\sigma}_1^2(\cdot)$ by our marginal density estimate $\hat{\pi}(\cdot)$, and differentiating we obtain

$$\frac{d}{dr}\left[\hat{\sigma}_1^2(r)\hat{\pi}(r)\right] = \frac{1}{Th^2\Delta}\sum_{t=1}^{T-1} K'\left(\frac{r-r_{t\Delta}}{h}\right)\left(r_{(t+1)\Delta}-r_{t\Delta}\right)^2 \tag{2.21}$$

because $\frac{dK(x/h)}{dx} = h^{-1}K'(x/h)$. Now, using (2.7) and $K'(x) = -xK(x)$, we have

$$\begin{aligned}
\hat{\mu}_2(r) &= \frac{1}{2\hat{\pi}(r)}\frac{d}{dr}\left[\hat{\sigma}_1^2(r)\hat{\pi}(r)\right] \\
&= \frac{1}{2\hat{\pi}(r)\Delta Th^2\sqrt{2\pi}h}\sum_{t=1}^{T-1}\exp\left(-\frac{(r-r_{t\Delta})^2}{2h^2}\right)(r_{t\Delta}-r)\left(r_{(t+1)\Delta}-r_{t\Delta}\right)^2. \qquad (2.22)
\end{aligned}$$

As can be seen from (2.22) with (2.20), the form of $\hat{\mu}_2(r)$ is much simpler than that of $\hat{\mu}_{JK}(r)$. Now to estimate the diffusion function, we utilize (2.8) and $\hat{\mu}_1(\cdot)$, as well as the information contained in the marginal density, $\hat{\pi}(\cdot)$. So

$$\begin{aligned}
\hat{\sigma}_2^2(r) &= \frac{2}{\hat{\pi}(r)}\int_0^r \hat{\mu}_1(u)\hat{\pi}(u)du \\
&= \frac{2}{\hat{\pi}(r)\Delta Th}\sum_{t=1}^{T-1}\left[r_{(t+1)\Delta}-r_{t\Delta}\right]\int_0^r K\left(\frac{u-r_{t\Delta}}{h}\right)du \\
&= \frac{2}{\hat{\pi}(r)\Delta T}\sum_{t=1}^{T-1}\left[r_{(t+1)\Delta}-r_{t\Delta}\right]\int_0^r K\left(\frac{u-r_{t\Delta}}{h}\right)d\left(\frac{u-r_{t\Delta}}{h}\right) \\
&= \frac{2}{\hat{\pi}(r)\Delta T}\sum_{t=1}^{T-1}\left[\Phi\left(\frac{r-r_{t\Delta}}{h}\right)-\Phi\left(-\frac{r_{t\Delta}}{h}\right)\right]\cdot\left[r_{(t+1)\Delta}-r_{t\Delta}\right], \qquad (2.23)
\end{aligned}$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard Normal random variable.

Equation (2.23) provides an explicit and computationally straightforward estimator for $\sigma^2(r)$. Let $\hat{m}_2(r) = \hat{\mu}_2(r)\hat{\pi}(r)$ and $\hat{V}_2(r) = \hat{\sigma}_2^2(r)\hat{\pi}(r)$. We now have the following propositions and their proofs are relegated to the Appendix.

Proposition 2.4. *Assume that Assumptions A.1, A.2 and A.3(iii) listed in the appendix hold. Then*

$$\begin{aligned}
E\left\{\int\left[\hat{m}_2(r)-m(r)\right]^2dr\right\} &= \frac{\Delta^2}{4Th^3}\cdot\frac{1}{4\sqrt{\pi}}\int\left(\mu^4(r)+\frac{3\sigma^4(r)}{\Delta^2}+\frac{6\mu^2(r)\sigma^2(r)}{\Delta}\right)\pi(r)dr + O\left(\frac{1}{Th}\right) \\
&+ \int\left(\frac{\Delta}{2}\frac{d}{dr}[\mu^2(r)\pi(r)] + \frac{h^2}{4}\frac{d^3}{dr^3}[\Delta\mu^2(r)\pi(r)+\sigma^2(r)\pi(r)] + o(h^2)\right)^2 dr.
\end{aligned}$$

(i) *When Δ is fixed but $h = d\ T^{-\frac{1}{5}}$ for some $d > 0$, we have*

$$E\left\{\int [\hat{m}_2(r) - m(r)]^2\,dr\right\} = \frac{C_1}{Th^3} + C_2\Delta^2 + C_3h^4 = D_1T^{-\frac{2}{5}} + D_2\Delta^2.$$

(ii) *When $\Delta = c\ h^2$ and $h = d\ T^{-\frac{1}{7}}$ for some $c > 0$ and $d > 0$, we have*

$$E\left\{\int [\hat{m}_2(r) - m(r)]^2\,dr\right\} = \frac{C_1}{Th^3} + C_2\Delta^2 + C_3h^4 = D_3T^{-\frac{4}{7}}. \tag{2.24}$$

Proposition 2.5. *Assume that Assumptions A.1, A.2 and A.3(iii) listed in the Appendix hold. Then*

$$\begin{aligned}E\left\{\int\left[\hat{V}_2(r) - V(r)\right]^2 dr\right\} &= \frac{4}{T}\cdot\int\left(\int\left[\mu^2(s) + \frac{\sigma^2(s)}{\Delta}\right]^2\left[\Phi\left(\frac{r-s}{h}\right) - \Phi\left(-\frac{s}{h}\right)\right]^2\pi(s)ds\right)dr\\ &+\ h^4\int\left(\int_0^r \frac{d^2}{dx^2}[\mu(x)\pi(x)]dx\right)^2 dr + o(h^4).\end{aligned}$$

(i) *When Δ is fixed but $h = d\ T^{-\frac{1}{5}}$ for some $d > 0$, we have*

$$E\left\{\int\left[\hat{V}_2(r) - V(r)\right]^2 dr\right\} = \frac{C_1}{T\Delta^2} + C_2h^4 = C_1T^{-\frac{4}{5}} + o\left(T^{-\frac{4}{5}}\right). \tag{2.25}$$

(ii) *When $\Delta = c\ h^2$ and $h = d\ T^{-\frac{1}{7}}$ for some $c > 0$ and $d > 0$, we have*

$$E\left\{\int\left[\hat{V}_2(r) - V(r)\right]^2 dr\right\} = \frac{C_1}{T\Delta^2} + C_2h^4 = D_1T^{-\frac{3}{7}} + o\left(T^{-\frac{3}{7}}\right). \tag{2.26}$$

Overall, we suggest using

- $h = d\ T^{-\frac{1}{5}}$ when Δ is fixed; and
- $\Delta = c\ h^2$ and $h = d\ T^{-\frac{1}{7}}$ when $\Delta \to 0$.

However, one may need to consider the pairs $(\hat{m}_1, \hat{V}_2)$ and $(\hat{m}_2, \hat{V}_1)$ separately, since $\hat{V}_2$ is constructed using $\hat{m}_1$, and $\hat{m}_2$ is based on $\hat{V}_1$.

Therefore, we may also suggest using

- $\Delta = c\ h$ and $h = d\ T^{-\frac{1}{6}}$ for Propositions 2.2 and 2.5. In this case, the resulting rates are

$$E\left\{\int [\hat{m}_1(r) - m(r)]^2\,dr\right\} = C_1T^{-\frac{4}{6}}, \tag{2.27}$$

$$E\left\{\int\left[\hat{V}_2(r) - V(r)\right]^2 dr\right\} = C_2T^{-\frac{4}{6}}; \tag{2.28}$$

- $\Delta = c\,h^2$ and $h = d\,T^{-\frac{1}{9}}$ for Proposition 2.3, and the resulting rate is

$$E\left\{\int \left[\hat{V}_1(r) - V(r)\right]^2 dr\right\} = C_3 T^{-\frac{4}{9}}; \tag{2.29}$$

- $\Delta = c\,h^2$ and $h = d\,T^{-\frac{1}{7}}$ for Proposition 2.4, and the resulting rate is

$$E\left\{\int \left[\hat{m}_2(r) - m(r)\right]^2 dr\right\} = C_4 T^{-\frac{4}{7}}. \tag{2.30}$$

Theoretically, we suggest using the pair $(\hat{\mu}_1(r), \hat{\sigma}_2^2(r))$. The empirical comparisons in Section 3 show that the pairs $(\hat{\mu}_1(r), \hat{\sigma}_1^2(r))$ and $(\hat{\mu}_2(r), \hat{\sigma}_2^2(r))$ are both appropriate for the two data sets. By considering both the theoretical properties and empirical comparisons of the proposed estimators, however, we would suggest using the pair $(\hat{\mu}_1(r), \hat{\sigma}_2^2(r))$ for the two sets of data.

2.3 *Method 3*

The estimators of the previous two methods rely on various alternative interpretations of the diffusion process. Neither of them place any restrictions on the drift nor the diffusion functions. For this method, we construct the diffusion function after placing the common mean-reverting parameterization $\mu(r;\theta) = \beta(\alpha - r)$ on the drift, similar to the approach taken by Aït-Sahalia (1996a). This restriction will allow us to see how the diffusion function is affected when compared to the previous methods' estimators.

The parameters β and α are estimated by ordinary least squares (OLS) estimation, and denoted by $\hat{\beta}$ and $\hat{\alpha}$, respectively. This suggests estimating $\mu(r;\theta)$ by

$$\hat{\mu}_3(r) = \mu(r;\hat{\theta}) = \hat{\beta}(\hat{\alpha} - r). \tag{2.31}$$

The estimated mean-reverting drift term can now be substituted into (2.8) together with the normal kernel density estimator for $\pi(\cdot)$ to construct our diffusion term $\hat{\sigma}_3^2(\cdot)$. Below we state the final estimator. More complete mathematical details of this derivation are relegated to the appendix.

From (2.8) and (2.31), we define a semiparametric estimator of $\sigma^2(r)$ of the form

$$\begin{aligned}
\hat{\sigma}_3^2(r) &= \frac{2}{\hat{\pi}(r)} \int_0^r \hat{\mu}_3(u)\hat{\pi}(u)du = \frac{2}{\hat{\pi}(r)} \int_0^r \mu(u;\hat{\theta})\hat{\pi}(u)du \\
&= \frac{2}{T\hat{\pi}(r)} \left\{ \hat{\beta} \sum_{t=1}^{T} (\hat{\alpha} - r_{t\Delta}) \left[\Phi\left(\frac{r - r_{t\Delta}}{h}\right) - \Phi\left(\frac{-r_{t\Delta}}{h}\right) \right] \right. \\
&+ \left. \frac{h}{\sqrt{2\pi}} \hat{\beta} \sum_{t=1}^{T} \left[\exp\left(-\frac{(r - r_{t\Delta})^2}{2h^2}\right) - \exp\left(-\frac{r_{t\Delta}^2}{2h^2}\right) \right] \right\}.
\end{aligned} \tag{2.32}$$

As can be seen, $\hat{\sigma}_3^2(r)$ has an explicit and computationally straightforward expression due to the use of the standard Normal kernel function. Let $\hat{V}_3(r) = \hat{\sigma}_3^2(r)\hat{\pi}(r)$. For the nonparametric estimator $\hat{V}_3^2(r)$, we establish the following proposition. Its proof is relegated to the Appendix.

Proposition 2.6. *Assume that Assumptions A.1, A.2 and A.3(ii) listed in the appendix hold. Then*

$$\begin{aligned} E\left\{\int \left[\hat{V}_3(r) - V(r)\right]^2 dr\right\} &= \frac{4\beta^2}{Th}\int \left(\int_0^r \int_0^r (\alpha-u)(\alpha-v)L\left(\frac{u-v}{h}\right)\pi(v)dvdu\right)dr \\ &+ 4h^4 \int \left(\int_0^r \beta(\alpha-u)\pi''(u)du\right)^2 dr. \end{aligned}$$

When $h = c\,T^{-\frac{1}{5}}$ for some $c > 0$, we have

$$E\left\{\int \left[\hat{V}_3(r) - V(r)\right]^2 dr\right\} = \frac{C_1}{Th} + C_2 h^4 = D_1 T^{-\frac{4}{5}} + o\left(T^{-\frac{4}{5}}\right). \tag{2.33}$$

Propositions 2.1–2.6 show that the MISEs of all the estimators except $\hat{m}_2(r)$ attain the minimum when Δ is fixed but h is proportional to $T^{-\frac{1}{5}}$. Therefore, it is justified to use $h = c \cdot T^{-\frac{1}{5}}$ with c to be specified later in the empirical comparisons in Section 3 below. In order to compare our empirical results with existing results, we also consider using the theoretical optimum value $h_a = \frac{c_a}{\ln(T)}T^{-\frac{1}{5}}$ used in Aït-Sahalia (1996a) for the Euro data.

3. TESTING FOR LINEARITY

To formally determine whether the assumption on linearity in the drift in Method 3 is appropriate for a given set of data, we consider testing

$$\mathcal{H}_0: \ \mu(r) = \mu(r;\theta_0) = \beta_0(\alpha_0 - r) \quad \text{versus} \quad \mathcal{H}_1: \ \mu(r) \neq \beta(\alpha - r) \tag{3.1}$$

for some $\theta_0 = (\alpha_0, \beta_0) \in \Theta$ and all $\theta = (\alpha, \beta) \in \Theta$, where Θ is a parameter space in R^2.

We approximate the semiparametric continuous–time diffusion model $dr_t = \beta(\alpha - r_t)dt + \sigma(r_t)dB_t$ by a semiparametric time series model of the form

$$Y_t = \beta(\alpha - r_{t\Delta}) + \sigma(r_{t\Delta})e_t, \tag{3.2}$$

where $Y_t = \frac{r_{(t+1)\Delta} - r_{t\Delta}}{\Delta}$, $\sigma(\cdot) > 0$ is unknown nonparametrically, and $e_t = \frac{B_{(t+1)\Delta} - B_{t\Delta}}{\Delta} \sim N\left(0, \Delta^{-1}\right)$.

We initially suggest a specification test of the form:

$$L_T = L_T(h) = \frac{\sum_{s=1}^T \sum_{t=1, t\neq s}^T K\left(\frac{r_{t\Delta} - r_{s\Delta}}{h}\right)\hat{\epsilon}_s\hat{\epsilon}_t}{\sqrt{2\sum_{s=1}^T \sum_{t=1}^T K^2\left(\frac{r_{t\Delta} - r_{s\Delta}}{h}\right)\hat{\epsilon}_s^2\hat{\epsilon}_t^2}}, \tag{3.3}$$

where $K(\cdot)$ is the standard normal density function, h is the bandwidth parameter, and $\hat{\epsilon}_t = Y_t - \hat{\mu}_3(r_{t\Delta})$.

As argued in Li and Wang (1998), and Gao and King (2003), the test statistic L has two main features: (i) the test appears to be more straightforward computationally, and (ii) it is not required to get a consistent estimator of the conditional variance involved. This implies that the applicability of the test for testing the drift does not depend on the structure of the conditional variance. Therefore, it is proper to apply the test to our case study once the bandwidth is appropriately chosen.

As shown in the appendix, $L_T(h)$ converges in distribution to the standard normality when $T \to \infty$. Our own experience and others show that the finite sample performance of $L_T(h)$ is not good in particular when h is chosen based on an optimal estimation procedure, such as the cross–validation criterion. The main reasons are as follows: (a) the use of an estimation based optimal value may not be optimal for testing purposes; and (b) the rate of convergence of $L_T(h)$ to the asymptotic normality is quite slow even when $\{e_t\}$ is now a sequence of normally distributed errors. In order to improve the finite sample performance of $L_T(h)$, we impose the following two measurements. We first propose using an adaptive version of $L_T(h)$ over a set of all possible bandwidth values. Second, we use a simulated critical value for computing the size and power values of the adaptive version of $L_T(h)$ instead of using an asymptotic value of $l_{0.05} = 1.645$ at the 5% level. To the best of our knowledge, both the measurements are novel in this kind of testing for linearity in drift under such a semiparametric setting.

We then propose using an adaptive test of the form

$$L^* = \max_{h \in H_T} L_T(h), \tag{3.4}$$

where $H_T = \left\{h = h_{\max} a^k : \ h \geq h_{\min}, \ k = 0, 1, 2, \ldots\right\}$, in which $0 < h_{\min} < h_{\max}$, and $0 < a < 1$. Let J_T denote the number of elements of H_T. In this case, $J_T \leq \log_{1/a}(h_{\max}/h_{\min})$.

Simulation Scheme: We now discuss how to obtain a simulated critical value for L^*. The exact α–level critical value, l_α^* $(0 < \alpha < 1)$ is the $1 - \alpha$ quantile of the exact finite-sample distribution of L^*. Because both the distribution of $\{e_t\}$ and θ_0 are unknown, l_α^* cannot be evaluated in practice. We therefore suggest approximating l_α^* by a simulated α–level critical value, l_α, using the following simulation procedure:

1. For each $t = 1, 2, \ldots, T$, generate $Y_t^* = \hat{\mu}_3(r_{t\Delta}) + \hat{\sigma}_3(r_{t\Delta}) e_t^*$, where $\{e_t^*\}$ is sampled randomly from $N(0, \Delta^{-1})$ for Δ to be specified as $\Delta = \frac{20}{250}$ for the monthly data and $\Delta = \frac{1}{250}$ for the daily data, which $\hat{\mu}_3(r)$ and $\hat{\sigma}_3(r)$ are as defined in (2.31) and (2.32), respectively. In practice, a kind of truncation procedure may be needed to ensure the positivity of $\hat{\sigma}_3(\cdot)$.

2. Use the data set $\{Y_t^* : t = 1, 2, \ldots, T\}$ to re-estimate θ_0. Denote the resulting estimate by $\widehat{\theta}^*$. Compute the statistic $\widehat{L}^*$ that is obtained by replacing Y_t and $\hat{\theta}$ with Y_t^* and $\widehat{\theta}^*$ on the right–hand side of (3.4).

3. Repeat the above steps M times and produce M versions of $\hat{L}^*$ denoted by $\hat{L}_m^*$ for $m = 1, 2, \ldots, M$. Use the M values of $\hat{L}_m^*$ to construct their empirical distribution function, that is, $F^*(u) = \frac{1}{M}\sum_{m=1}^{M} I(\hat{L}_m^* \leq u)$. Use the $1-\alpha$ quantile of the empirical distribution function to estimate the simulated α–level critical value, l_α.

We now state the following results and their proofs are relegated to the appendix.

Proposition 3.1. *Assume that Assumptions A.1, A.2 and A.4 hold. Then under* $\mathcal{H}_0$

$$\lim_{T\to\infty} P(L^* > l_\alpha) = \alpha.$$

The main result on the behavior of the test statistic L^* under $\mathcal{H}_0$ is that l_α is an asymptotically correct α–level critical value under any model in $\mathcal{H}_0$.

Proposition 3.2. *Assume that the conditions of Proposition 3.1 hold. Then under* $\mathcal{H}_1$

$$\lim_{T\to\infty} P(L^* > l_\alpha) = 1.$$

Proposition 3.2 shows that a consistent test will reject a false $\mathcal{H}_0$ with probability approaching one as $T \to \infty$.

To implement Propositions 3.1 and 3.2 to real data analysis, we need to compute the P–value of the test for each given set of data as follows:

1. For each data set, compute
$$L^* = \max_{h\in H_T} L_T(h), \tag{3.5}$$
where $H_T = \left\{h = h_{\max}a^k : \ h \geq h_{\min}, \ k = 0, 1, 2, \ldots\right\}$, in which $T^{-\frac{1}{5}} = h_{\min} < h_{\max} = 1.1\,(\text{loglog}T)^{-1}$, and $a = 0.8$ based on some preliminary calculations of the size and power values of $L_T(h)$ for a range of bandwidth values.

2. Compute $\hat{\epsilon}_t = Y_t - \hat{\mu}_3(r_{t\Delta})$ and then generate a sequence of bootstrap resamples $\{\widehat{\epsilon}_t^*\}$ using the wild bootstrap method (see Härdle and Mammen 1993) from $\{\hat{\epsilon}_t\}$.

3. Generate $\widehat{Y}_t^* = \hat{\mu}_3(r_{t\Delta}) + \widehat{\epsilon}_t^*$. Compute the corresponding version $\widehat{L}^*$ of L^* based on $\{\widehat{Y}_t^*\}$.

4. Repeat the above steps N times to find the bootstrap distribution of $\widehat{L}^*$ and then compute the proportion that $L^* < \widehat{L}^*$. This proportion is a simulated P–value of L^*.

With the three methods now constructed, it is of interest to see how they compare. More precisely, how the restriction placed on the drift function affects the estimation of the diffusion. A number of studies note that the prices of derivatives are crucially dependent on the specification of the diffusion function (see Aït-Sahalia 1996a), therefore qualifying parametric restrictions on the drift function. The test statistic L^* is also applied to formally test linearity in the drift using the simulated P–value.

4. EMPIRICAL COMPARISONS

4.1 *The data*

We now apply the three pairs of estimators constructed previously to two different financial data. A conclusion regarding which method best fits each data set will be offered. Also suggested here is an optimal bandwidth, based on both the theoretical properties of the MISEs in Propositions 2.1–2.4 and a comparison a number of common forms used in the literature.

To analyze the effect the sampling frequency (interval) has on the results, we use both monthly (low frequency) and daily (high frequency) sampled data. The three-month Treasury Bill rate data given in Figure 1 below are sampled monthly over the period from January 1963 to December 1998, providing 432 observations (i.e. $T = 432$; source: H-15 Federal Reserve Statistical Release). The number of working days in a year (excluding weekends and public holidays) are assumed to be 250 (and 20 working days per month). This gives $\Delta = \frac{20}{250}$. Chan, *et al.* (1992) offer evidence that the Fed rates are stationary by showing that the autocorrelations of month-to-month changes are neither large nor consistently positive or negative.

Figure 1 near here

Figure 2 near here

The second data set used in this analysis to compare and contrast the primary results is the high frequency seven–day Eurodollar deposit rate. The data are sampled daily from 1 June 1973 to 25 February 1995. This provides us with $T = 5505$ observations. Just as for the Fed data, holidays have not been treated and Monday is taken as the first day after Friday as there are no obvious weekend effects (Aït-Sahalia 1996b). Thus, our sampling interval $\Delta = \frac{1}{250}$. The data are plotted in Figure 2 above. Aït-Sahalia (1996a) rejects the null hypothesis of nonstationarity on this data based on results of an augmented Dickey-Fuller nonstationarity test.

4.2 *Bandwidth selection*

The choice of bandwidth or smoothing parameter is critical in any application of nonparametric kernel density and regression estimation. Propositions 2.1–2.6 provide some kind of guidance on how to choose the bandwidth in practice. Overall, we suggest using $h = d\, T^{-\frac{1}{5}}$ when Δ is fixed; and $\Delta = c\, h^2$ and $h = d\, T^{-\frac{1}{7}}$ when $\Delta \to 0$. As we deal with the fixed Δ in this paper, the forms of the bandwidth selectors used are listed below.

$$\begin{aligned} h_1 &= SS \times T^{-\frac{1}{5}} \\ h_2 &= \frac{1}{10} \times T^{-\frac{1}{5}} \\ h_3 &= \frac{1}{4} \times T^{-\frac{1}{5}} \\ h_4 &= 1.06 \times SS \times T^{-\frac{1}{5}} \end{aligned} \tag{4.1}$$

where T is the number of observations and SS is the standard deviation of the data. Thus h_1 and h_4 can, in this sense, be regarded as 'data-driven' bandwidth choices. Pritsker (1998) states that h_4 is the MISE–minimizing bandwidth assuming the data came from a normal distribution with variance SS^2. As can be seen from Propositions 2.1–2.4, the second and third bandwidths can be written as $h = c \cdot T^{-\frac{1}{5}}$, where c is a constant chosen to minimize the asymptotic MISE of the estimator involved. For example, when $\pi(r) = \frac{1}{\sqrt{2\pi}\sigma_\pi} e^{-\frac{(r-\mu_\pi)^2}{2\sigma_\pi^2}}$, it follows from Proposition 2.1 that the theoretically optimum value of h is $2^{1/5}\sigma_\pi T^{-\frac{1}{5}}$. Thus, both h_2 and h_3 are attainable with a suitable choice of σ_π. We then selected the bandwidth that gave, not only a smooth and informative marginal density estimate (see Figure 3 below), but one that possessed the greatest consistency between the three drift and diffusion estimators (see Figures 4 and 5 below for the Fed data and Figures 6 and 7 below for the Euro data). For the Euro data, we also borrowed bandwidth choices used by Aït-Sahalia (1996a, 1996b), as he has also used this data set.

4.3 *Results and comparisons*

Figure 3 near here

For the Fed data, we 'plugged-in' the bandwidths h_1, h_2, h_3, h_4 of $0.00949, 0.0297, 0.0743$ and 0.01 respectively, and estimated the marginal density, drift and diffusion functions. It was found that the optimal bandwidth refers to h_2 (0.0297). The density estimate produced for h_2 shown in Figure 3(A), appears to contain sufficient information. It is apparent with this choice of bandwidth, even though the high rate period of 1980–82 is included in the sample, the amount of information retained has produced a less accentuated right tail. Its shape and symmetry about 0.055 closely resembles that of a Gaussian density. The densities produced with the smaller bandwidths were overly informative while larger bandwidths resulted in smooth, quadratic like curves. An in-depth comparison for the bandwidths h_1 and h_3 has also been done and plots are available upon request.

$\boxed{\text{Figures 4 and 5 near here}}$

Comparisons of the drift and diffusion estimators give similar results. The three drift and diffusion estimators constructed using our optimal bandwidth choice are superimposed for comparative purposes in Figure 4. The best estimators for the Fed data are given in Figure 5. The drift functions $\hat{\mu}_1(\cdot)$ and $\hat{\mu}_2(\cdot)$ inherit similar nonlinearity for interest rates over the entire range of r. The best linear mean-reverting drift estimate is plotted in the bottom of Figure 4 and then Figure 5. The ordinary least squares method gave estimates for the parameters α and β of $\hat{\mu}_3(\cdot)$, of 0.07170 and 0.2721, respectively.

Looking now at the diffusion estimators we see $\hat{\sigma}_1^2(\cdot)$ and $\hat{\sigma}_2^2(\cdot)$ especially, are very similar. They closely resemble one another over the entire range of r in both shape and magnitude (see Figures 4 and 6). The curvature of $\hat{\sigma}_1^2(\cdot)$ and $\hat{\sigma}_2^2(\cdot)$ is close to that of a quadratic. This gives some support for the process of Brennan and Schwartz (1980) whose instantaneous variance increased at a rate proportional to r^2 and to Chan, *et al.* (1992) who found $\sigma \propto r^{1.49}$.

The best estimator $\hat{\sigma}_3^2(\cdot)$ (see the bottom of Figure 4 and then Figure 7) is comparable to $\hat{\sigma}_1^2(\cdot)$ and $\hat{\sigma}_2^2(\cdot)$ for low to moderate rates (i.e. rates below 12%). It lies above $\hat{\sigma}_1^2(\cdot)$ and $\hat{\sigma}_2^2(\cdot)$ for a greater (negative) mean–reverting force ($\hat{\mu}_3(\cdot) < \hat{\mu}_1(\cdot),\ \hat{\mu}_2(\cdot)$). It appears to be a linearly increasing function of the level of r (as in CIR 1985) for rates below 14% and this is apparent in the bottom of Figure 4 and then Figure 5.

Given that the two nonparametric drift estimators are unlike the linear mean-reverting specification (with their respective diffusion estimates $\hat{\sigma}_1^2(\cdot)$ and $\hat{\sigma}_2^2(\cdot)$ differing from $\hat{\sigma}_3^2(\cdot)$), we suggest here that the mean-reverting function is not appropriate for this data. Thus, $\hat{\mu}_3(\cdot)$ does indeed affect the estimation of the diffusion function and hence the pricing of derivative securities. Based on the above, for the monthly sampled Federal funds rate data, we believe that $\hat{\mu}_3(\cdot)$ imposes an unnecessary restriction which results in the misspecification of the diffusion function. Either of the pair $(\hat{\mu}_1(\cdot), \hat{\sigma}_1^2(\cdot))$ or $(\hat{\mu}_2(\cdot), \hat{\sigma}_2^2(\cdot))$ is recommended for this set of data.

The specification test L^* proposed in Section 3 was then applied in order to formally reject linearity in the drift. The null hypothesis $H_0 : \mu(r) = \beta(\alpha - r)$ of linearity is rejected at the 5% significance level. We obtain a simulated p–value of $P \leq 0.001$, which is much smaller than the 5% significance level.

$\boxed{\text{Figures 6 and 7 near here}}$

Now to the Euro data. The forms of $h_1, \ldots, h_4$ were applied to this data. In this case, because we have 5505 observations, the bandwidths were respectively $0.006413, 0.01786, 0.044$, and 0.0068. We also consider the bandwidth $h_a = 0.01347$ used by Aït-Sahalia (1996a) for the same data. The best estimators for the Euro data are given in Figures 6 and 7. Surprisingly, our optimal bandwidth for the Euro data also corresponds to $h_2 = \frac{1}{10} \times T^{-\frac{1}{5}}$. Similar to the

Fed data analysis where we concluded h_1 and h_4 severely undersmooth the density estimate, we infer similar results for the Euro data. With the Euro data consisting of 5505 observations, it is clear we would obtain a much smaller sample variance than the Fed data (consisting of 432 observations). Our best marginal density estimate, drift and diffusion estimators for the Euro data are reported in Figures 3(B) and 6 above. The drift and diffusion estimators are superimposed for comparative purposes in Figure 7 above. It is apparent that the two unrestricted drift estimators, $\hat{\mu}_1(\cdot)$ and $\hat{\mu}_2(\cdot)$ inherit very similar nonlinearity over the entire range of r (see Figure 7). Both estimators seem to exhibit mean reversion for $r > 15\%$, while our linear mean–reverting drift estimator $\hat{\mu}_3(\cdot)$ (see also Figure 7) is unexpectedly comparable to $\hat{\mu}_1(\cdot)$ and $\hat{\mu}_2(\cdot)$ for $r < 20\%$. The diffusion functions constructed using the unrestricted drift closely resemble one another, and are practically indistinguishable for the entire range of r. They both increase somewhat linearly for $r < 11\%$, both increase at a greater rate than r for $r > 11\%$ and possess a 'hump' at $r = 15\%$ where the instantaneous variance jumps (see Figure 7).

For the Euro data, the best OLS estimates of α and β are 0.08308 and 1.596 respectively, which are analogous to the first step OLS estimates computed by Aït-Sahalia (1996a). We see the Euro data has stronger mean-reversion than the Fed data ($\beta = 0.2721$) which is most likely the result of more frequent sampling. Aït-Sahalia (1996a) also found β to be larger for shorter-maturity proxies (seven-day Eurodollar vs. three-month T-bill). We see from Figure 7 above that the similarity of the three drift estimators may suggest the mean-reverting specification for drift is applicable (at least for $r < 20\%$). The similarity of the two diffusion functions $\hat{\sigma}_1^2(\cdot)$ and $\hat{\sigma}_2^2(\cdot)$ and the deviation of $\hat{\sigma}_3^2(\cdot)$ from these two estimators may however, suggest otherwise.

The proposed test L^* was run on the data. Our simulation returns a simulated p–value of $P \leq 0.001$, which directs us to strongly reject the null hypothesis of linearity at the 5% significance level. A likely explanation for this result is that as we have a long and frequently sampled data set, the use of even the slightest deviant from the actual drift will result in a compounded error effect or deviation of the specified model from the actual process. Thus, we suggest the mean-reverting drift function specification is not appropriate for high frequency data (more strongly than for the monthly sampled Fed data). To determine whether the high rate period of 1980-82 was responsible for the strong rejection of linearity, we also ran the test on the sub-sample and calculated p–value. The result suggests that linearity in the drift is also rejected for the sub-sample. Not only did we run the linearity test on the sub-sample of the Euro data, but we also estimated the drift and diffusion estimators for this set. Here, we'plugged-in' the bandwidth value of $h_s = 0.01653$ which Aït-Sahalia (1996a) reported was optimal for this sub-sample. The two unrestricted drift estimators exhibit similar nonlinearities for $r < 10\%$ with mean-reversion for $r > 10\%$ while their corresponding diffusion estimators resemble the quadratic diffusion specification of Brennan and Schwartz (1980). The diffusion estimator $\hat{\sigma}_3^2(\cdot)$ appears to be comparable to the constant volatility specification of Vasicek (1977) for $r < 12\%$. Such a difference in form is evidence against the linear mean-reverting drift function.

5. CONCLUSION

Many different specifications for the drift and diffusion functions of the common Itô diffusion process have opened up a new area of research. Some recently published papers empirically compare the plethora of proposed models.

In this study, we adopted a similar nonparametric approach as Aït-Sahalia (1996a) and Jiang and Knight (1997) to estimate the diffusion process. We used two popular short rates: the three–month Treasury Bill and the seven–day Eurodollar deposit rates. Based on our analysis, we suggested the use of the bandwidth $h = \frac{1}{10} \times T^{-\frac{1}{5}}$ for both sets of data. We then demonstrated how the bandwidth choice can have a dramatic effect on the drift and diffusion estimates. We rejected linearity in the drift despite theoretic economic justification, both empirically and more formally with the specification test L. Overall, we would suggest using the pair $(\hat{\mu}_1(r), \hat{\sigma}_2^2(r))$ for the two sets of data.

The nonparametric specification of the diffusion function of Method 3 is seen to differ significantly from the diffusion function estimates of Methods 1 and 2. We thus conclude that restrictions on the drift have a greater effect on the volatility than what is suggested in the literature. Our empirical comparisons suggest $\hat{\sigma}_3^2(\cdot)$ is misspecified primarily as a result of the assumptions imposed on it. That is, the assumption of a linear mean–reverting drift has a substantial effect on the final form of the diffusion. We suggest relaxing the drift assumption. The unrestricted drift estimates indicate that the fitting of a second order polynomial may be more appropriate. It would be interesting to regenerate these estimators but with a quadratic restriction on the drift. The results could then be compared with the diffusion function estimates generated by Method 1 and 2. Such an exercise is deferred to future research.

Extended research in this area should include a comparison of the nonparametric density, drift and diffusion estimators with those implied by some of the popular parametric models (see Aït-Sahalia 1999). In particular, to consider our nonparametrically estimated marginal density with say, the Gamma density of CIR (1985) applied to the Eurodollar data. Or, to review how the unrestricted diffusion estimators actually compare to the diffusion specifications of CIR (1985), Brennan and Schwartz (1980) and Chan, *et al.* (1992). Additionally, it would be useful to apply both the popular existing parametric models and our nonparametric estimates to price derivatives (e.g. bond options) in an attempt to determine the accuracy of the prices computed.

In this study, we selected our optimal bandwidth based on comparing the density, drift and diffusion estimators computed for a number of different bandwidths. That is, we used the 'plug-in' method and as such only approximated the optimum value. This procedure was selected because other interval based error-minimizing techniques requires one to discretize our continuous–time model. However, this is impractical for our case for numerous reasons. Further work is needed to select a bandwidth based on some minimization function.

ACKNOWLEDGEMENTS

We acknowledge comments from conference and seminar participants at The University of Western Australia in 2002–2004, Monash University in 2002, The Australasian Meeting of The Econometric Society in Sydney in 2003, London School of Economics, Cornell University, Northwestern University, Yale University and Princeton University in 2003, in particular, Yacine Aït–Sahalia, Torben Andersen, Heather Anderson, Donald Andrews, Songxi Chen, Jianqing Fan, Yongmiao Hong, Maxwell King, Oliver Linton, Michael McAleer, Adrian Pagan, Peter Phillips, Peter Robinson, Howell Tong and Qiwei Yao. We would like to thank the Australian Research Council for its financial support.

APPENDIX

A.1 Assumptions

Assumption A.1. (i) Assume that the process $\{r_t\}$ is strictly stationary and α-mixing with the mixing coefficient $\alpha(t) = C_\alpha \alpha^t$ defined by

$$\alpha(t) = \sup\{|P(A \cap B) - P(A)P(B)| : A \in \Omega_1^s, B \in \Omega_{s+t}^\infty\}$$

for all $s, t \geq 1$, where $0 < C_\alpha < \infty$ and $0 < \alpha < 1$ are constants, and Ω_i^j denotes the σ-field generated by $\{r_t : i \leq t \leq j\}$.

(ii) The bandwidth parameter h satisfies that

$$\lim_{T\to\infty} h = 0 \quad \text{and} \quad \lim_{T\to\infty} Th^2 = \infty.$$

Assumption A.2. (i) The density function $\pi(r)$ is three times continuously differentiable in r.

(ii) The drift and the diffusion functions $\mu(r)$ and $\sigma^2(r)$ are three times continuously differentiable in $r \in R^+ = [0, \infty)$, and $\sigma(r) > 0$ on R^+.

(iii) The integral of $\overline{\mu}(v) = \frac{1}{\sigma^2(v)} \exp\left(-\int_v^{\bar{v}} 2\frac{\mu(x)}{\sigma^2(x)} dx\right)$ converges at both boundaries of R^+, where $\bar{v}$ is fixed in R^+.

(iv) The integral of $s(v) = \exp\left(\int_v^{\bar{v}} 2\frac{\mu(x)}{\sigma^2(x)} dx\right)$ diverges at both boundaries of R^+.

Assumption A.3. (i) The second derivative of $\pi(r)$, $\pi''(r)$, is square integrable.

(ii) The following functions are integrable:

$$\left(r\pi'(r)\right)^2 \quad \text{and} \quad \left(\int_0^r \pi''(u)du\right)^2.$$

for $i = 1, 2$.

(iii) The following functions are integrable:

$$\mu^4(r)\pi(r),\ \sigma^4(r)\pi(r),\ \left(\frac{d^2}{dr^2}[\mu(r)\pi(r)]\right)^2,\ \left(\frac{d^i}{dr^i}[\mu^2(r)\pi(r)]\right)^2,\ \left(\frac{d^i}{dr^i}[\sigma^2(r)\pi(r)]\right)^2$$

for $i = 1, 2, 3$, and

$$\left(\int_0^r \frac{d^2}{dx^2}[\mu(x)\pi(x)]dx\right)^2.$$

Assumption A.4. (i) For $\psi(r) = \sigma^2(r)$ or $\sigma^4(r)$, $\psi(r)$ satisfies the Lipschitz type condition: $|\psi(r+v) - \psi(r)| \leq \Psi(r)|v|$ for $v \in S$ (any compact subset of R^1) and $E\left[\Psi^2(r)\right] < \infty$.

(ii) Assume that the set H_T has the structure of (3.4) with $c_{\max}\left(\text{loglog}T\right)^{-1} = h_{\max} > h_{\min} \geq T^{-\gamma}$ for some constant γ such that $0 < \gamma < \frac{1}{3}$.

Remark A.1. Assumptions A.1–A.4 are quite natural in this kind of problem. Assumption A.1(i) assumes the α–mixing condition, which is weaker than the β–mixing condition. Assumption A.1(ii) ensures that the theoretically optimum value of $h = c \cdot T^{-1/5}$ can be used. Assumption A.2 is equivalent to Assumption A1 of Aït-Sahalia (1996a), requiring the existence and uniqueness of a strong solution to model (1.1). Assumption A.3 basically requires that all the integrals involved in Propositions 2.1–2.6 do exist. Assumption A.4 is only used for the establishment and proof of Propositions 3.1 and 3.2. Similar conditions have been assumed in Assumptions 2 and 6 of Horowitz and Spokoiny (2001), and Assumption A.2 of Chen and Gao (2004).

A.2 Proof of Propositions 2.1–2.6

As there are some similarities among the proofs of Propositions 2.1–2.6, we only provide the proof of Propositions 2.4 and 2.5 in some detail. However, the details of the other proofs are available upon request.

Recall that $Y_t = \frac{r_{(t+1)\Delta} - r_{t\Delta}}{\Delta}$ and observe that

$$Y_t = \mu(X_t) + \sigma(X_t)\epsilon_t,$$

where $X_t = r_{t\Delta}$ and $\epsilon_t = \frac{B_{(t+1)\Delta} - B_{t\Delta}}{\Delta}$.

We now have

$$\begin{aligned}
\hat{m}_2(r) &= \hat{\mu}_2(r)\hat{\pi}(r) = \frac{\Delta}{2Th^2}\sum_{t=1}^{T-1}\frac{(X_t - r)}{h}K\left(-\frac{(X_t - r)}{h}\right)Y_t^2 \\
&= \frac{\Delta}{2Th^2}\sum_{t=1}^{T-1}\frac{(X_t - r)}{h}K\left(-\frac{(X_t - r)}{h}\right)\left[2\mu(X_t)\sigma(X_t) + \mu^2(X_t) + \sigma^2(X_t)\epsilon_t^2\right] \\
&= \frac{\Delta}{Th^2}\sum_{t=1}^{T-1}\frac{(X_t - r)}{h}K\left(-\frac{(X_t - r)}{h}\right)\mu(X_t)\sigma(X_t)\epsilon_t \\
&+ \frac{\Delta}{2Th^2}\sum_{t=1}^{T-1}\frac{(X_t - r)}{h}K\left(-\frac{(X_t - r)}{h}\right)\left[\mu^2(X_t) + \sigma^2(X_t)\epsilon_t^2\right].
\end{aligned}$$

Thus, a Taylor expansion implies that as $h \to 0$

$$
\begin{aligned}
E\left[\hat{m}_2(r)\right] &= \frac{-\Delta}{2h^2} \int \frac{r-s}{h} K\left(\frac{r-s}{h}\right) [\mu^2(s) + \sigma^2(s)\sigma_0^2]\pi(s) ds \\
&= \frac{-\Delta}{2h} \int xK(x) \left[\mu^2(r-xh) + \sigma^2(r-hx)\sigma_0^2\right] \pi(r-hx) dx \\
&= \frac{-\Delta}{2h} \left(-hp'(r) + \frac{h^3}{6} \int x^4 K(x) p^{(3)}(\xi) dx\right) \\
&= \frac{\Delta}{2} p'(r) + \frac{\Delta}{4} p^{(3)}(r) h^2 + o(h^2), \qquad \text{(A.1)}
\end{aligned}
$$

where $\sigma_0^2 = E[\epsilon_t^2] = \Delta^{-1}$, $p(r) = \mu^2(r) + \sigma^2(r)\sigma_0^2$, and ξ lies between $r - hx$ and r.

This implies

$$
E\left[\hat{m}_2(r)\right] = m(r) + \frac{\Delta}{2} \frac{d}{dr}[\mu^2(r)\pi(r)] + \frac{\Delta}{4} p^{(3)}(r) h^2 + o(h^2). \qquad \text{(A.2)}
$$

Let

$$
Z_t = \left(\frac{X_t - r}{h}\right) K\left(-\frac{X_t - r}{h}\right) \mu(X_t)\sigma(X_t)\epsilon_t
$$

and

$$
W_t = \left(\frac{X_t - r}{h}\right) K\left(-\frac{X_t - r}{h}\right) \left[\mu^2(X_t) + \sigma^2(X_t)\epsilon_t^2\right].
$$

Similarly,

$$
\begin{aligned}
\hat{m}_2(r) - E\left[\hat{m}_2(r)\right] &= \frac{\Delta}{Th^2} \sum_{t=1}^{T-1} Z_t + \frac{\Delta}{2Th^2} \sum_{t=1}^{T-1} (W_t - E[W_t]) \\
&\equiv I_{1T} + I_{2T}.
\end{aligned}
$$

Analogous to (A.1), we obtain that as $h \to 0$

$$
\begin{aligned}
E[I_{1T}^2] &= \frac{\Delta^2}{T^2 h^4} (1 + o(1)) \sum_{t=1}^{T-1} E\left[Z_t^2\right] + \sum_{t=1}^{T-1} \sum_{s=1, \neq t}^{T-1} E\left[Z_s Z_t\right] \\
&= (1 + o(1)) \frac{\Delta^2 \sigma_0^2}{Th^4} \int \left(\frac{s-r}{h}\right)^2 K^2\left(\frac{r-s}{h}\right) \mu^2(s)\sigma^2(s)\pi(s) ds \\
&= (1 + o(1)) \frac{\Delta^2 \sigma_0^2}{Th^3} q(r) \int x^2 K^2(x) dx + \frac{\Delta^2 \sigma_0^2}{Th} q''(r) \int x^4 K^2(x) dx, \qquad \text{(A.3)}
\end{aligned}
$$

where $q(r) = \mu^2(r)\sigma^2(r)\pi(r)$.

Similarly, we can show that as $h \to 0$

$$
E[I_{2T}^2] = \frac{\Delta^2}{4Th^3} (1 + o(1)) \left[\mu^4(r) + 3\sigma^4(r)\sigma_0^4 + 2\mu^2(r)\sigma^2(r)\sigma_0^2\right] \int x^2 K^2(x) dx. \qquad \text{(A.4)}
$$

Equations (A.1)–(A.4) then imply Proposition 2.4.

Observe that

$$
\begin{aligned}
E\left[\hat{V}_2(r)\right] &= \frac{2}{h}E\left[\mu(X_t)\int_0^r K\left(\frac{u-X_t}{h}\right)du\right] \\
&= \frac{2}{h}\int_o^r \left[\int K\left(\frac{u-v}{h}\right)\mu(v)\pi(v)dv\right]du \\
&= 2\int_0^r \mu(u)\pi(u)du + h^2\int_0^r \frac{d^2}{du^2}[\mu(u)\pi(u)]du \\
&= V(r) + h^2\int_0^r \frac{d^2}{du^2}[\mu(u)\pi(u)]du + o(h^2) \qquad \text{(A.5)}
\end{aligned}
$$

using a Taylor expansion.

Let $\Psi(X_t) = \Phi\left(\frac{r-X_t}{h}\right) - \Phi\left(\frac{-X_t}{h}\right)$. Similar to (A.3) and (A.4), one can obtain that as $T \to \infty$

$$
\begin{aligned}
E\left(\hat{V}_2(r) - E\left[\hat{V}_2(r)\right]\right)^2 &= E\left\{\frac{2}{T}\sum_{t=1}^{T-1}\left(Y_t\Psi(X_t) - E\left[Y_t\Psi(X_t)\right]\right)\right\}^2 \\
&= \frac{4}{T}\int\left[\mu^2(s) + \Delta^{-1}\sigma^2(s)\right]\Psi^2(s)\pi(s)ds + o\left(\frac{1}{T}\right). \qquad \text{(A.6)}
\end{aligned}
$$

Proposition 2.5 then follows from (A.5), (A.6) and

$$
E\left(\hat{V}_2(r) - V(r)\right)^2 = E\left(\hat{V}_2(r) - E\left[\hat{V}_2(r)\right]\right)^2 + \left(E\left[\hat{V}_2(r)\right] - V(r)\right)^2.
$$

A.3 Proofs of Propositions 3.1 and 3.2

To prove Propositions 3.1 and 3.2, we need only to verify certain conditions of Theorems 3.1 and 3.2 of Chen and Gao (2004). As this paper is only concerned with the specification of linearity in the conditional mean in the form of (3.2), Assumption A.1 of Chen and Gao (2004) holds automatically. As the estimate $\hat{\theta}$ involved in (2.31) is the least-squares estimator, Assumption A.1 of Chen and Gao (2004) also holds automatically. In addition, various parts of Assumption A.2 of Chen and Gao (2004) also hold trivially, since $\{e_t\}$ in (3.2) is a sequence of independent normal errors and also independent of $\{r_{t\Delta}\}$. Note that Assumption A.1(i) of this paper is the same as Assumption A.1(ii) of Chen and Gao (2004). Thus, all the corresponding conditions of Theorems 3.1 and 3.2 of Chen and Gao (2004) can be satisfied. We therefore omit the detailed proofs of Propositions 3.1 and 3.2. However, the details are available from the authors upon request.

A.4 Proof of Equation (2.32)

Keeping in mind that $K(x/h) = \frac{1}{\sqrt{2\pi}}\exp\{-\frac{x^2}{2h^2}\}$ and $\hat{\pi}(r) = \frac{1}{Th}\sum_{t=1}^{T} K\left(\frac{r-r_{t\Delta}}{h}\right)$, we now derive our estimator, $\hat{\sigma}_3^2(\cdot)$. Recall equation (2.8)

$$
\hat{\sigma}_3^2(r) = \frac{2}{\hat{\pi}(r)}\int_0^r \hat{\mu}(u;\hat{\theta})\hat{\pi}(u)du.
$$

Now, evaluating the integral on the right of this identity,

$$
\int_0^r \mu(u;\hat{\theta})\hat{\pi}(u)du = \int_0^r \hat{\beta}\left(\hat{\alpha} - u\right)\frac{1}{Th}\sum_{t=1}^{T} K\left(\frac{u-r_{t\Delta}}{h}\right)du
$$

$$
\begin{aligned}
&= \frac{\hat{\beta}}{Th}\sum_{t=1}^{T}\int_0^r (\hat{\alpha}-u)\,K\left(\frac{u-r_{t\Delta}}{h}\right)du \\
&= \frac{\hat{\beta}}{Th}\sum_{t=1}^{T}\int_0^r (\hat{\alpha}-r_{t\Delta}+r_{t\Delta}-u)\,K\left(\frac{u-r_{t\Delta}}{h}\right)du \\
&= \frac{\hat{\beta}}{Th}\sum_{t=1}^{T}\left(\int_0^r (\hat{\alpha}-r_{t\Delta})\,K\left(\frac{u-r_{t\Delta}}{h}\right)du+\int_0^r (r_{t\Delta}-u)\,K\left(\frac{u-r_{t\Delta}}{h}\right)du\right) \\
&= \frac{\hat{\beta}}{Th}\sum_{t=1}^{T}\left((\hat{\alpha}-r_{t\Delta})\int_0^r K\left(\frac{u-r_{t\Delta}}{h}\right)du-\int_0^r [u-r_{t\Delta}]\,K\left(\frac{u-r_{t\Delta}}{h}\right)du\right) \\
&= \frac{\hat{\beta}}{Th}\sum_{t=1}^{T}\left((\hat{\alpha}-r_{t\Delta})\,h\int_{-\frac{r_{t\Delta}}{h}}^{\frac{r-r_{t\Delta}}{h}} K(v)dv-h^2\int_{-\frac{r_{t\Delta}}{h}}^{\frac{r-r_{t\Delta}}{h}} vK(v)dv\right) \\
&= \frac{\hat{\beta}}{T}\sum_{t=1}^{T}\left((\hat{\alpha}-r_{t\Delta})\int_{-\frac{r_{t\Delta}}{h}}^{\frac{r-r_{t\Delta}}{h}} \frac{\exp\left\{-\frac{v^2}{2}\right\}}{\sqrt{2\pi}}dv-h\int_{-\frac{r_{t\Delta}}{h}}^{\frac{r-r_{t\Delta}}{h}} \frac{v}{\sqrt{2\pi}}\exp\left\{-\frac{v^2}{2}\right\}dv\right) \\
&= \frac{\hat{\beta}}{T}\sum_{t=1}^{T}\left((\hat{\alpha}-r_{t\Delta})\left[\Phi\left(\frac{r-r_{t\Delta}}{h}\right)-\Phi\left(-\frac{r_{t\Delta}}{h}\right)\right]-hI_1\right),
\end{aligned}
$$

where $\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}\exp\{-\frac{u^2}{2}\}du$ and $I_1=\frac{1}{\sqrt{2\pi}}\int_{-\frac{r_{t\Delta}}{h}}^{\frac{r-r_{t\Delta}}{h}} v\exp\{-\frac{v^2}{2}\}dv$.

Now,

$$
\begin{aligned}
I_1 &= \int_{-\frac{r_{t\Delta}}{h}}^{\frac{r-r_{t\Delta}}{h}} \frac{v}{\sqrt{2\pi}}\exp\left\{-\frac{v^2}{2}\right\}dv=-\frac{1}{\sqrt{2\pi}}\int_{-\frac{r_{t\Delta}}{h}}^{\frac{r-r_{t\Delta}}{h}} d\Big(\exp\left\{-\frac{v^2}{2}\right\}\Big) \\
&= -\frac{1}{\sqrt{2\pi}}\left[\exp\left\{-\frac{(r-r_{t\Delta})^2}{2h^2}\right\}-\exp\left\{-\frac{r_{t\Delta}^2}{2h^2}\right\}\right].
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\int_0^r \mu(u;\hat{\theta})\hat{\pi}(u)du &= \frac{\hat{\beta}}{T}\sum_{t=1}^{T}\Big\{\Big[\hat{\alpha}-r_{t\Delta}\Big]\Big[\Phi\Big(\frac{r-r_{t\Delta}}{h}\Big)-\Phi\Big(-\frac{r_{t\Delta}}{h}\Big)\Big] \\
&+ \frac{h}{\sqrt{2\pi}}\Big[\exp\left\{-\frac{(r-r_{t\Delta})^2}{2h^2}\right\}-\exp\left\{-\frac{r_{t\Delta}^2}{2h^2}\right\}\Big]\Big\} \\
&= \frac{\hat{\beta}}{T}\sum_{t=1}^{T}\Big[\hat{\alpha}-r_{t\Delta}\Big]\Big[\Phi\Big(\frac{r-r_{t\Delta}}{h}\Big)-\Phi\Big(-\frac{r_{t\Delta}}{h}\Big)\Big] \\
&+ \frac{\hat{\beta}\,h}{T\sqrt{2\pi}}\sum_{t=1}^{T}\Big[\exp\left\{-\frac{(r-r_{t\Delta})^2}{2h^2}\right\}-\exp\left\{-\frac{r_{t\Delta}^2}{2h^2}\right\}\Big].
\end{aligned}
$$

This gives,

$$
\begin{aligned}
\hat{\sigma}_3^2(r) &= \frac{2\hat{\beta}}{T\hat{\pi}(r)}\cdot\sum_{t=1}^{T}\Big[\hat{\alpha}-r_{t\Delta}\Big]\left[\Phi\left(\frac{r-r_{t\Delta}}{h}\right)-\Phi\left(\frac{-r_{t\Delta}}{h}\right)\right] \\
&+ \frac{2\hat{\beta}}{T\hat{\pi}(r)}\cdot\frac{h}{\sqrt{2\pi}}\sum_{t=1}^{T}\left[\exp\left(-\frac{(r-r_{t\Delta})^2}{2h^2}\right)-\exp\left(-\frac{r_{t\Delta}^2}{2h^2}\right)\right].
\end{aligned}
$$

REFERENCES

AHN, D. H. AND GAO, B. (1999). "A Parametric Nonlinear Model of Term Structure Dynamics." *Review of Financial Studies* 12, 721–762.

AÏT-SAHALIA, Y. (1996a). "Nonparametric Pricing of Interest Rate Derivative Securities." *Econometrica* 64, 527–560.

AÏT-SAHALIA, Y. (1996b). "Testing Continuous–Time Models of the Spot Interest Rate." *Review of Financial Studies* 9, 385–426.

AÏT-SAHALIA, Y. (1999). "Transition Densities for Interest Rate and Other Nonlinear Diffusions," *Journal of Finance* 54, 1361–1395.

BACHELIER, L. (1900). "Théorie e la Speculation," Reprinted in Cootner (ed.), 17–78.

BLACK, F., AND SCHOLES, M. (1973). "The Pricing of Options and Corporate Liabilities." *Journal of Political Economy* 3, 637–654.

BRENNAN, M. AND SCHWARTZ, E. (1980). "A Continuous–Time Approach to the Pricing of Bonds." *Journal of Banking and Finance* 3, 133–145.

CHAN, K., KAROLYI, F., LONGSTAFF, F. AND SANDERS, A. (1992). "An Empirical Comparison of Alternative Models of the Short–Term Interest Rate." *Journal of Finance* 47, 1209–1227.

CHAPMAN, D. AND PEARSON, N. (2000). "Is The Short Rate Drift Actually Nonlinear ?" *Journal of Finance* 54, 355–388.

CHEN, S. X. AND GAO, J. (2004). "An Adaptive Empirical Likelihood Goodness–of–Fit Test for Conditional Mean of Nonlinear Time Series Models." Available from www.maths.uwa.edu.au/~jiti/cg04.pdf.

CHEN, S. X. AND GAO, J. (2005). "On the Use of the Kernel Method for Specification Tests of Diffusion Models." Working paper.

CHEN, S. X., HÄRDLE, W. AND LI, M. (2003). "An Empirical Likelihood Goodness–of–Fit Test for Time Series." *Joural of the Royal Statistical Society Series B* 65, 663–678.

COX, J., INGERSOLL, E. AND ROSS, S. (1985). "An Intertemporal General Equilibrium Model of Asset Prices." *Econometrica* 53, 363–384.

DURHAM, G. B. (2004). "Likelihood–Based Specification Analysis of Continuous–Time Models of the Shoer–Term Interest Rate." Forthcoming in the *Journal of Financial Economics.*

FAN, J., AND ZHANG, C. M. (2003). "A Re-examination of Stanton's Diffusion Estimators with Applications to Financial Model Validation." *Journal of the American Statistical Association* 461, 118–134.

GAO, J., AND KING, M. (2003). "Model Estimation and Specification Testing in Nonparametric and Semiparametric Regression." Working paper 2004/02, School of Mathematics and Statistics, The University of Western Australia. Available from www.maths.uwa.edu.au/~jiti/jems.pdf.

GAO, J., AND KING, M. (2004). "Adaptive Testing in Continuous–Time Diffusion Models." *Econometric Theory* 20, 844–882.

HONG, Y. AND LI, H. (2004). "Nonparametric Specification Testing for Continuous–Time Models with Applications to Term Structure of Interest Rates." *Review of Financial Studies* (forthcoming).

HOROWITZ, J. L. AND SPOKOINY, V. (2001). "An Adaptive, Rate–Optimal Test of a Parametric Mean–Regression Model Against a Nonparametric Alternative." *Econometrica* 69, 599–632.

JIANG, G., AND KNIGHT, J. (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.

JONES, C. S. (2004). "Nonlinear Mean Reversion in the Short–Term Interest Rate." Forthcoming in the *Review of Financial Studies.*

LI, M., PEARSON, N. D. AND POTESHMAN, A. M. (2004). "Facing Up to Conditioned Diffusions." Forthcoming in the *Journal of Financial Economics.*

MERTON, R. C. (1973). "The Theory of Rational Option Pricing." *Bell Journal of Economics* 4, 141–183.

PRITSKER, M. (1998). "Nonparametric Density Estimation and Tests of Continuous Time Interest Rate Models." *Review of Financial Studies* 11, 449–487.

STANTON, R. (1997). "A Nonparametric Model of Term Structure Dynamics and the Market Price of Interest Rate Risk." *Journal of Finance* 52, 1973–2002.

SUNDARESAN, S. (2001). "Continuous–time Methods in Finance: A Review and an Assessment." *Journal of Finance* 55, 1569–1622.

VASICEK, O. (1977). "An Equilibrium Characterization of the Term Structure. *Journal of Financial Economics* 5, 177–188.

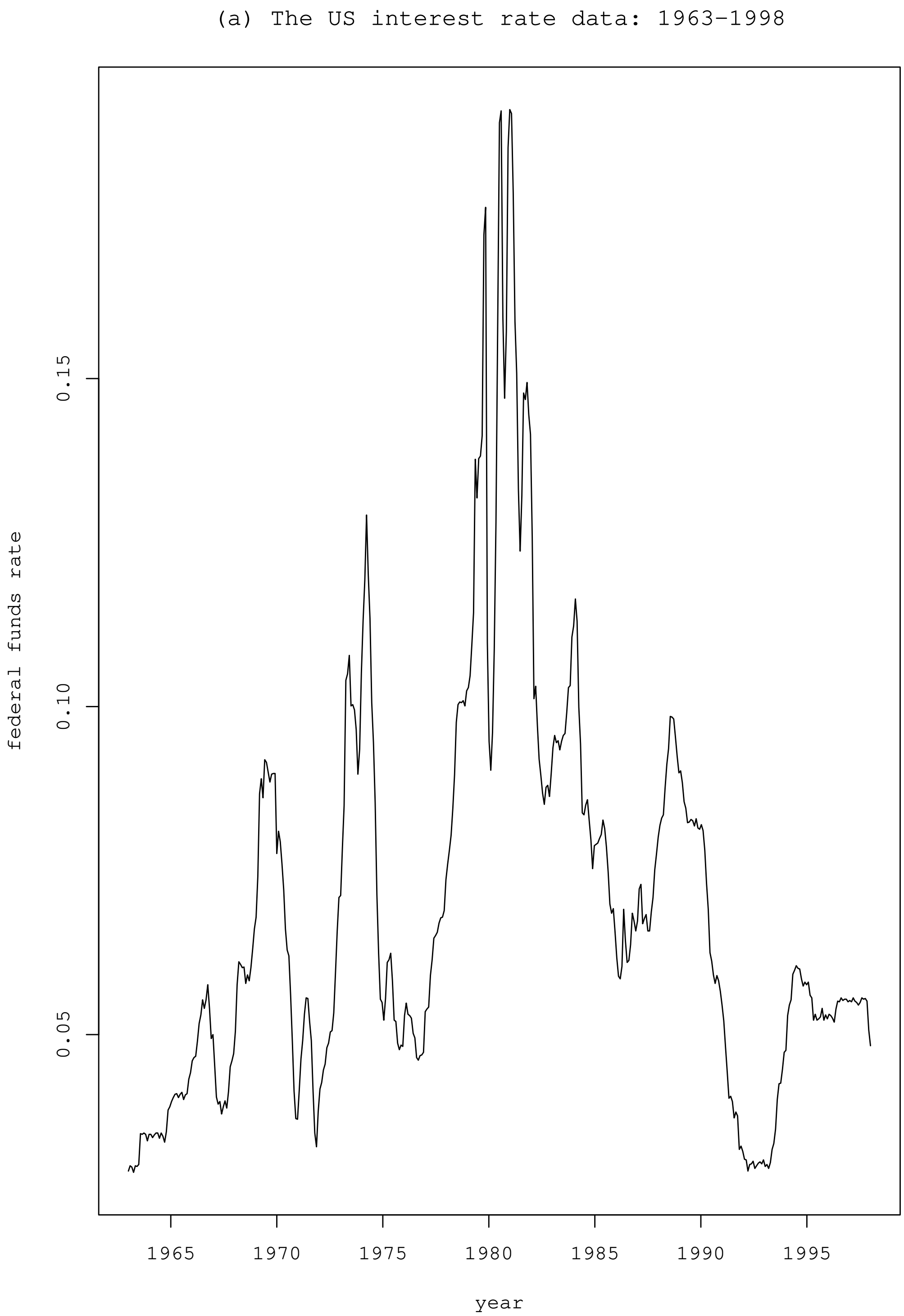


Figure 1: Three-month T-Bill rate, January 1963 to December 1998.

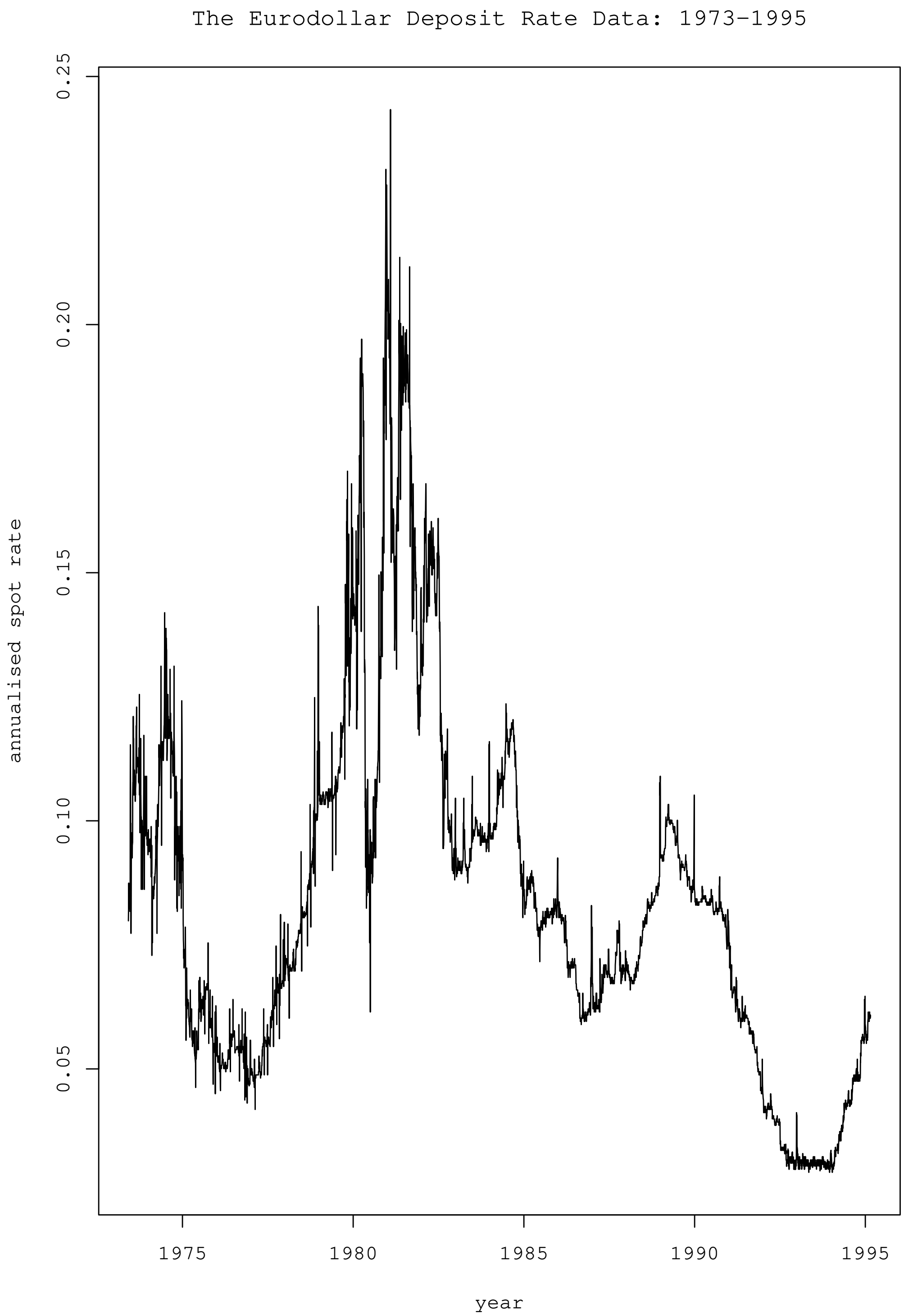
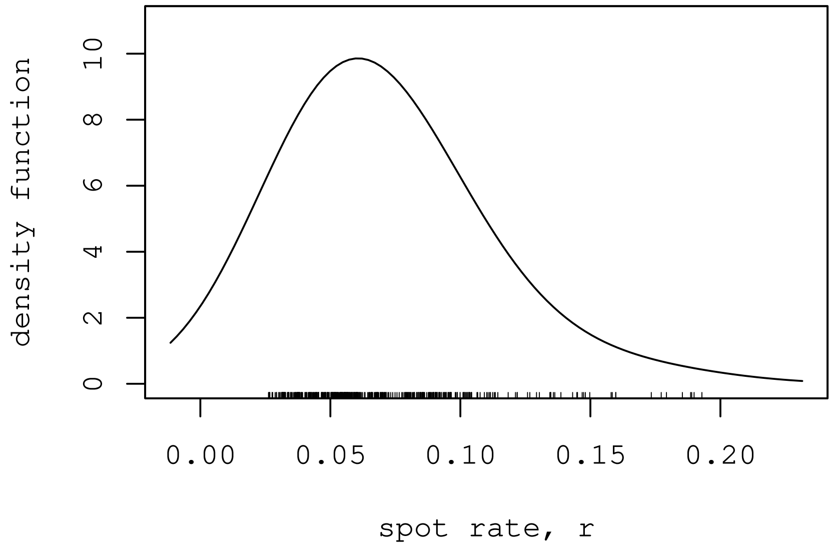
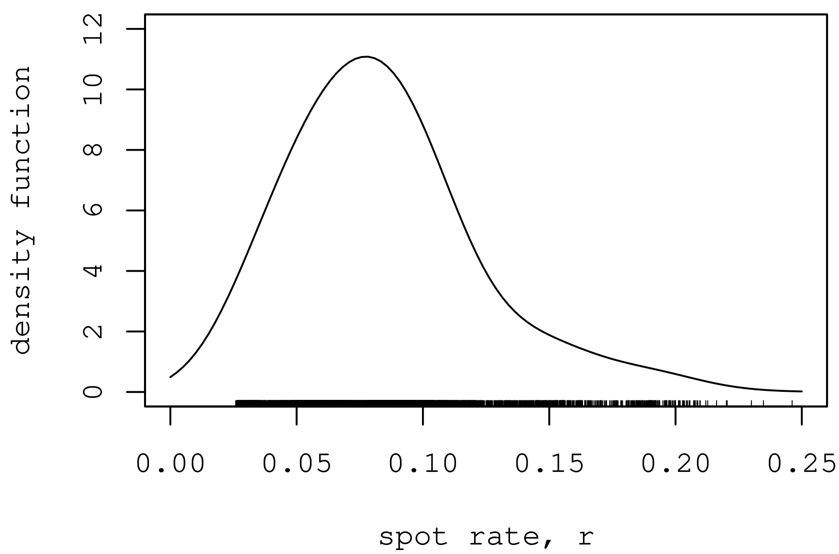


Figure 2: Seven-Day Eurodollar Deposit rate, 1 June 1973 to 25 February 1995.



(A)



(B)

Figure 3: A: Nonparametric kernel density estimator for the Fed data with $h_2 = 0.0297$. B: Nonparametric kernel density estimator for the Euro data with $h_2 = 0.01786$.

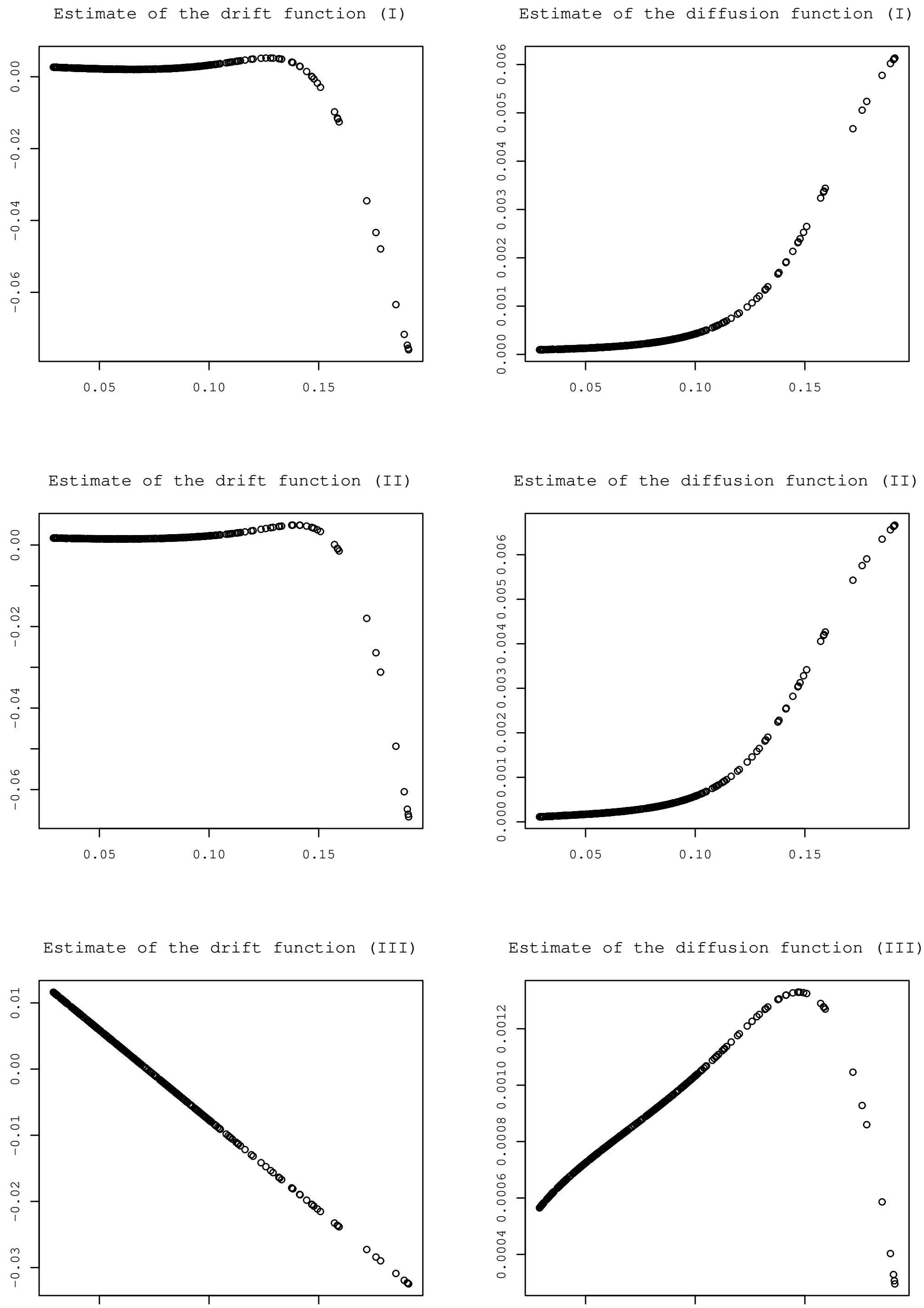


Figure 4: Nonparametric drift and diffusion estimators for the Fed data with $h_2 = 0.0297$.

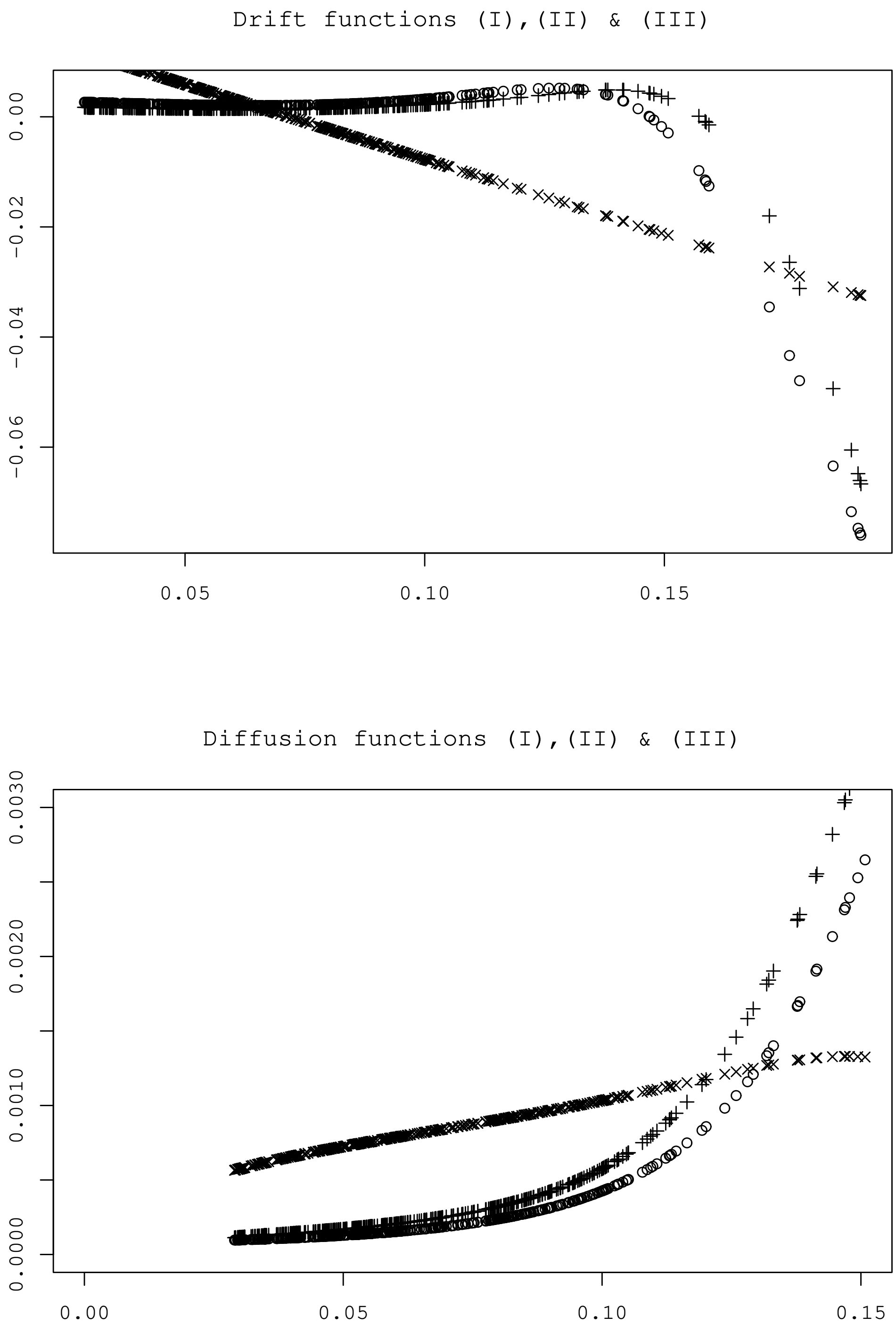


Figure 5: The best estimators for the Fed data with $h_2 = 0.0297$. The ($\circ$) refers to $\hat{\mu}_1$ and $\hat{\sigma}_1^2$, ($+$) to $\hat{\mu}_2$ and $\hat{\sigma}_2^2$ and ($\times$) to $\hat{\mu}_3$ and $\hat{\sigma}_3^2$.

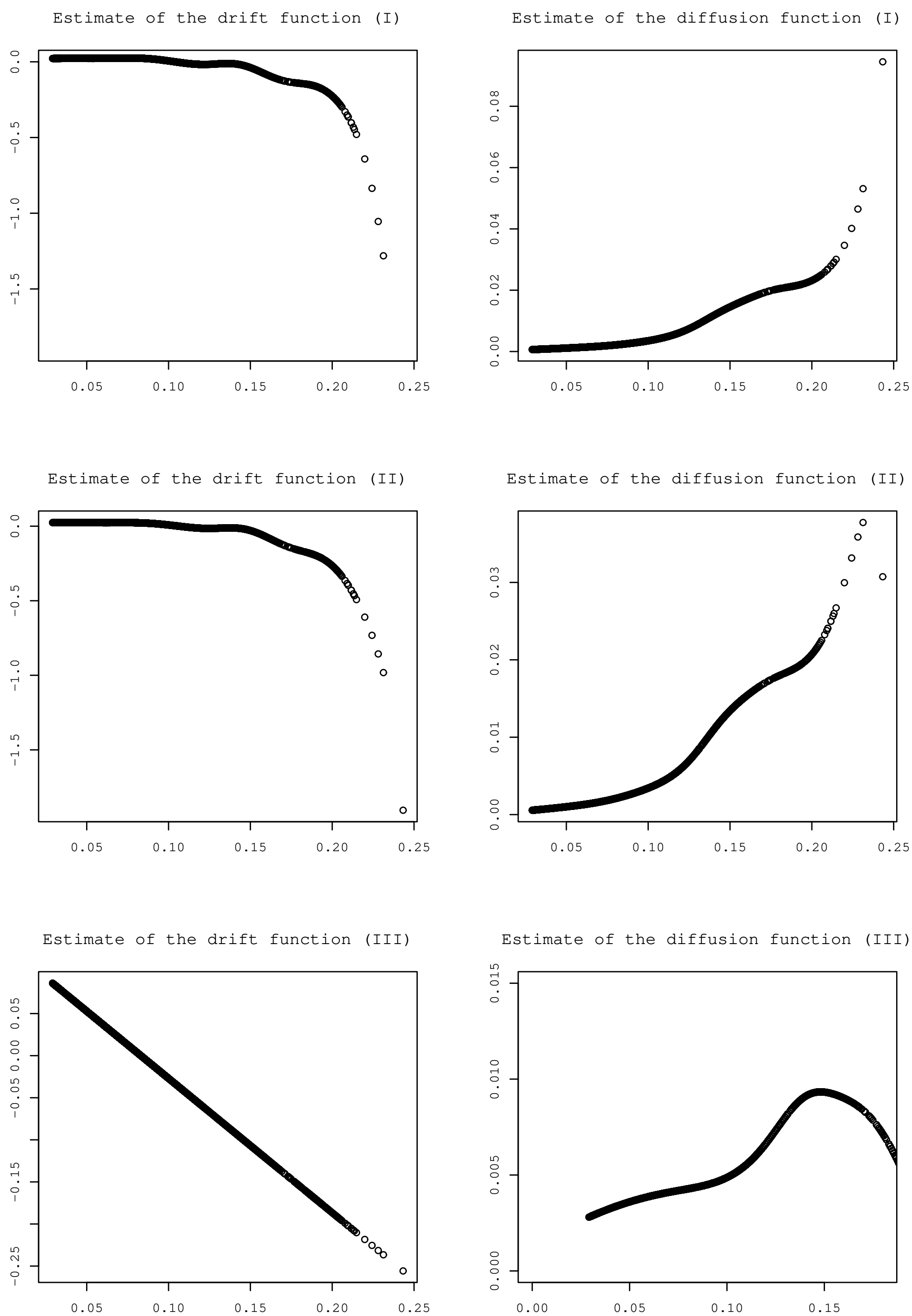


Figure 6: Nonparametric drift and diffusion estimators for the Euro data with $h_2 = 0.01786$.

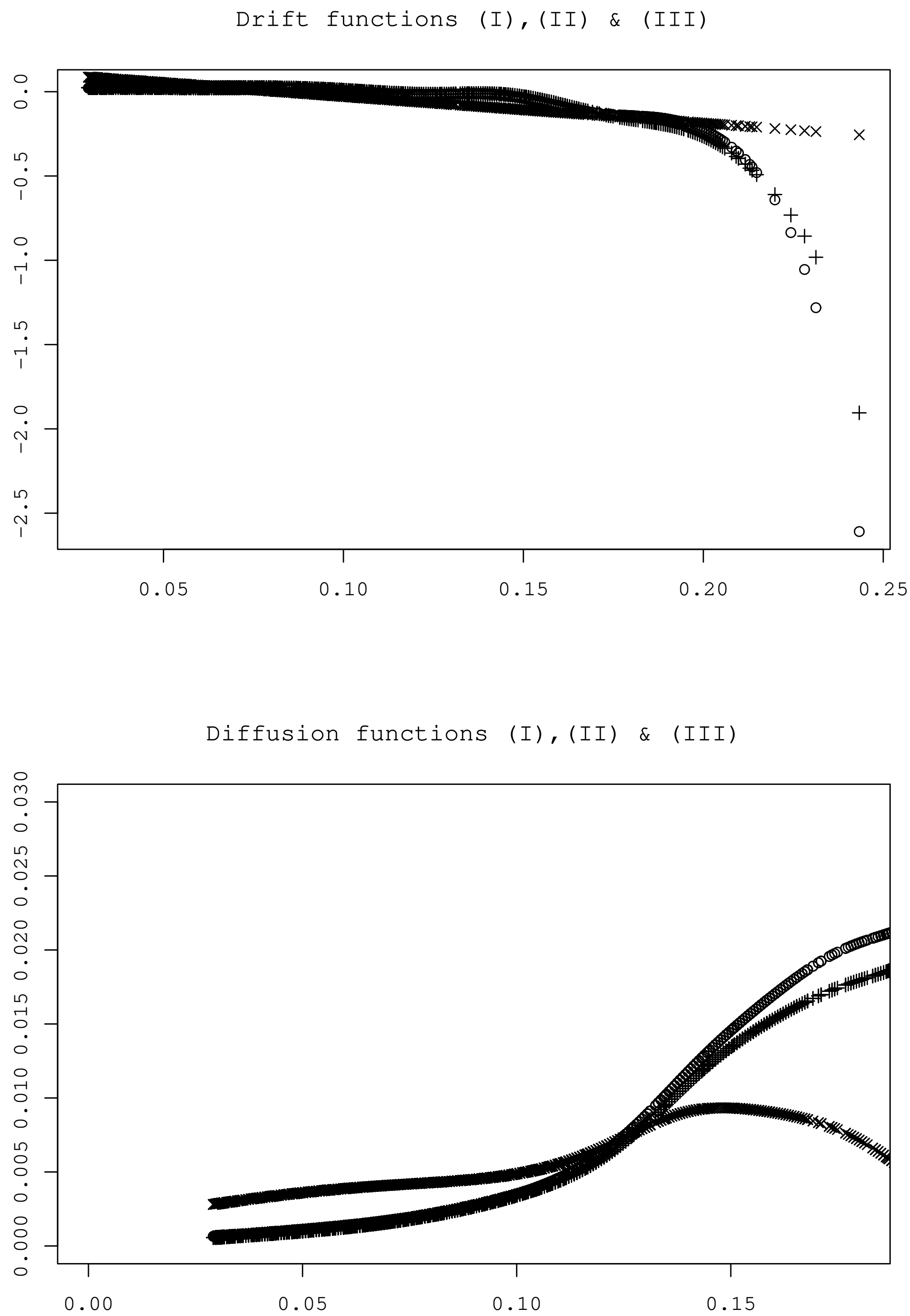


Figure 7: The best estimators for the Euro data with $h_2 = 0.01786$. The ($\circ$) refers to $\hat{\mu}_1$ and $\hat{\sigma}_1^2$, ($+$) to $\hat{\mu}_2$ and $\hat{\sigma}_2^2$ and ($\times$) to $\hat{\mu}_3$ and $\hat{\sigma}_3^2$.