Semiparametric Estimation for Time-Inhomogeneous Diffusions

Yan Yu

University of Cincinnati, USA

Keming Yu

Brunel University, UK

Hua Wang

Yahoo, USA

Min Li

California State University, Sacramento, USA

Abstract. We develop two likelihood based approaches to semiparametrically estimate the time-inhomogeneous diffusion process: log penalized splines (P-splines) and the local log-linear method. Positive volatility is naturally embedded and this positivity is not guaranteed in most existing diffusion models. We investigate different smoothing parameter selection methods. Separate bandwidths are used for drift and volatility estimation. In the log P-splines approach, different smoothness for different time varying coefficients is feasible by assigning different penalty parameters. We also provide accompanying theorems for both approaches. Finally, we present a case study using the weekly three-month Treasury bill data from 1954 to 2004. We find that the log P-splines approach seems to capture the volatility dips in mid-1960s and mid-1990s the best.

Keywords: Bandwidth Selection; Kernel Smoothing; Local Linear; Penalized likelihood; Variance Estimation; Volatility.

1. Introduction

Modern asset pricing theory offers valuable guidance for pricing contingent claims and risk management. Continuous-time diffusion processes are important tools to model the stochastic behavior of a range of economic variables, such as interest rates and stock prices. For example, the famous option pricing model of Black and Scholes (1973), term structure models of Vasicek (1977), Cox, Ingersoll, and Ross (CIR, 1985), Hull and White (1990), Heath, Jarrow, and Morton (1992), Chan, Kayolyi, Longstaff, and Sanders

(CKLS, 1992), all assume that the underlying state variables follow diffusions. A nice overview can be found in Merton (1992) and Duffie (2001).

All of the above mentioned diffusions are simple time-homogeneous parametric models taking the form: $dX_t = \mu(X_t;\theta)dt + \sigma(X_t;\theta)dW_t$, where X_t is an economic state variable depending on time t, W_t is the standard Brownian motion, θ is a parameter, $\mu(X_t;\theta)$ is the drift function, and $\sigma(X_t;\theta)$ is the diffusion or volatility function. The term volatility or diffusion in Finance is the same as the standard deviation in Statistics. Volatility is a key concept because it is a measure of uncertainty about future price movements. Volatility is directly related to the risk associated with holding financial securities and hence affects consumption/investment decisions and portfolio choice. Volatility is also the key parameter in the pricing of options and other derivative securities.

Unfortunately, empirical tests of these different parametric diffusion models yielded mixed results (Stanton 1997). This is not too surprising since they are neither derived from any economic theory nor offering guidance in choosing the correct model. With the availability of high-quality data of many financial assets, recent researchers have considered nonparametric techniques for diffusion models to avoid possible model misspecification. For example, Ait-Sahalia (1996) estimates the time-homogeneous diffusion $\sigma(X_t)$ nonparametrically using the kernel method, given a linear specification for the drift. Stanton (1997) estimates both the drift $\mu(X_t)$ and diffusion $\sigma(X_t)$ nonparametrically using the kernel method.

While nonparametric estimation of diffusion models is promising, mostly they consider only time-homogeneous diffusions. There are a variety of reasons to believe that the underlying process for many economic variables might change from time to time, due to changes in business cycles, general economic conditions, monetary policy, etc. One example is the volatility of interest rates at all maturities on the days of FOMC (Federal Open Market Committee) meeting increases. The so-called “calendar effects” on stock prices that the prices behave differently on different days of the week, month, and year, are often observed. Prices of many fixed-income securities and options change over time as the maturities of the contracts approach (see Egorov, Li, and Xu 2003, and the references therein).

This motivates researchers to consider time-inhomogeneous diffusion models

$$dX_t = \mu(t, X_t)\,dt + \sigma(t, X_t)\,dW_t,$$

where both the drift $\mu(t, X_t)$ and diffusion, or often called volatility $\sigma(t, X_t)$ depend on time t. Figure 1 (a) is a plot of weekly 3-month Tresaury bill rates during the period from 1954 to 2004. The differenced rates are ploted in (b). Visually the differenced rates seem to behave randomly with small volatility around mid 1960s and mid 1990s while with larger volatility during late 1950s, mid 1970s and early 1980s. This visual observation seems to be well represented in our log P-splines fit of volatility $\sigma(t, X_t)$ in Figure 1 (c), which catches the low volatility period around 1964. The two local fits (local log-linear and Fan, Jiang, Zhang, and Zhou 2003) of volatility seem dominated by the overall trend of the original rates and keep increasing in the 1960s. Especially in the period from 1961 to 1966, the contradiction is evident. A key point is that the differenced yield seems to exhibit time inhomogeneous variation, which is the main focus of this paper.

[Insert Figure 1 here]

In fact, some parametric time-inhomogeneous diffusion models have been developed in the finance literature and have been widely used in practice. For example, to capture the “smiles” (in contrast to the constant volatility assumption of geometric Brownian Motion in Black and Scholes (1973)) observed in the implied volatility from option prices, Rubinstein (1994) models stock return volatility as a deterministic function of stock price and time. Hull and White (1990) develop models where the short rate follows a parametric time-inhomogeneous diffusion process. A recent work by Fan, Jiang, Zhang, and Zhou (FJZZ, 2003) find that there is not sufficient information to determine the bivariate functions nonparametrically and forcing all coefficients in the drift and diffusion to be time-dependent may cause over-parameterization. Hence, a semiparametric time-inhomogeneous model is considered in this paper.

We focus on the following semiparametric time-inhomogeneous model

$$dX_t = \left(\alpha(t) + \beta(t) X_t\right) dt + \sigma(t)\left(X_t\right)^{\gamma} dW_t, \tag{1}$$

where γ is a scalar parameter independent of time t, $\alpha(t)$ and $\beta(t)$ are time-dependent coefficients of the drift, $\sigma(t)$ is a time-dependent coefficient of diffusion (volatility). Model (1) includes most of the well-known diffusion models. For example, when $\alpha(t)$, $\beta(t)$, and $\sigma(t)$ are constants (time independent), (1) yields to CKLS model. Among CKLS framework, $\gamma = 1$ corresponds to the famous Black-Scholes model; $\gamma = 0$ corresponds to Vasicek model; and $\gamma = 0.5$ corresponds to CIR model. A more general model with γ depending on time t has been considered by FJZZ. However, they note that there may be over-parameterization and unreliable estimates due to high collinearity.

The semiparametric time-inhomogeneous model (1) is continuous but the data sampled in the financial markets are usually discrete. Therefore in estimation, the discretized version of (1) based on the Euler scheme is used as an approximation. Suppose the data $\left\{X_{t_i},\ i = 1, \cdots, n+1\right\}$ are sampled at discrete time points, $t_1 < \cdots < t_{n+1}$. For weekly data when the time unit is a year, $t_i = t_0 + i/52\ \left(i = 1, \cdots, n\right)$, where t_0 is the initial time point. Denote $y_{t_i} = X_{t_{i+1}} - X_{t_i}$, $Z_{t_i} = W_{t_{i+1}} - W_{t_i}$, and $\Delta_i = t_{i+1} - t_i$. Z_{t_i} are independent and normally distributed with mean zero and variance Δ_i due to the independent increment property of Brownian motion W_{t_i}. The discretized version of (1) becomes

$$y_{t_i} \approx \left(\alpha(t_i) + \beta(t_i) X_{t_i}\right)\Delta_i + \sigma(t_i)\left(X_{t_i}\right)^{\gamma} \sqrt{\Delta_i}\varepsilon_{t_i}, \tag{2}$$

where $\left\{\varepsilon_{t_i}\right\}$ are independent and standard normal. According to Stanton (1997) and further studied in Fan and Zhang (2003), the first-order discretized approximation error to the continuous-time diffusion model is extremely small, as long as data are sampled monthly or more frequently. This finding simplifies the estimation procedure significantly.

This paper contains a statistical finance application on the short term Treasury bill data as well as some methodological development possibly for broader interest. In particular, we contribute to the literature of diffusion model estimation in the following aspects. First, we provide two practical tools to estimate the time dependent diffusion process semiparametrically. Two likelihood based approaches are developed: log P-splines maximizing penalized likelihood and the local log-linear method maximizing kernel-weighted likelihood. The necessary feature of positive volatility is naturally embedded in both log P-splines and local log-linear methods and this positivity is not guaranteed in most existing diffusion models. In addition, compared to the local constant method, the local log-linear approach in general can give lower bias and variance of estimates with more appealing properties at the boundary (Fan and Gijbels 1996; Yu and Jones 2004; FJZZ 2003).

Secondly, we investigate different smoothing parameter (bandwidth) selection criteria: generalized cross validation (GCV) and EBBS of Ruppert (1997) criteria for log P-splines and the Rule-of-thumb bandwidth (ROT) for the local log-linear method. Separate bandwidths are used for drift and volatility estimation. In addition, in the log P-splines approach, different smoothness for different time varying coefficients $\alpha(t)$ and $\beta(t)$ of drift is feasible by assigning different penalty parameters. Consistent with most literature (e.g. Jarrow, Ruppert, and Yu 2004; Yu and Jones 2004), small simulation studies (not reported) show that EBBS for log P-splines approach is more robust to possible autocorrelations and less prone to undersmoothing as often observed with generalized cross validation (GCV). Our ROT bandwidth is simple and works almost as well as the unavailable optimal bandwidths. The ROT local log-linear approach performs better than that of both the local linear and local constant approaches. The proposed approach is as good as or better than that of Ruppert, Wand, Holst, Hössjer (1997), even when the latter uses its optimal bandwidths.

Thirdly, we provide asymptotic results for both approaches so that inference is readily available. The asymptotic result also enables the proposal of our rule of thumb (ROT) bandwidth estimator in the local log-linear approach. Comparing two proposed approaches in the time-inhomogeneous diffusion estimation, we find that the log P-splines approach is computationally expedient and efficient, which is also often observed in complicated nonlinear regression context (Yu and Ruppert 2002; Jarrow, Ruppert, and Yu 2004). However, the theory from the local log-linear approach is more complete in the sense of “truely nonparametric.” We give a large sample property based on fixed knot P-splines, which often serves well in application. Other spline methods could also be used though direct adoption might be complicated and we expect the fit would be similar.

Finally, we present a case study using the weekly three-month Treasury bill data from 1954 to 2004. We find that both log P-splines and local log-linear approaches as well as the FJZZ kernel method catch the major trend of volatility well. Volatility is the highest during early 1980s (Figure 1b). This was in agreement with the economic situation then. During that period, the Federal Researve chairman Paul Volcker sharply increased the interest rate to combat the inflation crisis in U.S. Inflation decreased from 9% in 1980 to 3.2% in 1983. The interest rate (the yield on 3-month Treasury bills) also dropped dramatically by 1983. The swing in the interest rate during that period is reflected in the volatility plot. However, similar to what we have observed in Figure 1c for the dip in mid 1960s, the log P-splines fit seems to catch the relative low variation period in mid 1990s better, whereas the local log-linear and FJZZ methods are more prone to the domination by the original series. We will further explore details of the case study in Section 4. The rest of this paper is organized as follows. Section 2 investigates the log P-splines approach of the time-inhomogeneous diffusion model. Section 3 presents the local log-linear estimation.

2. Log Penalized Splines Diffusion Estimation

We develop a log penalized splines method for diffusion estimation. Log is necessary to guarantee that volatility is positive. P-splines are described in Eilers and Marx (1996), Ruppert and Carroll (1997), and Ruppert, Wand, and Carroll (2003). P-splines estimate fewer parameters than smoothing splines. The location of the knots in P-splines is considered not as crucial as that in regression splines such as MARS (Friedman 1991). Smoothness is achieved through a roughness penalty measure. In Eilers and Marx (1996), quadratic penalties are placed on finite differences of adjacent B-splines coefficients. An appealing feature of Ruppert and Carroll (1997) is to allow multiple smoothing parameters and also a variety of penalties, quadratic or nonquadratic, on the spline coefficients.

2.1 Maximum Penalized Likelihood Estimation

We model time dependent functions $\alpha\left(t_i\right)$, $\beta\left(t_i\right)$ of drift and $\log\sigma^2(t_i)$ of volatility in model (2) by splines:

$$\alpha(t_i) = \mathbf{B}_\alpha(t_i)\boldsymbol{\delta}_\alpha;\ \beta(t_i) = \mathbf{B}_\beta(t_i)\boldsymbol{\delta}_\beta,$$

$$\log\sigma^2(t_i) = 2\mathbf{B}_\sigma(t_i)\boldsymbol{\delta}_\sigma, \tag{3}$$

where $\mathbf{B}(t_i)$ is a vector of spline basis functions and $\boldsymbol{\delta}$ are vectors of spline coefficients. Different basis functions can be used for different coefficient functions and basis using truncated power function, B-splines, or natural cubic splines can also be adopted. Our experience shows that they yield similar fits. This is not surprising since the critical tuning parameter in P-splines is the penalty parameter. Hence, for notational simplicity, we will present P-splines using a truncated power basis function $\mathbf{B}(t_i) = \left[1,\ t_i,\ \cdots\ t_i^p,\ \left(t_i-\kappa_1\right)_+^p, \cdots, \left(t_i-\kappa_k\right)_+^p\right]$, where p is the spline polynomial degree, $\left(t_i-\kappa_k\right)_+ = \max\left(0, t-\kappa_k\right)$, $\kappa_1 < \kappa_2 < \cdots < \kappa_K$ are spline knots often located at equal-spaced sample quantiles for simplicity. Then we can write

$$\log\sigma(t_i) = \mathbf{B}(t_i)\boldsymbol{\delta}_\sigma = \delta_0^\sigma + \delta_1^\sigma t_i + \cdots + \delta_p^\sigma t_i^p + \delta_{p+1}^\sigma\left(t_i-\kappa_1\right)_+^p + \cdots + \delta_{p+k}^\sigma\left(t_i-\kappa_k\right)_+^p.$$

The log likelihood function, excluding constants, is negative

$$\sum\left(\frac{1}{\Delta_i}\left\{y_{t_i} - \left(\mathbf{B}_\alpha(t_i)\boldsymbol{\delta}_\alpha + \mathbf{B}_\beta(t_i)\boldsymbol{\delta}_\beta X_{t_i}\right)\Delta_i\right\}^2 \exp\left\{-\left(2\mathbf{B}_\sigma(t_i)\boldsymbol{\delta}_\sigma + \gamma\log X_{t_i}^2\right)\right\} + 2\mathbf{B}_\sigma(t_i)\boldsymbol{\delta}_\sigma + \gamma\log X_{t_i}^2\right).$$

For notational consistency, we will reserve the subscript 1 for drift and 2 for volatility. Denote parameter vectors $\boldsymbol{\delta}_1 = \left(\boldsymbol{\delta}_\alpha^T, \boldsymbol{\delta}_\beta^T\right)^T$ for drift and $\boldsymbol{\delta}_2 = \left(\boldsymbol{\delta}_\sigma^T, \gamma\right)^T$ for volatility. Denote the extended design matrix for drift $\mathbf{B}_1(t_i) = \left[\mathbf{B}_\alpha(t_i),\ \ \mathbf{B}_\beta(t_i)X_{t_i}\right]$ and the extended design matrix for volatility $\mathbf{B}_2(t_i) = \left[\mathbf{B}_\sigma(t_i),\ \ \log X_{t_i}\right]$. Further denote the parameter vector $\boldsymbol{\theta} = \left(\boldsymbol{\delta}_1{}^T, \boldsymbol{\delta}_2{}^T\right)^T = \left(\boldsymbol{\delta}_\alpha^T, \boldsymbol{\delta}_\beta^T, \boldsymbol{\delta}_\sigma^T, \gamma\right)^T$. Let the smoothing parameter vector $\boldsymbol{\lambda} = \left(\lambda_\alpha, \lambda_\beta, \lambda_2\right)^T$ and $\boldsymbol{\lambda}_1 = \left(\lambda_\alpha, \lambda_\beta\right)^T$, where λ_α, λ_β, and λ_2 are three smoothing parameters for $\alpha\left(t\right)$, $\beta\left(t\right)$, and $\log\sigma^2\left(t\right)$ respectively.

The penalized likelihood estimator of $\boldsymbol{\theta}$ maximizes the following penalized log likelihood function

$$Q_{n,\lambda}(\boldsymbol{\theta}) = L_n(\boldsymbol{\theta}) - \frac{n}{2}\boldsymbol{\lambda}\boldsymbol{\theta}^T\mathbf{D}\boldsymbol{\theta}, \tag{4}$$

Where

$$L_n(\boldsymbol{\theta}) := \sum l_n(\boldsymbol{\theta}, t_i) = -\sum\left(\frac{1}{\Delta_i}\left\{y_{t_i} - \mathbf{B}_1(t_i)\boldsymbol{\delta}_1\Delta_i\right\}^2 \exp\left\{-2\mathbf{B}_2(t_i)\boldsymbol{\delta}_2\right\} + 2\mathbf{B}_2(t_i)\boldsymbol{\delta}_2\right). \tag{5}$$

Here $\mathbf{D}$ is an appropriate positive semi-definite symmetric matrix. A common choice of $\mathbf{D}$ is given by Ruppert, Wand, and Carroll (2003) that penalizes jumps at the knots in the *p*th derivative of the spline. We will use this penalty. Other common penalties, for example, as in smoothing spline, are the quadratic penalty on the (second) derivatives of functions. Like the choice of basis functions, we found the choice of $\mathbf{D}$ is relatively unimportant that different $\mathbf{D}$ gives similar fits.

Now squared volatility estimate, using notation $V(t_i)$, suppressing X_{t_i}, can be obtained by

$$\hat{V}(t_i) := \hat{\sigma}^2\left(t_i, X_{t_i}\right) = \hat{\sigma}\left(t_i\right)^2 X_{t_i}^{\ 2\hat{\gamma}} = \exp\left(2\mathbf{B}_2(t_i)\hat{\boldsymbol{\delta}}_2\right). \tag{6}$$

And the volatility estimate is

$$\hat{\sigma}\left(t_i, X_{t_i}\right) = \sqrt{\hat{\sigma}\left(t_i\right)^2 X_{t_i}^{\ 2\hat{\gamma}}} = \exp\left(\mathbf{B}_2(t_i)\hat{\boldsymbol{\delta}}_2\right). \tag{7}$$

2.2 Asymptotic Properties and Inference

As is virtually always the case, theoretical results for the P-splines approach are not as readily obtainable as for local methods. Indeed, it is still an open question for a simple univariate P-splines regression (Hall and Opsomer 2005). Nevertheless, we give the results for the log P-splines estimator using a fixed number of knots, which is basically from a flexible but parametric model. We find that the fixed-knot P-splines analysis is useful for developing a practical methodology, which is also noted in the literature, e.g. Gray (1994), Carroll, Maca, and Ruppert (1999).

Theorem 1. Under mild regularity conditions, if the smoothing parameter vector $\boldsymbol{\lambda}_n = o\left(n^{-1/2}\right)$, then a sequence of penalized likelihood estimators $\hat{\boldsymbol{\theta}}$ exists, is consistent, and is asymptotically normally distributed. That is,

$$n^{1/2}\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right) \to_D \mathrm{N}\left(0, \mathbf{I}^{-1}\left(\boldsymbol{\theta}\right)\right), \tag{8}$$

where $\mathbf{I}\left(\boldsymbol{\theta}\right)$ is the usual Fisher information matrix.

The proof of Theorem 1 is standard with ordinary (no penalty) maximum likelihood estimates (Lehmann 1983); and is similar to Fan and Li (2001) and Yu (2005) with penalty function.

The result given in (8) does not involve penalty parameter, which is assumed to vanish sufficiently fast as n tends to infinity. For finite sample inference, this tends to overestimate the variance of $\hat{\boldsymbol{\theta}}$ and one would prefer the following asymptotic distribution with fixed penalty parameter derived from the estimating equation approach using the “sandwich formula.” (For details see Carroll, Ruppert, Stefanski 1995; Yu and Ruppert 2002.)

$$n^{1/2}\left(\hat{\boldsymbol{\theta}}(\boldsymbol{\lambda}) - \boldsymbol{\theta}(\boldsymbol{\lambda})\right) \to_D \mathrm{N}\left\{0, \mathbf{H}^{-1}\left(\boldsymbol{\theta}(\boldsymbol{\lambda})\right)\mathbf{G}\left(\boldsymbol{\theta}(\boldsymbol{\lambda})\right)\mathbf{H}^{-T}\left(\boldsymbol{\theta}(\boldsymbol{\lambda})\right)\right\}, \tag{9}$$

where $\mathbf{H}\left(\boldsymbol{\theta}\right) = \sum \frac{\partial}{\partial \boldsymbol{\theta}^T} \boldsymbol{\psi}_{t_i}\left(\boldsymbol{\theta}\right); \mathbf{G}\left(\boldsymbol{\theta}\right) = \sum \boldsymbol{\psi}_{t_i}\left(\boldsymbol{\theta}\right)\boldsymbol{\psi}_{t_i}^{\ T}\left(\boldsymbol{\theta}\right)$, $\boldsymbol{\psi}_{t_i}\left(\boldsymbol{\theta}\right) = -\frac{\partial}{\partial \boldsymbol{\theta}^T} l_n\left(\boldsymbol{\theta}; t_i\right) + n\boldsymbol{\lambda}\mathbf{D}\boldsymbol{\theta}$.

A standard error of the estimated volatility function $\hat{\sigma}\left(t_i, X_{t_i}\right) = \exp\left(\mathbf{B}_2(t_i)\hat{\boldsymbol{\delta}}_2\right)$ can be easily derived from a delta method calculation

$$sd\left\{\hat{\sigma}\left(t_i, X_{t_i}\right)\right\} = \sqrt{\mathbf{B}_2(t_i)\hat{V}ar(\hat{\boldsymbol{\delta}}_2)\mathbf{B}_2^{\ T}(t_i)}\exp\left(\mathbf{B}_2(t_i)\hat{\boldsymbol{\delta}}_2\right) \tag{10}$$

where $\hat{V}ar(\hat{\boldsymbol{\delta}}_2)$ is given by (9), recommended for finite sample with fixed penalty parameter. Note that as $\boldsymbol{\lambda}$ goes to zero, $\hat{V}ar(\hat{\boldsymbol{\delta}}_2(\boldsymbol{\lambda}))$ in (9) converges to the corresponding matrix of inverse of Fisher information given in (8).

2.3 An Algorithm

One could do one-step maximization on the penalized likelihood function (4). However, in our P-splines approach, the number of parameters could be large and this estimation algorithm may not be efficient. We then implemented an iterative algorithm by reweighting drift estimation using the inverse of the estimated volatility with several iterations as suggested by Carroll, Wu, and Ruppert (1988). We find in our simulation and case studies that two or three iterations are sufficient. Moreover, volatility estimation is our primary focus of interest. Hence, we advocate two-step estimation in practice.

Step 1: Drift Estimation.

The time-inhomogeneous drift $\mu\left(t_i, X_{t_i}\right) = \left(\alpha\left(t_i\right) + \beta\left(t_i\right) X_{t_i}\right)$ is estimated by minimizing

$$\sum_{i=1}^{n} \left\{ \frac{y_{t_i}}{\Delta_i} - \mathbf{B}_1(t_i)\boldsymbol{\delta}_1 \right\}^2 + \frac{n}{2}\boldsymbol{\lambda}_1 \boldsymbol{\delta}_1{}^T \mathbf{D}_1 \boldsymbol{\delta}_1 . \tag{11}$$

This can be achieved by a simple ridge regression $\hat{\boldsymbol{\delta}}_1 = \left(\mathbf{B}_1^T \mathbf{B}_1 + n\boldsymbol{\lambda}_1 \mathbf{D}_1\right)^{-1} \mathbf{B}_1^T \mathbf{Y}$, where vector $\mathbf{Y}$ has the i*th* element y_{t_i} / Δ_i. The smoothing parameter can be chosen by GCV or EBBS etc., which we will discuss in more detail in Section 2.4.

Step 2: Log P-splines Volatility Estimation.

Denote the residual from the previous drift estimation

$$e_{t_i} = \frac{1}{\sqrt{\Delta_i}} \left(y_{t_i} - \hat{\mu}(t_i, X_{t_i}) \Delta_i \right). \tag{12}$$

Then we have $e_{t_i} \approx \sigma\left(t_i\right) \left(X_{t_i}\right)^{\gamma} \varepsilon_{t_i}$.

Remark: Stanton (1997) pointed out that this approximation holds even if $\hat{\mu}(t_i, X_{t_i}) = 0$ is assumed, though the approximation error using (12) is of smaller order. This observation further confirms the validity of our two-step approach with primary focus on the volatility estimation.

We estimate the parameter $\boldsymbol{\delta}_2$ for volatility by minimizing the negative penalized likelihood

$$\sum\left(e_{t_i}{}^{2}\exp\left\{-2\mathbf{B}_2(t_i)\boldsymbol{\delta}_2\right\}+2\mathbf{B}_2(t_i)\boldsymbol{\delta}_2\right)+\frac{n}{2}\boldsymbol{\lambda}_2\boldsymbol{\delta}_2{}^{T}\mathbf{D}_2\boldsymbol{\delta}_2 . \tag{13}$$

The usual Newton-Raphson procedure can be applied. We used the nonlinear optimization routine *lsqnonlin()* from Matlab's optimization toolbox. A preliminary parameter estimate $\hat{\boldsymbol{\delta}}_{2,pre}$ for volatility can be obtained by a simple ridge regression $\hat{\boldsymbol{\delta}}_{2,pre}=\left(\mathbf{B}_2^T\mathbf{B}_2+n\lambda_2\mathbf{D}_2\right)^{-1}\mathbf{B}_2^T\mathbf{E}$, where vector $\mathbf{E}$ has the i*th* element $\log\left|e_{t_i}\right|$. The volatility estimate is calculated by

$$\hat{\sigma}\left(t_i,X_{t_i}\right)=\sqrt{\hat{\sigma}\left(t_i\right)^2 X_{t_i}{}^{2\hat{\gamma}}}=\exp\left(\mathbf{B}_2(t_i)\hat{\boldsymbol{\delta}}_2\right).$$

Remark: We have also implemented a generalized least squares variance estimation algorithm as outlined in Ruppert and Carroll (1997) for our diffusion model, where the main step of iteration involves minimizing $\sum_{i=1}^{n}\left\{\frac{e_{t_i}^2-\exp\left(2\mathbf{B}_2(t_i)\boldsymbol{\delta}_2\right)}{\exp\left(2\mathbf{B}_2(t_i)\hat{\boldsymbol{\delta}}_{2,old}\right)}\right\}^2+\frac{n}{2}\lambda_2\boldsymbol{\delta}_2^T\mathbf{D}_2\boldsymbol{\delta}_2$. The fit is also promising in our case study.

2.4 Selection of Smoothing Parameter

2.4.1 GCV

Generalized cross validation (GCV) is a common smoothing parameter selection criterion in spline literature. In step 1 of drift estimation, the GCV smoothing parameter $\boldsymbol{\lambda}_1$ minimizes

$$GCV\left(\boldsymbol{\lambda}_1\right)=\frac{ASR\left(\boldsymbol{\lambda}_1\right)}{\left[1-trace\left\{\mathbf{B}_1\left(\mathbf{B}_1^T\mathbf{B}_1+n\boldsymbol{\lambda}_1\mathbf{D}_1\right)^{-1}\mathbf{B}_1^T\big/n\right\}\right]^2}, \tag{14}$$

where $ASR(\boldsymbol{\lambda}_1)=\sum_{i=1}^{n}\left\{y_{t_i}\big/\Delta_i-\mathbf{B}_1(t_i)\hat{\boldsymbol{\delta}}_1(\boldsymbol{\lambda}_1)\right\}^2$ is the usual average squared residuals from linear ridge regression.

The GCV smoothing parameter λ_2 for step 2 of volatility estimates minimizes

$$GCV(\lambda_2) = \frac{Deviance(\lambda_2)}{\left[1 - trace\left\{\mathbf{B}_2\left(\mathbf{B}_2^T\mathbf{B}_2 + n\lambda_2\mathbf{D}_2\right)^{-1}\mathbf{B}_2^T\right\}\Big/ n\right]^2}, \tag{15}$$

where the numerator is the deviance (McCullagh and Nelder 1989) of the model for a fixed value of the smoothing parameter λ_2.

2.4.2 EBBS

EBBS (empirical bias bandwidth selection) has been proposed for local polynomial variance function estimation with a number of advantages (Ruppert et al. 1997). Jarrow, Ruppert, and Yu (2004) also observe that in interest rate term structure estimation EBBS seems more robust to autocorrelations and smoothing on derivatives, whereas GCV is more prone to undersmoothing even with an artificial hyperparameter introduced. We extend EBBS for use with log P-splines diffusion estimation.

EBBS minimizes the average MSE (mean squared error) of the estimated values, which is a function of smoothing parameter λ. In log P-splines volatility estimation, the variance of the volatility fit $\hat{\sigma}\left(t_i, X_{t_i}\right) = \exp\left(\mathbf{B}_2(t_i)\hat{\boldsymbol{\delta}}_2\right)$ can be estimated by (10). EBBS models the bias of the volatility fit as a function of the penalty parameter λ at any fixed t_i. The estimated MSE of $\hat{\sigma}\left(t_i, X_{t_i}\right)$ at t_i and λ, $MSE(\hat{\sigma}; t_i; \lambda)$, is then calculated as the estimated squared bias plus the estimated variance. $MSE(\hat{\sigma}; t_i; \lambda)$ is averaged over t_i and then minimized over λ. The bias at any fixed t_i is obtained by a fit at t_i for a range of values of the smoothing parameter λ and a curve is then fitted to model bias. Our implementation is similar to Jarrow, Ruppert, and Yu (2004). Also see Ruppert (1997) and Ruppert et al. (1997) for details.

We need to point out that we by no means recommend against GCV choice of smoothing parameters in general. Indeed, we prefer the GCV criterion when in most cases GCV and EBBS perform similarly and GCV is usually simpler to compute. If a relatively small number of knots are used, then GCV and EBBS give virtually the same fit unsurprisingly. The number of knots for α, β, and $\sigma\left(t_i, X_{t_i}\right)$ in our simulation and case studies found to give the most stable results is only around 10. GCV tends to undersmooth and EBBS seems to be more appropriate in the applications when autocorrelations are evident, or when derivative function is of interest, or when local smoothing parameter is preferred. As discussed in Section 4, autocorrelations of the residuals are very mild in the data we used. Therefore, GCV is the preferred method in our case.

2.4.3 Multiple Smoothing Parameter for Drift Estimation

Different levels of smoothness are sometimes desired for different coefficient functions. A particular nice feature for the P-splines approach is that different smoothing parameters can be easily adopted. For example, in the drift estimation different smoothing parameters λ_α and λ_β can be readily implemented for coefficient functions $\alpha(t_i)$ and $\beta(t_i)$ respectively. It is not obvious to us how to incorporate multiple bandwidths for drift optimally in local approaches. Computationally, one could use a two-dimensional grid search. We suggest a simple calculation as in Ruppert and Carroll (2000). First obtain a common smoothing parameter λ by GCV or EBBS, which is chosen from a trial sequence of grid values. Starting with this common smoothing parameter with λ_β fixed, we select λ_α by GCV or EBBS. We then fix the selected λ_α, select λ_β by GCV or EBBS.

2.5 Discussion

One computational advantage in the above log P-splines approach for diffusion models is that the power term γ is naturally embedded in the spline estimation with the extended spline basis for volatility. γ has interesting implications. For example, as discussed in Section 1, model (1) with the usual linear drift and $\gamma = 1$ gives the Black-Scholes model. As we will see in Section 3 next, the estimation of γ in the local method is not trivial in that some complicated iterative algorithm is involved. Our experience from the case study and a limited simulation study suggests that the log P-splines approach is more stable and efficient in practice, though asymptotic theorems of local methods may be more complete.

Another appealing feature for P-splines is that the nonquadratic penalty function can be readily implemented. A general penalty can take the form $\sum \|\boldsymbol{\delta}\|^{q}$ with $q=2$, corresponding to the usual quadratic penalty. Using absolute deviation penalty $q=1$ as in LASSO (Tibshrani 1996) or $q=0$, for example, the fitted splines can take changes, accommodating possible jumps. Accurate estimation of volatility with jumps is an ongoing research topic of rising interest in finance. We wish to further explore it in the future.

3. Local Log-Linear Diffusion Estimation

We also consider local log-linear volatility estimation for the time-inhomogeneous diffusion model (2) based on a kernel-weighted likelihood method. Some other kernel-based variance function estimation methods (e.g., Ruppert et al. 1997, Fan and Yao 1998) could also be used for the aim here. However, these methods are based on residuals and do not take advantage of information from the likelihood function. Also, these methods do not always give non-negative estimators due to possible negativity of the local linear weight function. The local log-linear approach may give smaller bias than kernel-weighted residuals estimation for a class of variance function (Yu and Jones 2004). Usually, local linear can achieve both lower bias and variance of estimates with nicer properties at the boundary than local constant in FJZZ (2003).

3.1 Maximum Kernel-Weighted Likelihood Estimation

An appropriate localized normal log-likelihood for model (2) is given by minus

$$\sum_{i=1}^{n}\frac{1}{h}K\left(\frac{t_i-t}{h}\right)\left(\frac{1}{\Delta_i}\left\{Y_{t_i}-\mu\left(t_i,X_{t_i}\right)\Delta_i\right\}^2\Big/\sigma^2\left(t_i,X_{t_i}\right)+\log\left(\sigma^2\left(t_i,X_{t_i}\right)\right)\right),$$

where $\mu(t,X_t)=\alpha(t)+\beta(t)X_t$, $\sigma^2(t,X_t)=\sigma^2(t)X_t^{2\gamma}$, and $\alpha(t)$, $\beta(t)$ of drift and $\log\sigma^2(t)$ of volatility are functions to be fitted locally. Similar to the log P-splines approach and that of Yu and Jones (2004), we find the natural shortcut of the two-step procedure gives very good fits. Hence, we focus on volatility estimation and advocate a two-step procedure in practice. Different bandwidths h_1 and h_2 are desirable for drift and volatility estimation respectively. The same bandwidth h_1 for $\alpha\left(t\right)$ and $\beta\left(t\right)$ is used.

In Step 1, a standard local approach of the least-square routine of regression mean function estimation can be adopted. Let $K_i = h^{-1}K\left(h^{-1}\left(t_i - t\right)\right)$ be the short-hand kernel function. We note that under kernel (local constant) smooth, we can minimize $\sum_i \left(Y_{t_i}/\Delta_i - \alpha - \beta X_{t_i}\right)^2 K_{i1}$ with respect to α and β. This gives

$$\hat{\alpha} = \left(A_0B_2 - A_1B_1\right)\Big/\left(B_0B_2 - B_1^2\right),\ \hat{\beta} = \left(A_1B_0 - A_0B_1\right)\Big/\left(B_0B_2 - B_1^2\right),$$

where $A_j = \sum_i \left(Y_{t_i}/\Delta_i\right)\left(X_{t_i}\right)^j K_{i1}$ and $B_j = \sum_i \left(X_{t_i}\right)^j K_{i1}$, $j = 0,\ 1,\ 2$. Then $\hat{\alpha}$ (same for $\hat{\beta}$) can also be written as $\hat{\alpha} = \sum_i K_i \left(B_2 - X_{t_i}B_1\right)Y_{t_i}\Big/\left(B_0B_2 - B_1^2\right)$, a same formula as to the local linear regression mean function estimation (Wand and Jones 1995). This indicates that those existing bandwidth selection rules for kernel smoothing mean could be modified and adapted for use in drift estimation.

Let $\hat{\mu}\left(t, X_t\right)$ be the time-inhomogeneous drift estimator from Step 1. As in (12), denote $e_{t_i} = \left(1\Big/\sqrt{\Delta_i}\right)\left(y_{t_i} - \hat{\mu}(t_i, X_{t_i})\Delta_i\right)$. We then model $\log\sigma^2\left(t\right)$ as a local linear function. This leads to the following local kernel weighted likelihood estimation equation in Step 2 volatility estimation:

$$\sum_{i=1}^{n} K_{i2}\left(e_{t_i}{}^2 \exp\left\{-\left(\upsilon_0 + \upsilon_1\left(t_i - t\right)\right)\right\} X_{t_i}^{-2\gamma} + \upsilon_0 + \upsilon_1\left(t_i - t\right) + \gamma \log X_{t_i}^2\right), \tag{16}$$

where υ_0 and υ_1 are local linear parameter functions. The scale parameter γ is estimated via global minimization of the following equation

$$\sum_{i=1}^{n}\left(e_{t_i}{}^2 \exp\left(-\hat{\upsilon}_0(t_i)\right) X_{t_i}^{-2\gamma} + \gamma \log X_{t_i}^2\right). \tag{17}$$

Once we have estimates for υ_0 and γ, which are denoted as $\hat{\upsilon}_0\left(t\right)$ and $\hat{\gamma}$ respectively, we can estimate the volatility by

$$\hat{\sigma}\left(t, X_t\right) = \exp\left(\hat{\upsilon}_0(t)/2\right) X_t^{\hat{\gamma}} .$$

Note that given γ, (16) is similar to the estimating equation in Yu and Jones (2004). We outline an algorithm via setting the partial derivatives of localized normal log-likelihood function to zero. Take derivatives of (16) with respect to υ_0 and υ_1, we obtain

$$\exp(\upsilon_0) = \sum_{i=1}^{n} K_{i2} e_{t_i}{}^2 \exp\left(-\upsilon_1 \left(t_i - t\right)\right) X_{t_i}^{-2\gamma} \Big/ \sum_{i=1}^{n} K_{i2} \,, \tag{18}$$

$$\exp(\upsilon_0) = \sum_{i=1}^{n} K_{i2} e_{t_i}{}^2 \left(t_i - t\right) \exp\left(-\upsilon_1 \left(t_i - t\right)\right) X_{t_i}^{-2\gamma} \Big/ \sum_{i=1}^{n} K_{i2} \left(t_i - t\right). \tag{19}$$

Equating equation (18) to (19) provides a single equation to solve a single function υ_1. Once we obtain υ_1, we can get υ_0 via equation (18) or (19). Alternatively, an iterative algorithm via equation (18) and (19) can be used.

3.2 Rule-of-Thumb (ROT) Bandwidth Selection

Two independent data-based bandwidths are used for estimating drift and volatility respectively. Basically, bandwidths could be selected based on minimization of the integrated version of asymptotic mean squared errors or the residual squares criterion. Typically, the bandwidth for estimating drift could use many existing rules for smoothing regression mean functions. An example is the RSW rule (Ruppert et al. 1995).

For volatility estimation, we suggest a simple rule of thumb (ROT) bandwidth selection h_2 similar as in Yu and Jones (2004). It is based on minimizing the asymptotic mean integrated squared errors (MISE) using the results from Theorems in Section 3.3. In particular, a simple rule-of-thumb bandwidth selector is:

$$h_2 = \left\{ \frac{2R\left(K\right)V_1}{a_2^2\left(K\right)Bn} \right\}^{1/5},$$

where

$$a_2\left(K\right) = \int_{-1}^{1} z^2 K\left(z\right) dz \,, R\left(K\right) = \int_{-1}^{1} K^2\left(z\right) dz$$

$$B = \left(4/n\right) \sum_{i=1}^{n} \left(\hat{c}_2 + 3\hat{c}_3 t_i\right)^2 \exp\left(2\left(\hat{c}_0 + \hat{c}_1 t_i + \hat{c}_2 t_i^2 + \hat{c}_3 t_i^2\right)\right),$$

$$V_1 = \int_a^b \exp\left(2\left(\hat{c}_0 + \hat{c}_1 t + \hat{c}_2 t^2 + \hat{c}_3 t^3\right)\right) dt \,,$$

and the latter being obtained numerically. $\hat{c}_i$ $(i=1,2,3)$ are obtained via fitting a cubic function globally to the logged squared residuals arising from an initial fitting of drift (see Yu and Jones 2004 for details).

3.3 Asymptotic Properties

The asymptotic properties of estimating squared volatility $V(t)=\sigma^2(t,X_t)$, volatility $\sigma(t,X_t)$, and power γ are given by Theorems 2, 3, and 4 respectively under the following conditions:

(1) drift $\mu(t,X_t)$ and volatility $\sigma(t,X_t)$ are second-differentiable functions.

(2) kernel function K is a Lipschitz continuous symmetric density on $[-1,1]$.

(3) bandwidths $h_j=h_j(n)\to 0$ and $nh_j^{2+\delta}\to\infty$ for some $\delta>0$, $j=1,\ 2$.

Let $g(t)$ be the density function of time, which is usually a uniform distribution on time interval $[a,b]$. Then we have the following theorems.

Theorem 2. Under the foregoing regularity conditions, as $n\to\infty$, the estimator $\hat{V}(t)$ from (16) satisfies

$$\sqrt{nh_2}\,s(V(t))\times\left(\hat{V}(t)-V(t)-\frac{1}{2}a_2(K)b(t)h_2^2\{1+\mathrm{O}(h_2)\}\right)\to_D \mathrm{N}(0,1),$$

where $a_2(K)=\int_{-1}^{1}z^2K(z)dz$, $R(K)=\int_{-1}^{1}K^2(z)dz$, $b(t)=V(t)(\log V)''(t)$, and $s^2(V(t))=\left[2V^2(t)/nh_2 g(t)\right]R(K)$.

The proof is a combination of Taylor series expansion of normalized function of (16) and Gramer-Wold rule. The proof is long and is included in a working paper version downloadable from http://statqa.cba.uc.edu/~yuy/YYWL2.pdf.

In terms of estimating volatility by (16) via $\hat{\sigma}(t,X_t)$, we may derive the asymptotic property of $\hat{\sigma}(t,X_t)-\sigma(t,X_t)$ by Taylor expansion $\sqrt{\hat{V}(t)}-\sqrt{V(t)}\sim\left[1/2\sigma(t,X_t)\right]\left(\hat{V}(t)-V(t)\right)$.

Theorem 3. Under the same conditions of *Theorem 2*,

$$\sqrt{nh_2}\, s^*(t)\times\left(\hat{\sigma}(t,X_t)-\sigma(t,X_t)-\frac{1}{2}a_2(K)b(t)h_2^2\left\{1+\mathrm{O}(h_2)\right\}\right)\to_D \mathrm{N}(0,1)$$

where $s^*(t)^2=\left[\sigma^2(t,X_t)/2nh_2 g(t)\right]R(K)$.

By applying the likelihood estimation property to the log-likelihood equation (17) over parameter γ, we have another theorem.

Theorem 4. When (17) is a second continuous differentiable function on $(0,\infty)$ over γ and $n\to\infty$, the estimator $\hat{\gamma}$ from (20) is consistent and satisfies

$$\sqrt{n}I(\gamma)^{1/2}(\hat{\gamma}-\gamma)\to_D \mathrm{N}(0,1),$$

where the Fisher information

$$I(\gamma)=\mathrm{E}\left(-\gamma\sum_i\left(\frac{\frac{1}{\Delta_i}\left\{Y_{t_i}/\Delta_i-\mu(t_i,X_{t_i})\right\}^2}{\sigma(t_i)^2\left(X_{t_i}^2\right)^{\gamma+1}}+\log X_{t_i}^2\right)\right)^2.$$

Remark: From the asymptotic analysis, the optimal bandwidth is of the usual $O\left(n^{-1/5}\right)$ size and the optimal mean integrated squared errors are of the order $O\left(n^{-4/5}\right)$. An analogous theorem near the boundary can be easily obtained, which verifies the theoretical advantage of local linear (and local log-linear) approach over local constant (kernel) method at the boundary.

4. Treasury Bill Case Study

4.1 Data

We compare log P-splines, local log-linear, and FJZZ, in a case study with the weekly 3-month Treasury bill secondary market rate (weekly averages of business days) obtained from Federal Reserve Bank of St. Louis. The data set contains 2,638 observations from January 8, 1954 to July 23, 2004. The yields and their changes are plotted in Figure 1 (a) and (b). The volatility of changes in yield is clearly time-inhomogeneous. High volatility (Figure 1b) corresponds to high levels of interest rates (Figure 1a). During the high interest rate period from 1979 to 1982, the volatility was also large. These are confirmed by the descriptive statistics in Table 1. It displays the mean and standard deviation of both the weekly yields X_{t_i} and their changes y_{t_i} ($X_{t_i} - X_{t_{i-1}}$) during three periods: 1954 to 1978, 1979 to 1982, 1983 to 2004. Both the level of the yields (mean 11.52445) and the volatility of the yield changes (standard deviation 0.556564) were particularly high from 1979 to 1982.

Table 1. Descriptive statistics of yields and their Changes, artificially divided into three periods for illustration purpose.

		Mean		
Variables	Sample Size	1954-1978	1979-1982	1983-2004
Yield	2638	4.259%	11.524%	5.243%
Change	2637	0.006%	-0.006%	-0.006%

		Standard Deviation		
Variables	Sample Size	1954-1978	1979-1982	1983-2004
Yield	2638	1.892%	2.559%	2.269%
Change	2637	0.157%	0.557%	0.111%

In the next section, we report the estimation results from the three methods.

4.2 Estimation Results

For the log P-splines method, we focus on the two-step estimation method outlined in Section 2.3. A combination of degree of 1 and around 10 equally spaced quantile knots in the power basis for $\alpha(t_i)$, $\beta(t_i)$, and $\log\sigma(t_i)$ is found to give stable results. In log P-splines, the choice of smoothing parameter, as discussed in Section 2.4, is more critical than the degree or the number of knots. The smoothing parameter can be chosen using either GCV or EBBS but the results are similar. As shown in Figure 6, the autocorrelation in the residuals is mild and thus EBBS may not be necessary. GCV is certainly simpler to compute and the results reported here are from using GCV. Both the local log linear method and FJZZ's local constant method also estimate the drift and the volatility in an iterative fashion as in log P-splines. However, the parameter γ is not naturally embedded in volatility estimation as in log P-splines. In the local log linear method as described in Section 3.1, γ is estimated by minimizing equation (17). FJZZ maximizes a profile pseudo likelihood of γ to obtain an estimate. The local log linear method selects the bandwidth using the rule of thumb (ROT) while FJZZ minimizes the average prediction error (a function of the bandwidth) to choose the bandwidth.

Figure 1(b) clearly indicates that the volatility is much lower during the mid-

1960s than other periods. This means a drop of the fitted volatility during that period. Figure 1(c) shows the volatility estimates from the three methods. The volatility plot from log P-splines shows the decrease clearly while the other two methods show instead an increase of volatility. Log P-splines appear to model the volatility better than the other two methods. We further explore this issue with a small simulation study next, focusing on comparison of log P-splines and FJZZ.

The drift in the semiparametric inhomogeneous diffusion model (2) is set to 0 and the (true) inhomogeneous volatility $\sigma(t)$ follows the nonlinear trend in Figure 2(c). 1,000 simulations of sample size 2,000 are generated and estimated. A typical sample path and its difference are shown in Figures 2(a) and 2(b). Figure 2(c) shows that the log P-splines estimate of $\sigma(t)$ is very close to the true $\sigma(t)$. The estimate from FJZZ is very different from the true $\sigma(t)$. However, the volatility estimates $\hat{\sigma}(t)X_t^{\hat{\gamma}}$ from both log P-splines and FJZZ are close to the true volatility (see Figure 3). It appears that X_t^{γ} plays a significant role in the FJZZ estimate. Table 1 reports the median of both MSE (Mean Squared Error) and MAD (Mean Absolute Deviation) from the two methods. Figure 4 displays the boxplots of MSE and MAD from 1,000 simulations. Both the boxplots and Table 2 clearly indicate that the log P-splines method gives smaller MSE and MAD for $\sigma(t)$, γ, and the volatility.

Table 2. MSE and MAD Comparison: Median of 1,000 simulations.

SIGMA	LOG P-SPLINES	FJZZ
MSE	2.40E-03	2.28E-02
MAD	3.53E-02	1.28E-01
GAMMA	LOG P-SPLINES	FJZZ
MSE	6.25E-04	1.30E-03
MAD	1.69E-02	2.37E-02
VOLATILITY	LOG P-SPLINES	FJZZ
MSE	4.95E-04	3.59E-03
MAD	1.55E-02	3.96E-02

[Insert Figure 2 here]

[Insert Figure 3 here]

[Insert Figure 4 here]

We now go back to the case study. To assess the accuracy of the estimators, confidence intervals can be constructed. There are two ways to do this. When the assumption of independent Y_{t_i} holds, confidence intervals based on the asymptotic theorems in Sections 2 and 3 can be computed. When Y_{t_i} are dependent, the regression bootstrap (see Franke, Kreiss, and Mammen 2002) can be adopted. The basic idea is to generate the bootstrap samples $Y_{t_i}^*$ in (2) using the estimates of $\alpha(t_i)$, $\beta(t_i)$, $\sigma(t_i)$, γ, and normal errors ε_{t_i}. Then estimate the drift, volatility and other estimators of interest. Repeat this process a number of times to generate samples and find the confidence intervals. Figure 5 displays the estimate and bootstrap confidence band for volatility based on 1,000 bootstrap samples from the log P-splines method. Volatility is highest in early 1980s. During that period, the bootstrap confidence band is widest and the volatility is the most inhomogeneous. This is in agreement with the economic situation then, as described in Section 1. Asymptotic theorems in Sections 2 and 3 can also be applied to construct the confidence bands and the results are similar.

[Insert Figure 5 here]

4.3 Diagnostics

Diagnostics are performed to check the adequacy of the three methods. Figure 6 plots the autocorrelation functions of the standardized residuals from the three methods. There was mild autocorrelation at lag 1 for all three methods. We observe that two smoothing parameter selection criteria EBBS and GCV give similar results for the log P-splines approach. This is not surprising when the autocorrelation is mild. Since GCV is computationally more efficient, we recommend GCV for this case study. In other situations when autocorrelation is severe, EBBS might be more desirable.

The predictive power for the drift and volatility is compared using the correlation coefficient. The following correlation coefficients ρ_1 and ρ_2 are computed: ρ_1 is the correlation coefficient between the yield change and the estimated drift from the three methods. ρ_2 is the correlation coefficient between the squared yield change and the estimated volatility. The local log-linear method gives the highest ρ_1 and thus has the highest predictive power for the drift. The log P-splines method gives the highest ρ_2 and models the volatility best.

[Insert Figure 6 here]

Table 3. Correlation Coefficients.

Models	Log P-splines	Local Log-linear	FJZZ
ρ_1	0.1329	0.2347	0.0353
ρ_2	0.4071	0.3964	0.3099

4.4 Discussion

From this case study of the weekly three-month Treasury bill data and some limited simulation study, we find that the proposed log P-spline and local log-linear approaches can be successfully applied to time-inhomogeneous diffusion models. Our experience shows that log P-splines seem to be able to model the volatility best in our case study data. Log P-splines are also computationally efficient, and thus are recommended in practice. Both approaches guarantee that the volatility to be positive, an important appealing feature in practice. Inference is also readily available via either asymptotic theorems presented or regression bootstrap.

References

Ait-Sahalia, Y. (1996), "Nonparametric Pricing of Interest Rate Derivative Securities," *Econometrica*, 64, 527-560.

Black, F., and Scholes, M. (1973), "The Pricing of Options and Corporate Liabilities," *Journal of Political Economy*, 81, 637-654.

Carroll, R. J., Maca, J. D., and Ruppert, D. (1999), "Nonparametric regression in the presence of measurement error," *Biometrika*, 86, 541-554.

Carroll, R. J., Ruppert, D., and Stefanski, L. A. (1995), Measurement Error in Nonlinear Models, London: Chapman & Hall.

Carroll, R. J., Wu, C. J. F., and Ruppert, D. (1988), "The Effect of Estimating Weights in Weighted Least Squares," *Journal of the American Statistical Association*, 83, 1045-1054.

Chan, K. C., Kayolyi, G. A., Longstaff, F. A., and Sanders, A. B. (1992), "An Empirical Comparison of Alternative Models of the Short-Term Interest Rate," *Journal of Finance*, Vol. XLVII, NO. 3, 1209-1227.

Cox, J. C., Ingersoll J. E., and Ross, S. A. (1985), "A Theory of the Term Structure of

Interest Rates," *Econometrica*, 61, 929-952.

Duffie, D. (2001), *Dynamic Asset Pricing Theory* (3rd ed.), Princeton University Press.

Eilers, P.H. C., and Marx, B.D. (1996), "Flexible Smoothing with B-splines and Penalties (with Discussion)," *Statistical Science*, 11, 89-121.

Egorov, A. V., Li, H., Xu, Y. (2003), "Maximum Likelihood Estimation of Time-Inhomogeneous Diffusions," *Journal of Econometrics*, 114, 107-39.

Fan, J. and Gijbels, I. (1996), *Local Polynomial Modelling and Its Applications*, Chapman and Hall, London.

Fan, J., Jiang, J., Zhang, C., and Zhou, Z. (2003), "Time-inhomogeneous Diffusion Models for Term Structure Dynamics and the Stock Price Volatility," *Statistica Sinica*, 13, 965-992.

Fan, J. and Li, R. Z. (2001), "Variable selection via penalized likelihood," *Journal of the American Statistical Association*, 1348-1360.

Fan, J., and Yao, Q (1998), "Efficient Estimation of Conditional Variance Functions in Stochastic Regression," *Biometrika*, 85, 645-660.

Fan, J. and Zhang, C. (2003), "A re-examination of Stanton's diffusion estimations with applications to financial model validation," *Journal of the American Statistical Association*, 98, 118-134.

Franke, J., Kreiss, J. P., and Mammen, E. (2002), "Bootstrap of kernel smoothing in nonlinear time series," *Bernoulli*, 8, 1-37.

Friedman, J. H. (1991), Multivariate Adaptive Regression Splines (with discussion), *Annals of Statistics*, 19, 1-141.

Gray, R. J. (1994), "Spline-based tests in survival analysis," Biometrics, 50, 640-652.

Hall, P. and Opsomer, J. D. (2005), "Theory for penalized spline regression," *Biometrika*, 92, 105-118.

Heath, D., Jarrow, R., and Morton, A. (1992), Bond Pricing and the Term Structure of Interest Rates: a New Methodology for Contingent Claims Valuation," *Econometrica*, 60, 77-105.

Hull, J. and White, A. (1990), "Pricing Interest-rate Derivative Securities," *Review of Financial Studies*, 3, 573-592.

Jarrow, R., Ruppert, D., and Yu, Y. (2004). Estimating the Term Structure of Corporate Debt with a Semiparametric Penalized Spline Model. *Journal of the American Statistical Association*, 99, 57-66.

Lehmann, E. L. (1983), *Theory of Point Estimation*, Wiley, New York.

McCulloch, P. and Nelder, J. A. (1989), *Generalized Linear Models*, 2nd ed. London: Chapman & Hall.

Merton, R. C. (1992), *Continuous-Time Finance*, Oxford: Blackwell.

Rubinstein, M. (1994), "Implied Binomial Trees," *Journal of Finance*, 49, 771-818.

Ruppert, D. (1997), "Empirical-bias bandwidths for local polynomial nonparametric regression and density estimation," *Journal of the American Statistical Association*, 92, 1049-62.

Ruppert, D. and Carroll, R. J. (1997), "Penalized Regression Splines," Working Paper, School of Operations Research and Industrial Engineering, Cornell University.

Ruppert, D. and Carroll, R. J. (2000), "Spatially-adaptive penalties for spline fitting," Australian and New Zealand Journal of Statistics, 42, 205-253.

Ruppert, D., Sheather, S. J., and Wand, M. P. (1995), "An Effective Bandwidth Selector for Local Least Square Regression," *Journal of the American Statistical Association*, 90, 1257-1270.

Ruppert, D., Wand, M. P., Holst, U., and Hössjer, O. (1997), "Local Polynomial Variance Function Estimation," *Technometrics*, 39, 262-273.

Ruppert D., Wand, M. P., and Carroll, R. J. (2003), *Semiparametric Regression*, Cambridge University Press.

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

Tibshirani, R. (1996), "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society*, *Series B* 58: 267-88.

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

Wand, M. P. and Jones, M. C. (1995), *Kernel Smoothing*, London: Chapman & Hall.

Yu, Y. (2005), “Penalized Spline Estimation for Generalized Partially Linear Single-Index Models,” working paper.

Yu, K. and Jones, M. C. (2004), “Likelihood-based Local Linear Estimation of the Conditional Variance Function,” *Journal of the American Statistical Association*, 99, 139-144.

Yu, Y. and Ruppert, R. (2002), “Penalized Spline Estimation for Partially Linear Single-Index Models,” *Journal of the American Statistical Association*, 97, 1042-1054.

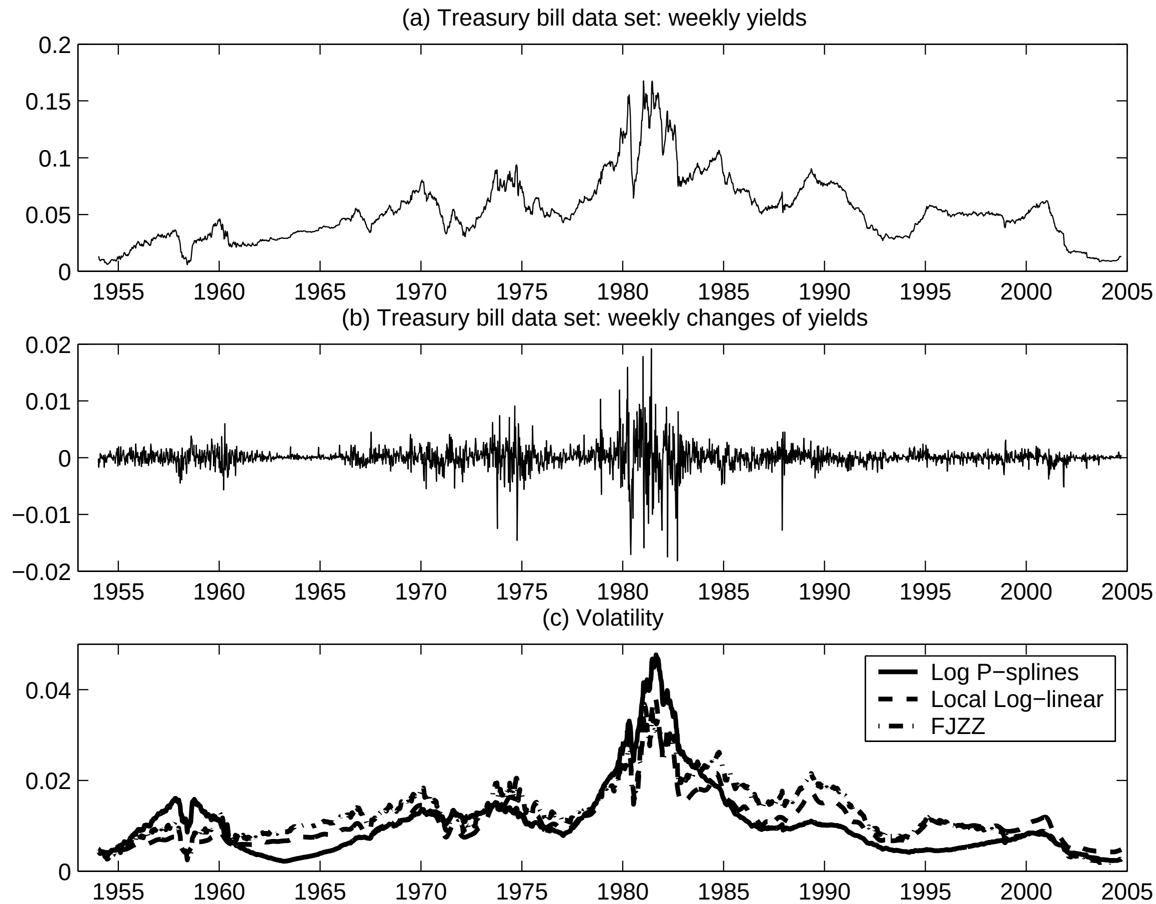


Figure 1. Weekly Treasury bill yields from 1954 to 2004. The yields, their changes, squared residuals from the drift estimation, and estimated squared volatility are plotted.

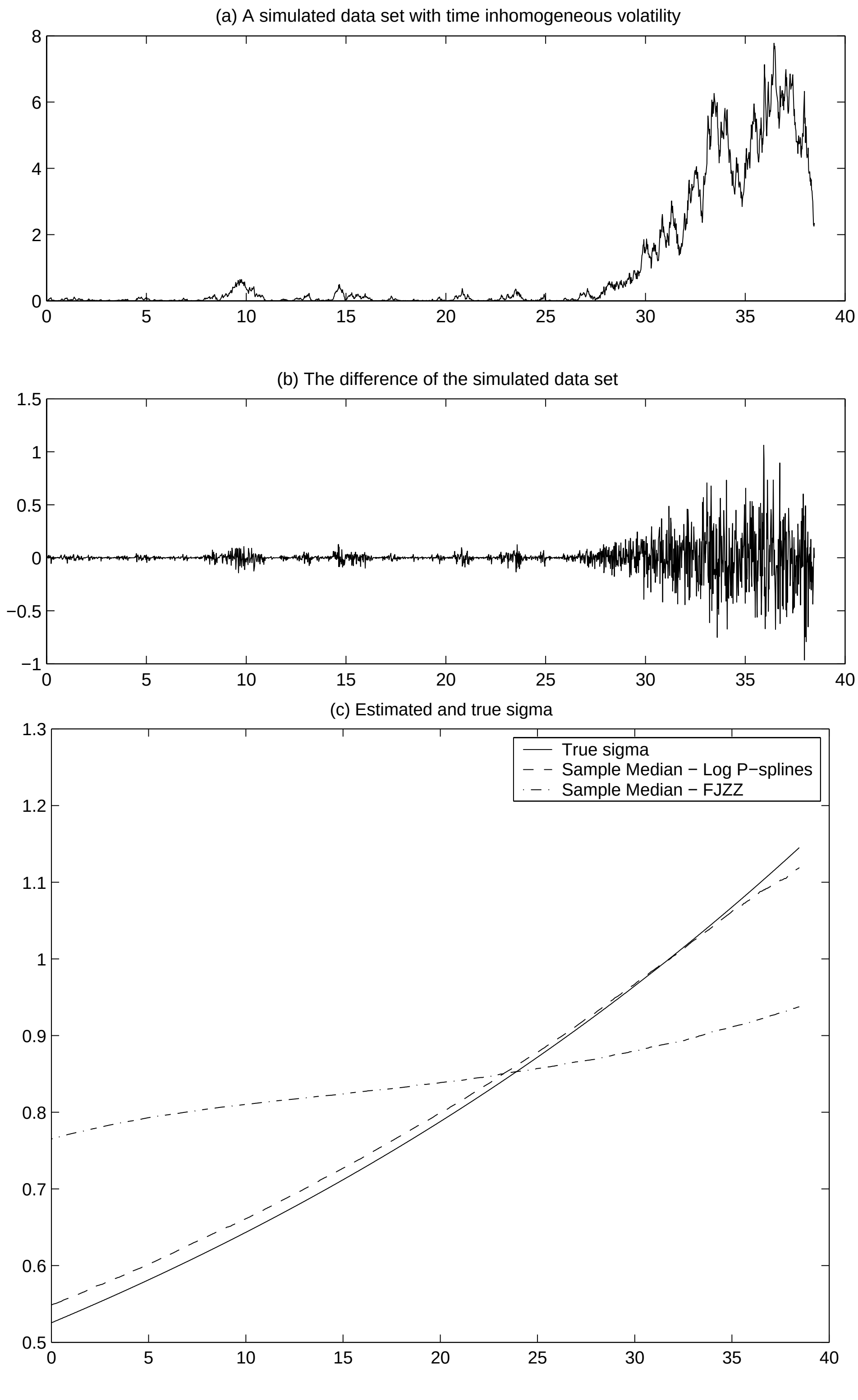
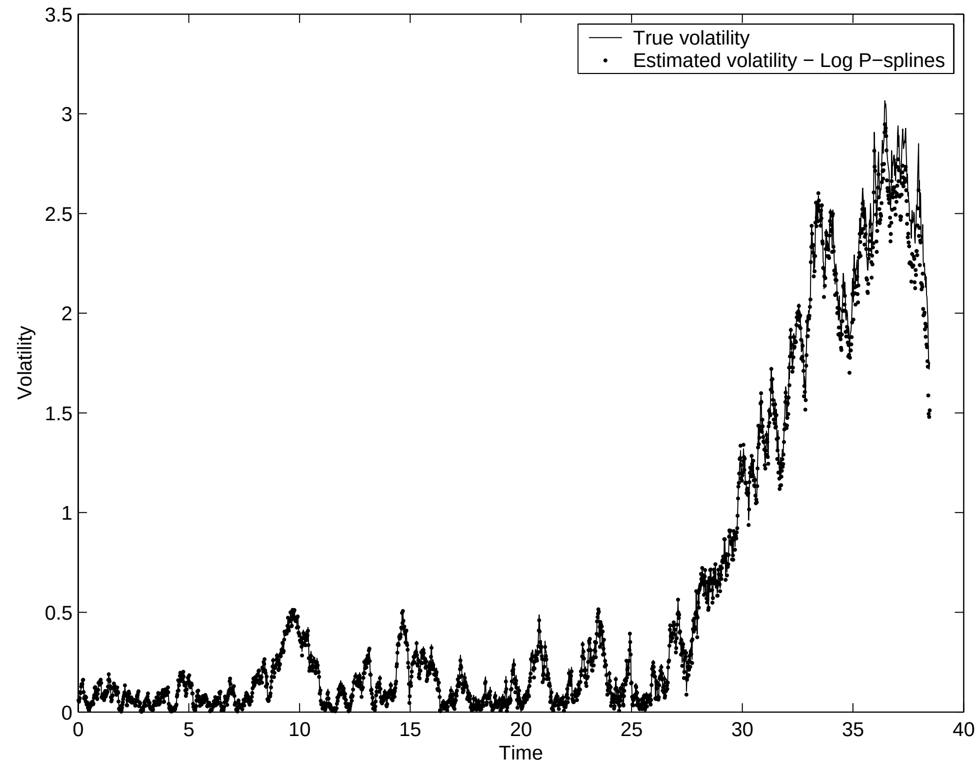


Figure 2. Simulated data and the estimated Sigma.



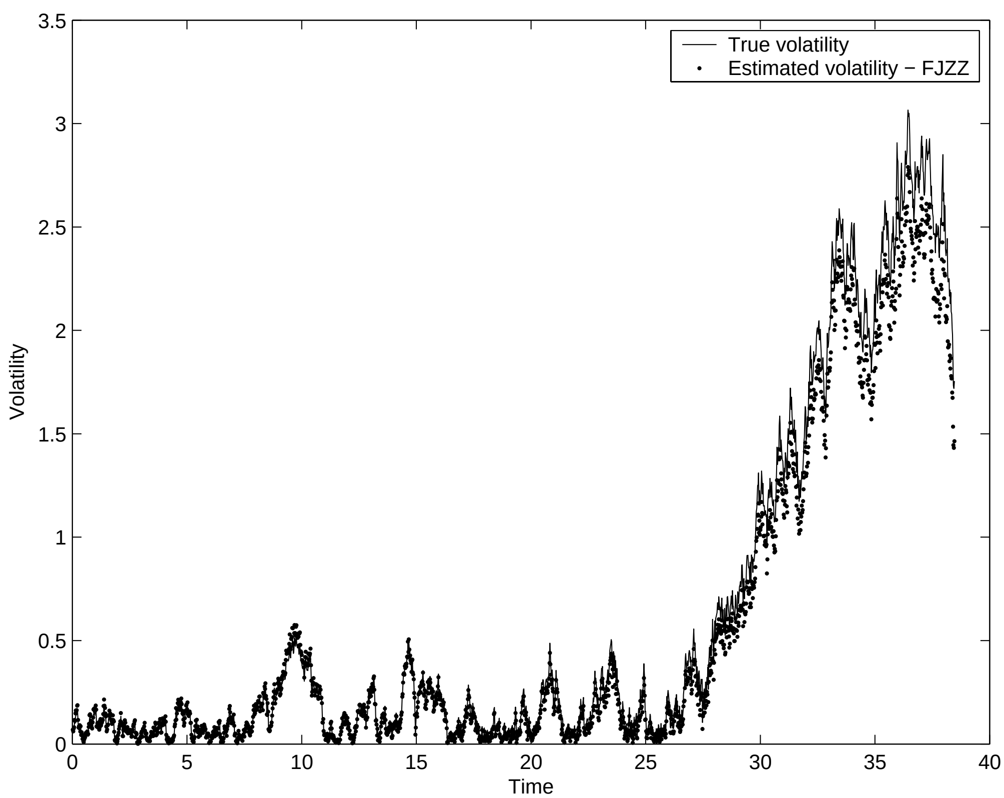


Figure 3. True and the estimated volatility (median) from 1,000 simulations.

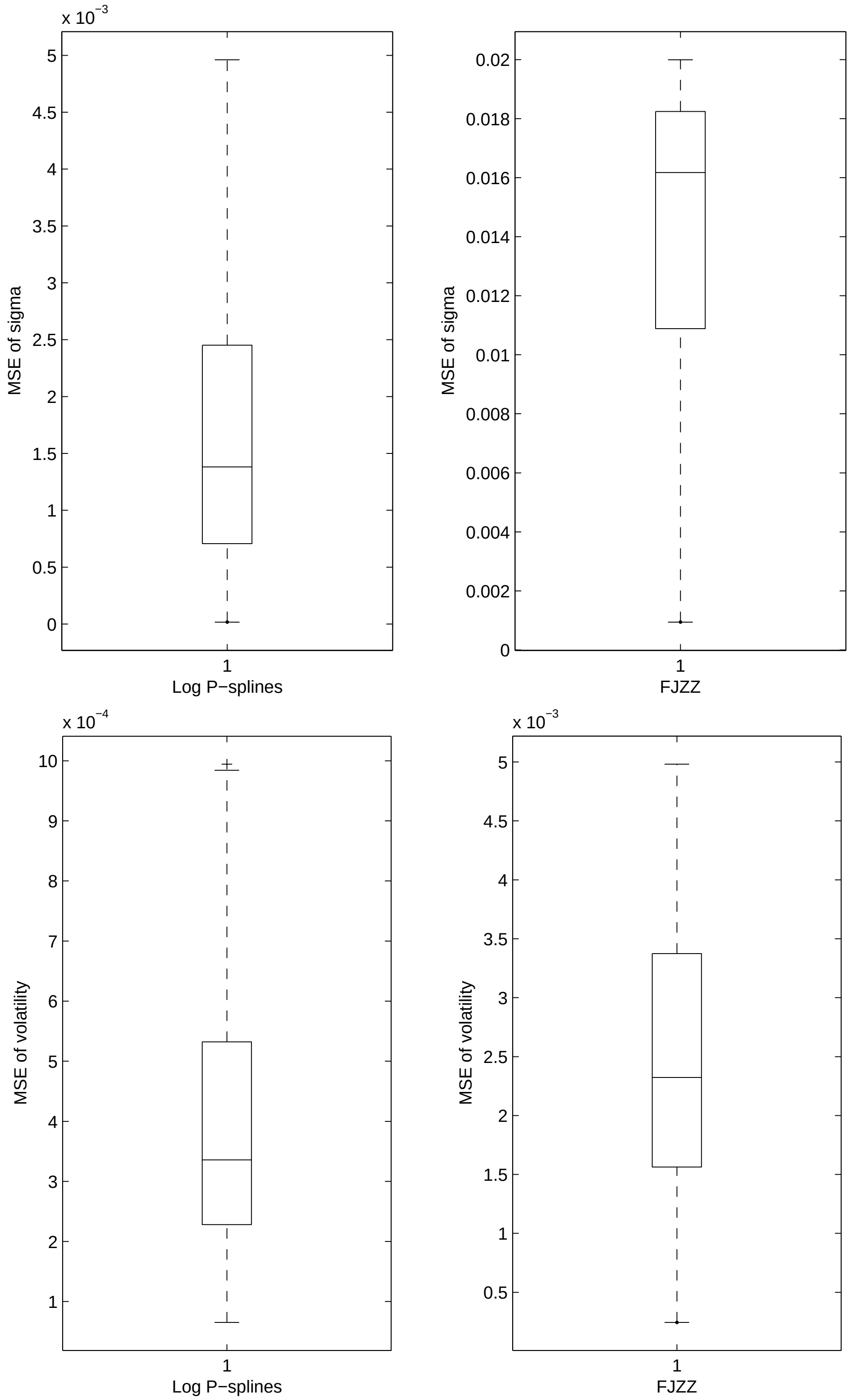


Figure 4. Boxplots of MSE and MAD from 1,000 simulations.

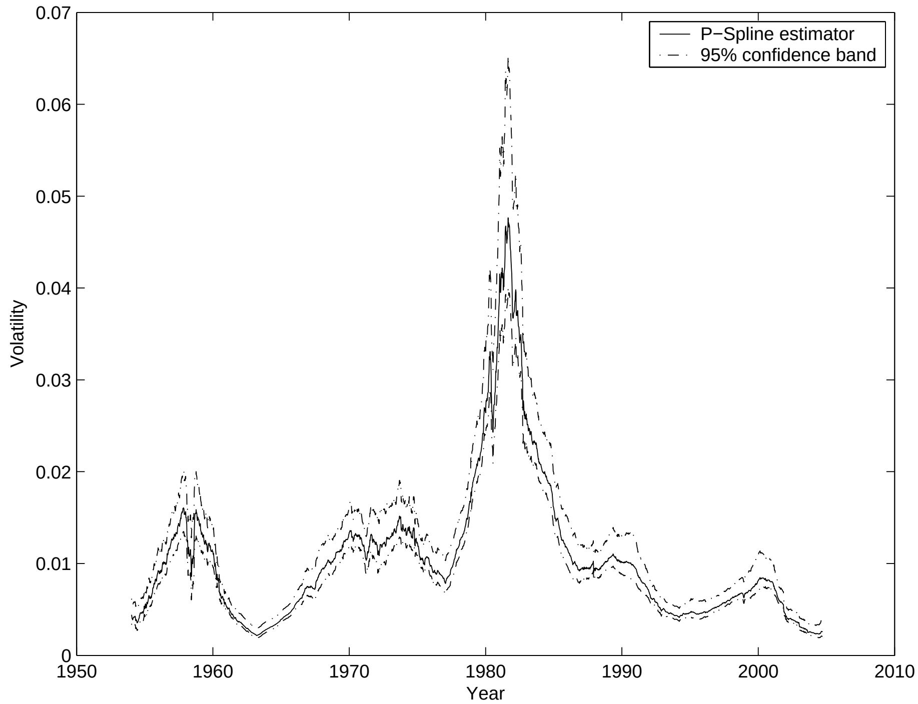


Figure 5. Time-inhomogeneous log P-splines diffusion estimates with the regression bootstrap confidence interval.

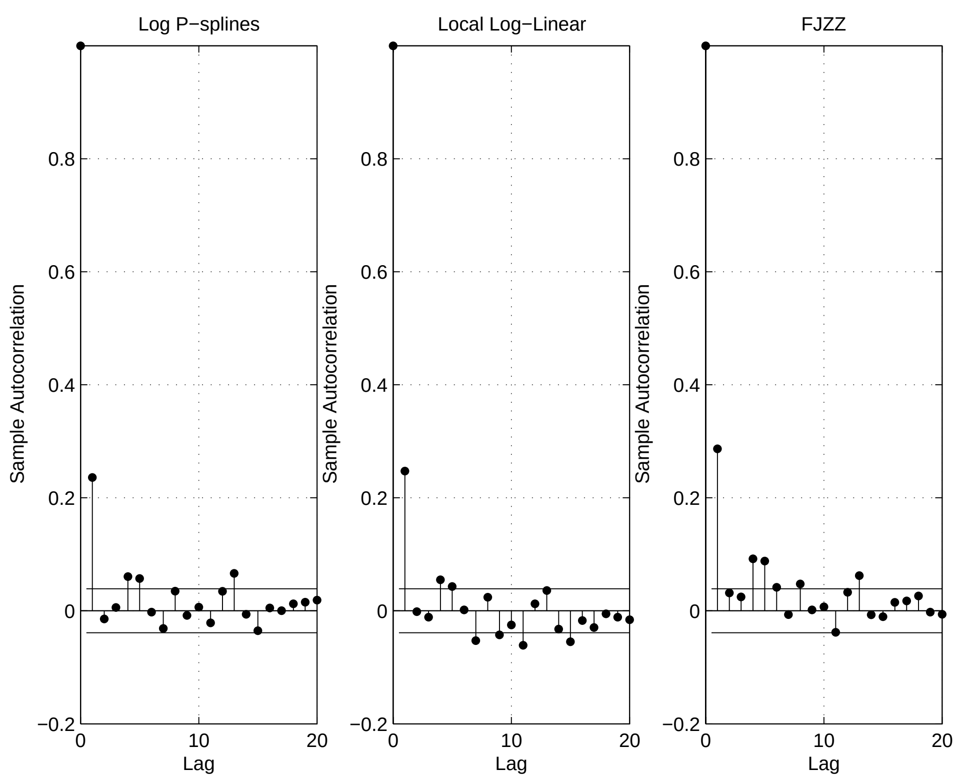


Figure 6. Diagnostic check: residual autocorrelation functions and normal QQ-plots using three model, log P-splines, local log-linear, FJZZ.