Statistica Sinica **19** (2009), 843-867

SEMIPARAMETRIC ESTIMATION FOR A CLASS OF TIME-INHOMOGENEOUS DIFFUSION PROCESSES

Yan Yu, Keming Yu, Hua Wang and Min Li

University of Cincinnati, Brunel University, Yahoo! Inc. and California State University, Sacramento

Abstract: We develop two likelihood-based approaches to semiparametrically estimate a class of time-inhomogeneous diffusion processes: 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 selections. 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 theorems for both approaches and report statistical inference results. 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 dip in mid-1960s the best. We also present an application to calculate a financial market risk measure called Value at Risk (VaR) using statistical estimates from log P-splines.

Key words and phrases: Bandwidth selection, kernel smoothing, local linear, option pricing, penalized likelihood, VaR, variance estimation, volatility.

1. Introduction

Diffusion processes are important tools for modeling the stochastic behavior of a range of economic variables, such as interest rates and stock prices. They are essential building blocks for pricing options (derivatives whose price depends on the price of another underlying asset) and risk management. The growth of the derivative market, which has existed for only about 35 years, is astonishing. To put this in perspective, the size of the derivatives market grew to \$415 trillion in 2006 (as measured in notional amounts outstanding; Source: Bank for International Settlements). On the other hand, due to the volatility of financial variables, risk management has become critical to corporations, especially after such institutions as Orange County and Long Term Capital Management lost billions of dollars in financial markets when senior management poorly managed risk exposure (Jorion (2000)).

We propose two semiparametric likelihood-based approaches, log P-splines and the local log-linear method, to modeling a class of time-inhomogeneous diffusion processes. Asymptotic properties are developed for inference. A case study of weekly three-month Treasury bill data from 1954 to 2004 is presented, where we further investigate derivative (bond) pricing and provide risk measures such as Value at Risk (VaR).

Most continuous-time asset pricing models assume that the underlying state variables follow diffusions, for example, the famous option pricing model of Black and Scholes (1973) (Scholes won the Nobel Prize in Economics in 1997), interest rate term structure models of Vasicek (1977), Cox, Ingersoll and Ross (1985, CIR), Hull and White (1990), Heath, Jarrow and Morton (1992) and Chan, Kayolyi, Longstaff and Sanders (1992, CKLS)). A nice overview can be found in Merton (1992) and Duffie (2001).

All of these diffusion processes 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 (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 option pricing. Finally, volatility itself is so important that the volatility index (VIX) of a market has recently become a financial instrument. On March 26, 2004 the VIX compiled by the Chicago Board of Option Exchange began trading in futures.

Empirical tests of the different parametric diffusion models mentioned above have yielded mixed results (Stanton (1997)). This is not too surprising since they are neither derived from any economic theory nor have offered guidance in choosing the correct model. With the availability of high-quality data on many financial assets, researchers have recently 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) and Jiang and Knight (1997) estimate both the drift $\mu(X_t)$ and diffusion $\sigma(X_t)$ nonparametrically using the kernel method. Jiang (1998) develops a nonparametric approach to model the interest rate term structure dynamics based on a spot rate process. Bandi and Phillips (2003) propose a nonparametric scalar diffusion model without assuming stationarity. Ait-Sahalia, Fan and Peng (2006) develop a specification test for the diffusion process to compare nonparametric and parametric estimates. Hong and Li (2005) propose a transition density based validation approach.

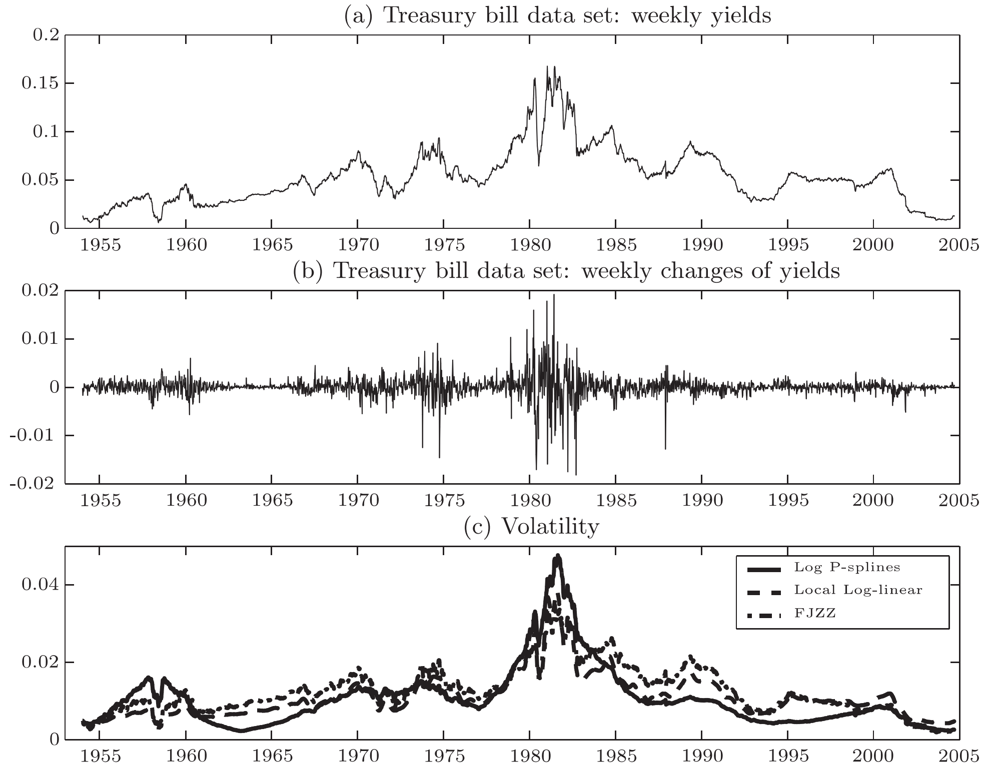

Figure 1. Weekly Treasury bill yields from 1954 to 2004. The yields, their changes, squared residuals from the drift estimation, and estimated squared volatility are plotted.

While nonparametric estimation of diffusion models is promising, mostly they posit 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 increase for interest rates at all maturities on the days of FOMC (Federal Open Market Committee) meetings. 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 the diffusion or volatility $\sigma(t, X_t)$ depend on time t. Figure 1 (a) is a plot of weekly 3-month Treasury bill rates during the period from 1954 to 2004. The differenced rates are plotted in (b). Visually the differenced rates seem to behave randomly with small volatility around the mid-1960s and mid-1990s, with larger volatility during the 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), hereafter, FJZZ) 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.

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 FJZZ finds that there is not sufficient information to determine the bivariate functions nonparametrically, and that forcing all coefficients in the drift and diffusion to be time-dependent may cause over-parameterization.

Here we focus on the semiparametric time-inhomogeneous model

$$dX_t = (\alpha(t) + \beta(t)X_t)dt + \sigma(t)(X_t)^{\gamma} dW_t, \tag{1.1}$$

where γ is a scalar parameter independent of time t, $\alpha(t)$ and $\beta(t)$ are time-dependent coefficients of the drift, and $\sigma(t)$ is a time-dependent coefficient of diffusion (volatility). Model (1.1) includes most of the well-known diffusion models. For example, when $\alpha(t)$, $\beta(t)$, and $\sigma(t)$ are constants (time independent), (1.1) yields to the CKLS model. In the CKLS framework, $\gamma = 1$ corresponds to the famous Black-Scholes model; $\gamma = 0$ corresponds to the Vasicek model; and $\gamma = 0.5$ corresponds to the 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 model proposed here, with γ as a parameter, has the advantage of allowing testing parametric restrictions to determine which model fits the data adequately using formal tests such as the Wald test, as discussed in Section 2.2.

This paper contains a statistical finance application to the short term Treasury bill data, as well as some methodological developments for perhaps 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 can in general give lower bias and variance of estimates with more appealing properties at the boundary (Fan and Gijbels (1996), Yu and Jones (2004) and Fan, Jiang, Zhang and Zhou (2003)).

Second, we investigate different smoothing parameter (bandwidth) selection criteria: generalized cross validation (GCV) and the 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) and 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). This empirical selection of smoothing parameters provides an alternative and efficient new method in this field. Our ROT bandwidth is simple and works almost as well as the unavailable optimal bandwidths. The ROT local log-linear approach performs better than the local linear and local constant approaches. The proposed approach is as good as or better than that of Ruppert, Wand, Holst and Hössjer (1997), even when the latter uses its optimal bandwidths.

Third, 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 the 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 contexts (Yu and Ruppert (2002) and Jarrow, Ruppert and Yu (2004)). However, the theory from the local log-linear approach is more complete in the sense of being "truely nonparametric." We give a large sample property based on fixed knot P-splines, which often serve well in application. Other spline methods could also be used, though their direct adoption might be complicated and we expect the fit would be similar. This semiparametric model encompasses many well-known asset pricing models such as Black-Scholes, and asymptotic results can be applied to test the adequacy of these models via simple parametric restrictions.

Based on the asymptotic results, confidence intervals of the volatility estimate can be obtained. Estimating volatility is generally challenging and difficult, and the confidence interval provides additional information about the accuracy and variation of underlying volatility.

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 discusses local log-linear estimation. A case study of 3-month Treasury bill data is presented in Section 4.

2. Log Penalized Splines Diffusion Estimation

The semiparametric time-inhomogeneous model (1.1) is continuous, but the data sampled in the financial markets are usually discrete. Therefore, in estimation, the discretized version of (1.1) based on the Euler scheme is used as an approximation. Suppose the data $\{X_{t_i},\ i = 1, \ldots, n+1\}$ are sampled at discrete time points, $t_1 < \cdots < t_{n+1}$. For weekly data when the time unit is a year, $t_i = t_0 + i/52$ $(i = 1, \ldots, n)$, 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 $\Delta_i = t_{i+1} - t_i$. The 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 (1.1) becomes

$$y_{t_i} \approx (\alpha(t_i) + \beta(t_i) X_{t_i})\Delta_i + \sigma(t_i)(X_{t_i})^\gamma \sqrt{\Delta_i}\varepsilon_{t_i}, \tag{2.1}$$

where $\{\varepsilon_{t_i}\}$ 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.

We first 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) and Ruppert, Wand and Carroll (2003). They 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)) and smoothness is achieved through a roughness penalty measure. An appealing feature of P-splines is that they allow multiple smoothing parameters and a variety of penalties, quadratic or nonquadratic, on the spline coefficients.

2.1. Maximum penalized likelihood estimation

We model time dependent functions $\alpha(t_i)$, $\beta(t_i)$ of drift, and $\log \sigma^2(t_i)$ of volatility, in model (2.1) by splines:

$$\begin{gathered} \alpha(t_i) = \mathbf{B}_\alpha(t_i)\boldsymbol{\delta}_\alpha; \quad \beta(t_i) = \mathbf{B}_\beta(t_i)\boldsymbol{\delta}_\beta, \\ \log \sigma^2(t_i) = 2\mathbf{B}_\sigma(t_i)\boldsymbol{\delta}_\sigma, \end{gathered} \tag{2.2}$$

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 bases using truncated power functions, 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 present P-splines using a truncated power basis function $\mathbf{B}(t_i) = [1, t_i, \dots t_i^p, (t_i - \kappa_1)_+^p, \dots, (t_i - \kappa_k)_+^p]$, where p is the spline polynomial degree, $(t_i - \kappa_k)_+ = \max(0, t - \kappa_k)$, $\kappa_1 < \cdots < \kappa_K$ are spline knots often located at equal-spaced sample quantiles for simplicity. Then we can write

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

The log likelihood function, excluding constants, is negative

$$\sum \left(\Big(\frac{1}{\Delta_1}\Big) \Big\{ y_{t_i} - \Big(\mathbf{B}_\alpha(t_i)\boldsymbol{\delta}_\alpha + \mathbf{B}_\beta(t_i)\boldsymbol{\delta}_\beta X_{t_i}\Big)\Delta_i \Big\}^2 \times \exp\Big\{ -\Big(2\mathbf{B}_\sigma(t_i)\boldsymbol{\delta}_\sigma + \gamma \log X_{t_i}^2\Big)\Big\} + 2\mathbf{B}_\sigma(t_i)\boldsymbol{\delta}_\sigma + \gamma \log X_{t_i}^2 \right).$$

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

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

$$Q_{n,\lambda}(\boldsymbol{\theta}) = L_n(\boldsymbol{\theta}) - \Big(\frac{n}{2}\Big)\boldsymbol{\lambda}\boldsymbol{\theta}^T\mathbf{D}\boldsymbol{\theta}, \tag{2.3}$$

where

$$L_n(\boldsymbol{\theta}) = \sum l_n(\boldsymbol{\theta}, t_i) = -\sum \left(\Big(\frac{1}{\Delta_i}\Big)\Big\{ y_{t_i} - \mathbf{B}_1(t_i)\boldsymbol{\delta}_1\Delta_i \Big\}^2 \times \exp\{-2\mathbf{B}_2(t_i)\boldsymbol{\delta}_2\} + 2\mathbf{B}_2(t_i)\boldsymbol{\delta}_2 \right). \tag{2.4}$$

Here $\mathbf{D}$ is an appropriate positive semi-definite matrix. We choose $\mathbf{D}$ as in Ruppert, Wand and Carroll (2003) that penalizes jumps at the knots in the pth derivative of the spline. Like the choice of basis functions, we found the choice of $\mathbf{D}$ to be relatively unimportant, and different $\mathbf{D}$'s give similar fits.

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

$$\hat{V}(t_i) := \hat{\sigma}^2(t_i, X_{t_i}) = \hat{\sigma}(t_i)^2 X_{t_i}^{2\hat{\gamma}} = \exp(2\mathbf{B}_2(t_i)\hat{\boldsymbol{\delta}}_2), \tag{2.5}$$

so the volatility estimate is

$$\hat{\sigma}(t_i, X_{t_i}) = \sqrt{\hat{\sigma}(t_i)^2 X_{t_i}^{2\hat{\gamma}}} = \exp(\mathbf{B}_2(t_i)\hat{\sigma}_2). \tag{2.6}$$

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, there are open questions about 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, as has been noted in the literature, e.g., Gray (1994) and Carroll, Maca and Ruppert (1999).

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

$$n^{\frac{1}{2}}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \rightarrow_D N(0, \mathbf{I}^{-1}(\boldsymbol{\theta})), \tag{2.7}$$

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

The proof of Theorem 1 is standard with ordinary (no penalty) maximum likelihood estimates (Lehmann and Casella (1998)), and is similar to Fan and Li (2001) with penalty function.

The result given in (2.7) 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 asymptotic distribution with fixed penalty parameter derived from the estimating equation approach using the "sandwich formula"

$$n^{\frac{1}{2}}(\hat{\boldsymbol{\theta}}(\boldsymbol{\lambda}) - \boldsymbol{\theta}(\boldsymbol{\lambda})) \rightarrow_D N\Big\{0, \mathbf{H}^{-1}(\boldsymbol{\theta}(\boldsymbol{\lambda}))\mathbf{G}(\boldsymbol{\theta}(\boldsymbol{\lambda}))\mathbf{H}^{-T}(\boldsymbol{\theta}(\boldsymbol{\lambda}))\Big\}, \tag{2.8}$$

where $\mathbf{H}(\boldsymbol{\theta}) = \sum(\partial/\partial\boldsymbol{\theta}^T)\boldsymbol{\psi}_{t_i}(\boldsymbol{\theta})$, $\mathbf{G}(\boldsymbol{\theta}) = \sum \boldsymbol{\psi}_{t_i}(\boldsymbol{\theta})\boldsymbol{\psi}_{t_i}^T(\boldsymbol{\theta})$, $\boldsymbol{\psi}_{t_i}(\boldsymbol{\theta}) = -(\partial/\partial\boldsymbol{\theta}^T)$ $l_n(\boldsymbol{\theta}; t_i) + \boldsymbol{\lambda}\mathbf{D}\boldsymbol{\theta}$, (see Carroll, Ruppert, Stefanski and Crainiceanu (2006) and Yu and Ruppert (2002)).

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

$$sd\Big\{\hat{\sigma}(t_i, X_{t_i})\Big\} = \sqrt{\mathbf{B}_2(t_i)\hat{Var}(\hat{\boldsymbol{\delta}}_2)\mathbf{B}_2^T(t_i)}\exp(\mathbf{B}_2(t_i)\hat{\boldsymbol{\delta}}_2), \tag{2.9}$$

where $\hat{Var}(\hat{\boldsymbol{\delta}}_2)$ is given by (2.8). Note that as $\boldsymbol{\lambda}$ goes to zero, $\hat{Var}(\hat{\boldsymbol{\delta}}_2(\boldsymbol{\lambda}))$ in (2.8) converges to the corresponding matrix of inverse of Fisher information given in (2.7).

The asymptotic results can be readily used to construct confidence bands or perform hypothesis testing. Confidence bands for $\alpha(t_i)$, $\beta(t_i)$, and the volatility estimate $\hat{\sigma}(t_i)X_{t_i}^{\hat{\gamma}}$ are reported in Section 4.2. Hypotheses tests such as the Wald test can be applied to test certain restrictions on $\boldsymbol{\theta}$. For example, the restrictions that $\alpha(t_i) = 0$, $\beta(t_i)$ and $\sigma(t_i)$ only containing the intercepts, and $\gamma = 1$ give the Black-Scholes option pricing model. The Wald test can test the null hypothesis $H_0 : R\boldsymbol{\theta}_0 - q_0 = 0$ where R represents the restrictions on the parameter vector $\boldsymbol{\theta}$ and is of size $d_1 \times \dim(\boldsymbol{\theta})$ with $d_1 \leq \dim(\boldsymbol{\theta})$. The Wald statistic $W = (R\hat{\boldsymbol{\theta}} - q_0)^T(R\hat{\mathrm{Var}}\,(\hat{\boldsymbol{\theta}})R^T)^{-1}(R\hat{\boldsymbol{\theta}} - q_o)$ has a Chi-squared limiting distribution with d_1 degrees of freedom, where $\hat{\mathrm{Var}}\,(\hat{\boldsymbol{\theta}})$ can be computed using (2.8).

2.3. An algorithm

One could do one-step maximization on the penalized likelihood function (2.3). However, in our P-splines approach, the number of parameters could be large and this estimation algorithm may not be efficient. We then implement an iterative algorithm by reweighing 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. Volatility estimation is our primary focus, and we advocate two-step estimation in practice.

Step 1: Drift Estimation.

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

$$\sum_{i=1}^{n} \left\{ \left(\frac{y_{t_i}}{\Delta_i}\right) - \mathbf{B}_1(t_i)\boldsymbol{\delta}_i \right\}^2 + \left(\frac{n}{2}\right)\boldsymbol{\lambda}_1\boldsymbol{\delta}_1^T\mathbf{D}_1\boldsymbol{\delta}_1. \tag{2.10}$$

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

Step 2: Log P-splines Volatility Estimation.

Denote the residual from the previous drift estimation by

$$e_{t_i} = \left(\frac{1}{\Delta_i^{1/2}}\right)(y_{t_i} - \hat{\mu}(t_i, X_{t_i})\Delta_i). \tag{2.11}$$

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

Remark. Stanton (1997) pointed out that this last approximation holds even if $\hat{\mu}(t_i, X_{t_i}) = 0$ is assumed, though the approximation error using (2.11) is of smaller order. This observation further confirms the validity of our two-step approach with primary focus on volatility estimation.

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

$$\sum \Big(e_{t_i}^2 \exp\{-2\mathbf{B}_2(t_i)\boldsymbol{\delta}_2\} + 2\mathbf{B}_2(t_i)\boldsymbol{\delta}_2\Big) + \Big(\frac{n}{2}\Big)\boldsymbol{\lambda}_2\boldsymbol{\delta}_2^T\mathbf{D}_2\boldsymbol{\delta}_2. \tag{2.12}$$

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 $\hat{\boldsymbol{\delta}}_{2,pre} = (\mathbf{B}_2^T\mathbf{B}_2 + n\lambda_2\mathbf{D}_2)^{-1}\mathbf{B}_2^T\mathbf{E}$, where vector $\mathbf{E}$ has the ith element $\log|e_{t_i}|$. The volatility estimate is given by

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

2.4. Selection of smoothing parameter

It is well-known that smoothing parameter selection is very important in nonparametric methods. The typical method is cross-validation and minimization of mean square error (or asymptotic MSE). Empirical selection of smoothing parameters provides an alternative and efficient new method in this field.

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

$$GCV(\boldsymbol{\lambda}_1) = \frac{ASR(\boldsymbol{\lambda}_1)}{\Big[1 - trace\Big\{\mathbf{B}_1(\mathbf{B}_1^T\mathbf{B}_1 + n\boldsymbol{\lambda}_1\mathbf{D}_1)^{-1}\mathbf{B}_1^T\Big/n\Big\}\Big]^2}, \tag{2.13}$$

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

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

$$GCV(\lambda_2) = \frac{Deviance(\lambda_2)}{\Big[1 - trace\Big\{\mathbf{B}_2(\mathbf{B}_2^T\mathbf{B}_2 + n\lambda_2\mathbf{D}_2)^{-1}\mathbf{B}_2^T\Big\}\Big/n\Big]^2}, \tag{2.14}$$

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 for a number of reasons (Ruppert, Wand, Holst and Hössjer (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 $\hat{\sigma}(t_i, X_{t_i}) = \exp(\mathbf{B}_2(t_i)\hat{\boldsymbol{\delta}}_2)$ can be estimated by (2.9). 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 $\hat{\sigma}(t_i, X_{t_i})$ at t_i and λ, $\text{MSE}(\hat{\sigma}; t_i; \lambda)$, is then calculated as the estimated squared bias plus the estimated variance. $\text{MSE}(\hat{\sigma}; t_i; \lambda)$ 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 that of Jarrow, Ruppert and Yu (2004). See also Ruppert (1997) and Ruppert, Wand, Holst and Hössjer (1997).

We by no means recommend against GCV choice of smoothing parameters in general. Indeed, we prefer the GCV criterion in most cases where GCV and EBBS perform similarly. 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 $\sigma(t_i, X_{t_i})$ 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 the derivative function is of interest, or when the local smoothing parameter is preferred. As discussed in Section 4, autocorrelations of the residuals are very mild in the data we used, and the derivative function is not of interest in our estimation. Therefore, GCV is the preferred method in our case study.

2.4.3. Multiple smoothing parameter for drift estimation

Different levels of smoothness are sometimes desired for different coefficient functions. A particular nice feature of 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, 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 and 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. The parameter γ has interesting implications. For example, as discussed in Section 1, model (1.1) with the usual linear drift and $\gamma = 1$ gives the Black-Scholes model. As we will see in Section 3, 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.

3. Local Log-Linear Diffusion Estimation

We also consider local log-linear volatility estimation for the time-inhomogeneous diffusion model (2.1) based on a kernel-weighted likelihood method. Some other kernel-based variance function estimation methods (e.g., Ruppert, Wand, Holst and Hössjer (1997) and Fan and Yao (1998)) could also be used 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, and 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 the local constant in Fan, Jiang, Zhang and Zhou (2003).

3.1. Maximum kernel-weighted likelihood estimation

An appropriate localized normal log-likelihood for model (2.1) is given by minus

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

where $\mu(t, X_t) = \alpha(t) + \beta(t)X_i$, $\sigma^2(t, X_t) = \sigma^2(t)X_t^{2\gamma}$, and $\alpha(t)$, $\beta(t)$ for drift and $\log\sigma^2(t)$ for 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 $\alpha(t)$ and $\beta(t)$ is used.

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

$$\hat{\alpha} = \frac{(A_0B_2 - A_1B_1)}{(B_0B_2 - B_1^2)}, \quad \hat{\beta} = \frac{(A_1B_0 - A_0B_1)}{(B_0B_2 - B_1^2)},$$

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

Let $\hat{\mu}(t, X_t)$ be the time-inhomogeneous drift estimator from Step 1. As in (2.11), denote $e_{t_i} = (1/\sqrt{\Delta_i})(y_{t_i} - \hat{\mu}(t_i, X_{t_i})\Delta_i)$. We then model $\log\sigma^2(t)$ as a local linear function. This leads to the local kernel weighted likelihood estimation equation in Step 2 volatility estimation:

$$\sum_{i=1}^{n} K_{i2}\bigg(e_{t_i}^2 \exp\Big\{-(v_0 + v_1(t_i - t))\Big\} X_{t_i}^{-2\gamma} + v_0 + v_1(t_i - t) + \gamma \log X_{t_i}^2\bigg), \quad (3.1)$$

where v_0 and v_1 are local linear parameter functions. The scale parameter γ is estimated via global minimization of

$$\sum_{i=1}^{n} \Big(e_{t_i}^2 \exp(-\hat{v}_0(t_i)) X_{t_i}^{-2\gamma} + \gamma \log X_{t_i}^2\Big). \quad (3.2)$$

Once we have estimates for v_0 and γ, denoted by $\hat{v}_0$ and $\hat{\gamma}$, respectively, we can estimate the volatility by

$$\hat{\sigma}(t, X_t) = \exp\Big(\frac{\hat{v}_0(t)}{2}\Big) X_t^{\hat{\gamma}}.$$

Note that given γ, (3.1) 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 (3.1) with respect to v_0 and v_1 to find

$$\exp(v_0) = \frac{\sum_{i=1}^n K_{i2} e_{t_i}^2 \exp(-v_1(t_i - t)) X_{t_i}^{-2\gamma}}{\sum_{i=1}^n K_{i2}}, \tag{3.3}$$

$$\exp(v_0) = \frac{\sum_{i=1}^n K_{i2} e_{t_i}^2 (t_i - t) \exp(-v_1(t_i - t)) X_{t_i}^{-2\gamma}}{\sum_{i=1}^n (t_i - t) K_{i2}}. \tag{3.4}$$

Equating (3.3) and (3.4) provides a single equation to solve for v_1. Once we obtain v_1, we can get via equation (3.3) or (3.4). Alternatively, an iterative algorithm via equations (3.3) and (3.4) 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, Sheather and Wand (1995)).

For volatility estimation, we suggest a simple rule of thumb (ROT) bandwidth selection h_2 similar to that 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 $h_2 = \{2R(K)V_1/a_2^2(K)Bn\}^{1/5}$, where

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

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

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

with the latter being obtained numerically. $\hat{c}_i$ $(i = 1, 2, 3)$ is 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 , volatility $V(t) = \sigma^2(t, X_t)$, 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) the 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 be the density function of time, usually a uniform distribution on time interval $[a, b]$.

Theorem 2. *Under the foregoing regularity conditions, as* $n \to \infty$*, the estimator* $\hat{V}(t)$ *from* (3.1) *satisfies*

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

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

The proof is a combination of Taylor series expansion of a normalized function of (3.1) and the Cramer-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 (3.1) via $\hat{\sigma}(t, X_t)$, we derive the asymptotic property of $\hat{\sigma}(t, X_t) - \sigma(t, X_t)$ by using the Taylor expansion $\sqrt{\hat{V}(t)} - \sqrt{V(t)} \sim [1/2\sigma(t, X_t)](\hat{V}(t) - V(t))$.

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

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

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

By applying the likelihood estimation property to the log-likelihood equation (3.2) in parameter γ, we have another theorem.

Theorem 4. *When* (3.2) *is a second continuous differentiable function on* $(0, \infty)$ *in* γ *and* $n \to \infty$*, the estimator* $\hat{\gamma}$ *from* (4.1) *is consistent and satisfies* $\sqrt{n}I(\gamma)^{1/2}(\hat{\gamma} - \gamma) \to_D N(0, 1)$*, where*

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

Remark. From the asymptotic analysis, the optimal bandwidth is of the usual $O(n^{-1/5})$ size, and the optimal mean integrated squared errors are of the order $O(n^{-4/5})$. An analogous theorem near the boundary can be easily obtained,

Table 1. Descriptive statistics of interest rate yields and their changes.

		Mean		
Variables	Sample Size	1954-1978	1979-1982	1983-2004
Yield	2,638	4.250%	11.524%	5.243%
Change	2,637	0.006%	-0.006%	-0.006%

		Mean		
Variables	Sample Size	1954-1978	1979-1982	1983-2004
Yield	2,638	1.892%	2.559%	2.269%
Change	2,637	0.157%	0.557%	0.111%

which verifies the theoretical advantage of the local linear (and local log-linear) approach over the local constant (kernel) method at the boundary.

4. Treasury Bill Case Study

4.1. Data and preliminary results

We compare log P-splines, local log-linear, and FJZZ in a case study using weekly 3-month Treasury bill secondary market rates (weekly averages of business days) obtained from the Federal Reserve Bank of St. Louis. The secondary market rates are annualized using a 360-day year or bank interest. A Treasury bill is a financial contract issued by U.S. government with value (price) $P_t(T)$ at time t that yields a known amount on a future date, the maturity date T. Thus, 3-month Treasury bills mature on date ($T = t + 3$ months). $P_t(T)$ is determined by the rate evolution, and they have an inverse relationship (see (4.2) in Section 4.3). The market rates usually change every day. The data set contains 2,638 observations from January 8, 1954 to July 23, 2004. The interest rate yields and their changes are plotted in Figures 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 large. These are confirmed by the descriptive statistics (artificially divided into three periods) in Table 1.

The table 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.

Estimation results show that both the log P-splines and the local log-linear approaches, as well as the FJZZ kernel method, catch the major trend of volatility well. Volatility is highest during early 1980s (Figure 1b). This was in agreement with the economic situation then. During that period, the Federal Reserve chairman Paul Volcker sharply increased the interest rate to combat the inflation crisis in the 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. Similar to what we have observed in Figure 1c for the dip in the mid-1960s, the log P-splines fit seems to catch the relative low variation period in the mid-1990s better, whereas the local log-linear and FJZZ methods are more prone to be dominated by the original series. We next report the estimation results from the three methods in detail.

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 1 and around 10 equally spaced quantile knots in the power basis for $\alpha(t_1)$, $\beta(t_i)$, 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. The autocorrelation in the residuals is found to be 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 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 it is in log P-splines. In the local log-linear method, as described in Section 3.1, γ is estimated by minimizing equation (3.2). 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 an increase of volatility. We further explore this issue with a small simulation study, focusing on a comparison of log P-splines and FJZZ.

The drift in the semiparametric inhomogeneous diffusion model (2.1) is set to 0 and the (true) inhomogeneous volatility $\sigma(t) = \exp(t)/2$ follows the non-linear trend in Figure 2(c). One thousand simulations of sample size 2,000 were generated and estimated. Fixed sampling interval and finite sample period were used in simulation. 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 $\sigma(t)$ while the estimate from FJZZ is very different. However,

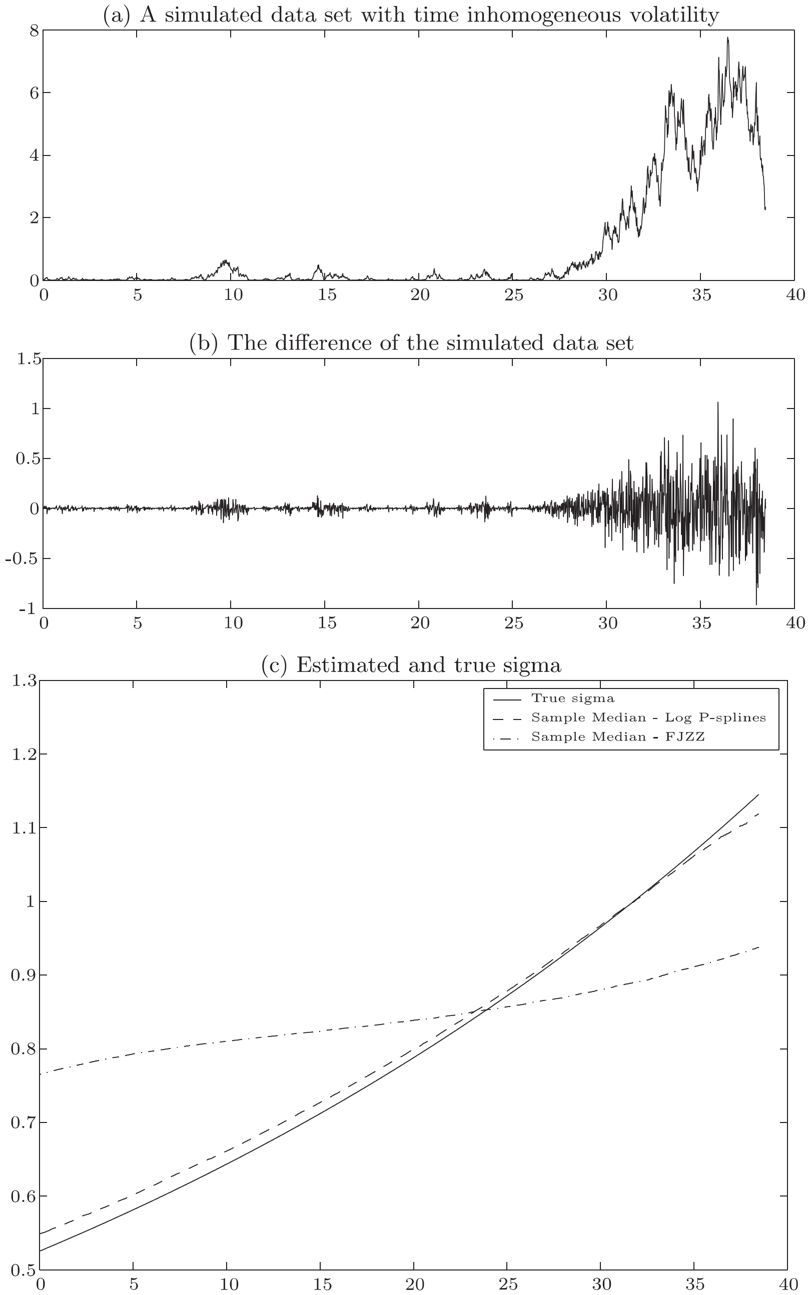

Figure 2. Simulated data and the estimated Sigma.

the volatility estimates $\hat{\sigma}(t)X_t^{\hat{\gamma}}$ from both log P-splines and FJZZ are close to the true volatility (the plot is not shown here but in a longer version of the paper: `http://statqa.cba.uc.edu/~yuy/YYWL2.pdf`). It appears that X_t^{γ} plays

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	2.96E-02

a significant role in the FJZZ estimate. Table 2 reports the median of both MSE (Mean Squared Error) and MAD (Mean Absolute Deviation) from the two methods. Table 2 (and boxplots of MSE and MAD from 1,000 simulations reported in the longer version of this paper) clearly indicates that the log P-splines method gives smaller MSE and MAD for $\sigma(t)$, γ, and the volatility.

We now go back to the case study. To assess the accuracy of the estimators, confidence intervals can be constructed. Asymptotic theorems in Sections 2 and 3 can be applied to construct the confidence bands and perform hypothesis testing. Figures 3 and 4 display the confidence bands for both the drift and volatility estimate using the asymptotic results from Section 2.2. The confidence band from the local linear method (see Section 3.3) is very similar and thus not plotted here. It appears that volatility is highest in the early 1980s when the economy experienced high inflation and interest rates, as described in Section 4.1. During that period, the confidence band is widest and the volatility is the most inhomogeneous. The confidence bands for $\alpha(t_i)$ and $\beta(t_i)$ are shown in Figure 5. The estimation results of $\beta(t_1)$ clearly show mean reversion of the drift estimate.

We next apply the Wald test of Section 2.2 to test whether some of the sub-models, such as Black-Scholes or CIR, are valid. When all coefficients for $\alpha(t_i)$, $\beta(t_i)$, and $\sigma(t_i)$, except the intercepts for $\beta(t_i)$ and $\sigma(t_i)$, are 0, $\gamma = 1$ gives the familiar geometric Brownian motion (GBM) process of Black and Scholes (1973). When all coefficients for $\alpha(t_i)$, $\beta(t_i)$, and $\sigma(t_i)$, except the intercepts are 0, $\gamma = 0.5$ gives the CIR term structure model (Cox, Ingersoll and Ross (1985)). When these restrictions are tested using the Wald test of Section 2.2., all p-values are less than 0.0001. Thus both the Black-Scholes and CIR models are rejected.

4.3. Estimating value at risk

One application of the diffusion estimates is to calculate extreme quantiles, called Value at Risk (VaR) in finance. VaR is the maximum loss or risk (credit

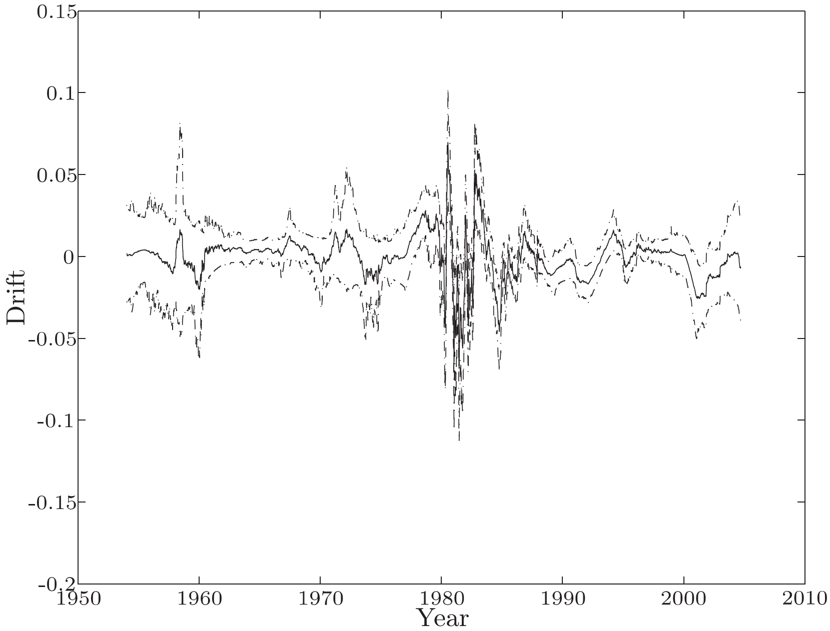

Figure 3. Time-inhomogeneous log P-spline drift estimates with confidence bands constructed using the asymptotic results in Section 2.2.

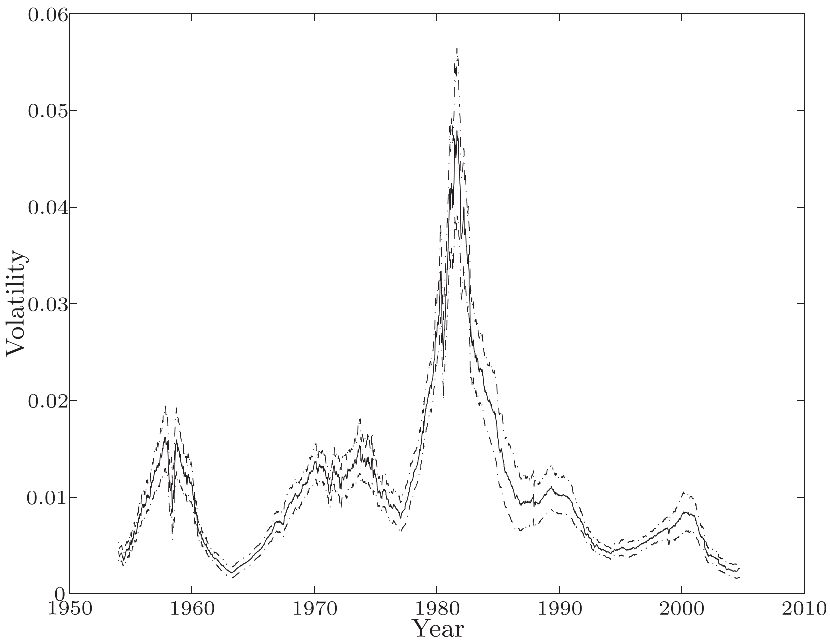

Figure 4. Time-inhomogeneous log P-spline volatility estimates with confidence bands constructed using the asymptotic results in Section 2.2.

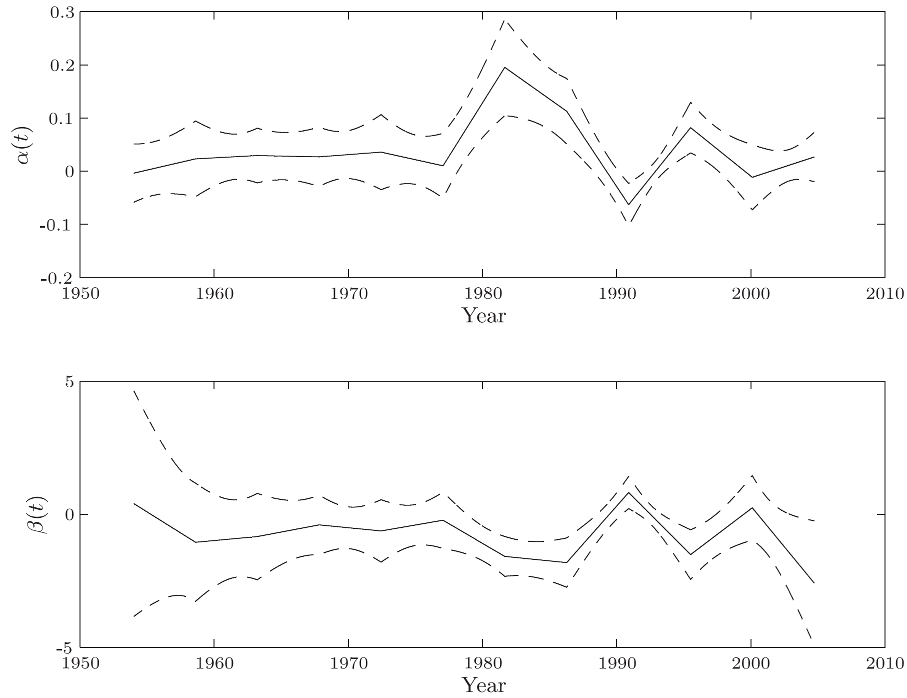

Figure 5. Time-inhomogeneous log P-splines drift component estimates $\alpha(t)$ and $\beta(t)$ with confidence bands constructed using the asymptotic results in Section 2.2.

risk, liquidity risk, market risk, etc.) during a fixed time period given a fixed probability a financial institution is expected to incur after an extreme event or a shock to the financial system (see Jorion (2000) for a detailed description of VaR). It is of great value to a financial institution or a regulatory committee, as it can be used to ensure the financial institution will still be solvent after an extreme event. More formally, if $\Delta P(\Delta t)$ is the change in value (price) of the assets from time t to $t + \Delta t$, VaR is defined as following:

$$p = \text{Prob}(\Delta P(\Delta t) \leq \text{VaR}). \tag{4.1}$$

It says that with probability $(1-p)$, the maximum loss over the next time period Δt is VaR (see Tsay (2002, Chap. 7)).

Suppose we are interested in the VaR for the weekly three-month Treasury bill in two weeks. The price of these Treasury bills is determined from the interest rates which can be estimated using the semiparametric diffusion model (2.1). The price of a zero-coupon bond that pays \$1 at maturity ($t = T$) is

$$P_t(T) = E_t\Big[\exp\Big(-\int_t^T \overline{r}_v dv\Big)\Big], \tag{4.2}$$

Table 3. Treasury bill price and Value at Risk for a 2-week horizon.

Inerest rate	0.0132	0.05	0.1
VaR	0.04832	0.15792	0.30591
Price	99.95	99.85	99.70

where $d\bar{r}_v = [\alpha(v) + \beta(v)\bar{r}_v]dv + \sigma(v)\bar{r}_v^{\gamma} dW_v$ (the market price of interest rate risk is assumed to be 0 for simplicity). A zero-coupon bond has no cash dividend (coupon) during its life. To compute the price of the bond, $\alpha(v)$, $\beta(v)$, $\sigma(v)$, and γ can be replaced with their estimates at time t, and the expectation in (4.1) can be computed using Monte Carlo simulation as in FJZZ.

We use three different current interest rate levels to estimate the price and VaR: low (0.0132), median (0.05), and high (0.1). The data set ends in 2004 and the interest rate yields during that year are unusually low (0.0132 is the last observation in this data set). The low interest rates were due to relentless rate cuts by the Federal Reserve after the dot com crash and the September 11, 2001 terrorist attacks. We performed 10,000 simulation runs and report the price of Treasury bills which pay \$100 at maturity and the 95% (confidence) level VaR ($p = 0.05$) for a 2-week horizon in Table 3, where $\alpha(v)$, $\beta(v)$, $\sigma(v)$, and γ are estimated using log P-splines. The interest rates (with drift and volatility), price of the bonds, and VaR are then obtained from these estimates.

The price in Table 3 is the average price of a Treasury bill that pays \$100 at maturity from 10,000 simulation runs. The 95% (confidence) level VaR is the difference between the maturity price (\$100) and the 5% quantile of the price from 10,000 simulation runs. Thus, with probability 0.95, the maximum loss over the next two weeks is 4.8 cents on \$100 at the current interest level of 0.0132, while the maximum loss increases to 30.59 cents on \$100 if the current interest level is 0.1. Financial institutions could have a significant amount of bond holdings, such as Treasury bills, and this maximum loss (VaR) could be substantial, in the millions or billions of dollars.

4.4. Discussion

From the case study of the weekly three-month Treasury bill data and some limited simulations study, we find that the proposed log P-splines 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. Log P-splines are also computationally efficient, and thus are recommended in practice. Both approaches guarantee that the volatility is positive, an important appealing feature in practice. Inference is also readily available via either the asymptotic theorems presented or regression bootstrap.

We also need to point out that, when data is sampled at very high frequency, extra attention needs to be taken to incorporate issues such as high serial correlation and restrictions on the underlying diffusion process. These may be considered in future research.

References

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

Ait-Sahalia, Y., Fan, J. and Peng, H. (2006). Nonparametric transition-based tests for diffusions. Working paper.

Bandi, F. M. and Phillips, P. C. B. (2003), Fully nonparametric estimation of scalar diffusion models. *Econometrica* **71**, 241-283.

Black, F. and Scholes, M. (1973). The pricing of options and corporate liabilities. *J. Polit. 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., Stefanski, L. A. and Crainiceanu, C. M. (2006). *Measurement Error in Nonlinear Models*. 2nd edition. Chapman and Hall, London.

Carroll, R. J., Wu, C. J. F. and Ruppert, D. (1988). The effect of estimating weights in weighted least squares. *J. Amer. Statist. Assoc.* **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. *J. Finance* **47**, 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 edition. Princeton University Press, Princeton, New Jersey.

Eilers, P. H. C. and Marx, B. D. (1996). Flexible smoothing with B-splines and penalties (with discussion). *Statist. Sci.* **11**, 89-121.

Egorov, A. V., Li, H. and Xu, Y. (2003). Maximum likelihood estimation of time-inhomogeneous diffusions. *J. 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. *Statist. Sinica* **13**, 965-992.

Fan, J. and Li, R. Z. (2001). Variable selection via penalized likelihood. *J. Amer. Statist. Assoc.* **96**, 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. *J. Amer. Statist. Assoc.* **98**, 118-134.

Friedman, J. H. (1991). Multivariate adaptive regression splines (with discussion). *Ann. Statist.* **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 mew methodology for contingent claims valuation. *Econometrica* **60**, 77-105.

Hong, Y. and Li, H. (2005). Nonparametric specification testing for continuous-time models with applications to term structure of interest. *Rev. Finan. Stud.* **18**, 37-84.

Hull, J. and White, A. (1990). Pricing interest-rate derivative securities. *Rev. Finan. Stud.* **3**, 573-592.

Jarrow, R., Ruppert, D. and Yu, Y. (2004). Estimating the term structure of corporate debt with a semiparametric penalized spline model. *J. Amer. Statist. Assoc.* **99**, 57-66.

Jiang, G. J. (1998). Nonparametric modeling of U.S. interest rate term structure dynamics and implications on the prices of derivative securities. *J. Finan. Quant. Anal.* **33**, 465-497.

Jiang, G. J. and Knight, J. L. (1997). A nonparametric approach to the estimation of diffusion processes, with an application to a short-term interest rate model. *Econom. Theory* **13**, 615-45.

Jorion, P. (2000). *Value at Risk: The New Benchmark for Controlling Market Risk*. McGraw-Hill, Chicago.

Lehmann, E. L. and Casella, G. (1998). *Theory of Point Estimation*. 2nd edition. Springer-Verlag, New York.

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*. 2nd edition. Chapman & Hall/CRC, Boca Raton.

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

Rubinstein, M. (1994). Implied binomial trees. *J. Finance* **49**, 771-818.

Ruppert, D. (1997). Empirical-bias bandwidths for local polynomial nonparametric regression and density estimation. *J. Amer. Statist. Assoc.* **92**, 1049-62.

Ruppert, D. and Carroll, R. J. (2000). Spatially-adaptive penalties for spline fitting. *Austral. N. Z. J. Statist.* **42**, 205-253.

Ruppert, D., Sheather, S. J. and Wand, M. P. (1995). An effective bandwidth selector for local least square regression. *J. Amer. Statist. Assoc.* **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, New York.

Stanton, R. (1997). A nonparametric model of term structure dynamics and the market price of interest rate risk. *J. Finance* **52**, 1973-2002.

Tsay, L. (2002). *Analysis of Financial Time Series*. Wiley, New York.

Vasicek, O. A. (1977). An equilibrium characterization of the term structure. *J. Finan. Econom.* **5**, 177-188.

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

Yu, K. and Jones, M. C. (2004). Likelihood-based local linear estimation of the conditional variance function. *J. Amer. Statist. Assoc.* **99**, 139-144.

Yu, Y. and Ruppert, R. (2002). Penalized spline estimation for partially linear single-index models. *J. Amer. Statist. Assoc.* **97**, 1042-1054.

Department of Quantitative Analysis and Operations Management, College of Business, University of Cincinnati, PO BOX 210130, Cincinnati, OH, 45221, U.S.A.

E-mail: yan.yu@uc.edu

Department of Mathematical Sciences, Brunel University, Uxbridge, UB8 3PH, U.K.

E-mail: Keming.Yu@brunel.ac.uk

Yahoo! Inc., 3333 Empire Avenue, Burbank, CA 91504, U.S.A.

E-mail: wangde@yahoo-inc.com

College of Business Administration, California State University, Sacramento, 6000 J Street, Sacramento, CA 95819-6088, U.S.A.

E-mail: limin@csus.edu

(Received September 2006; accepted November 2007)