Statistica Sinica **13**(2003), 965-992

TIME-DEPENDENT DIFFUSION MODELS FOR TERM STRUCTURE DYNAMICS

Jianqing Fan, Jiancheng Jiang, Chunming Zhang and Zhenwei Zhou

Chinese University of Hong Kong, Peking University,
University of Wisconsin Madison and University of California Los Angeles

Abstract: In an effort to capture the time variation on the instantaneous return and volatility functions, a family of time-dependent diffusion processes is introduced to model the term structure dynamics. This allows one to examine how the instantaneous return and price volatility change over time and price level. Nonparametric techniques, based on kernel regression, are used to estimate the time-varying coefficient functions in the drift and diffusion. The newly proposed semiparametric model includes most of the well-known short-term interest rate models, such as those proposed by Cox, Ingersoll and Ross (1985) and Chan, Karolyi, Longstaff and Sanders (1992). It can be used to test the goodness-of-fit of these famous time-homogeneous short rate models. The newly proposed method complements the time-homogeneous nonparametric estimation techniques of Stanton (1997) and Fan and Yao (1998), and is shown through simulations to truly capture the heteroscedasticity and time-inhomogeneous structure in volatility. A family of new statistics is introduced to test whether the time-homogeneous models adequately fit interest rates for certain periods of the economy. We illustrate the new methods by using weekly three-month treasury bill data.

Key words and phrases: Diffusion model, kernel regression, nonparametric goodness-of-fit.

1. Introduction

The theory of pricing contingent claims is one of the most celebrated mathematical results in finance. It offers valuable practical guidance for asset valuation and risk managements. An excellent introductory treatment of this is in Hull (1997), and more rigorous accounts can be found in Merton (1992), Duffie (1996), among others. The short-term riskless interest rates are fundamental and important in financial markets. They are directly related to consumer spending, corporate earnings, asset pricing, inflation and the overall economy. See Mishkin (1997) for further discussions. Many useful short-rate models have been proposed to explain term-structure dynamics and other issues in finance. See for example Merton (1973), Vasicek (1977), Dothan (1978), Brennan and Schwartz (1979, 1980), Cox, Ingersoll and Ross (1980, 1985), Constantinides and Ingersoll

(1984), Schaefer and Schwartz (1984), Feldman (1989), Longstaff (1989), Hull and White (1990), Black and Karasinski (1991), Longstaff and Schwartz (1992), Chan, Karolyi, Longstaff and Sanders (1992), Hansen and Scheinkman (1995), Andersen and Lund (1996), Gallant and Tauchen (1997, 1998), Gallant, Rossi and Tauchen (1997), Aït-Sahalia (1996a, 1996b), Stutzer (1996), Stanton (1997), Aït-Sahalia and Lo (1998), among others. These models provide useful insights into the term structure dynamics.

Modern asset pricing theory allows one to value and hedge contingent claims, once a model for the dynamics of an underlying state variable is given. Many such models have been developed, such as the geometric Brownian motion (Black and Scholes (1973)) and the interest-rate models mentioned in the last paragraph. Most of these are simple and convenient time-homogeneous parametric models, attempting to capture certain salient features of observed dynamic movements. However, they are not fully derived from any economic theory and cannot be expected to fit all financial data well. Thus, while the pricing theory gives us spectacularly beautiful formulas when an underlying dynamic is correctly modeled, it offers little guidance in choosing a correct model or validating a specific parametric model. Hence there is a possibility that misspecification of a model leads to erroneous valuation and hedging. This motivates us to consider a large class of nonparametric and semiparametric models. An advantage of such models is that they reduce possible modeling biases and can be used to build and validate a parametric model. This allows us to better explore the explanatory power of parametric approaches by means of nonparametric validation methods.

Economic conditions change from time to time. Thus, it is reasonable to expect that the instantaneous expected return and volatility depend on both time and price level for a given state variable, such as stock prices and bond yields. It is difficult, however, to precisely describe how the bivariate functions for the expected return and volatility vary over time and price level. Restrictive functional forms of expected return and volatility can create large biases for different assets within certain period. The most flexible model is not to assume any specific forms of the bivariate functions, but to let the data themselves determine appropriate forms that describe the dynamics. Such a data-analytic approach is called nonparametric regression in the statistical literature. For an overview of nonparametric methods, see recent books by Hastie and Tibshirani (1990), Härdle (1990), Scott (1992), Green and Silverman (1995), Simonoff (1995) and Fan and Gijbels (1996). However, as we explain in Section 2, there is not sufficient information to determine nonparametrically the bivariate functions. Hence, some form of the instantaneous return and volatility functions should be imposed.

In an attempt to capture time variation on the instantaneous return and price volatility, we expand the interest rate model of Chan et al. (1992) in two important aspects: we allow coefficients to change smoothly over time and we permit a transform of the state variable to enter into the equation. This semiparametric model enables one to simultaneously capture the time effect and reduce modeling bias. These functions in the semiparametric model can be estimated with reasonable accuracy, because of the availability of data, such as the yields of treasury bills and stock price indices over a long period. In other words, by using semiparametric and nonparametric models, we reduce model bias without excessively inflating the variance of the estimated functions.

The nonparametric techniques that we employ here are based on local constant fit (which is for simplicity, one can also use the popular local linear fit or local polynomial fit) with left-sided kernels. While the local linear fit has some theoretical advantages (Fan and Gijbels (1996)), our experience shows that it can create some artificial, statistically insignificant time trend. Compared with the traditional two-sided kernel methods, the one-sided kernel allows one to estimate a function, at any point in its support, using only historical observations. This modification makes prediction much easier.

As in all nonparametric approaches, our techniques require selection of the bandwidth. Popular methods include cross-validation (Stone (1974)), generalized cross-validation (Wahba (1977)), the pre-asymptotic substitution method (Fan and Gijbels (1995)) and the plug-in method (Ruppert, Sheather and Wand (1995)). Our bandwidth selection method is to minimize overall prediction errors, thanks to the one-sided kernel methods which facilitate the prediction.

Our time-dependent semiparametric model contains most of the well-known parametric models for interest rate dynamics. This allows us to test whether a particular parametric model fits a given dataset, regarding the semiparametric model as an alternative. Our testing procedure is based on a generalized pseudo-likelihood ratio test. It is shown in Fan, Zhang and Zhang (2001) that this test possesses a number of good statistical properties. In our current applications, a bootstrap method is used to estimate the null distribution of the test statistic. We apply the techniques to test various parametric models. Similar to the conclusions of Chan et al. (1992) and Gallant, Long and Tauchen (1997), all these parametric models have very small P-values and strong evidence for lack of fit.

2. Method of Estimation

In valuing contingent claims, it is frequently assumed that an underlying state variable, X_t, satisfies a time-dependent continuous-time stochastic differential equation:

$$dX_t = \mu(t, X_t)\,dt + \sigma(t, X_t)\,dW_t. \tag{1}$$

Here W_t denotes the standard Brownian motion and the bivariate functions $\mu(t, X_t)$ and $\sigma(t, X_t)$ are called the drift and diffusion of the process $\{X_t\}$, respectively (Wong (1970) and Duffie (1996)). Note that

$$\mu(t, X_t) = \lim_{\Delta \to 0} \frac{1}{\Delta} E(X_{t+\Delta} - X_t | X_t), \text{ and } \sigma^2(t, X_t) = \lim_{\Delta \to 0} \frac{1}{\Delta} E\{(X_{t+\Delta} - X_t)^2 | X_t\}. \tag{2}$$

Examples of (1) include geometric Brownian motion (GBM) for stock prices, and the interest rate models of Merton (1970), Vasicek (VAS) (1977), Cox, Ingersoll and Ross (CIR VR) (1980), Cox, Ingersoll and Ross (CIR SR) (1985), Chan Karolyi, Longstaff and Sanders (CKLS) (1992), among others. Different models postulate different forms of μ and σ, for instance,

$$\begin{aligned}
\text{GBM:}\ & dX_t = \mu X_t\, dt + \sigma X_t\, dW_t, \\
\text{VAS:}\ & dX_t = (\alpha_0 + \alpha_1 X_t)\, dt + \sigma\, dW_t, \\
\text{CIR VR:}\ & dX_t = \sigma {X_t}^{3/2}\, dW_t, \\
\text{CIR SR:}\ & dX_t = (\alpha_0 + \alpha_1 X_t)\, dt + \sigma \sqrt{X_t}\, dW_t, \\
\text{CKLS:}\ & dX_t = (\alpha_0 + \alpha_1 X_t)\, dt + \sigma X_t^{\gamma}\, dW_t.
\end{aligned}$$

These time-homogeneous models are a specific family of the nonparametric models,

$$dX_t = \mu(X_t)\, dt + \sigma(X_t)\, dW_t, \tag{3}$$

studied by Stanton (1997), Fan and Yao (1998) and Chapman and Pearson (2000), where the functional forms of μ and σ are unspecified.

2.1. Time-dependent diffusion models

It is reasonable to expect that the instantaneous return and volatility slowly evolve with time. Time-homogeneous models, while useful, are not capable of capturing this kind of feature. In fact, it is common practice to apply parametric models to a window of time series (e.g., in 1999, one uses only data between 1995 and 1999 to estimate parameters in the model), and this window of series moves as time evolves (e.g., in 2002, one would now use the data between 1998 and 2002 to estimate parameters). The resulting estimates are time-dependent. This in essence utilizes time-dependent parametric techniques with a prescribed time window. Various efforts have been made to explicitly express the dependence of parameters on time. These include the time-dependent models of Ho and Lee (HL) (1986), Hull and White (HW) (1990), Black, Derman and Toy (BDT) (1990) and Black and Karasinski (BK) (1991). They assume, respectively, the following forms:

$$\begin{aligned}
\text{HL: } & dX_t = \mu(t)\,dt + \sigma(t)\,dW_t,\\
\text{HW: } & dX_t = \{\alpha_0(t) + \alpha_1(t)X_t\}\,dt + \sigma(t){X_t}^i\,dW_t, \quad i = 0 \text{ or } 0.5,\\
\text{BDT: } & dX_t = \{\alpha_1(t)X_t + \alpha_2(t)X_t\log(X_t)\}\,dt + \beta_0(t)X_t\,dW_t,\\
\text{BK: } & dX_t = \{\alpha_1(t)X_t + \alpha_2(t)X_t\log(X_t)\}\,dt + \beta_0(t)X_t\,dW_t,\\
& \text{where } \alpha_2(t) = \frac{d\log\{\beta_0(t)\}}{dt}.
\end{aligned}$$

These forms are specific examples of the following time-dependent model:

$$dX_t = \{\alpha_0(t) + \alpha_1(t)g(X_t)\}\,dt + \beta_0(t)h(X_t)^{\beta_1(t)}\,dW_t, \tag{4}$$

for some functions α_0, α_1, g, β_0, β_1 and h whose forms are not specified. (The BDT and BK models can be included in (4) if one uses the transformed variable $X_t^* = \log(X_t)$.) Indeed, the time-inhomogeneous nonparametric model (4) includes all of the time-homogeneous and time-inhomogeneous models mentioned above. For example, the nonparametric model of Stanton (1997) corresponds to (4) with $\alpha_0(t) = 0$, $\alpha_1(t) = 1$, $\beta_0(t) = 1$ and $\beta_1(t) = 1$. Model (4) also allows one to check whether a particular model is valid or not, via either formal statistical tests or visual comparisons between parametric and nonparametric fits. It reduces degrees of danger on model misspecification and permits one to choose a parametric model from nonparametric analyses. This provides a useful integration of parametric and nonparametric approaches.

One notable distinction between the models at (1) and (4) is the estimability of the parameters in the expected return and volatility. The widest possible one-factor model is the one with the forms of the drift and diffusion in (1) completely unspecified. However, the drift and diffusion functions are then inestimable, since only a trajectory in the time and state domains is observed. In contrast, the expected return and the volatility functions in the nonparametric model (4) are estimable.

A useful class of (4) specifies the functions g and h. An example of this is

$$dX_t = \{\alpha_0(t) + \alpha_1(t)X_t\}\,dt + \beta_0(t)X_t^{\beta_1(t)}\,dW_t. \tag{5}$$

This submodel is an extension of the CKLS model that allows the coefficients to depend arbitrarily on time. It includes all of the aforementioned parametric models, in both time-homogeneous and time-dependent settings. One can also specify other forms of g and h, and our techniques continue to apply. Thus, this paper concentrates mainly on model (5). An interesting probabilistic question is the conditions under which model (5) is arbitrage-free. Sandmann and Sondermann (1997) have studied some aspect of this issue.

A further restriction of (5) is to assume that $\beta_1(t) = \beta_1$ is time-independent. While this imposes certain restrictions, it avoids some collinearity problems in estimating $\beta_0(t)$ and $\beta_1(t)$. The parameter β_1 and the coefficient function $\beta_0(t)$ in this parsimonious model can be estimated more reliably. We discuss this issue in Section 6 after we have introduced some necessary tools.

2.2. Estimation of instantaneous return

Assume that the coefficient functions in model (5) are twice continuously differentiable and we are given the data $\{X_{t_i}, i = 1, \ldots, n+1\}$ sampled at discrete time points, $t_1 < \cdots < t_{n+1}$. In many applications, the time points are equally spaced. For example, when the time unit is a year, weekly data are sampled at $t_i = t_0 + i/52$, $i = 1, \ldots, n+1$, for a given initial time point t_0. Let $Y_{t_i} = X_{t_{i+1}} - X_{t_i}$, $Z_{t_i} = W_{t_{i+1}} - W_{t_i}$, and $\Delta_i = t_{i+1} - t_i$. According to the independent increment property of Brownian motion, the Z_{t_i} are independent and normally distributed with mean zero and variance Δ_i. Thus the discretized version of (5) can be expressed as

$$Y_{t_i} \approx \{\alpha_0(t_i) + \alpha_1(t_i) X_{t_i}\}\Delta_i + \beta_0(t_i) X_{t_i}^{\beta_1(t_i)} \sqrt{\Delta_i}\, \varepsilon_{t_i}, \quad i = 1, \ldots, n, \tag{6}$$

where $\{\varepsilon_{t_i}\}_{i=1}^n$ are independent and standard normal. As pointed out in Chan et al. (1992) and demonstrated by Stanton (1997), the discretized approximation error to the continuous-time model is of second order when the data are observed over a short time period. See also the recent work of Aït-Sahalia (1999, 2002). Indeed, according to Stanton (1997), as long as data are sampled monthly or more frequently, the errors introduced by using approximations rather than the true drift and diffusion are extremely small when compared with the likely size of estimation errors. Higher order differences, such as those elaborated by Stanton (1997), are also possible. While higher order approximations lead to lower order approximation errors, they significantly increase the variance of nonparametric estimators (Fan and Zhang (2003)). An asymptotic analysis in Fan and Zhang (2003) shows that, in the time-homogeneous nonparametric model studied by Stanton (1997), the variance inflation factors for estimating the instantaneous return (denoted by $V_1(k)$) and squared volatility (denoted by $V_2(k)$) using k-th order differences are very substantial. Details are excerpted in Table 1. This makes higher order approximations less attractive. Thus, for simplicity and for variance reduction, we opt for the first order difference.

Recall that the forms of the functions $\alpha_0(t)$ and $\alpha_1(t)$ are not specified. We can only use their qualitative features: the functions are smooth so that they can be *locally* approximated by a constant. That is, at a given time point t_0, we use the approximation $\alpha_i(t) \approx \alpha_i(t_0)$, $i = 0,\ 1$, for t in a small neighborhood of t_0. Let h denote the size of the neighborhood and K be a nonnegative weight function.

These are called a bandwidth parameter and a kernel function, respectively. Following the local regression technique (see Fan and Gijbels (1996)), we can find estimates of $\alpha_i(t_0)$, $i = 0, 1$, via the following (locally) weighted least-squares criterion: Minimize

$$\sum_{i=1}^{n} \left[\frac{Y_{t_i}}{\Delta_i} - a - bX_{t_i} \right]^2 K_h(t_i - t_0) \tag{7}$$

with respect to parameters a and b, where $K_h(\cdot) = K(\cdot/h)/h$. Note that when K has a one-sided support, such as $[-1, 0)$, for example the one-sided Epanechnikov kernel $3/4(1 - t^2)I(t < 0)$ (see Epanechnikov (1969)), the above local constant regression only uses the data observed in the time interval $[t_0 - h, t_0)$. This amounts to using only historical data and is useful for forecasting. It also facilitates data-driven bandwidth selection. Our experience shows that there is not significant difference between nonparametric fitting with one-sided and two-sided kernels.

Table 1. Variance inflation factors in using higher order differences (from Fan and Zhang (2003)).

Order k	1	2	3	4	5	6	7	8	9	10
$V_1(k)$	1.00	2.50	4.83	9.25	18.95	42.68	105.49	281.65	798.01	2364.63
$V_2(k)$	1.00	3.00	8.00	21.66	61.50	183.40	570.66	1837.28	6076.25	20527.22

Let $\hat{a}$ and $\hat{b}$ be the minimizers of the weighted least-squares regression (7). Then, the local estimators of $\alpha_0(t_0)$ and $\alpha_1(t_0)$ are $\hat{\alpha}_0(t_0) = \hat{a}$ and $\hat{\alpha}_1(t_0) = \hat{b}$. To obtain the estimated functions, $\hat{\alpha}_0(\cdot)$ and $\hat{\alpha}_1(\cdot)$, we usually evaluate the estimates at hundreds of grid points. Note that we have ignored the heteroscedasticity at (7). In principle, we can incorporate heteroscedasticity via minimizing

$$\sum_{i=1}^{n} \left[\frac{Y_{t_i}}{\Delta_i} - a - bX_{t_i} \right]^2 \hat{\beta}_0^{-2}(t_i) X_{t_i}^{-2\hat{\beta}_1(t_i)} K_h(t_i - t_0), \tag{8}$$

where $\hat{\beta}_0$ and $\hat{\beta}_1$ are obtained from the procedure described in the next section. However, we do not experience any substantial gains, due to the large stochastic noise contaminating the expected return function.

2.3. Estimation of volatility

Let $\widehat{\mu}(t, X_t) = \widehat{\alpha}_0(t) + \widehat{\alpha}_1(t)X_t$ stand for the estimated mean function and set $\widehat{E}_t = \{Y_t - \widehat{\mu}(t, X_t)\Delta_t\}/\sqrt{\Delta_t}$. Then, by (6), we have

$$\widehat{E}_t \approx \beta_0(t) X_t^{\beta_1(t)} \varepsilon_t. \tag{9}$$

Note that this approximation also holds when $\widehat{E}_t$ is replaced by $Y_t/\sqrt{\Delta_t}$, as pointed out by Stanton (1997). However, with $E_t = \{Y_t - \mu(X_t, t)\Delta_t\}/\sqrt{\Delta_t}$, the approximation error is of smaller order. The conditional log-likelihood of $\widehat{E}_t$ given X_t is, up to an additive constant, approximately expressed as

$$-\frac{1}{2}\log\{\beta_0^2(t)X_t^{2\beta_1(t)}\} - \frac{\hat{E}_t^2}{2\beta_0^2(t)X_t^{2\beta_1(t)}}.$$

Summing it up with respect to t, we obtain the logarithm of the pseudo-likelihood. By using the local constant approximation and introducing the kernel weight we obtain, at a time point t_0, the local pseudo-likelihood

$$\ell(\beta_0, \beta_1; t_0) = -\frac{1}{2}\sum_{i=1}^{n} K_h(t_i - t_0)\left(\log(\beta_0^2 X_{t_i}^{2\beta_1}) + \frac{\hat{E}_{t_i}^2}{\beta_0^2 X_{t_i}^{2\beta_1}}\right). \qquad (10)$$

Maximizing (10) with respect to the local parameters β_0 and β_1, we obtain the estimates $\hat{\beta}_0(t_0) = \hat{\beta}_0$, and $\hat{\beta}_1(t_0) = \hat{\beta}_1$. The whole functions $\beta_0(\cdot)$ and $\beta_1(\cdot)$ can be estimated by repeatedly maximizing (10) over a grid of time points. This type of local pseudo-likelihood method is related to the generalized method of moments of Hansen (1982), but is used now in a local neighborhood. See also Florens-Zmirou (1993) and Genon-Catalot and Jacod (1993).

Note that for given β_1, the maximization of $\ell(\beta_0, \beta_1; t_0)$ is obtained at

$$\hat{\beta}_0^2(t_0; \beta_1) = \sum_{i=1}^{n} K_h(t_i - t_0)\hat{E}_{t_i}^2 |X_{t_i}|^{-2\beta_1} \Big/ \sum_{i=1}^{n} K_h(t_i - t_0). \qquad (11)$$

Thus, at a point t_0, we only need to maximize the one-dimensional function $\ell(\hat{\beta}_0(t_0; \beta_1), \beta_1; t_0)$ with respect to β_1. The whole function $\beta_1(t)$ can be obtained by repeatedly optimizing this one-dimensional function at a grid of time points. Using the estimate $\hat{\beta}_1(t_j)$ as the initial value for maximizing the target function at the next grid point t_{j+1}, the maximizer can be found within only a few iterations. Thus the computational burden is not much heavier than that for estimating the drift function. The estimated function $\hat{\beta}_0(t)$ can be obtained by using (11) at each grid point.

An alternative approach is to use the local least-squares method by noting that (9) implies

$$\log(\widehat{E}_t^2) \approx \log\{\beta_0^2(t)\} + \beta_1(t)\log(X_t^2) + \log(\varepsilon_t^2). \qquad (12)$$

This is again a semi-parametric model and the method in Section 2.2 for estimating the drift function can be applied to estimate the parameters $\log\{\beta_0^2(t)\}$ and $\beta_1(t)$. We implemented this method but did not get satisfactory results. This is mainly due to the exponentiation operation used in the estimation of $\beta_0^2(t)$, which inflates (deflates) the estimate substantially.

The time-dependent model (5) is also related to a GARCH(1,1) model. To see this relation, take $\beta_1 \equiv 0$, so that (5) becomes $dX_t = \{\alpha_0(t) + \alpha_1(t)X_t\}\,dt + \beta_0(t)\,dW_t$, and (11) reduces to $\hat\beta_0^2 = \sum_{i=1}^n K_h(t_i - t_0)\hat E_{t_i}^2 / \sum_{i=1}^n K_h(t_i - t_0)$. Assume $t_i = \Delta i$ and let $r_i = \hat E_{t_i}$, $i = 1, \ldots, n$. To stress the dependency of $\hat\beta_0$ on the time point t_0, we write it as $\hat\sigma_{t_0}$. If we take $K(x) = a^x I(x < 0)$ for some parameter $a > 1$, then $K_h(x) = b^x I(x < 0)/h$ with $b = a^{1/h}$, which is greater than 1. Consequently, it follows that $\hat\sigma_t^2 = \sum_{i>0} \lambda^i r_{t-i}^2 / \sum_{i>0} \lambda^i$, where $\lambda = b^{-\Delta}$. Note that $\hat\sigma_t^2 = \lambda\hat\sigma_{t-1}^2 + (1-\lambda) r_t^2$. This is indeed the J. P. Morgan (1996) estimator for volatility. This estimator can be regarded as the one from a GARCH(1,1) model. In other words, our time-dependent model and the GARCH(1,1) model have some intrinsic connections: both of them use the volatility in the recent history.

2.4. Bandwidth selection

The bandwidth h in the kernel regression can be tuned to optimize the performance of the estimated functions. It can be subjectively tuned by users to trade off the bias and variance of the nonparametric estimates by visual inspection, or chosen by data to minimize some criteria that are related to the prediction error.

The criteria that we proposed here take advantage of the fact that the one-sided kernel is employed so that only historical data are used in the construction of estimators. The bandwidth for the expected instantaneous return can be chosen to minimize the average prediction error (APE) as a function of the bandwidth:

$$\text{APE} = m^{-1} \sum_{i=1}^m \frac{(Y_{t_i^*} - \hat Y_{t_i^*})^2}{\sigma_{t_i^*}^2},$$

with $\hat Y_{t_i^*} = \{\hat\alpha_0(t_i^*) + \hat\alpha_1(t_i^*) X_{t_i^*}\}\Delta_{t_i^*}$ and $\sigma_{t_i^*} = \hat\beta_0(t_i^*) X_{t_i^*}^{\hat\beta_1(t_i^*)}$. In the above definition, the prediction errors are computed at the prescribed time points t_i^*, $i = 1, \ldots, m$. This kind of idea has also been used by Hart (1994, 1996).

For estimation of the volatility, since the local pseudo-likelihood is used to construct nonparametric estimators, the bandwidth will be chosen to maximize the pseudo-likelihood of $\hat E_t$ given X_t. More specifically, for a given bandwidth h, the pseudo-likelihood function is defined as

$$-\frac{1}{2} \sum_{i=1}^m \left(\log\{\hat\beta_0^2(t_i^*) X_{t_i^*}^{2\hat\beta_1(t_i^*)}\} + \frac{\hat E_{t_i^*}^2}{\hat\beta_0^2(t_i^*) X_{t_i^*}^{2\hat\beta_1(t_i^*)}} \right).$$

2.5. Standard errors

Standard errors of statistical estimators are useful for assessing accuracy. The kernel regression estimators in (7) and (8) are actually weighted least-squares estimators. Thus, traditional linear regression techniques continue to apply. When the data are independent, the formulas for standard errors of the local linear estimators are given on page 115 of Fan and Gijbels (1996). For dependent data, we can use the regression bootstrap (see Franke, Kreiss and Mammen (2002)) to assess the sampling variability. The idea is to generate the bootstrap responses $\{Y_{t_i}^*\}$ from (6), using the estimated parameter functions and $\{X_{t_i}\}$, but with the new random shocks $\{\varepsilon_{t_i}^*\}$. Based on the bootstrap sample $\{(X_{t_i}, Y_{t_i}^*), i = 1, \ldots, n\}$, the coefficient functions are estimated and sampling variability can be evaluated. In our simulations, the bootstrap confidence intervals are calculated based on 1,000 bootstrap samples.

3. Applications and Simulations

In this section, we first apply our proposed techniques to the treasury bill data. After that, we verify our techniques by using two simulated data sets, which are similar to short-rate dynamics.

3.1. Treasury bill data

To understand interest rate dynamics, we use the yields of the three-month treasury bill from the secondary market rates on Fridays. The secondary market rates are annualized using a 360-day year of bank interest and are quoted on a discount basis. The data consist of 1461 weekly observations, from January 2, 1970 to December 26, 1997.

The annualized three-month yields and their weekly rate changes are presented in Figure 1. Figure 2 shows the estimated coefficient functions $\alpha_0(t)$, $\alpha_1(t)$, $\beta_0(t)$ and $\beta_1(t)$ along with the 95% pointwise confidence bands produced by the bootstrap method. Figure 3 describes how the expected instantaneous return and volatility change over time, and also displays their 95% pointwise confidence bands. The bandwidths are selected by the methods described in Section 2.4. The heteroscedasticity and time effect on volatility are evident. A careful inspection of Figure 2 suggests a time effect after 1980, but not before. This is probably due to the fact that the Federal Reserve changed its monetary policy on October 6, 1979 when its newly appointed chairman, Paul Volcker, initiated money supply targeting and abandoned interest rate targeting.

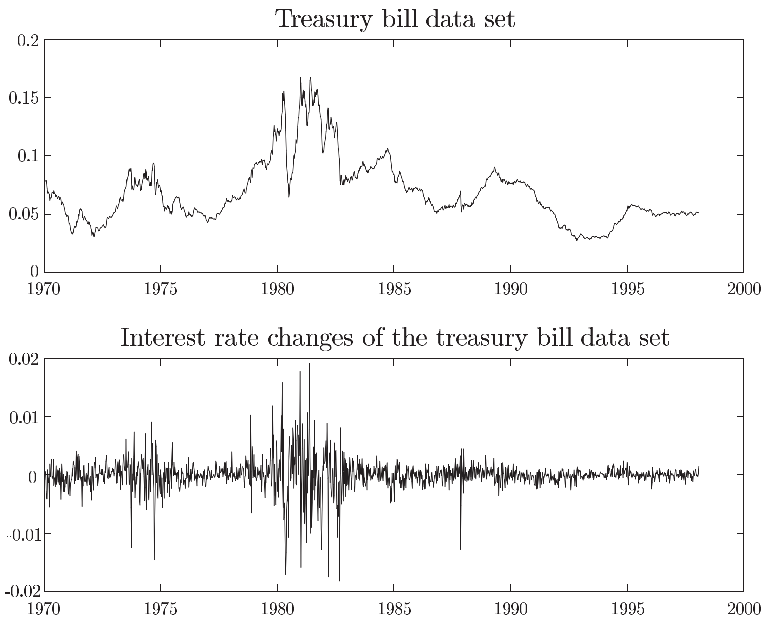


Figure 1. Three month treasury Bill rate, from January 2, 1970 to December 26, 1997. Annualized yield on three-month treasury bills, January 2, 1970 to December 26, 1997. Top panel: Weekly yields. Bottom panel: Changes of weekly yields.

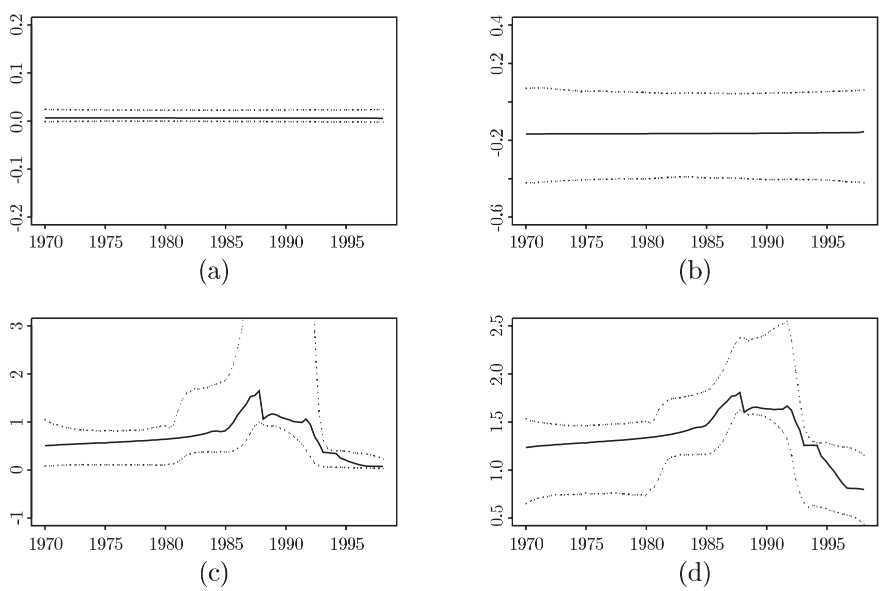
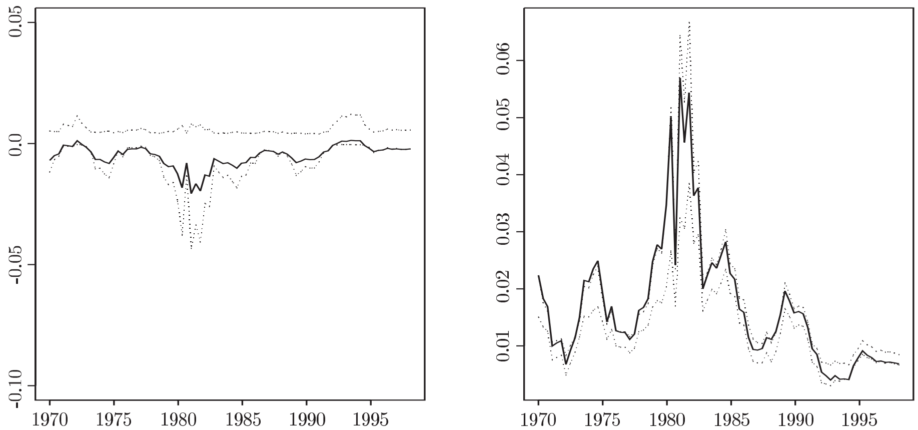


Figure 2. Three month treasury Bill rate. (a) Estimated $\alpha_0(t)$ with 95% bootstrap confidence band. (b) Estimated $\alpha_1(t)$ with 95% bootstrap confidence band. (c) Estimated $\beta_0(t)$ with 95% bootstrap confidence band. (d) Estimated $\beta_1(t)$ with 95% bootstrap confidence band. Solid – estimator, dotted – confidence band.



(a) (b)

Figure 3. Three month treasury Bill rate. (a) Estimated drift with 95% bootstrap confidence band. (b) Estimated volatility with 95% bootstrap confidence band. Solid – estimator, dotted – confidence band.

3.2. Verification of the proposed techniques

We now test our techniques by simulating two data sets from specific models of (5). The first one is the time-homogeneous model

$$dX_t = (0.0408 - 0.5921X_t)\,dt + \sqrt{1.6704}\,X_t^{1.4999}\,dW_t. \tag{13}$$

The values of the parameters are given in Chan et al. (1992), based on one-month treasury bill yields. We generate 1,735 weekly observationsn from January 5, 1962 to March 31, 1995. Based on 400 simulations, we get the estimators of $\alpha_0(t_i)$, $\alpha_1(t_i)$, $\beta_0(t_i)$ and $\beta_1(t_i)$. Figure 4 reports the pointwise 2.5th, 12.5th, 87.5th and 97.5th sample percentiles. Figure 5 gives the typical estimated drift and volatility of the rate change, where the typical estimated curves presented have median performance in terms of mean squared errors among 400 simulations. We can see that even for the time-independent CKLS model, our techniques can capture the true structure of the model without "false alarms", namely, reporting time-homogeneous models correctly to be time-homogeneous models.

Next, we test our methodology on a time-inhomogeneous model. For simplicity we consider model (5) with drift zero, and assume that the coefficient functions for volatility contain nonlinear trends as depicted in Figure 7 (see the solid lines). In order to visually display characteristics of this artificial model, we present in Figure 6 a simulated sample path consisting of 780 weekly observations from the model. Figure 7 displays the fitted coefficient functions for the data set. In 400 simulations, each containing 780 observations from the model, we compute the nonparametric estimators of the coefficient for each sample.

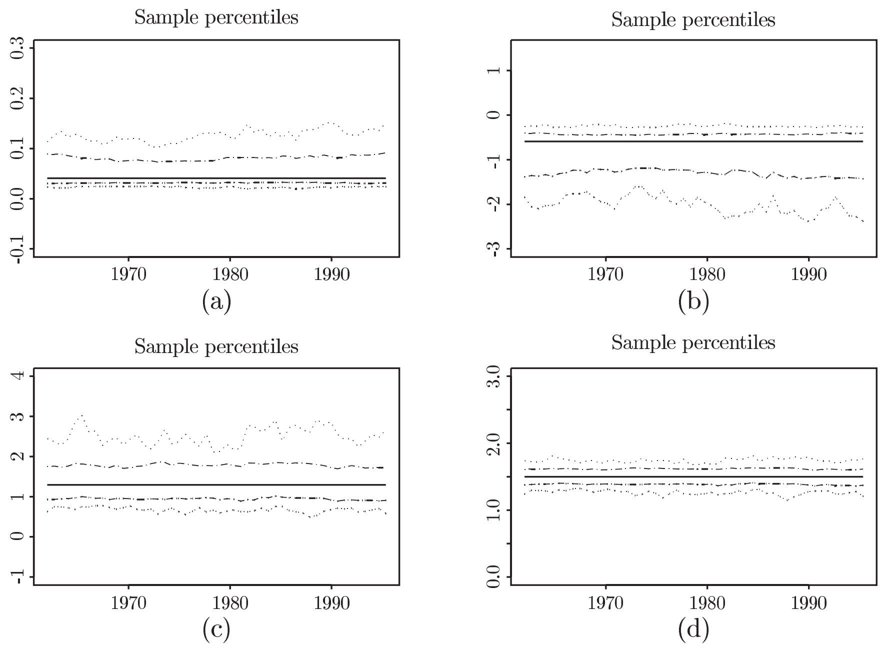


Figure 4. Envelopes formed via pointwise 2.5th, 12.5th, 87.5th and 97.5th sample percentiles among 400 simulations of model (13). Solid – true curve, dash-dotted – 75% envelopes, dotted – 95% envelopes. (a) $\alpha_0(t)$. (b) $\alpha_1(t)$. (c) $\beta_0(t)$. (d) $\beta_1(t)$.

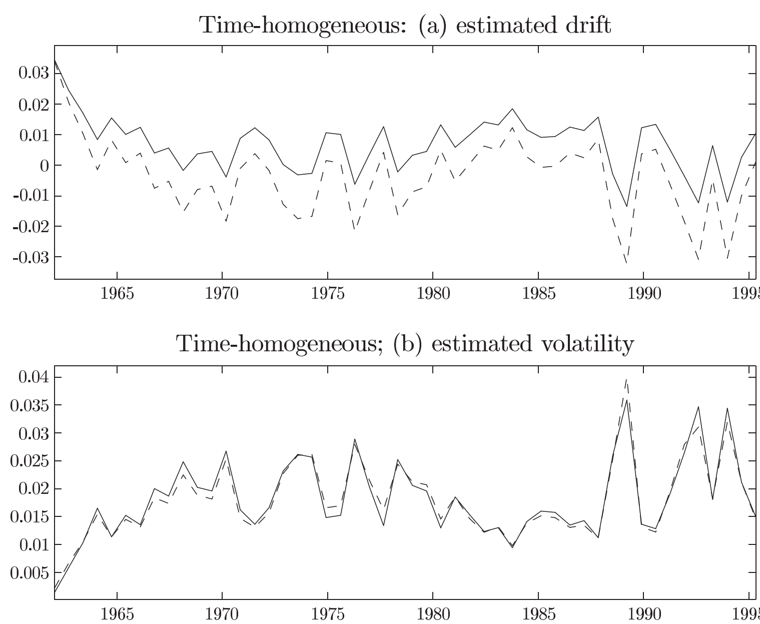


Figure 5. Typical estimated drift and volatility of the rate change among 400 simulations for model (13). Solid – true, dashed – our estimators.

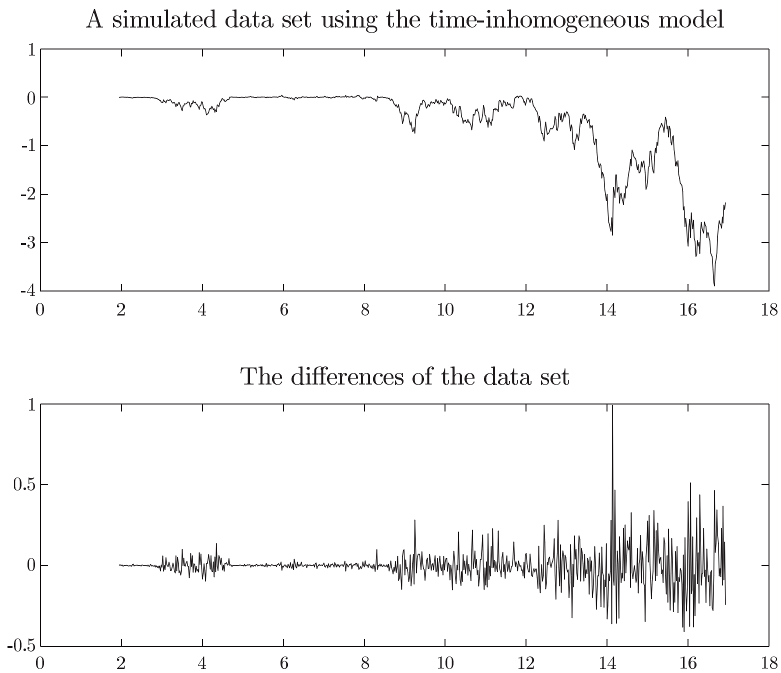


Figure 6. A simulated data set from model (5) with drift zero and time-inhomogeneous volatility with coefficient functions given in Figure 7.

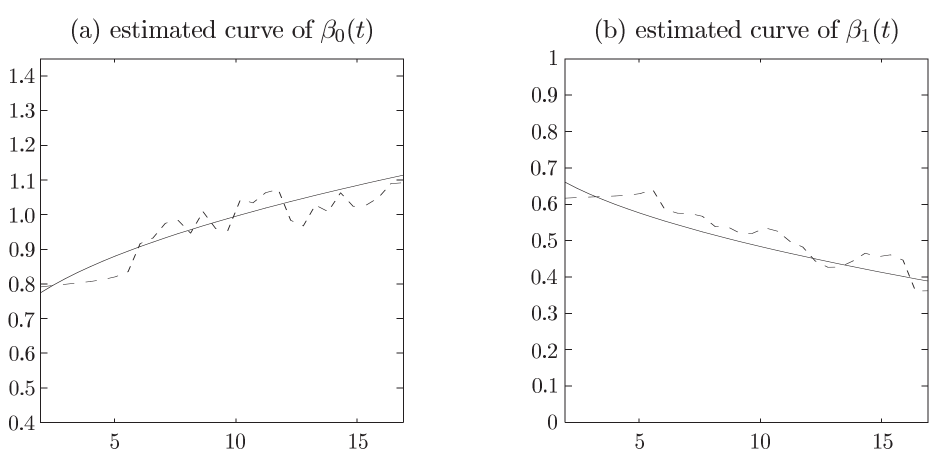


Figure 7. Estimated coefficient functions for the simulated data set in Figure 6. Solid – true curves, dashed – our estimators.

Figure 8 summarizes the simulation results by plotting the pointwise 5th, 12.5th, 87.5th and 95th percentiles among 400 simulations. Clearly our time-inhomogeneous model (5) does well at capturing the time effect in volatility.

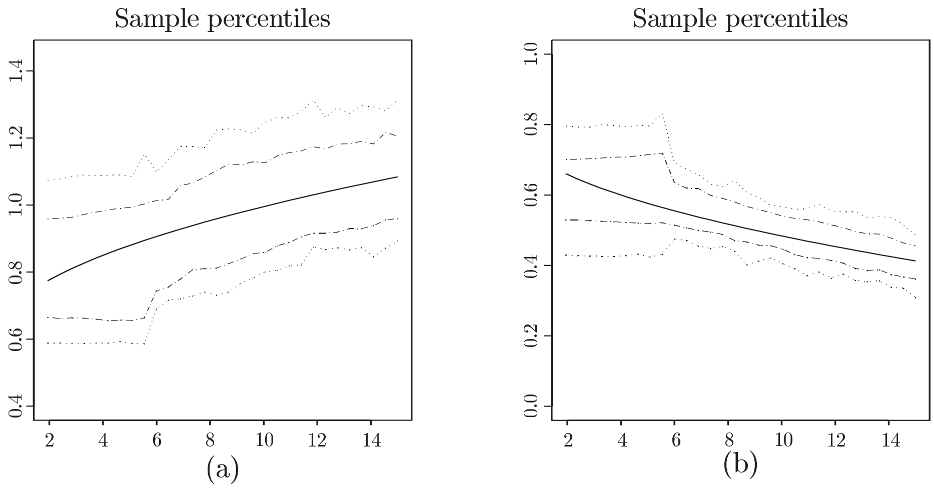


Figure 8. Envelopes for volatility formed via pointwise 5th, 12.5th, 87.5th, and 95th sample percentiles among 400 simulations for model (5). Solid – the true, dash-dotted – 75% envelopes, dotted – 90% envelopes.

3.3. Empirical comparisons

In order to gauge the relative performance among several models, we test their forecasting power for the interest rate changes considered in Section 3.1. In addition, we test their forecasting power for the squared interest rate changes and for the logarithm of the squared interest rate changes (the logarithm transforms the multiplicative model into the additive model as in (12)). This provides simple measures of how well each interest rate model captures the expected return and volatility. The predictive powers are measured by the correlation coefficient between the rate changes and their conditional expected return, and the correlation coefficient between the squared rate changes and their conditional expected volatility. Denote by ρ_1 and ρ_2 the two correlations, respectively. They are related to the coefficient of determination R^2 used in Chan et al. (1992) via the simple relation $R^2 = \rho^2$. We also denote by ρ_2^* the correlation coefficient between the logarithm of the squared interest rate changes and the logarithm of the conditional expected volatility. These correlation coefficients provide alternative measures for model comparisons.

Table 2. Correlation coefficients.

Models	Stanton	CKLS	CIR SR	CIR VR	VAS	GBM	Model (5)	Model (4)
ρ_1	0.1005	0.0479	0.0479	0.0479	0.0479	0.0479	0.0726	0.1018
ρ_2	0.3181	0.3801	0.3411	0.3840	0.0000	0.3656	0.4057	0.3229
ρ_2^*	0.3012	0.2723	0.2321	0.2723	0.0000	0.2370	0.3215	0.3213

Model (4) with both $g(\cdot)$ and $h(\cdot)$ taken as the estimate from Stanton's model.

The results are presented in Table 2. It is evident that, from measures ρ_2 and ρ_2^*, our time-dependent model (5) does a better job of capturing the short-rate volatility than its time-homogeneous counterpart, the CKLS model. In light of ρ_1, the measure of predictive power for the drift, model (4) performs the best, followed by Stanton's model and our model (5). The overall performance of our nonparametric model (5) is the best among all competing models.

4. Goodness-of-Fit Test

There are a number of stimulating interest rate models that capture different aspects of the term structure dynamics. A question arises naturally: are these models statistically different? That is, given the amount of information in the data, are they distinguishable? Here, we take the advantage of the fact that the time-dependent model (5) includes most of the popular parametric models for interest rates, therefore it can be treated as the alternative hypothesis. For example, we may wish to test whether the coefficient functions depend on time. This amounts to testing $H_0 : \alpha_0(t) = \alpha_0,\ \alpha_1(t) = \alpha_1,\ \beta_0(t) = \beta_0,\ \beta_1(t) = \beta_1$ under model (5). One can also test whether the interest rate data follows the CIR model by checking $H_0 : \alpha_0(t) = \alpha_0,\ \alpha_1(t) = \alpha_1,\ \beta_0(t) = \beta_0,\ \beta_1(t) = 0.5$ in model (5). Assessing the adequacy of the geometric Brownian motion can be formulated in a similar manner.

4.1. Generalized pseudo-likelihood ratio test

Due to large stochastic errors in the estimation of the instantaneous return functions, most reasonable models for the drift function will be accepted. For this reason, we focus on testing the functional forms of the volatility function, though the technique applies to problems of testing the instantaneous return function.

For brevity, we outline a procedure for testing the CKLS model against the time-dependent model (5). The technique applies equally to testing other forms of parametric models. Consider testing $H_0 : \beta_0(t) = \beta_0,\ \beta_1(t) = \beta_1$. Under model (5), the logarithm of the pseudo-likelihood is represented by

$$\ell(H) = -\frac{1}{2}\sum_{i=1}^{n}\left(\log\{\hat\beta_0^2(t_i)X_{t_i}^{2\hat\beta_1(t_i)}\} + \frac{\hat E_{t_i}^2}{\hat\beta_0^2(t_i)X_{t_i}^{2\hat\beta_1(t_i)}}\right),$$

where $\hat\beta_0(t_i)$ and $\hat\beta_1(t_i)$ are the kernel estimates outlined in Section 2.3. Similarly, under the hypothesis, one can estimate the parameters β_0 and β_1 by maximizing the corresponding pseudo-likelihood to obtain

$$\ell(H_0) = -\frac{1}{2}\sum_{i=1}^{n}\left(\log(\hat\beta_0^2 X_{t_i}^{2\hat\beta_1}) + \frac{\hat E_{t_i}^2}{\hat\beta_0^2 X_{t_i}^{2\hat\beta_1}}\right).$$

The plausibility of the hypotheses can then be evaluated by

$$\lambda(\mathbf{X}, h) = 2\{\ell(H) - \ell(H_0)\}, \tag{14}$$

where $\mathbf{X}$ denotes the observed data and h is the bandwidth used for the nonparametric estimates. The null hypothesis is rejected when $\lambda(\mathbf{X}, h)$ is too large.

To compute the P-value of the test statistic, we need to find the distribution of $\lambda(\mathbf{X}, h)$ under the hypothesis. The analytic form of the distribution is hard to find, but the distribution can be estimated by the parametric bootstrap (simulation) procedure. Under the hypothesis, the observed data are generated from the model

$$dX_t = \{\alpha_0(t) + \alpha_1(t)X_t\}\, dt + \beta_0 X_t^{\beta_1}\, dW_t. \tag{15}$$

Set the parameters (β_0, β_1) at their estimated values $(\hat{\beta}_0, \hat{\beta}_1)$, and set the functions $\alpha_0(t)$ and $\alpha_1(t)$ at their estimated values or even at the global least-squares estimates $\hat{\alpha}_0$ and $\hat{\alpha}_1$, because of their insignificant influence. Simulate a pseudo-sample path $\{X^*_{t_i}, i = 1, \ldots, n+1\}$ from (15), and obtain the test statistic $\lambda(\mathbf{X}^*, h)$. Repeating this procedure 1000 times (say), we obtain 1000 statistics of $\lambda(\mathbf{X}^*, h)$. The estimated P-value of the test statistic $\lambda(\mathbf{X}, h)$ is simply the percentage of the simulated statistics $\{\lambda(\mathbf{X}^*, h)\}$ exceeding the observed value of $\lambda(\mathbf{X}, h)$.

The theoretical justification of the above parametric bootstrap method is the so-called Wilks phenomenon (Fan, Zhang and Zhang (2001)). There it is shown that, in somewhat different settings, the asymptotic null distribution of the generalized likelihood ratio statistic often does not depend on, to first order, the nuisance parameters under the null hypothesis. Translating this property into our setting, the null distribution of $\lambda(\mathbf{X}, h)$ does not depend heavily on the values of $\alpha_0(t)$, $\alpha_1(t)$, β_0 or β_1. Setting them at reasonable estimates, such as those suggested above, the distribution of $\lambda(\mathbf{X}, h)$ is known and can be simulated.

The above technique applies readily to other hypothesis testing problems. For example, to test the CIR model with $\beta_1 = 1/2$, one can compute the pseudo-likelihood under the null hypothesis using the known value $\beta_1 = 1/2$. In the bootstrap estimation of the null distribution, the value $\beta_1 = 1/2$ should also be used directly.

4.2. Power simulation

The purposes of the simulation are two-fold. One is to demonstrate that our bootstrap method gives the right estimate of the null distribution, and the other is to show the power of our proposed test. For simplicity, we only consider the simulations for the volatility part. We use the models that are relevant to the term structure dynamics. The CKLS model is taken as the null hypothesis. Power is evaluated at a sequence of alternatives, ranging from the CKLS model to reasonably far away from it. Let $\alpha_i(t)$ and $\beta_i(t)$, $i = 0, 1$, be the known functions defined as in Figure 2 (see the solid lines). Let $\bar{\beta}_i$ be the corresponding estimators of the functions $\beta_i(t)$, $i = 0, 1$, under the null hypothesis, for the treasury bill data studied in Section 3.1. We evaluate the power of the pseudo-likelihood ratio test at a sequence of alternative models indexed by θ:

$$dX_t = \{\alpha_0(t) + \alpha_1(t)X_t\}\,dt + [\bar{\beta}_0 + \theta\{\beta_0(t) - \bar{\beta}_0\}]X_t^{[\bar{\beta}_1 + \theta\{\beta_1(t) - \bar{\beta}_1\}]}\,dW_t. \quad (16)$$

For each given value of θ, we simulate weekly data from model (16) with length 1461, the same as the treasury data used in Section 3.1. Based on 1000 simulations, we compute the rejection rate by using the test statistic (14). Note that when $\theta = 0$, model (16) becomes the CKLS model so that the power should be roughly 5% (or 10%) at the nominal significance level 0.05 (or 0.10), if the bootstrap estimate of the null distribution is reasonable. That is, the chance of falsely rejecting the null hypothesis is approximately 5% (or 10%). This is indeed the case as shown in Table 3. (The simulated powers are only nearly monotonic, which may result from sampling variability.) As the index θ increases, the alternative hypothesis deviates further away from the null and one would hope that the rejection rates increase. In fact, our simulation confirms that the test is very powerful. Even when $\theta = 0.40$, we already reject approximately 98% of the time (correct decision). This means that we make few mistakes of falsely accepting the null model. When $\theta = 1$, model (16) is similar to the term dynamics that we estimated for the three month treasury bill data. The power of the test against this alternative is close to 1. This in turn suggests that we have a high discriminating power for differentiating model (5) from the CKLS model.

Table 3. Simulated powers of the proposed test at significance level 5%. Similar results at level 10% are shown in brackets.

θ	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
Power	0.051 (0.101)	0.312 (0.401)	0.815 (0.851)	0.943 (0.958)	0.979 (0.980)	0.985 (0.987)	0.985 (0.986)	0.984 (0.987)	0.984 (0.987)	0.983 (0.985)	0.979 (0.983)

4.3. Testing commonly-used short-rate models

After verifying our proposed procedure, we apply the test statistic (14) to see whether the commonly-used short-rate models fit the treasury data. We tested the volatility components and now report the observed levels of significance (P-value, computed from 1000 bootstrap samples)-the smaller, the stronger evidence we have against a given parametric model. General statistics practice is to reject the null hypothesis if the P-value is less than 5%. If the P-value is below 1%, the results are interpreted as highly statistically significant, namely, we have very strong evidence against the null hypothesis.

The P-values for testing various forms of the volatility function are shown in Table 4. For Stanton's model, we test it against model (4) with $h(\cdot)$ taken as the estimate from Stanton's model. (For Stanton's model (3), let $\hat{E}_t = \{Y_t - \hat{\mu}(X_t)\}/\sqrt{\Delta_t}$. Then $\hat{E}_t \approx \sigma(X_t)\varepsilon_t$. Therefore, $E(\hat{E}_t^2|X_t) \approx \sigma^2(X_t)$. Naturally, $\sigma^2(x)$ can be estimated via local constant regression of $\hat{E}_{t_i}^2$ on X_{t_i}.)

Table 4. P-values for testing the forms of volatility function.

Form	GBM	VAS	CIR SR	CIR VR	CKLS	Stanton
λ-statistic	276.58	1201.76	613.98	205.74	211.58	40.27
P-value	0	0	0	0	0	0.145

5. Valuation of Interest-Rate Derivatives

Given the time-inhomogeneous interest rate model (5), the price $P_t(T)$ of a zero coupon bond with a payoff of \$1 at time T, given the current interest rate r_t, is of the form

$$P_t(T) = E_t\left[\exp\left(-\int_t^T \overline{r}_u du\right)\right], \tag{17}$$

where $\overline{r}_t = r_t$,

$$d\overline{r}_u = \{\alpha_0(u) + \alpha_1(u)\overline{r}_u - \lambda(r_u, u)\}du + \beta_0(u)\overline{r}_u^{\beta_1(u)}\, dW_u, \tag{18}$$

and $\lambda(r_u, u)$ is the market price of interest rate risk. The stochastic differential equation (18) involves parameter functions $\alpha_0(u)$, $\alpha_1(u)$, $\beta_0(u)$ and $\beta_1(u)$ at a future time. They are not estimable. However, they are slowly evolving with time. Thus, they can be reasonably replaced by their estimated values at time t. This leads to an approximate time-homogeneous model

$$d\overline{r}_u = \{\hat{\alpha}_0(t) + \hat{\alpha}_1(t)\overline{r}_u - \lambda(r_u, u)\}du + \hat{\beta}_0(t)\overline{r}_u^{\hat{\beta}_1(t)}\, dW_u. \tag{19}$$

The expectation in (17) can be computed via Monte Carlo simulation. Repeatedly simulate sample paths from the dynamics (19) with the initial value $\overline{r}_t = r_t$, calculate the realization of the quantity inside the expectation in equation (17) for each sample path, and then average over the values obtained for each sample path to obtain an estimate of $P_t(T)$. The standard deviation of the values obtained for each sample path can be used to monitor the accuracy of convergence.

In particular, the standard error of the sample average for M independent realizations is the standard deviation divided by $\sqrt{M}$. The market risk function $\lambda(r_u, u)$ can be estimated by using an approach similar to that of Stanton (1997). To illustrate how to use our time-dependent model to value the price of bonds, we take $\lambda(r_u, u) = 0$. Moreover, we use the assumption of zero market price of risk to illustrate the similarity and difference between the parametric and nonparametric approaches. The results for different maturities are reported in Table 5, by using 1000 simulations (the true current interest rate is 0.0512).

Table 5. Bond valuation with zero price of risk for different maturities (standard errors in parentheses).

Maturity\Interest rate	0.02	0.0512	0.08
One year	0.9769 (0.0022)	0.9503 (0.0027)	0.9260 (0.0033)
Two years	0.9493 (0.0056)	0.9031 (0.0069)	0.8627 (0.0075)
Three years	0.9177 (0.0093)	0.8585 (0.0112)	0.8072 (0.0122)

For valuation of bond price, the nonparametric time-dependent model is basically the same as the parametric time-independent model. In fact model (19) is the same as that of the CKLS model. However, an important difference is that the parameters $\hat{\alpha}_0(t)$, $\hat{\alpha}_1(t)$, $\hat{\beta}_0(t)$ and $\hat{\beta}_1(t)$ in model (19) are estimated differently. In the nonparametric approach, the window (bandwidth) over which the CKLS model should be fitted is determined automatically from historical data and a suitable weight has been introduced to reduce the contribution of historical data. Note that the bandwidth, ten years, selected by the data are reasonably large (over five years). This means that the constant approximation holds reasonably within a period of over five years. Hence, the extrapolation used in (19) for a period of up to five years is reasonable.

6. Semiparametric Time-Dependent Diffusion Models

Our previous model (5) specifies the coefficients as time-varying functions. One may question whether this can create an over-parameterization problem in some situations, and whether it is reasonable to let $\beta_1(\cdot)$ vary with the time. Nevertheless, our previous fitting techniques can be adapted to other models, for example, the semiparametric model

$$dX_t = \{\alpha_0(t) + \alpha_1(t)X_t\}\, dt + \beta_0(t)X_t^{\beta_1}\, dW_t. \tag{20}$$

This family of models is wide enough to cover the time-independent parametric models in Section 2 and the time-dependent parametric models in Section 2.1. Analogously, the discretized version of (20) can then be written as

$$Y_{t_i} \approx \{\alpha_0(t_i) + \alpha_1(t_i) X_{t_i}\}\Delta_i + \beta_0(t_i) X_{t_i}^{\beta_1} \sqrt{\Delta_i}\, \varepsilon_{t_i}, \quad i = 1, \ldots, n, \tag{21}$$

where $\{\varepsilon_{t_i}\}_{i=1}^n$ are defined as before. This model can provide more stable estimates of $\beta_0(t)$ and β_1 than their counterparts in model (5), because it avoids the local collinearity between the constant vector of ones and the vector $\{\log(X_{t_i})\}$ in a local time region.

Using (7) and (8), we get the estimators of the coefficients in the expected return. For given β_1, $\beta_0(t)$ can be estimated by the kernel estimator $\hat{\beta}_0(t;\beta_1)$ given in (11). Then β_1 can be estimated via maximizing the profile pseudo-likelihood of β_1:

$$\ell(\hat{\beta}_0(\cdot;\beta_1), \beta_1) = -\frac{1}{2}\sum_{i=1}^n \left(\log\{\hat{\beta}_0^2(t_i;\beta_1) X_{t_i}^{2\beta_1}\} + \frac{\hat{E}_{t_i}^2}{\hat{\beta}_0^2(t_i;\beta_1) X_{t_i}^{2\beta_1}} \right), \tag{22}$$

where the form of $\hat{\beta}_0^2(t_i;\beta_1)$ is similar to that in (11).

Note that all of the time-homogeneous models mentioned before are specific examples of model (20) above. Model (20) also enables one to check whether a particular time-homogeneous model is valid or not as described before. Our experience shows that the technique developed here is very helpful in some cases, in which the previous model (5) suffers from severe collinearity problems caused by over-parameterization.

For illustration, we consider the treasury bill data consisting of weekly observations, from January 8, 1954 to December 27, 1974. The total number of observations is 1095. Our analysis based on model (5) tells us that the time-inhomogeneity in volatility is significant, but the result is unreliable because we encounter severe collinearity and numerically unstable results. By using model (20), this problem disappears. Figure 9 depicts the data set. Figure 10 reports the estimates of the coefficients in model (20) along with the 95% bootstrap confidence bands using 1000 simulations, where the estimator for β_1 is 0.50. The estimates of the expected return and volatility are shown in Figure 11. Now let us check the goodness of fit of the CKLS and CIR (SR) models. Based on 1000 simulations, we obtain P-values of 0.078 for the CKLS model and 0.311 for the CIR model. (The bandwidths used for calculating P-values are the same as those for estimation.) Therefore, the generalized pseudo-likelihood ratio test reveals that the CKLS model captures reasonably the interest rate dynamics in this period, and the CIR model outperforms the CKLS model.

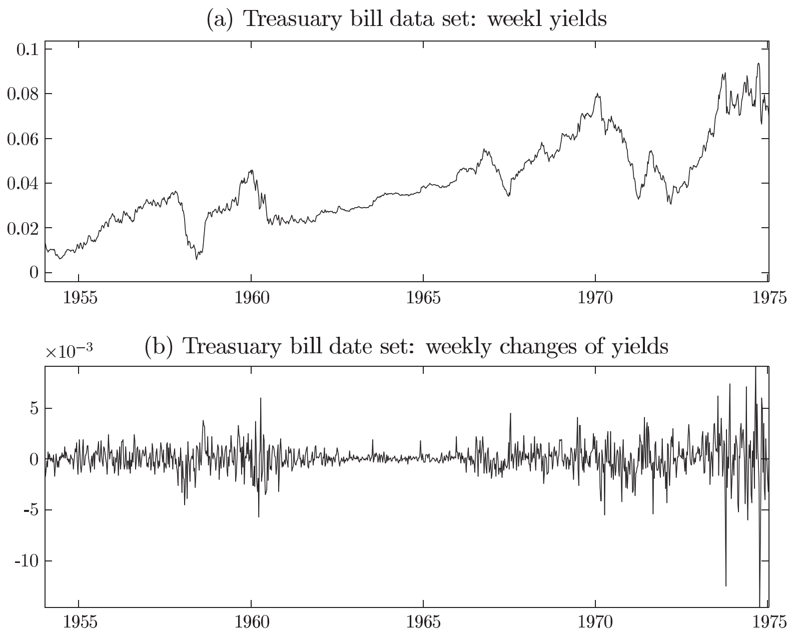


Figure 9. Treasury bill data set, from January 8, 1954 to December 27, 1974.

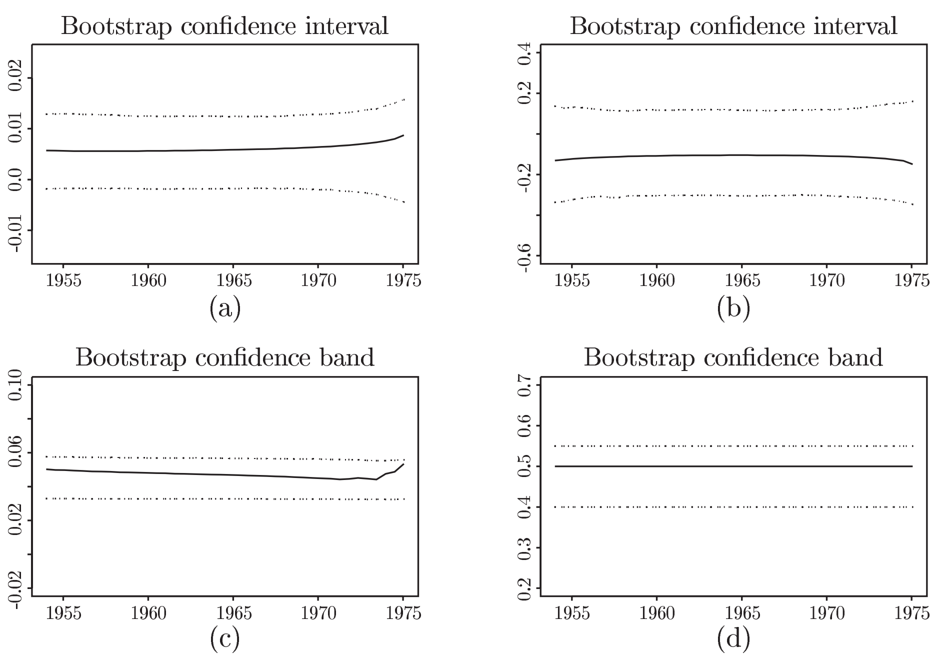


Figure 10. Estimated coefficient functions for the semiparametric model (20). Solid – our estimators, dotted – 95% confidence bands among 1000 simulations.

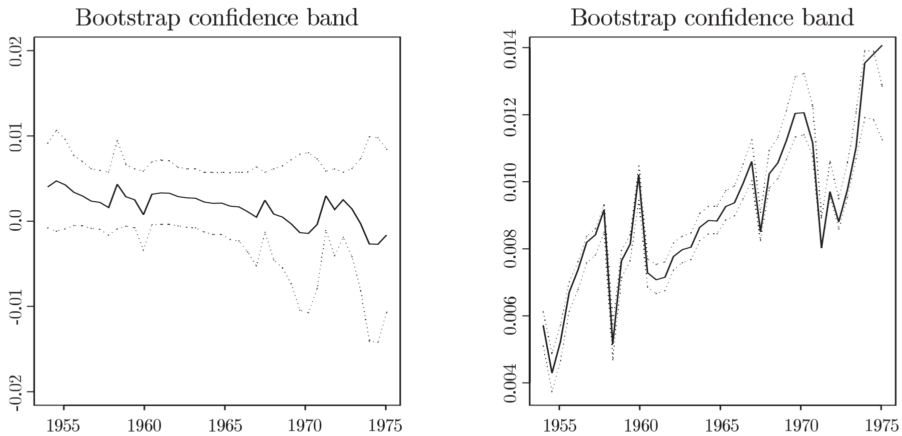


Figure 11. Estimated drift and volatility for the semiparametric model (20). Solid – our estimators, dotted – 95% confidence band among 1000 simulations.

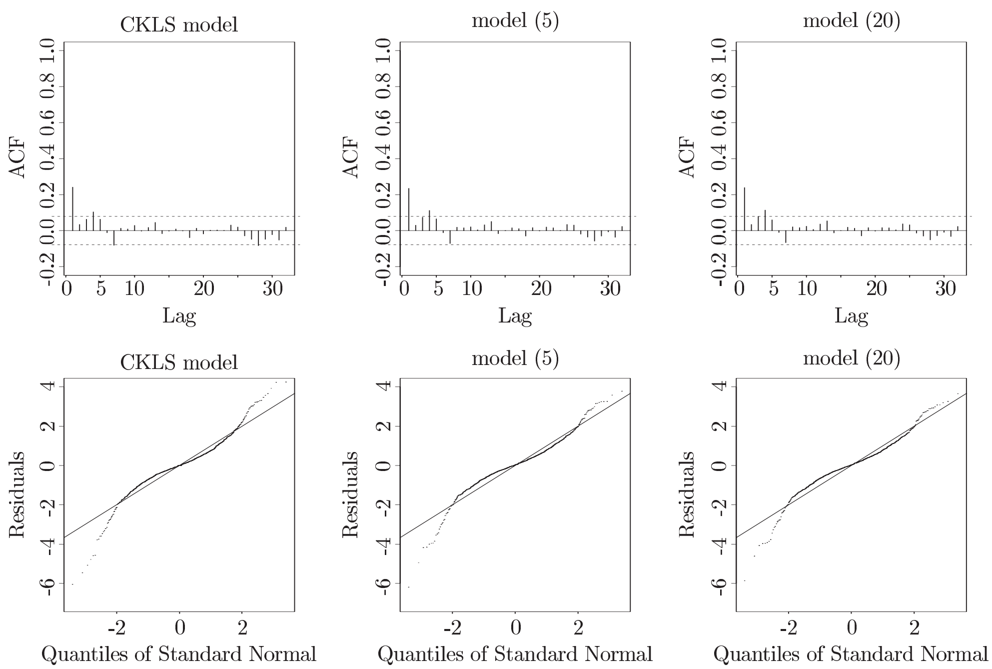


Figure 12. Diagnostic check: residual autocorrelation functions and Q-Q plots. Left panel: CKLS model; middle panel: nonparametric model (5); right panel: semiparametric model (20).

The generalized pseudo-likelihood ratio test checks one important aspect of model fitting. Certainly one can check other aspects of model fitting. In practice, one may also examine if the residuals behave like a Gaussian white noise. To compare the performances of the CKLS model, model (5) and model (20), we consider again the treasury bill data in Section 3.1. (For this set of data, the P-value for testing homogeneity in volatility under model (20) is 0.03.) The adequacy of the three models can also be assessed by their residual autocorrelation functions. Figure 12 gives the residual autocorrelation functions and the Q-Q plots for the residuals from the three models. It is evident that, from the Q-Q plots, our models (5) and (20) are more adequate than the CKLS model. Figure 13 reports the estimates of the coefficients in model (20) along with the 95% bootstrap confidence bands using 1000 simulations. The estimates of the expected return and volatility are shown in Figure 14. It seems that the time variation in volatility occurs after 1980 (admittedly the confidence interval is wide) and not before; this observation is similar to that made in Section 3.1.

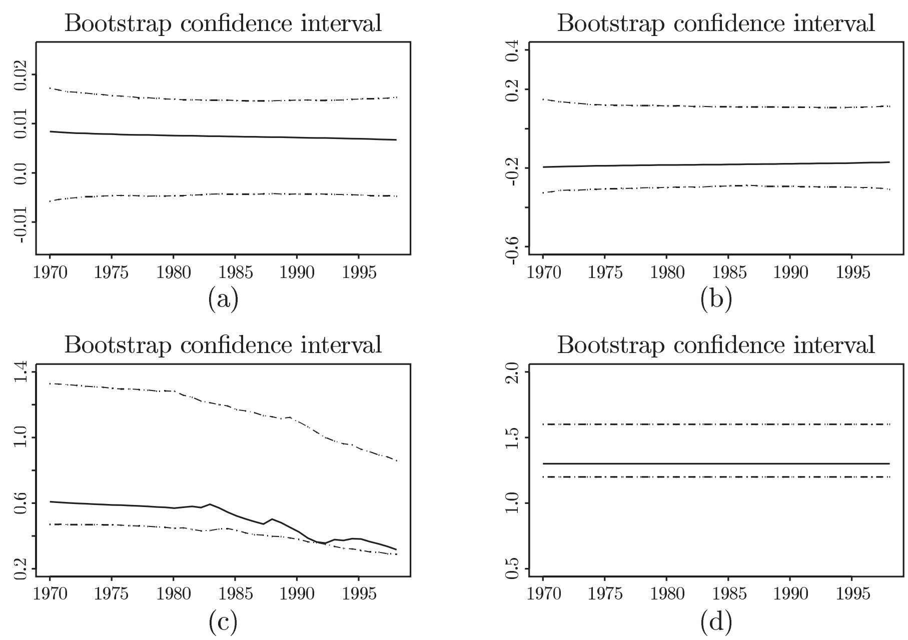
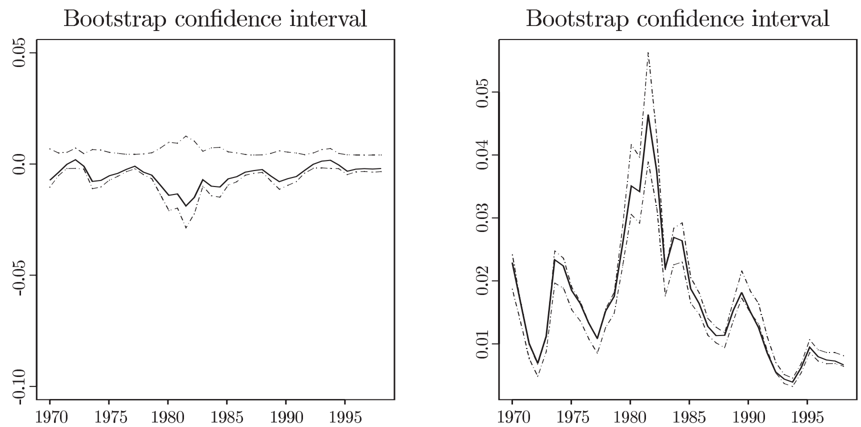


Figure 13. Three month treasury Bill rate, from January 2, 1970 to December 26, 1997: Estimated coefficient functions for the semiparametric model (20). Solid – our estimators, dash-dotted – 95% confidence band among 1000 simulations.

The sample skewness and kurtosis of the original series $\{Y_{t_i}\}$ and the residuals from the three models are reported in Table 6. The kurtoses of the residuals from the three models are much smaller in magnitude than that of $\{Y_{t_i}\}$, which reflects that the three models successfully reveal the phenomena of time-varying volatilities in the yields of the treasury bill. Note that the residuals from our model (20) have the smallest kurtosis in this example.



(a) (b)

Figure 14. Three month treasury Bill rate, from January 2, 1970 to December 26, 1997: Estimated drift and volatility for the semiparametric model (20). Solid – our estimators, dash-dotted – 95% confidence band.

Table 6. Skewness and kurtosis.

Model	Skewness	Kurtosis
Original series $\{Y_{t_i}\}$	-0.002	13.975
Residuals from CKLS model	0.021	4.620
Residuals from Model (5)	0.015	4.429
Residuals from Model (20)	0.009	3.973

7. Conclusion

The time-varying coefficient model (5) is introduced to better capture the time variation of short-term dynamics. It has been demonstrated to be an effective tool for modelling volatility and validating existing models. It arises naturally from various considerations and encompasses most of the commonly used models as special cases. Yet, due to limited independent data information, coefficients in model (5) may not be estimated very reliably. The semiparametric model (20) provides a useful alternative.

Acknowledgements

Fan's work was partially supported by the NSF Grant DMS-0204329, the grant CUHK 4299/00 of the Research Grant Council grant of Hong Kong and a direct grant from the Chinese University of Hong Kong. Jiang's research was supported in part by Chinese NSF Grants 10001004 and 79800017. Research of the third author was supported by Wisconsin Alumni Research Foundation. The authors wish to thank the guest Editor, Ruey Tsay, and two anonymous referees for their constructive comments which improved the presentation of this paper. We also gratefully acknowledge helpful comments and suggestions from Professors Ron Gallant and Jia He.

References

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

Aït-Sahalia, Y. (1996b). Testing continuous-time models of the spot interest rate. *Rev. Finan. Stud.* **9**, 385-426.

Aït-Sahalia, Y. and Lo, A. W. (1998). Nonparametric estimation of state-price densities implicit in financial asset prices. *J. Finance* **53**, 499-548.

Aït-Sahalia, Y. (1999). Transition densities for interest rate and other nonlinear diffusions. *J. Finance* **LIV**, 1361-1395.

Aït-Sahalia, Y. (2002). Maximum likelihood estimation of discretely sampled diffusions: A closed-form approximation approach. *Econometrica* **70**, 223-262.

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

Black, F., Derman, E. and Toy, W. (1990). A one-factor model of interest rates and its application to treasury bond options. *Finan. Analysts' J.* **46**, 33-39.

Black, F. and Karasinski, P. (1991). Bond and option pricing when short rates are lognormal. *Finan. Analysts' J.* **47**, 52-59.

Black, F. and Scholes, M. (1973). The pricing of options and corporate liabilities. *J. Polit. Economy* **81**, 637-654.

Brennan, M. J. and Schwartz, E. S. (1979). A continuous time approach to the pricing of bonds. *J. Banking Finance* **3**, 133-155.

Brennan, M. J. and Schwartz, E. S. (1982). An equilibrium model of bond pricing and a test of market efficiency. *J. Finan. Quant. Anal.* **17**, 210-239.

Bühlman, P. and Künsch, H. R. (1995). The blockwise bootstrap for general parameters of a stationary time series. *Scand. J. Statist.* **22**, 35-54.

Chan, K. C., Karolyi, A. G., 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.

Chapman, D. A. and Pearson, N. D. (2000). Is the short rate drift actually nonlinear? *J. Finance* **55**, 355-388.

Constantinides, G. M. and Ingersoll, J. E. (1984). Optimal bond trading with personal taxes. *J. Finan. Econom.* **13**, 299-335.

Cox, J. C., Ingersoll, J. E. and Ross, S. A. (1980). An analysis of variable rate loan contracts. *J. Finance* **35**, 389-403.

Cox, J. C., Ingersoll, J. E. and Ross, S. A. (1985). A theory of the term structure of interest rates. *Econometrica* **53**, 385-467.

Dothan, U. L. (1978). On the term structure of interest rates. *J. Finan. Econom.* **6**, 59-69.

Duffie, D. (1996). *Dynamic Asset Pricing Theory.* 2nd edition. Princeton University Press, Princeton, New Jersey.

Epanechnikov, V. A. (1969). Nonparametric estimation of a multidimensional probability density. *Theory Probab. Appl.* **13**, 153-158.

Fan, J. (1993). Local linear regression smoothers and their minimax efficiencies. *Ann. Statist.* **21**, 196-216.

Fan, J. and Gijbels, I. (1995). Data-driven Bandwidth selection in local polynomial fitting: variable bandwidth and spatial adaptation. *J. Roy. Statist. Soc. Ser. B* **57**, 371-394.

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

Fan, J. and Yao, Q. W. (1998). Efficient estimation of conditional variance functions in stochastic regression. *Biometrika* **85**, 645-660.

Fan, J., Zhang, C. M. and Zhang, J. (2001). Generalized likelihood ratio statistics and Wilks phenomenon. *Ann. Statist.* **29**, 153-193.

Fan, J. and Zhang, C. M. (2003). A re-examination of Stantan's diffusion estimators with applications to financial model validation. *J. Amer. Statist. Assoc.* **98**, 118-134.

Feldman, D. (1989). The term structure of interest rates in a partially observable economy. *J. Finance* **44**, 789-812.

Florens-Zmirou, D. (1993). On estimating the diffusion coefficient from discrete observations. *J. Appl. Probab.* **30**, 790-804.

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

Gallant, A. R. and Long, J. R. (1997). Estimating stochastic differential equations efficiently by minimum chi-squared. *Biometrika* **84**, 125-141.

Gallant, A. R., Rossi, P. E. and Tauchen, G. (1997). Nonlinear dynamic structures. *Econometrica* **61**, 871-907.

Gallant, A. R. and Tauchen, G. (1997). Estimation of continuous time models for stock returns and interest rates. *Macroecon. Dynam.* **1**, 135-168.

Gallant, A. R. and Tauchen, G. (1998). Reprojecting partially observed systems with application to interest rate diffusions. *J. Amer. Statist. Assoc.* **93**, 10-24.

Genon-Catalot, V. and Jacod, J. (1993). On the estimation of the diffusion coefficient for multi-dimensional diffusion processes. *Ann. Inst. H. Poincar Probab. Statist.* **29**, 119-51.

Green, P. J. and Silverman, B. W. (1994). *Nonparametric Regression and Generalized Linear Models: a Roughness Penalty Approach*. Chapman and Hall, London.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica* **50**, 1029-1054.

Hansen, L. P. and Scheinkman, J. A. (1995). Back to the future: generating moment implications for continuous-time Markov processes. *Econometrica* **63**, 767-804.

Hart, J. D. (1994). Automated kernel smoothing of dependent data by using time series cross-validation. *J. Roy. Statist. Soc. Ser. B* **56**, 529–542.

Hart, J. D. (1996). Some automated methods of smoothing time-dependent data. *J. Nonparametr. Statist.* **2-3**, 115-142.

Härdle, W. (1990). *Applied Nonparametric Regression*. Cambridge University Press, Boston.

Hastie, T. J. and Tibshirani, R. (1990). *Generalized Additive Models*. Chapman and Hall, London.

Hastie, T. J. and Loaders, C. (1993). Local regression: automatic kernel carpentry (with discussion). *Statist. Sci.* **8**, 120-143.

Ho, T. S. Y. and Lee, S. B. (1986). Term structure movements and pricing interest rate contingent claims. *J. Finance* **41**, 1011-1029.

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

Hull, J. C. (1997). *Options, Futures and Other Derivatives*. 3rd edition. Prentice Hall, Upper Saddle River, New Jersey.

Künsch, H. R. (1989). The jackknife and the bootstrap for general stationary observations. *Ann. Statist.* **17**, 1217-1241.

Longstaff, F. A. (1989). A nonlinear general equilibrium model of the term structure of interest rates. *J. Finan. Econom.* **23**, 195-224.

Longstaff, F. A. and Schwartz, E. S. (1992). Interest-rate volatility and the term structure: a two-factor general equilibrium model. *J. Finance* **47**, 1259-1282.

Merton, R. C. (1973). Theory of rational option pricing. *Bell J. Econom. Management Sci.* **4**, 141-183.

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

Mishkin, F. S. (1997). *The Economics for Money, Banking and Financial Markets*. 5th edition. Addison-Wesley, New York.

Morgan, J. P. (1996). *RiskMetrics Technical Document*. 4th edition. New York.

Ruppert, D. and Wand, M. P. (1994). Multivariate weighted least squares regression. *Ann. Statist.* **22**, 1346-1370.

Ruppert, D., Sheather, S. J. and Wand, M. P. (1995). An effective bandwidth selector for local least squares regression. *J. Amer. Statist. Assoc.* **90**, 1257-1270.

Sandmann, K. and Sondermann, D. (1997). A note on the stability of lognormal interest rate models and the pricing of Eurodollar futures. *Math. Finance* **7**, 119-125.

Scott, D. W. (1992). *Multivariate Density Estimation: Theory, Practice and Visualization*. John Wiley, New York.

Schaefer, S. and Schwartz, E. (1984). A two-factor model of the term structure: an approximate analytical solution. *J. Finan. Quantitative Anal.* **19**, 413-424.

Simonoff, J. S. (1996). *Smoothing Methods in Statistics*. Springer-Verlag, New York.

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

Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions (with discussion). *J. Roy. Statist. Soc. Ser. B* **36**, 111-147.

Stutzer, M. (1996). A simple nonparametric approach to derivative security valuation. *J. Finance* **51**, 1633-1652.

Wahba, G. (1977). A survey of some smoothing problems and the method of generalized cross-validation for solving them. In *Applications of Statistics* (Edited by P. R. Krisnaiah), 507-523. North Holland, Amsterdam.

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

Wong, E. (1971). *Stochastic Processes in Information and Dynamical Systems*. McGraw-Hill, New York.

Department of Statistics, Chinese University of Hong Kong, Shatin, Hong Kong.

E-mail: jfan@sta.cuhk.edu.hk

Department of Probability and Statistics, Peking University, Beijing, 100871, P.R.C.

E-mail: jiang@math.pku.edu.cn

Department of Statistics, University of Wisconsin Madison, WI 53706, U.S.A.

E-mail: cmzhang@stat.wisc.edu

Department of Statistics, University of California Los Angeles, CA 90095, U.S.A.

E-mail: zzw@stat.ucla.edu

(Received March 2001; accepted July 2003)