Modelling and Forecasting Short-Term Interest Rate Volatility: A Semiparametric Approach

HOU, Ai Jun; Suardi, Sandy

Published in:
Journal of Empirical Finance

DOI:
10.1016/j.jempfin.2011.05.001

2011

Modelling and Forecasting Short-Term Interest Rate Volatility: A Semiparametric Approach

Abstract

This paper employs a semiparametric procedure to estimate the diffusion process of short-term interest rates. The Monte Carlo study shows that the semiparametric approach produces more accurate volatility estimates than models that accommodate asymmetry, levels effect and serial dependence in the conditional variance. Moreover, the semiparametric approach yields robust volatility estimates even if the short rate drift function and the underlying innovation distribution are misspecified. Empirical investigation with the U.S. three-month Treasury bill rates suggests that the semiparametric procedure produces superior in-sample and out-of-sample forecast of short rate changes volatility compared with the widely used single-factor diffusion models. This forecast improvement has implications for pricing interest rate derivatives.

Keywords: Interest Rates; GARCH modelling; Nonparametric method; Volatility estimation; Forecasts

J.E.L. Reference Numbers: E43;C22; C53

1 Introduction

There is an extensive literature on the modelling of the short-term interest rate as this rate is fundamental to the pricing of fixed-income securities and for the measurement of the interest rate risk associated with holding portfolios of these securities. One of the earliest papers that formally compares a number of single-factor models is Chan et al. (1992). Based on U.S. data, their study controversially rejects the commonly adopted square root diffusion model of Cox et al. (1985), whereby the volatility of short rate changes is proportional to the square root of the interest rate levels. Instead, their model shows that volatility is more sensitive to interest rate levels, specifying an exponent for the commonly known level effect in the region of 1.5. A more recent study by Brenner et al. (1996) shows that models which parameterise volatility only as a function of interest rate levels tend to over emphasize the sensitivity of volatility to levels and do not take into consideration the serial correlation in conditional variances. They propose a new class of models which allows volatility to depend on both interest rate levels and information shocks. There is by now a general consensus in the literature that short rate models which account for both levels effect and serial correlation in the volatility processes perform better than models that solely parameterises levels effect or serial dependence in the conditional variances (Bali, 2000).

Unlike the diffusion process, the appropriate drift specification of short rate models remains highly controversial in the literature. While a large proportion of research reports a linear drift (such as Chan et al., 1992 and the models nested by it), others argue to the contrary (Ait Sahalia, 1996a,b; Conley et al., 1997; Jones, 2003) finding nonlinear mean reversion. Using a semi-nonparametric approach, Ait-Sahalia (1996a,b) constructs a general specification test of a short-rate model and rejects a linear drift in favour of models that imply no mean reversion for levels of the short rate between certain threshold levels, and strong mean reversion for extreme levels of the short rate. Stanton (1997) and Jiang (1998) estimate a model of the short rate nonparametrically using different data sets from Ait-Sahalia and find supporting evidence of nonlinearities in the drift function. Bali and Wu (2006) document evidence that the speeds of mean-reversion for short-term interest rates at extremely high interest rates such as in the Volker (1979-1982) regime are different than at normal times. They attribute the nonlinearity in short-rate drift to the differences in the degree of mean-reversion at different interest rate levels. Be that as it may, the robustness of the nonlinear drift function in short rate models has been questioned by some authors. Pritsker (1998) examines the finite-sample properties of Ait-Sahalia's nonparametric test, showing that upon adjusting for the high persistence in interest rates, the nonlinearity in the drift function becomes statistically insignificant. Chapman and Pearson (2000) perform simulation exercise and show that the evidence supporting the nonlinear drift function could be an artifact of the nonparametric estimation procedure rather than depicting the true data generating process. Using Bayesian estimation method, Jones (2003) shows that the determination of short-rate drift specification is dependent on the assumption of the prior distribution. In particular, under the assumption of a flat prior distribution and the imposition of stationarity in interest rate dynamics, he identifies a nonlinear drift. However, when he implements an approximate Jeffreys prior there is no mean-reverting evidence. Finally, Durham (2004) finds that the significance of nonlinearity in the drift function depends on the specification of the diffusion process, a finding which accords with Bali (2007).

This paper considers an alternative method for modelling short rate volatility. We apply a semiparametric smoothing technique to the generalised autoregressive conditional heteroskedasticity (GARCH) model of short rate volatility. This involves estimating a parametric form of the short rate drift function followed by estimating the hidden volatility process nonparametrically. On the basis that the literature is divided about the appropriate drift specification, we estimate both linear and nonlinear drift functions of the short rate. The estimation of a parametric drift specification qualifies this approach as a semiparametric method (Jiang and Knight, 1997). To estimate the latent volatility process, we use the algorithm developed by Bühlman and McNeil (2002) and apply it to a generalised additive model of Hastie and Tibshirani (1990). Bühlman and McNeil (2002) argue that estimating the volatility process with the nonparametric approach is less sensitive to model misspecification and does not require *a priori* knowledge of the innovation distribution. This feature makes the application of nonparametric method attractive for estimating short rate diffusion process given that short rates are known to possess distributions that depart from normality. We specify the latent volatility process as a general additive function of the lagged value of the conditional variance, innovations and interest rate levels. This specification is consistent with a class of single-factor short rate diffusion models where the volatility of short rate changes is serially dependent on past volatility, squared innovations and interest rate levels. In addition, the additive structure of the hidden volatility facilitates the use of a backfitting algorithm to estimate the diffusion process.

The potential usefulness of the semiparametric approach for estimating short rate volatility is examined by comparing its forecast performance with a variety of one-factor short rate diffusion models. Results from our Monte Carlo simulation illustrate the robustness of the semiparametric approach when estimating short rate changes volatility to misspecification in the short rate drift function and the underlying innovation distribution. Moreover, the in-sample forecast performance of the semiparametric approach is superior to the parametric models considered. The empirical application to U.S. three-month Treasury bill yields suggests that the semiparametric estimation procedure provides superior in-sample and out-of-sample volatility forecast than short rate volatility models of Brenner et al. (1996) which feature asymmetric and levels-dependent conditional variance. Although the semiparametric approach does not specify the asymmetric feature of the volatility process, this procedure improves upon the goodness of fit and the predictive power of the volatility estimates. We do not find any evidence of nonlinearities in short rate drift and conditional skewness in short rate changes distribution. Finally, we demonstrate that the semiparametric approach, which yields a greater degree of accuracy in modelling short rate changes volatility, has pertinent implications for pricing long-dated and path-dependent interest rate derivatives.[1] Using simulation method, we show that the semiparametric modelling approach gives rise to significantly different probability distributions of future interest rate levels compared with parametric short rate models. The confidence intervals of future interest rate levels are narrower than any of the parametric models considered, thereby leading to a lesser degree of price variability of interest rate derivatives.

The rest of the paper is organised as follows. Section 2 describes the short rate models and the semiparametric smoothing technique. Section 3 outlines the design of the Monte Carlo experiment to examine the in-sample predictive power of the semiparametric approach and its forecast property when subject to possible misspecifications in the drift function and the innovation distribution. This section also reports the results of the simulation study. Section 4 applies the semiparametric technique to the U.S. short-term interest rates to evaluate its in-sample and out-of-sample forecast performance relative to other short rate models. Implications of this forecast improvement on pricing interest rate derivatives are also discussed. Section 5 concludes.

2 The Short Rate Models and the Semiparametric Approach

2.1 The Short Rate Models

Chan et al. (1992) (CKLS, hereafter) propose the generalised continuous time short rate specification,

$$dr = (\mu + \lambda r)\, dt + \phi r^{\delta} dW \tag{1}$$

where r denotes the level of the short-term interest rate, W is a Brownian motion and μ, λ and δ are parameters. The drift component of short-term interest rates is captured by $\mu + \lambda r$ while the variance of unexpected changes in interest rates equals $\phi^2 r^{2\delta}$. The parameter ϕ is a scale factor and δ controls the degree to which the interest rate level influences the volatility of short-term interest rate changes. The CKLS model nests many of the existing interest rate models. For example, when $\delta = 0$ then (1) reduces to the Vasicek (1977) model, while $\delta = 1/2$ yields the Cox, Ingersoll and Ross (1985) model, see Chan *et al.* (1992) *inter alia* for further details. There is a dearth of literature that focuses on the univariate CKLS model. Czellar, Karolyi and Ronchetti (2007) study different estimation techniques for the CKLS short rate model. Bali and Wu (2006) investigate extensions of the mean specification of the CKLS model. On the other hand, Nowman and Sorwar (2005) use the CKLS model to price bonds and contingent claims.

It is common to consider the Euler-Maruyama discrete time approximation to (1) written as:

$$\Delta r_t = \mu + \lambda r_{t-1} + \varepsilon_t. \tag{2}$$

Let Ω_{t-1} represent the information set available at time $t-1$ and that $E(\varepsilon_t|\Omega_{t-1}) = 0$. Suppose h_t represent the conditional variance of the short-term interest rate changes then $E(\varepsilon_t^2|\Omega_{t-1}) \equiv h_t = \phi^2 r_{t-1}^{2\delta}$. It can be seen that the only source of conditional heteroscedasticity in (2) is through the level of the interest rate. Brenner et al. (1996) (BHK, hereafter) relaxes the assumption of a constant ϕ^2 by allowing it to vary according to information arrival process. One common approach to capturing the effect of unanticipated news is the GARCH(1,1) model:

$$h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 h_{t-1}. \tag{3}$$

The innovation ε_t denotes a change in the information set from time $t-1$ to t and can be treated as a collective measure of unanticipated news. In (3) only the magnitude of the innovation is important in determining h_t. BHK extend (2) to allow information from unanticipated news and the one-period lagged interest rate levels to govern the dynamics of short rate volatility in the following way:

$$\begin{aligned} \Delta r_t &= \mu + \lambda r_{t-1} + \varepsilon_t, \\ \varepsilon_t &= \sqrt{h_t} z_t,\ z_t \sim t(v) \text{ and} \\ h_t &= \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 h_{t-1} + b r_{t-1}^{2\delta}. \end{aligned} \tag{4}$$

Equation (4) is known as the GARCH-X process. Under the restriction $\alpha_0 = \alpha_1 = \alpha_2 = 0,$ (4) collapses to (2) where $b = \phi^2$ and that volatility depends on interest rate levels alone. Furthermore when $b = 0$ then there is no levels effect. The GARCH-X model does not permit short rate volatility to respond asymmetrically to interest rate innovations of different sign. BHK relax the assumption of a symmetric GARCH-X process by modelling the conditional variance specification as:

$$h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 h_{t-1} + b r_{t-1}^{2\delta} + \alpha_3 \xi_{t-1}^2 \quad (5)$$

where $\xi_{t-1} = \min(0, \varepsilon_{t-1})$. BHK refer to this model as the AsyGARCH-X. For simplicity, we refer to the symmetric (asymmetric) GARCH-X as the GARCHX (AGARCHX) model. For the purpose of this paper, we only consider the additive levels effect as the (A)GARCHX model is consistent with the generalised additive non-parametric GARCH model discussed in the next sub-section.[2] In practice, when estimating the (A)GARCHX model it is common to scale the interest rate level term in the variance equation with a factor (1/10) such that the levels dependence in the conditional variance is captured by $b(r_{t-1}/10)^{2\delta}$ (see Brenner et al., 1996).

The linear drift in equation (2) implies that the strength of mean reversion is the same for all levels of the short rate. Even though there is no *a priori* economic intuition that would suggest the existence of a nonlinear drift, empirical research has shown that there is evidence of nonlinear drift in short-term interest rates, that is, mean reversion is stronger for extreme low or high levels of short rate. Ait-Sahalia (1996a) advocates the use of a flexible functional forms to approximate the true unknown shapes of short rate process. He estimates a short rate model:

$$dr_t = (\mu + \lambda_1 r_t + \lambda_2 r_t^2 + \frac{\lambda_3}{r_t})dt + \sqrt{\beta_0 + \beta_1 r_t + \beta_2 r_t^{\beta_3}} dW_t. \quad (6)$$

and finds that the test rejects a linear drift in favour of models that imply no mean reversion for levels of the short rate between 4 and 22 percent, and strong mean reversion for extreme levels of the short rate. Conley et al. (1997) adopt the same drift parameterisation as Ait-Sahalia but keeps the constant elasticity variance diffusion used by CKLS:

$$dr_t = (\mu + \lambda_1 r_t + \lambda_2 r_t^2 + \frac{\lambda_3}{r_t})dt + \sigma r_t^{\gamma} dW_t. \quad (7)$$

They find that the drift function displays mean-reversion only for rates below 3% or above 11%. Bali (2007) also estimates the nonlinear drift specification of Ait-Sahalia in (6) but with a diffusion process that follows a GARCH(1,1) model and is dependent on interest rate levels. Bali and Wu (2006), estimate a variant of the drift specification in (6) which includes a fifth order polynomial. As is apparent from extensive research which adopts the nonlinear drift specification of Ait-Sahalia and the possible influence this nonlinear drift might exert on the conditional volatility of interest rate changes, we also estimate a discrete time approximation of the Ait-Sahalia (1996a) nonlinear drift specification:

$$\Delta r_t = \mu + \lambda_1 r_{t-1} + \lambda_2 r_{t-1}^2 + \frac{\lambda_3}{r_{t-1}} + \varepsilon_t \tag{8}$$

and the conditional variance of the short-term interest rate changes which follows equation (5).

Empirical studies on short-term interest rates have shown that the standardised residuals obtained from the GARCH models exhibit leptokurtosis. The assumption of normality is easily rejected by the Jarque-Bera test when applied to short rate data. Consequently, the Student's t distribution is commonly employed to capture the thicker tails in the empirical distribution of short rates. There are, however, other non-normal distributions which have been used to characterise the distribution of short rate changes. In particular, much attention has been paid in modeling the skewness of the distribution. Bali (2007) adopts the skewed generalised error distribution of Theodossiou (1998) as well as the Hansen's (1994) skewed t-distribution to capture the skewness in the empirical distribution of the U.S. 3-month Treasury bill yield. Following Bali (2007), we employ both the Student's t and Hansen's skewed t distributions in the Monte Carlo experiment and empirical application. For the skewed t distribution, we define the residuals in equation (4) as:

$$\varepsilon_t = \sqrt{h_t} z_t, \;\; z_t \sim Hansen's \; t(v, \eta) \; . \tag{9}$$

The parameters η and v control the direction of asymmetry and kurtosis of the distribution.

The Hansen's skewed t distribution is defined by

$$f(z_t;\Theta,v,\eta)=\begin{cases} bc\left(1+\frac{1}{v-2}\left(\frac{bz_t+a}{1-\eta}\right)^2\right)^{-\frac{v+1}{2}} & \text{if } z_t<-\frac{a}{b} \\ bc\left(1+\frac{1}{v-2}\left(\frac{bz_t+a}{1+\eta}\right)^2\right)^{-\frac{v+1}{2}} & \text{if } z_t\geq-\frac{a}{b} \end{cases} \tag{10}$$

where $z_t=\frac{\varepsilon_t}{\sqrt{h_t}}$, Θ is the set of parameters associated with the drift and diffusion specifications of the short rate model, and the constants a, b and c are given by

$$a=4\lambda c\left(\frac{v-2}{v-1}\right),\quad b^2=1+3\lambda^2-a^2,\quad c=\frac{\Gamma\left(\frac{v+1}{2}\right)}{\sqrt{\pi(v-2)}\Gamma\left(\frac{v}{2}\right)}.$$

For $\eta=0$ the Hansen's distribution reduces to the traditional standardised t distribution, while for $\eta=0$ and $v=\infty$, it reduces to a normal density. The density (10) is defined for $2<v<\infty$ and $-1<\eta<1$.

2.2 The Generalised Additive Semiparametric GARCH Model

Consider the short-rate model:

$$X_t = \sigma_t Z_t \tag{11}$$

$$\sigma_t^2 = f_1(X_{t-1})+f_2(\sigma_{t-1}^2)+f_3(r_{t-1}) \tag{12}$$

where $\{Z_t;t\in\mathbb{Z}\}$ is an i.i.d innovation with zero mean, unit variance and a finite fourth moment, and $X_t=\Delta r_t-(\mu+\lambda r_{t-1})$ for a linear drift and $X_t=\Delta r_t-(\mu+\lambda_1 r_{t-1}+\lambda_2 r_{t-1}^2+\frac{\lambda_3}{r_{t-1}})$ for a nonlinear drift. Let $f_1:\Re\rightarrow\Re_+$, $f_2:\Re_+\rightarrow\Re_+$ and $f_3:\Re_+\rightarrow\Re_+$ be strictly positive-valued functions. The conditional variance and volatility are denoted by σ_t^2 and σ_t, respectively. Further assume that X_t and r_t are stationary stochastic processes and $\{X_t;t\in\mathbb{Z}\}$ is adapted to the $\sigma-$filtration $\{F_t;t\in\mathbb{Z}\}$ with $F_t=\sigma\left(\{X_s;s\leq t\}\right)$. The assumption of stationarity in r_t is empirically verified by performing the Seo (1999) unit root test on the U.S. 3-month Treasury bill rate. To ensure comparability with the CKLS and BHK short rate models, we first estimate a linear drift function specified in equation

(2). However, given the vast literature on short rate models with nonlinear drift function, we also investigate the nonlinear drift specification of Ait-Sahalia (1996a) given by equation (8). The exact form of the functions f_1, f_2 and f_3 in (12) is left unspecified but it can be estimated by a nonparametric method in which X_t^2 is regressed on the lagged variables X_{t-1}, σ_{t-1}^2 and r_{t-1}. To show that this procedure is applicable for estimating the unobserved variable σ_t^2, we re-write the model (11 and 12) as:

$$\begin{aligned} X_t^2 &= f_1(X_{t-1}) + f_2(\sigma_{t-1}^2) + f_3(r_{t-1}) + V_t \\ V_t &= \left[f_1(X_{t-1}) + f_2(\sigma_{t-1}^2) + f_3(r_{t-1})\right]\left(Z_t^2 - 1\right) \end{aligned} \tag{13}$$

where V_t is a martingale difference series with $E(V_t) = E(V_t|F_{t-1}) = 0$ and $Cov(V_s, V_t) = Cov(V_s, V_t|F_{t-1}) = 0$ for $s < t$. Taking the conditional expectations of X_t^2 in (13) yields:

$$E(X_t^2|F_{t-1}) = f_1(X_{t-1}) + f_2(\sigma_{t-1}^2) + f_3(r_{t-1}) \tag{14}$$

and its conditional variance can be shown to be:

$$Var(X_t^2|F_{t-1}) = \left[f_1(X_{t-1}) + f_2(\sigma_{t-1}^2) + f_3(r_{t-1})\right]^2 \left[E(Z_t^4) - 1\right]. \tag{15}$$

To estimate the latent variable σ_t^2 in (12), we adopt the estimation algorithm of Bühlmann and McNeil (2002). For a given data sample we first calculate the volatility estimate $\hat{\sigma}_{t,0}$ by first estimating the linear drift specification (2) and the conditional variance (3) using the method of maximum likelihood. Further, we verify that the semiparametric approach is robust to possible nonlinear drift by estimating the volatility estimate $\hat{\sigma}_{t,0}$ with a nonlinear drift specification. Note that $\hat{\sigma}_{t,0}$ is used as the initial volatility estimate. In the first iteration, we regress $\{X_t^2; 2 \le t \le n\}$ against $\{X_{t-1}; 2 \le t \le n\}$, $\left\{\hat{\sigma}_{t-1,0}^2; 2 \le t \le n\right\}$ and $\{r_{t-1}; 2 \le t \le n\}$ using a nonparametric smoothing procedure with a backfitting algorithm to obtain an estimate $\hat{f}_{i,1}$ of f_i for $i = 1, 2$ and 3.[3] The regression is performed with regression weights $\left\{\hat{\sigma}_{t,0}^{-2}; 2 \le t \le n\right\}$ as this yields improved estimates of σ_t^2 (Bühlmann and McNeil, 2002).

Having estimated $f_{i,1}$, we then calculate $\hat{\sigma}^2_{t,1} = \hat{f}_{1,1}(X_{t-1}) + \hat{f}_{2,1}(\hat{\sigma}^2_{t-1,0}) + \hat{f}_{3,1}(r_{t-1})$. In the next iteration, we perform another regression to obtain $\hat{f}_{i,2}$ and $\hat{\sigma}^2_{t,2}$ which yields improved estimates of the conditional variance $\hat{\sigma}^2_{t,2}$. This iterative process is performed for a prespecified number of iterations, m. As shown by Bühlmann and McNeil (2002) and according to our estimation experience which is documented in the simulation results, the improvement over the parametric GARCH estimation of volatility can be attained in a small number of iterations which usually takes place in the first 4 iterations. The algorithm can be improved by averaging over the final K estimates to derive:

$$\hat{\sigma}_{t,*} = \frac{1}{K} \sum_{m=M-K+1}^{M} \hat{\sigma}_{t,m}. \tag{16}$$

Note that we average over the volatility rather than the conditional variance since $\hat{\sigma}_t$ is our proxy for volatility. In the final smoothing, we regress X_t^2 against X_{t-1}, $\hat{\sigma}^2_{t-1,*}$ and r_{t-1} to obtain $\hat{f}_i$ and $\hat{\sigma}^2_t = \hat{f}_1(X_{t-1}) + \hat{f}_2(\hat{\sigma}^2_{t-1,*}) + \hat{f}_3(r_{t-1})$. In our empirical application and simulation experiment we obtain the final smoothing based on $K = 4$ for eight iterations (i.e. $m = 8$).

3 Monte Carlo Study

3.1 Experimental Design

The purpose of the simulation experiment is to illustrate the superior volatility forecast performance of the semiparametric procedure compared with parametric short rate models. In addition, we show that the semiparametric method yields volatility forecasts that are invariant to the underlying distribution of the short rate and its drift specification.

The data generating process (DGP) for interest rates with a linear drift follows the

AGARCHX model (2) and (5). Specifically, the DGP with a linear drift is:

$$
\begin{aligned}
\Delta r_t &= 0.06 - 0.008 r_{t-1} + \varepsilon_t, && (17)\\
\varepsilon_t &= \sigma_t z_t, \;\; z_t \sim t\,(4), && (18)\\
\sigma_t^2 &= 0.24 + 0.1026\varepsilon_{t-1}^2 + 0.5595\xi_{t-1}^2 + 0.3282\sigma_{t-1}^2 + 0.015(r_{t-1}/10) && (19)
\end{aligned}
$$

where $\xi_{t-1} = \min(0, \varepsilon_{t-1})$. The use of Student's t distribution for interest rate innovation is consistent with the widely observed non-normal short-term interest rate distribution. Moreover, for the purpose of examining the effects of nonlinear drift function on forecasts generated by the semiparametric approach, we consider the following DGP:

$$
\begin{aligned}
\Delta r_t &= 0.06 + 0.008 r_{t-1} - 0.01 r_{t-1}^2 + 0.0002/r_{t-1} + \varepsilon_t, && (20)\\
\varepsilon_t &= \sigma_t z_t, \;\; z_t \sim \; t\,(6), && (21)\\
\sigma_t^2 &= 0.05 + 0.1026\varepsilon_{t-1}^2 + 0.5595\xi_{t-1}^2 + 0.3282\sigma_{t-1}^2 + 0.045(r_{t-1}/10) && (22)
\end{aligned}
$$

where $\xi_{t-1} = \min(0, \varepsilon_{t-1})$. The parameter values used in the DGPs are typical of short rate empirical research. We discard the initial 50 observations to mitigate the effect of start-up values yielding samples of 1000 observations, drawn with 50 replications. The small number of replications does not bias the results in any way. In fact, this is consistent with the number of replications performed in the simulation experiment conducted by Bühlmann and McNeil (2002). Upon generating the data, we estimate the parametric models of short-term interest rates with linear and nonlinear drifts, symmetric and asymmetric GARCHX models, and with three different innovation distributions, namely normal, Student's t and Hansen's skewed t distributions. In addition, we estimate the latent volatility using the method of the generalised additive semiparametric GARCH model discussed in sub-section 2.2. For both DGPs we fit linear and nonlinear drift specifications before applying the nonparametric smoothing technique to the volatility estimates. The parametric models are estimated by maximizing the log-likelihood function using the Broyden, Fletcher, Goldfarb and Shanno

(BFGS) algorithm with the Bollerslev and Wooldridge (1992) robust standard error.

To compare the goodness of fit of the in-sample volatility estimates for the different models, we compute the absolute error (AE) and the squared error (SE) given by:

$$AE(\hat{\sigma}_{.,m}) = \frac{1}{1000-r}\sum_{t=r+1}^{1000}|\hat{\sigma}_{t,m}-\sigma_t| \qquad \text{and} \qquad SE(\hat{\sigma}_{.,m}) = \frac{1}{1000-r}\sum_{t=r+1}^{1000}(\hat{\sigma}_{t,m}-\sigma_t)^2 \tag{23}$$

where $r = 50$ is the semiparametric estimates of volatility at the first fifty time points that are omitted as these estimates may be unreliable, and m applies only to the semiparametric approach and refers to the specific number of iterations. These measures are also computed at each iteration of the semiparametric procedure to show the degree of improvement in the goodness of fit of the volatility estimates. For the 50 independent realisations, we average our volatility estimation error statistics to provide an estimate of mean squared error (MSE) and mean absolute error (MAE), as well as the standard errors for the MSE and MAE estimates.

3.2 Simulation Results

Figures 1(a) and (b) present plots of volatility estimates from the DGP with linear and nonlinear drifts, respectively. Column three of Figures 1(a) and (b) shows volatility plots of the semiparametric method while columns one and two show volatility plots of the parametric GARCHX and AGARCHX models. Both the true and estimated volatility are plotted together so as to provide a visual impression of their goodness of fit. To conserve space, we only report an arbitrarily selected sample of 100 observations from one of the replications results. The plot of the volatility estimates produced by the semiparametric method is based on the final smoothed σ_t estimate. A cursory look at Figure 1(a) and (b) suggests that the semiparametric approach yields volatility estimates which match the true simulated volatility better than the parametric models. This result is robust to the innovation distributional assumption. The GARCHX and AGARCHX models fail to produce volatility estimates that can adequately capture the variation in the true volatility even though in some instances they capture the spikes relatively well. Another interesting observation which is commonly shared by Figures 1(a) and (b) is that the parametric volatility estimates tend to be higher than the actual volatility level. This is not the case with the semiparametric estimates; they trace the level of the true volatility well. When comparing the volatility estimates produced by the GARCHX and the AGARCHX models, we find that a model with asymmetric conditional variance produces estimates that better depict the actual volatility. This result is, perhaps, not surprising as the DGP possesses this asymmetric feature in the conditional variance. There is some indication that the volatility estimates generated by the same parametric model but with different innovation distributional assumptions are distinct. This distinction is less noticeable with estimates produced by the semiparametric approach.

- Figures 1a and b about here -

Tables 1(a) and (b) show the estimation error results for the in-sample volatility estimates of various short rate models for DGPs with linear and nonlinear drifts, respectively. For both DGPs, there is evidence that the standard GARCH model yields the largest MSE and MAE. This result is robust to the innovation distributional assumption and the drift specification that is estimated. On the other hand, amongst the different parametric GARCH models, the AGARCHX model produces the lowest MSE and MAE. There is evidence that fitting the correct conditional variance specification and using the appropriate innovation distribution give rise to significant improvement in the MSE and MAE. In the case of the linear drift DGP, the improvement in the MSE between a linear drift model with GARCH and AGARCHX specifications is about 16% for Normal distribution, 26% for Student's t and 16% for skewed t distribution. On the other hand, for a nonlinear drift DGP, the improvement is the MSE between a nonlinear drift model with GARCH and AGARCHX specifications is about 5% for Normal distribution, 11% for Student's t and 18% for skewed t distribution. We also observe that fitting an erroneous drift specification tends to increase the MSE and MAE of the in-sample fit.

-Tables 1a and b about here -

Turning to the estimation error of the semiparametric approach for both DGPs indicates that the MSE and MAE are substantially smaller than the parametric models. For the linear drift DGP, between the best fitting AGARCHX model with linear drift and the semiparametric approach with linear drift, the improvement in the MSE(MAE) is about 21%(6%) for both Normal and skewed t distributions, and 20%(6%) for the Student's t distribution. Similarly, for the nonlinear drift DGP, the improvement is about 9%(4%) for Normal distribution, 7%(4%) for Student's t and 12%(3%) for skewed t distribution. It can be inferred, therefore, that while there is gain to be made from using a semiparametric approach over parametric GARCH models in estimating latent volatility, the benefit is more substantial for the case of a short rate model with a linear drift. An interesting observation about the semiparametric approach which contrasts the parametric models is that the MSE and MAE produced by the final smoothed semiparametric approach tend to be very close to each other for the three different innovation distributions, as well as the different drift specifications. This result is interesting as it suggests that the semiparametric approach yields volatility estimates that are robust to the innovation distributional assumption and possible misspecification of the short rate drift function - a feature that is lacking in parametric models.

Last but not least, according to Bühlmann and McNeil (2002) the apparent improvement in the volatility estimates produced by the semiparametric technique should show up in the first four iterations of the smoothing procedure. Indeed we observe that the reduction in the estimation error (relative to a GARCH model) is largest at the first iteration of the procedure. However, this reduction is more substantial in the case of the linear drift model than the nonlinear drift model.

4 Empirical Application

4.1 Data Description

The empirical investigation is based on 1892 weekly observations on the U.S. 3-month Treasury bill rate, sampled from 9 February 1973 to 8 May 2009. The data are obtained from the Federal Reserve Bank of St. Louis (FRED) database. This period comprises a shift from historically high interest rates in the late 1970s to early 1980s during the Volcker monetary regime to low interest rate levels in the latter part of the sample period. The interest rate data and the first differenced series are presented in Figure 2. Summary statistics for the data set are provided in Table 2.

- Figure 2 about here -

From Figure 2 it is clear that there is indeed a tendency for the volatility in the interest rate series to be positively correlated with current interest rate levels. At the start of the sample period, the association between the level of interest rate and its volatility is visible. This feature becomes more apparent for the 1979-1983 period during which both the level and volatility of the rate are high. The level effect is not as obvious after the Volker monetary regime. These empirical features tally with those reported in Brenner et al. (1996). The time-varying nature of the volatility in the sample is indicative that unexpected 'news' might be equally important in explaining the volatility of interest rates, in addition to the level effect.

- Table 2 about here -

The time-varying nature of the volatility that is evident in Figure 2 is associated, in turn, with an empirical distribution for the first differenced data that exhibits excess kurtosis. The relevant kurtosis statistic reported in Table 2 is significantly greater than the value of 3 associated with the normal distribution. The negative skewness coefficient is also significantly less than the value of zero associated with the symmetric normal distribution. This is reflective of a 'leverage' effect of sorts, whereby interest rate falls are associated with higher volatility than increases of the same magnitude. The first differenced data exhibit strong correlation as shown by the Ljung-Box test statistic which overwhelmingly rejects the null hypothesis of no serial correlation at the fifth and tenth lag order. The interest rate series clearly possesses conditional heteroskedasticity as indicated by application of a formal fifth and tenth order LM test for ARCH to the residuals from an AR(10) regression of the interest rate data. The Jarque-Bera test strongly rejects the null of normality in the interest rate series.

The stationarity property of the interest rate data is less clear-cut. There is a lot of controversy in the literature surrounding the unit root property of interest rates. Short-rate diffusion models estimated by Marsh and Rosenfeld (1983), Chan et al. (1992), and Aquila et al. (2003) *inter alia* based on U.S. data document evidence that short-term interest rates behave like a random walk process. In contrast, Brenner et al. (1996) and Ball and Torous (1999) amongst others show supporting evidence that the U.S. short rates mean revert. As it is widely known the standard Dickey–Fuller test is subject to typically moderate size distortion in the presence of a neglected GARCH effect in the series (see Kim and Schmidt, 1993; Haldrup, 1994). To circumvent the problem of neglected GARCH effects in unit root testing, Seo (1999) suggests the unit root test equation and GARCH process should be estimated jointly when the series examined exhibits GARCH effects. We pursue this testing approach to ensure that the unit root test result is robust to the presence of GARCH effects. As is evident from Table 2, the mean level of interest rate is 5.8252. This suggests that the unit root tests should include an intercept in the mean equation.

Seo (1999) augments the standard Dickey-Fuller testing equations as follows:

$$\begin{aligned} \Delta y_t &= \alpha + \beta y_{t-1} + \varepsilon_t \\ \sigma_t^2 &= \phi_0 + \phi_1 \varepsilon_{t-1}^2 + \phi_2 \sigma_{t-1}^2 \\ \varepsilon_t &= \sigma_t v_t \ \ , \ \ v_t \sim N(0,1). \end{aligned} \tag{24}$$

The mean equation in (24) differs slightly from Seo's (1999) approach in which the intercept is excluded. Seo (1999) considers the use of a preliminary regression to demean or detrend the series prior to testing the series for a unit root. Cook (2007), however, presents an approach where the deterministic terms are explicitly included in the testing equation such as in (24). Moreover, he simulates a new set of critical values involving different ϕ_0, ϕ_1 and ϕ_2 parameter values which are more typical in empirical research. The unit root hypothesis is examined via the maximum likelihood t-ratio for β, which is denoted as t_β. Seo (1999) shows that the asymptotic distribution of t_β is a mixture of the non-standard Dickey–Fuller distribution and the standard normal. The extent to which the distribution moves towards the standard normal from the Dickey–Fuller depends upon the strength of the GARCH effect which is determined by a nuisance parameter, ρ. The null hypothesis of a unit root is rejected if t_β is less than the critical value at the conventional significance levels.

In addition to applying the Seo (1999) test, we also perform the augmented Dickey–Fuller (ADF) test and the higher powered GLS-based Dickey–Fuller test (Elliott et al., 1996). The optimal lag length, or degree of augmentation, of the testing equation is determined using the modified Akaike Information Criterion (MAIC) proposed by Ng and Perron (2001) following initial consideration of a maximum lag length given by $int[12(T/100)]^{0.25}$. Hayashi (2000) provides a justification of this upper bound. The appropriate degree of augmentation for both tests is found to be 25. The results obtained from the application of these tests, denoted as τ_μ and τ_μ^{GLS}, are given in Table 2. Using the 5% critical values obtained from Fuller (1996) and Pantula et al. (1994), the derived test statistics, respectively, show the unit root null hypothesis is not rejected by either of the tests. However, the interest rate series clearly possesses conditional heteroskedasticity as indicated by the application of a formal twelfth order LM test for ARCH to the residuals from the ADF test. Given the presence of conditional heteroskedasticity, Seo's (1999) approach outlined above is followed to test the unit root hypothesis. Accordingly, an ADF testing equation with 18 lags is estimated jointly with a GARCH(1,1) process using maximum likelihood estimation and the Bernt-Hall-Hall-Hausman (BHHH) algorithm. The test statistic which uses the Bollerslev and Wooldridge

(1992) standard errors is denoted as $t_\beta(BW)$.[4] We simulate the 5% critical value for the estimated GARCH parameters of $\{\hat{\phi}_1, \hat{\phi}_2\} = \{0.14, 0.85\}$ along with the effective sample size of 1892 observations since neither Seo (1999) nor Cook (2007) studies provide critical values that can be applied to our results.[5] The simulated critical value at the 5% level of significance is -1.9074 for the non-robust standard errors and -1.8891 for the Bolleslev-Wooldridge robust standard errors. The calculated test statistic for t_β and $t_\beta(BW)$ are -2.4301 and -2.5147 respectively. These results imply that the unit root hypothesis can be rejected comfortably in both cases. On the basis that Seo (1999) test incorporates the GARCH effects into the testing framework, we are more inclined to believe in its robust inference, that is the weekly U.S. 3-month Treasury bill interest rates is stationary.

4.2 Empirical Results

The data descriptive statistics indicate that an appropriate model of short rate volatility should account for its time-varying nature, its asymmetric response to shocks of different sign and its dependence on interest rate levels. For this reason, we estimate the GARCHX and AGARCHX models for the diffusion process. As for the drift specification, we estimate both linear and nonlinear drifts to determine the presence of nonlinearities. Given the evidence of unconditional skewness in short rate changes, we also estimate the models with three different distributional assumptions, namely normal, Student's t and skewed t distributions. All the models are estimated with the Bollerslev and Wooldridge (1992) quasi-maximum likelihood method, which gives robust standard errors. The in-sample and out-of-sample volatility forecasts of these parametric models are then compared with the semiparametric model. To produce the one-period ahead out-of-sample volatility forecasts, we exclude the last 100 observations from our sample and estimate the parametric and semiparametric models recursively over the remainder of the data. In other words, each time we produce a one-period ahead volatility forecast we estimate the model using all the data up until the period prior to that forecast. The estimation results for the parametric models with linear and nonlinear drifts are reported in Tables 3(a) and 3(b), respectively.

- Tables 3a and b about here -

It can be seen in Table 3(a) that the coefficients of the linear drift function are only statistically significant at the 5% significance level for the models fitted with a Student's t distribution. The estimate for the coefficient λ which captures the degree of mean reversion is very small implying that the degree of mean reversion is weak. The estimates of the interest rate levels sensitivity parameters, b and δ, the coefficients of last period's unexpected news, α_1, as well as the last period's volatility, α_2, and the coefficient of the asymmetric response of current volatility to last period's bad news, α_3, are found to be highly significant. Taken together, these results suggest that there is overwhelming evidence of GARCH, levels and asymmetric GARCH effects in the diffusion process. In terms of maximised log-likelihood values, the AGARCHX with Student's t distribution performs better than the other models. There is evidence that, independent of the underlying distribution, models that account for both asymmetric GARCH and levels effects perform better than models that do not account for asymmetric GARCH effects. The simple GARCH model performs the worst in terms of the log-likelihood values. This model fails to capture the asymmetry and level-dependence in the short rate volatility process. Moreover, the Ljung-Box test of the twelfth order serial correlation in the squared standardised residuals rejects the null of no serial correlation implying that the GARCH model does not adequately characterise the volatility dynamic of short rate changes. The skewness parameter η of the skewed t distribution turns out to be statistically insignificant at all conventional significance levels. Furthermore, the η estimate for the three short rate models is virtually zero, implying that a Student's t distribution is adequate in characterising the short rate distribution. Our finding which supports the use of Student's t instead of skewed t distribution is consistent with the results of Bali (2007).

In Table 3(b) we show the estimation results for nonlinear drift short rate models. Irrespective of the distributional assumption of the error, we find that the coefficients λ_2 and

λ_3 which govern the nonlinear dynamics in the drift function are statistically insignificant. Our results, which support the lack of evidence of nonlinearity in the 3-month T-bill data, concur with earlier findings by Bali (2007) who shows that the incorporation of the GARCH effects into the volatility process gives rise to no evidence of nonlinearity in the drift specification. The skewness parameter η of the skewed t distribution, again turns out to be statistically insignificant for all models, implying the lack of evidence for skewness asymmetry in the short rate changes distribution. Comparing models with linear and nonlinear drifts across similar distributional assumption indicates a substantial reduction in the log-likelihood value, thereby suggesting that a short rate model with linear drift is the preferred specification. Based on this result, we do not consider the in-sample and out-of-sample forecast performance of short rate models with nonlinear drift and skewed t distribution.

We use four different metrics to evaluate the in-sample and out-of-sample volatility forecast performance of the semiparametric approach compared to its parametric conterparts. In addition to the MAE and MSE measures given in equation (23), we also use the Akaike Information Criterion (AIC) which is a penalised negative log-likelihood criterion adjusted for the degree of parameters that are estimated, and the R^2_{vol} measure of Bali (2003). For the four metrics, we proxy the unobserved true volatility σ_t with $|r_t - r_{t-1}|$. The AIC is computed as:

$$AIC = 2K + T\left[\ln\left(\frac{2\pi \cdot RSS}{T}\right) + 1\right] \tag{25}$$

where K is the number of parameters that are estimated, T is the sample size and $RSS = \sum_{t=1}^{T}(\sigma_t - \hat{\sigma}_t)^2$. The R^2_{vol} measure essentially computes the total variation in the true volatility proxied by $|r_t - r_{t-1}|$ that can be explained by the estimated conditional volatilities. This is obtained from the coefficient of determination of an OLS regression of the form:

$$\sigma_t = a_0 + a_1\sigma_t^f + e_t \tag{26}$$

where σ_t and σ_t^f are the actual and forecasted volatility of $(r_t - r_{t-1})$, respectively. It should be highlighted that the R^2_{vol} measure is a crude measure and is subject to the following caveat. As pointed out Andersen and Bollerslev (1998), the idiosyncratic component of daily interest rate changes is large, thus the use of realised interest rate changes may not fully capture day-by-day movements in volatility. To circumvent this problem, we use a range based volatility proxy by adopting the Garman and Klass (1980) extreme value estimator to construct a minimum variance unbiased estimator that utilises the opening, closing, high and low prices. Due to the paucity of high frequency data, the use of the Garman and Klass (1980) (GK hereafter) extreme value estimator is deemed as a compromise to the preferred realised volatility measure derived from high frequency data (see Andersen, Bollerslev, Diebold and Labys, 2001).[6] Our choice of the GK estimator is also motivated by the findings of Bali and Weinbaum (2005) who perform a horse race among all the extreme-value estimators which have featured in the literature. They show that, in practice, the GK estimator is the least biased and most efficient estimator compared with other extreme value estimators. The GK minimum variance and unbiased estimator is $\widehat{\sigma}^2_{GK} = \frac{1}{n}\sum_{t=1}^{n}\left(0.511\left(\ln\frac{H_t}{L_t}\right)^2 - 0.019\left[\ln\left(\frac{C_t}{O_t}\right)\ln\left(\frac{H_tL_t}{O_t}\right) - 2\ln\left(\frac{H_t}{O_t}\right)\ln\left(\frac{L_t}{O_t}\right)\right] - 0.383\left[\ln\left(\frac{C_t}{O_t}\right)\right]^2\right)$, $n \geqslant 1$ where O_t, C_t, H_t and L_t denote, respectively, the opening, closing, high and low prices on day t and n is the number of days in the sample. The IRX index data are obtained from the Yahoo/finance web site.[7]

- Tables 4a and b about here -

Tables 4(a) and (b) report the results of the four metrics for evaluating the in-sample forecast performance of the models using the volatility benchmarks $|r_t - r_{t-1}|$ and $\widehat{\sigma}^2_{GK}$, respectively. Focusing on the results with volatility proxy $|r_t - r_{t-1}|$, we find that the AGARCHX model with Student's t distribution performs the best compared with other parametric models. Not only does it deliver the lowest MSE and MAE, it also gives the lowest AIC and highest R^2_{vol}. The GARCH model, which does not take into account the levels dependence and asymmetric response in the conditional variance of short rate changes, performs the worst. However, there is evidence that the semiparametric model yields the most superior in-sample volatility forecast performance as judged by the four metrics. When compared with the best fitting AGARCHX Student's t model, the reduction in the MSE and MAE based on the final smoothed semiparametric method is about 4% and 1%, respectively. The AIC shows a marked improvement in the goodness of fit falling from -498.52 to -601.80, while the R^2_{vol} increases by about 11%. Looking at the four metrics, we also find that in each of the iteration of the semiparametric smoothing procedure there is a significant improvement in the volatility forecast performance compared with the AGARCHX Student's t model. Interestingly, we find that for the semiparametric approach, fitting a nonlinear drift function erroneously to obtain an initial volatility estimate does not give rise to an inferior forecast performance. The difference in forecast performance results for the linear and nonlinear drifts semiparametric approach is negligible thus implying that the choice of the drift function is immaterial to the forecast performance of the semiparametric approach. This result corroborates the simulation results in which we find that neglecting to fit the correct drift function in the semiparametric approach does not bear any influence on its volatility forecast performance. Another important finding is that the choice of the innovation distribution, whether it is normal, Student's t or skewed t, does not have a considerable impact on the forecast performance of the semiparametric approach.[8] Taken together, these results highlight the robust forecast property of the semiparametric approach to possible misspecifications of the drift function and the innovation distribution. In Table 4(b), we show that these results are qualitatively unchanged even with the use of a more accurate volatility benchmark (i.e. $\widehat{\sigma}^2_{GK}$) to assess the forecast performance of the semiparametric approach relative to parametric models.

- Figure 3a and b about here -

Given the extensive results reported in Tables 4(a) and (b), we summarise these findings by presenting them in Figures 3(a) and (b). To interpret the plot, the four shaded bars represent the metric value of the parametric models: AGARCHX-T, AGARCHX-N, GARCHX-T and GARCHX-N in that order. The line that is plotted across the x-axis is a locus of the metric value for the GARCH model (represented by the first mark on the x-axis), the metric value for the eight iterations of the semiparametric approach (represented by the second to ninth marks on the x-axis) and the final smoothed stage of the semiparametric approach (represented by the tenth mark on the x-axis). It is evident from the plot that the semiparametric approach yields the best results based on all four forecast performance measures. The results are consistent whether we use the crude or more accurate volatility benchmark. To visually illustrate the superior performance of the semiparametric approach compared with the AGARCHX Student's t model, we plot in Figures 4(a) to (d) the in-sample volatility estimates of the two models for an arbitrarily selected period 1997/01/01 - 2000/01/01. Figures 4(a) and (b) employ $|r_t - r_{t-1}|$ as the true volatility proxy while Figures 4(c) and (d) are based on the more accurate volatility proxy given by $\widehat{\sigma}^2_{GK}$.

- Figures 4a, b, c and d about here -

By comparing the plots of the volatility estimates between the best fitting parametric AGARCHX Student's t model and the semiparametric model, we can see that the latter model is capable of capturing movements of the short rate volatility process better than the former model. The AGARCHX Student's t model tends to yield an overly smoothed volatility estimates of the true volatility process that is proxied by $|r_t - r_{t-1}|$ ($\widehat{\sigma}^2_{GK}$) in Figure 4b (Figure 4d) . There are two important features about the way the volatility estimates obtained from the semiparametric approach improves upon the estimates of the AGARCHX Student's t model. First, there are peaks or spikes in the volatility of short rate changes which are well captured by the semiparametric model but not by the AGARCHX Student's t model. For example, the peak that is observed on 1998/01/01 is clearly captured by the semiparametric approach but not by the AGARCHX Student's t model. Second, the volatility estimates produced by the semiparametric approach tend to match the rise and fall in interest rates better than the AGARCHX Student's t model. The most obvious of this point is the drop in interest rates between the two peaks that happened prior to 1999/01/01 (see Figure 4a). While the semiparametric approach does not fully capture the drop in rates, nonetheless it does a better job at capturing the fall in interest rates than the AGARCHX Student's t model.

- Tables 5a and b about here -

- Figures 5a, b, c and d about here -

Turning to the out-of-sample volatility forecast performance of the semiparametric model, we find that the volatility estimates obtained from the final smoothed stage have better predictive power than those produced by the parametric models. Using the volatility benchmark $\widehat{\sigma}^2_{GK}$ (see Table 5b), the improvement in the volatility forecast estimation error measured by the MSE and the MAE is 14% and 11%, respectively between the final smoothed semiparametric approach and the AGARCHX Student's t model. On the other hand, the reduction in the forecast estimation error based on the crude volatility proxy $|r_t - r_{t-1}|$ is more conservative; the MSE and MAE fall by 6% and 5% respectively (see Table 5a). Figures 5(a) to (d) provide plots of the volatility forecast estimates of the two contending models. Unlike the in-sample volatility estimates, we fail to find that the semiparametric approach is capable of capturing the observed peaks in interest rate volatility, particularly with the volatility benchmark $\widehat{\sigma}^2_{GK}$ and the sharp spike at the start of the forecast horizon (see Figure 5c). However, this does not diminish the out-of-sample forecast performance of the semiparametric approach compared with the AGARCHX Student's t model. The latter model continues to provide an overly smoothed out-of-sample volatility forecast of interest rates. In contrast, the semiparametric approach yields volatility forecasts that better capture fluctuations in the short rate thus leading to smaller estimation error than the AGARCHX Student's t model.

4.3 Implications for Pricing Interest Rate Derivatives

Given that the volatility processes of the semiparametric and parametric models are distinct, it is very likely that the two classes of models will generate different probability distribution of future interest rate levels. Predictions of future interest rates are essential for pricing long-dated, path-dependent interest rate derivatives such as the index amortizing rate (IAR) swaps, amongst others. For the purpose of illustration, we consider the IAR swaps. The notional value of the IAR swaps is reduced over time according to an amortization schedule based on the level of a reference interest rate on a particular fixed date in the future (usually every three or six months). The value of this swap is contingent on the probability distribution of the reference rate on each reset date. Since the amount of principal that remains on any reset date depends on past interest rate levels, the IAR swaps are considered as "path-dependent" securities. In other words, fluctuations in interest rates and hence the accuracy in modelling short rate volatility matters for the pricing of the IAR swaps. For a detailed discussion of the IAR swaps refer to Galaif (1993).

To examine how an improvement in the estimation accuracy of short rate volatility could affect the pricing of interest rate derivatives, we follow Brenner et al. (1996) and perform the following experiment. We simulate the semiparametric model and the parametric models 5000 times using the 3-month Treasury bill rates estimation results with June 8, 2007 as the starting date. The interest rate level is 4.67% on this date. Following Brenner et al (1996), we focus on the volatility process and employ the mean equation $r_t - r_{t-1} = -0.0015$ given that the average weekly change in the interest rate over the estimated sample period is -0.0015.[9] Figure 6 graphs the 5th, 25th, 50th, 75th and 95th percentiles of the 5000 simulated paths for each horizon up to 100 weeks for the different short rate models. The solid lines represent the confidence intervals for simulated interest rates based on the parametric models. The ordering from the outermost to innermost lines represents the resulting interest rate distributions for the AGARCHX-T, GARCHX-T, AGARCXH-N and GARCHX-N models.

The dotted lines denote the short rate distribution of the semiparametric model.

- Figure 6 about here -

Visual inspection of Figure 6 suggests that there are several interesting results. First, the distribution assumption in the parametric models does not seem to matter for derivative prices. The interest rate distributions are very similar when comparing between the same type of model with different distribution assumptions. Secondly, like Brenner et al. (1996) we find that whether we model asymmetries in the parametric models or not is immaterial for the paths of future interest rates and therefore this will not greatly affect interest rate derivative prices. Thirdly, amongst the different models considered, the confidence intervals of future short rate levels generated by the semiparametric model are narrower, particularly at the 5% and 95% levels. In other words, the semiparametric model predicts a narrower confidence band of extreme interest rate movements than the parametric models. For other confidence levels considered, we find that the future levels of short-term interest rates are comparable with the parametric models.

Based on these results, what can be said about the pricing of certain path-dependent interest rate derivative such as the IAR swaps mentioned above, mortgages and collateralised mortgage obligation? Given that parametric models produce larger upper tails, the average predicted amortization will be less for such models than the semiparametric model. In other words, the predicted lives of these securities and their cash flows will increase. Accordingly these securities would be overpriced for the parametric models relative to the semiparametric model. On the other hand, the larger lower tails of the parametric models would imply that these securities would be underpriced compared to the semiparametric model. Our results for the parametric models are consistent with that of Brenner et al. (1996) who find that the conditional variance model specification does not have an influence on the pricing of interest rate derivatives. In particular, they show that whether a model specifies an asymmetric conditional variance or an additive or multiplicative levels effect in the variance specification does not yield significant differences in the pricing of interest rate derivatives. Likewise, we demonstrate that the asymmetric specification of the diffusion process and the distributional assumption for parametric models do not affect the pricing of interest rate derivatives. More importantly, we find that the narrower confidence intervals of future interest rate levels produced by the semiparametric model relative to any of the parametric models suggests that our method would yield less price variation for long-dated and path-dependent interest rate derivatives.

Although the semiparametric model does not give rise to simple analytical solution for the pricing of derivatives, the estimation process sets up naturally for Monte Carlo evaluation. Thus, like the BHK models, the semiparametric model can be easily applied to the valuation of securities that already require Monte Carlo evaluation. These securities include those interest rate derivatives discussed above.

5 Conclusion

In this paper an application of a semiparametric GARCH approach to modelling short-term interest rate volatility has been proposed. The semiparametric smoothing technique uses a general additive function of lagged innovations, volatilities and past interest rate levels, and a backfitting algorithm to estimate the unobserved diffusion process. While the volatility model is estimated semiparametrically, it resembles the widely used short-rate volatility models of Brenner et al. (1996) which features interest rate levels dependence and asymmetric dynamic in the conditional variance. Consequently, we compare the performance of the semiparametric approach with this class of single-factor short rate diffusion models in terms of its ability to characterise short rate volatility. Our simulation study shows that the semiparametric model provides a superior fit of the in-sample volatility estimates to a GARCH model that exhibits asymmetry and levels effect. The volatility forecast performance of the semiparametric procedure, unlike the parametric GARCH models, is also robust to potential misspecification in the short rate drift and the innovation distribution. The empirical application to weekly U.S. 3-month Treasury bill rates between 1971 and 2009 further illustrates improvement in the in-sample and out-of-sample predictive power of the semiparametric model over models of Brenner et al. (1996). Finally, we show that the greater degree of accuracy in modelling short rate volatility offered by the semiparametric model is important for pricing long-dated and path-dependent interest rate derivatives.

For future research, we intend to apply this technique to the two-factor model of Black, Derman and Toy (1990) with stochastic volatility which was developed and estimated by Bali (2003). The two factors of the model are the short-term interest rate and the volatility of interest rate changes. This would involve performing a nonparametric estimation on both the drift and diffusion of the short rate process. The application of this technique to the two-factor arbitrage free model could be used to assess the importance of modelling short rate changes volatility accurately and its implications on default-free bond pricing.

References

[1] Aït Sahalia, Y. (1996a), 'Testing continuous-time models of the spot interest rates', *Review of Financial Studies,* 9, 385-426.

[2] Aït Sahalia, Y. (1996b), 'Nonparametric pricing of interest rate derivatives', *Econometrica,* 64, 527-560.

[3] Andersen, T.G. and Bollerslev, T. (1998), 'Answering the skeptics: Yes, standard volatility models do provide accurate forecasts', *International Economic Review*, 39, 885-905.

[4] Andersen, T.G., Bollerslev, T., Diebold, F.X. and Labys, P. (2001), 'The distribution of exchange rate volatility', *Journal of the American Statistical Association*, 96, 42-55.

[5] Aquila, R.D., Ronchetti, E. and Trojani, F. (2003), 'Robust GMM analysis of models for the short rate process', *Journal of Empirical Finance,* 10, 373-397.

[6] Ball, C.A. and Torous, W.N. (1999), 'The stochastic volatility of short-term interest rates: Some international evidence', *Journal of Finance,* 54, 2339-2359.

[7] Bali, T.G. (2000), 'Testing the empirical performance of stochastic volatility models of the short term interest rate', *Journal of Financial and Quantitative Analysis*, 35, 191-215.

[8] Bali, T.G. (2003), 'Modeling the stochastic behaviour of short-term interest rates: Pricing implications for discount bonds', *Journal of Banking and Finance*, 27, 201-228.

[9] Bali, T.G. (2007), 'Modeling the dynamics of interest rate volatility with skewed fat-tailed distributions', *Annals of Operations Research*, 151, 151-178.

[10] Bali, T.G. and Weinbaum, D. (2005), 'A comparative study of alternative extreme-value volatility estimators', *Journal of Futures Market*, 25, 873-892.

[11] Bali, T. and Wu, L. (2006), 'A comprehensive analysis of the short-term interest rate dynamics', *Journal of Banking and Finance*, 30(4), 1269-1290.

[12] Brenner, R.J., Harjes, R.H. and Kroner, K.F. (1996), 'Another look at models of the short-term interest rate', *Journal of Financial and Quantitative Analysis*, 31, 85-107.

[13] Black, F., Derman, E. and Toy, W. (1990), 'A one factor model of interest rates and its applications to treasury bond options', *Financial Analysts Journal*, 46, 33-39.

[14] Bollerslev, T. and Wooldridge, J. (1992), 'Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances', *Econometric Review*, 11, 143-172.

[15] Bühlmann, P. and McNeil, A. (2002), 'An algorithm for nonparametric GARCH modelling', *Computational Statistics and Data Analysis*, 40, 665-683.

[16] Chapman, D, and Pearson, N. (2000), 'Is the short rate drift actually nonlinear?', *Journal of Finance*, 55, 355-388.

[17] Chan, K.C., Karolyi, G.A., Longstaff, F.A. and Sanders, A.B. (1992), 'An empirical comparison of alternative models of the short-term interest rate', *Journal of Finance*, 47, 1209-1227.

[18] Conley, T.G., Hansen, L.P., Luttmer, E.G.Z., and Scheinkman, J.A. (1997), 'Short-term interest rates as subordinated diffusions', *Review of Financial Studies*, 10, 525-577.

[19] Cook, S. (2007), 'Joint maximum likelihood estimation of unit root testing equations and GARCH processes: Some finite-sample issues', *Mathematics and Computers in Simulation*, in press.

[20] Cox, J.C., Ingersoll, J.E., and Ross, S. (1985), 'A theory of the term structure of interest rates', *Econometrica*, 53, 385-407.

[21] Czellar, V., Karolyi, G. A., and Ronchetti, E. (2007), 'Indirect robust estimation of the short-term interest rate process', *Journal of Empirical Finance*, 14(4), 546-563.

[22] Durham, G.B. (2004), 'Likelihood-based specification analysis of continuous time models of the short-term interest rate', *Journal of Financial Econics*, 70, 463-487.

[23] Elliott, G., Rothenberg, T. and Stock, J. (1996), 'Efficient tests for an autoregressive unit root', *Econometrica*, 64, 813-836.

[24] Friedman, J.H. and Stuetzle, W. (1981), 'Projection pursuit regression', *Journal of the American Statistical Association*, 76, 817-823.

[25] Fuller, W. (1996), An Introduction Statistical Time Series, second edition, Wiley, New Jersey.

[26] Galaif, L.N. (1993), 'Index Amortizing Rate Swaps', Quarterly Review of the Federal Reserve Bank of New York.

[27] Garman, M.B. and Klass, M. (1980), 'On the estimation of security price volatilities from historical data', *Jounral of Business*, 53, 67-78.

[28] Gospodinov, N. (2005), 'Testing for threshold nonlinearity in short-term interest rates' *Journal of Financial Econometrics*, 3, 344-371.

[29] Haldrup, N. (1994), 'Heteroskedasticity in non-stationary time series: Monte Carlo evidence', *Statistical Papers*, 35, 287-307.

[30] Hansen, B. (1994), 'Autoregressive conditional density estimation', *International Economic Review*, 35, 705-730.

[31] Hastie, T. and Tibshirani, R. (1990), *Generalized Additive Models*, Chapman & Hall, New York.

[32] Hayashi, F. (2000), *Econometrics*, Princeton University Press, Princeton.

[33] Jiang, G.J. (1998), 'Nonparametric modeling of U.S. interest rate term structure dynamics and implications on the prices of derivative securities', *Journal of Financial and Quantitative Analysis*, 33, 465-497.

[34] Jiang, G. and Knight, J. (1997), 'A nonparametric approach to the estimation of diffusion processes with an application to a short-term interest rate model', *Econometric Theory*, 13, 615-645.

[35] Jones, C.S. (2003), 'Nonlinear mean reversion in the short-term interest rate', *Review of Financial Studies*, 16, 793-843.

[36] Kim, K. and Schmidt, P. (1993), 'Unit root tests with conditional heteroskedasticity', *Jounal of Econometrics*, 287-300.

[37] Marsh, T. and Rosenfeld, E. (1983), 'Stochastic processes for interest rates and equilibrium bond prices', *Journal of Finance*, 38, 635-646.

[38] Ng, S. and Perron, P. (2001), 'Lag length selection and the construction of unit root tests with good size and power', *Econometrica*, 69, 1519-1554.

[39] Nowman, K. B. and Sorwar, G. (2005), 'Derivative prices from interest rate models: Results for Canada, Hong Kong, and United States', *International Review of Financial Analysis*, 14(4), 428-438.

[40] Pantula, S. (1989), 'Estimation of autoregressive models with ARCH errors', *Sankhya B*, 50, 119-138.

[41] Pantula, S.G., Gonzales-Farias, G., and Fuller, W.A. (1994), 'A Comparison of Unit-Root Test Criteria', *Journal of Business and Economic Statistics,* 12, 449-459.

[42] Pritsker, M. (1998), 'Nonparametric density estimation and tests of continuous time interest rate models', *Review of Financial Studies*, 11, 449-487.

[43] Seo, B. (1999), 'Distribution theory for unit root tests with conditional heteroskedasticity', *Journal of Econometrics*, 91, 113-144.

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

[45] Theodossiou, P. (1998), 'Financial data and the skewed generalised t distribution', *Management Science*, 44, 1650-1661.

[46] Vasicek, O. (1977), 'An equilibrium characterization of the term structure', *Journal of Financial Economics*, 5, 177-188.

Appendix

To simulate the critical values for the Seo (1999) test, the following data generating process (DGP) is employed:

$$\begin{aligned} y_t &= y_{t-1} + \varepsilon_t, \;\; t = 1, ..., T \\ \sigma_t^2 &= \phi_0 + \phi_1 \varepsilon_{t-1}^2 + \phi_2 \sigma_{t-1}^2 \\ \varepsilon_t &= \sigma_t v_t \;\;,\;\; v_t \sim N(0,1). \end{aligned} \tag{27}$$

We set the parameters $\phi_1 = 0.14$, $\phi_2 = 0.85$ and $\phi_0 = 1 - \phi_1 - \phi_2$. These values are taken from estimates of our ADF testing equation with 18 lags which is estimated jointly with a GARCH(1,1) process. T is set to 1892 to match our sample size. Once the data are simulated, we perform Seo (1999) test by estimating:

$$\begin{aligned} \Delta y_t &= \alpha + \beta y_{t-1} + \varepsilon_t \\ \sigma_t^2 &= \phi_0 + \phi_1 \varepsilon_{t-1}^2 + \phi_2 \sigma_{t-1}^2 \\ \varepsilon_t &= \sigma_t v_t \;\;,\;\; v_t \sim N(0,1). \end{aligned} \tag{28}$$

with the maximum likelihood method using the BHHH algorithm. The resulting t-test for the null hypothesis of a unit root process in y_t (i.e. $\beta = 0$) which is denoted as t_β is computed. In addition, we compute the robust t-test, $t_\beta(BW)$, using the Bollerslev and Woodridge (1992) robust standard error. The experiment is repeated 25,000 times and each time the test statistic values for t_β and $t_\beta(BW)$ are saved. The resulting series of $\hat{t}_\beta$ and $\hat{t}_\beta(BW)$ are sorted and the 1%, 5% and 10% critical values are obtained accordingly. The critical values at the 1%, 5% and 10% significance levels for t_β are -2.4280,-1.9073 and -1.6701, and for $t_\beta(BW)$ are -2.4196, -1.8891 and -1.6454, respectively.

Notes

[1]These interest rate derivatives include index amortizing rate swaps, CMO swaps, swaptions, mortgages and adjustable rate preferred securities.

[2]BHK also consider the multiplicative levels effect in which ϕ^2 in $E\left(\varepsilon_t^2|\Omega_{t-1}\right) \equiv h_t = \phi^2 r_{t-1}^{2\delta}$ follows a GARCH(1,1) process.

[3]For a discussion on the backfitting algorithm refer to Friedman and Stutzle (1981) and Hastie and Tibshirani (1986).

[4]Cook (2007) shows that the maximum likelihood estimation of the Seo (1999) unit root test equations could employ the Bollerslev and Wooldridge (1992) robust standard errors. The t-test statistic for the slope coefficient β with robust standard error is given by $t_\beta(BW)$.

[5]The Appendix provides details of the simulation to obtain 1%, 5% and 10% critical values for t_β and $t_\beta(BW)$.

[6]Implied volatility can be obtained from the value of the 13-week Treasury index (IRX) which is based on the discount rate of the most recently auctioned 13-week U.S. T-bill. However, high frequency IRX data are only available from November 3, 1997. Our sample period, on the other hand, commences from February 9, 1973.

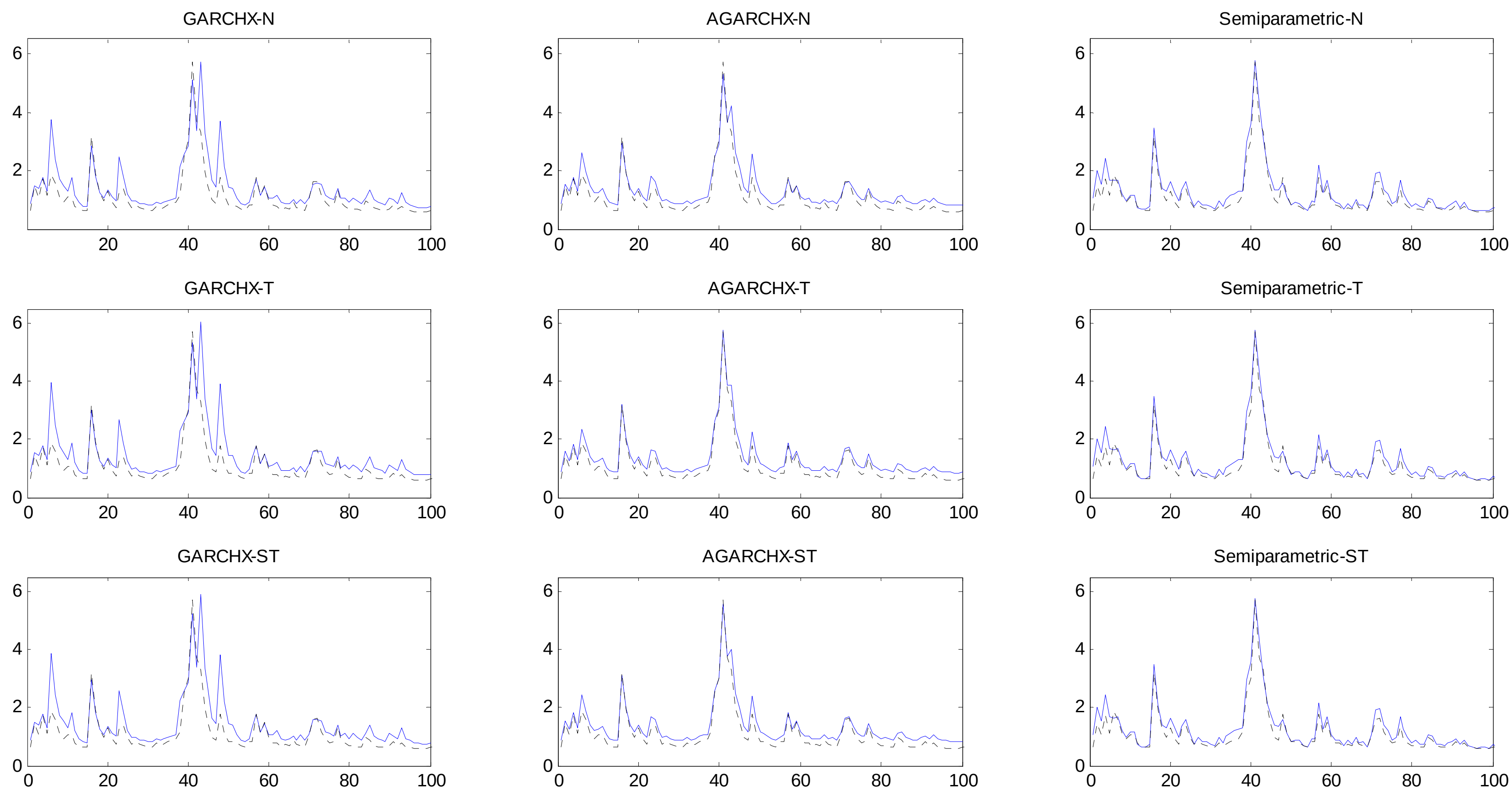
[7]The URL for the IRX data is http://finance.yahoo.com/q/hp?s=%5EIRX+Historical+Prices.

We thank the referee for directing us to this data source.

[8]To conserve space, we do not report the results for the semiparametric approach with a skewed t distribution. These results are available from the authors upon request.

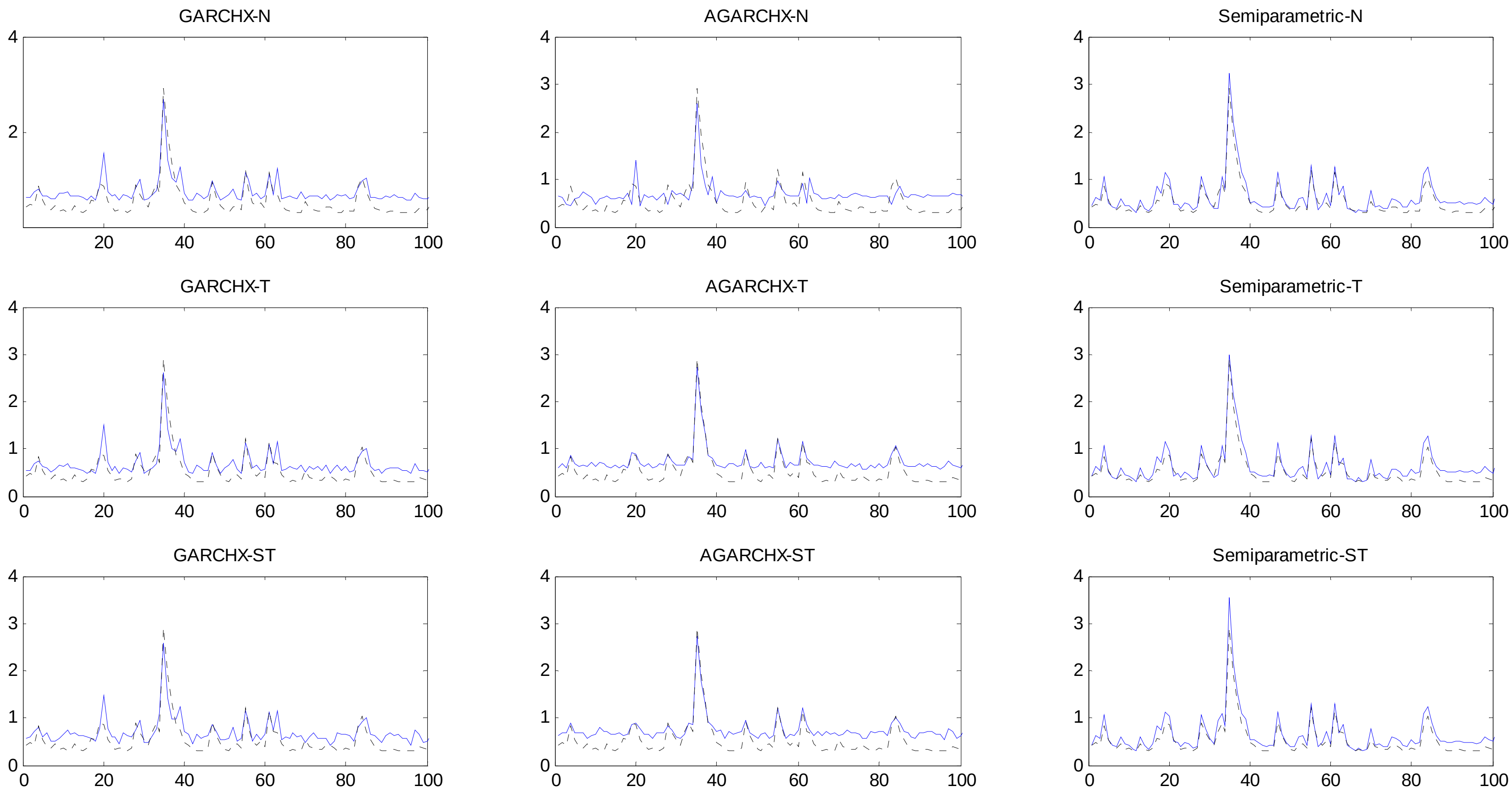
[9]Although the mean-reverting slope coefficient is significant, the coefficient estimate is very small and it is virtually close to zero. For this reason, ignoring the mean-reverting dynamics in the simulation is a reasonable assumption.

Figure 1 (a) Plots of Volatility Estimates of Various Models for Simulated Data with Linear Drift



Note:The dotted line represents simulated true volatility, while the solid line represents the estimated volatility derived from estimating a model with a linear drift. The estimated volatility for the semiparametric approach is for the final smoothed volatility. N, T and ST denote Normal, Student's t and Skewed t distributions, respectively.

Figure 1 (b) Plots of Volatility Estimates of Various Models for Simulated Data with Nonlinear Drift



Note:The dotted line represents simulated true volatility, while the solid line represents the estimated volatility derived from estimating a model with a nonlinear drift. The estimated volatility for the semiparametric approach is for the final smoothed volatility. N, T and ST denote Normal, Student's t and Skewed t distributions, respectively.

Figure 2 The U.S. Short Rates: Levels and First Differences

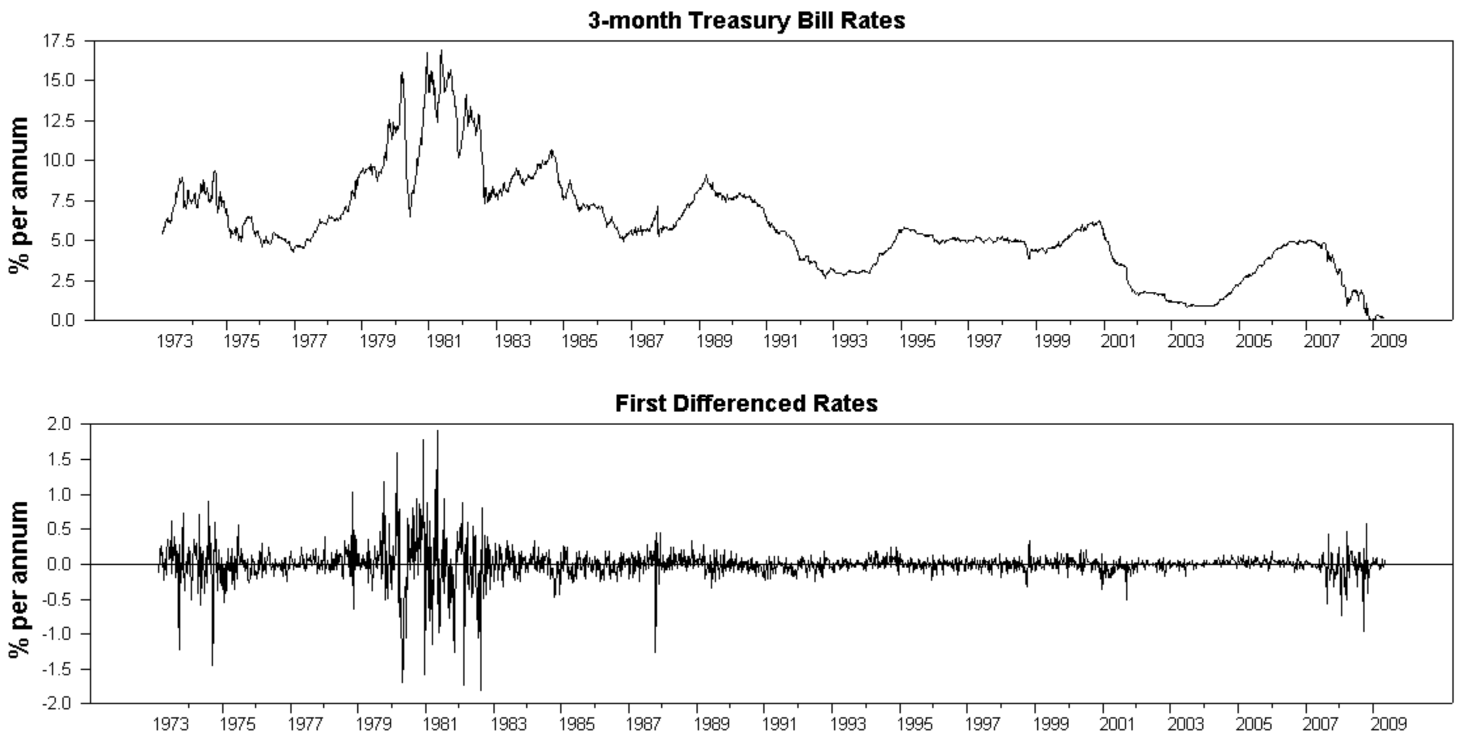
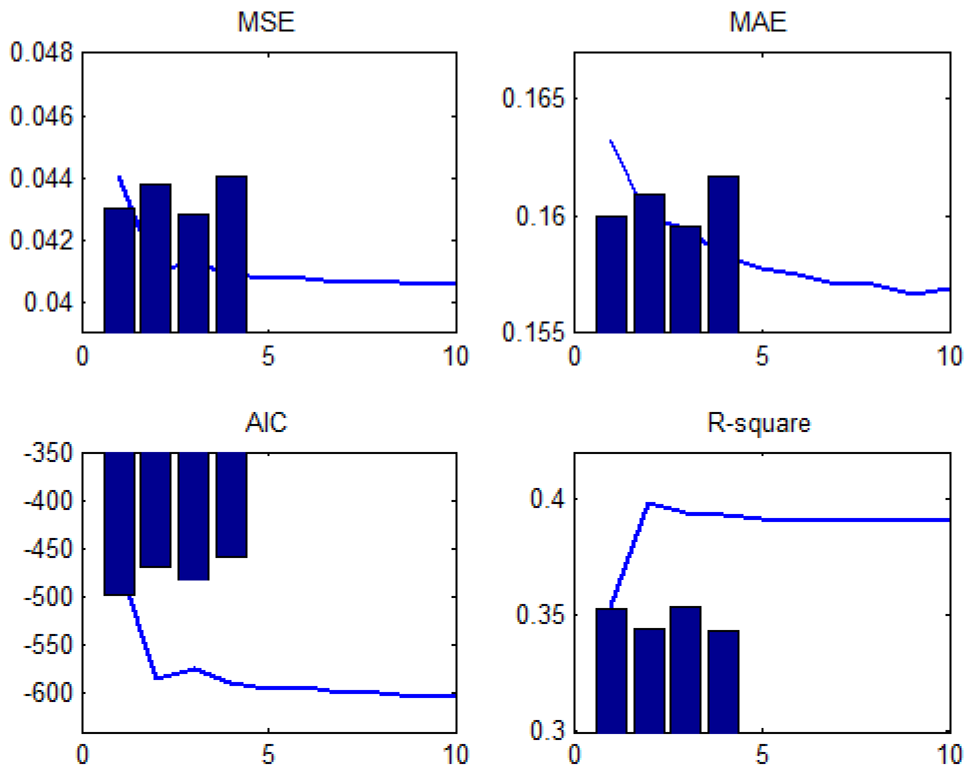
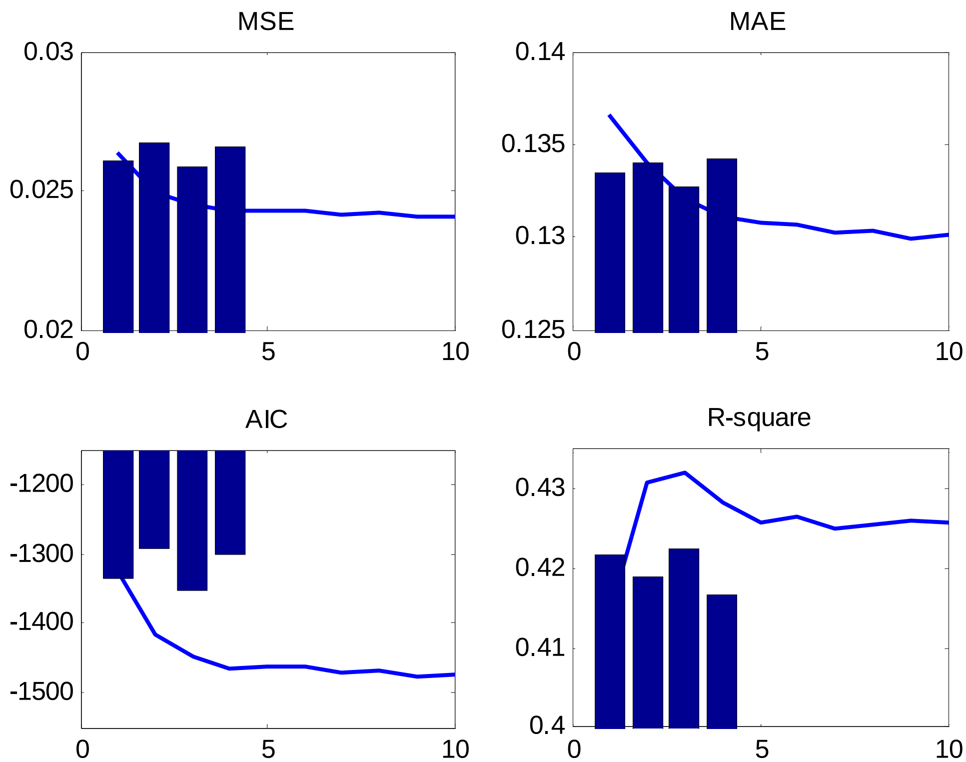


Figure 3 **Plots of MSE, MAE, AIC and R^2_{vol} for the In-Sample Volatility Forecast Performance of the Parametric and Semiparametric Short Rate Models**

(a) Volatility benchmark $|r_t - r_{t-1}|$



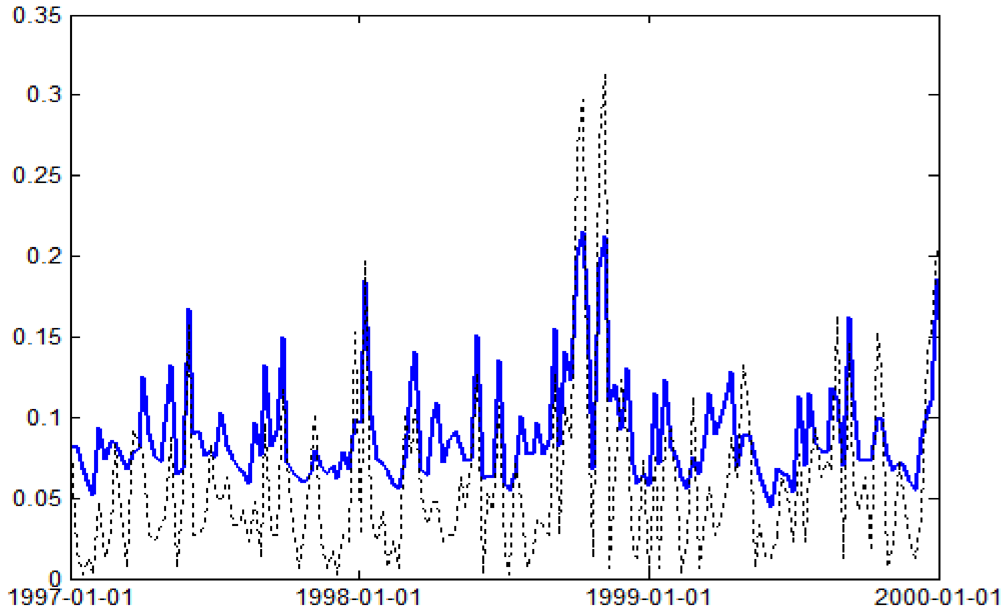
(b) Volatility benchmark $\hat{\sigma}^2_{GK}$



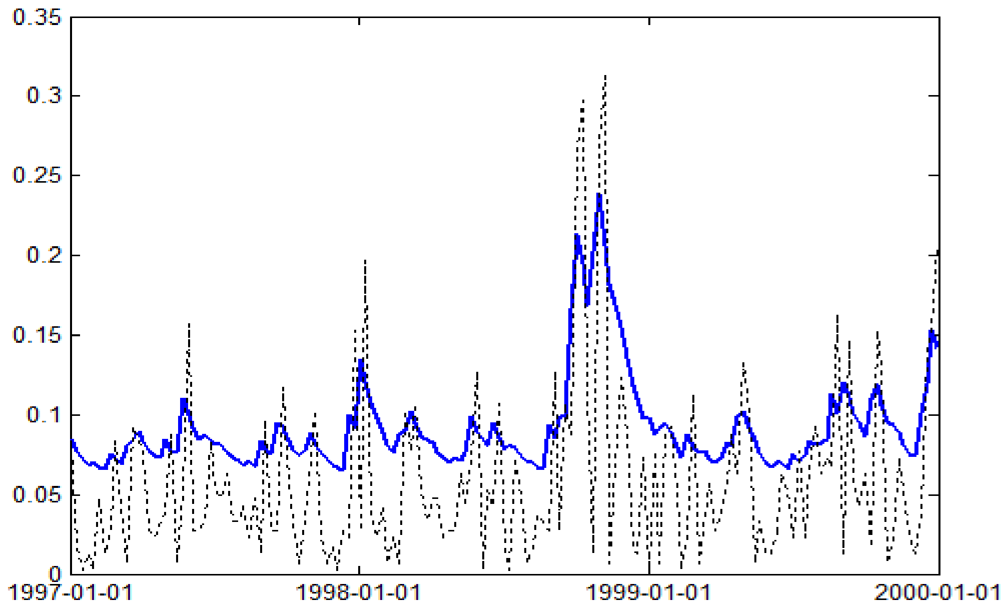
Note: The shaded bars represent the metric value for the parametric models in the following order: AGARCHX-T, AGARCHX-N, GARCHX-T, GARCHX-N. The 1 to 10 marks on the x-axis are to be interpreted in the following way. The first mark represents the metric value for the parametric GARCH model. The second to ninth marks represent the metric values for the eight iterations that are performed in the semiparametric procedure, while the tenth mark denotes the metric value for the final smoothed stage. R-square denotes R^2_{vol} . The results are for the sample period 1973/02/09 - 2007/06/08.

Figure 4 Plots of In-Sample Volatility Estimates for the U.S. Short Rates over the Sample Period 1997/01/01 - 2000/01/01

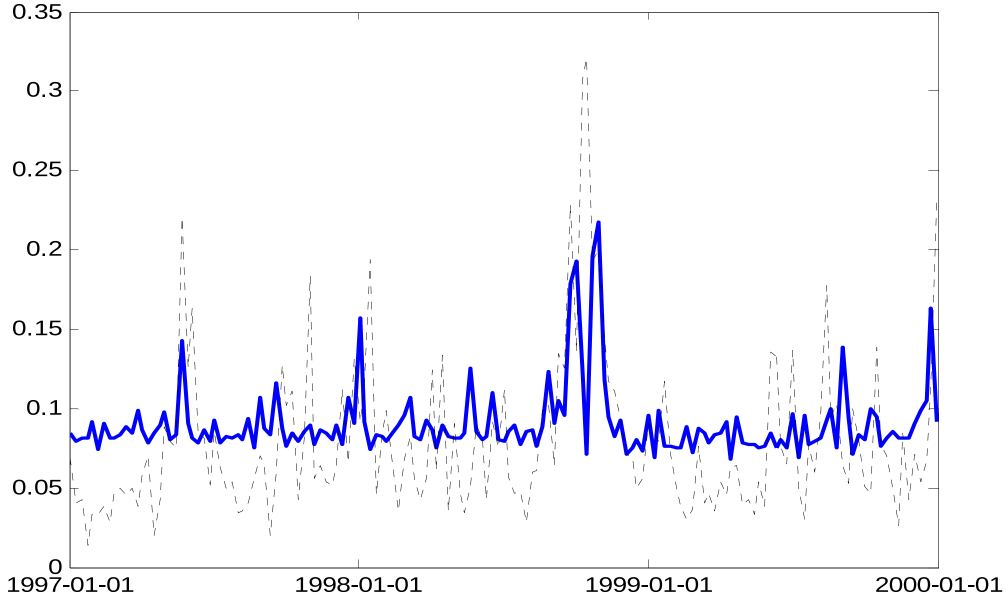
(a) Semiparametric approach



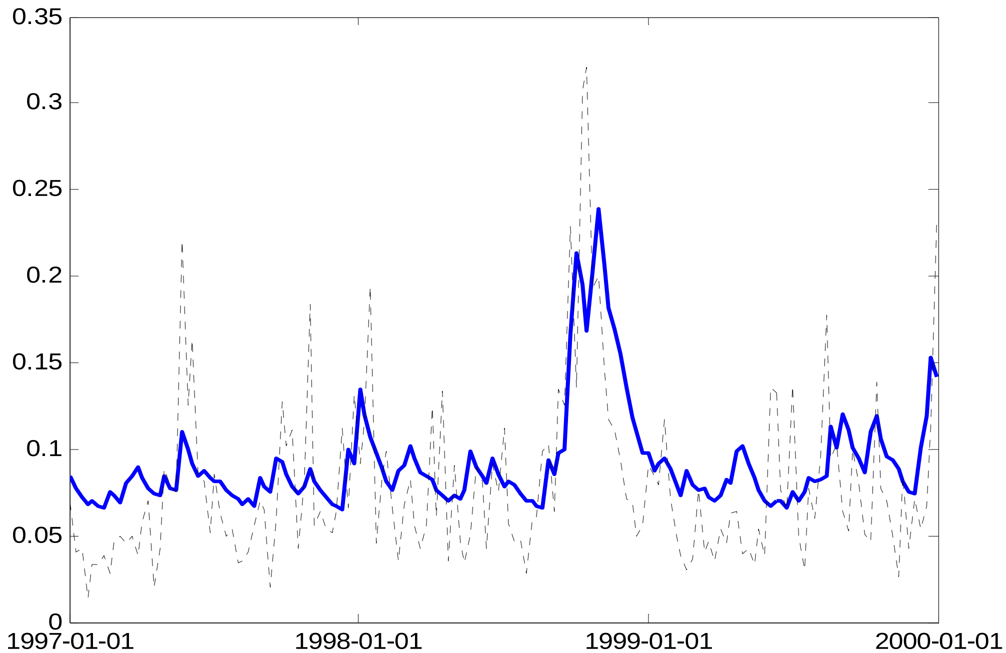
(b) AGARCHX-T model



(c) Semiparametric approach



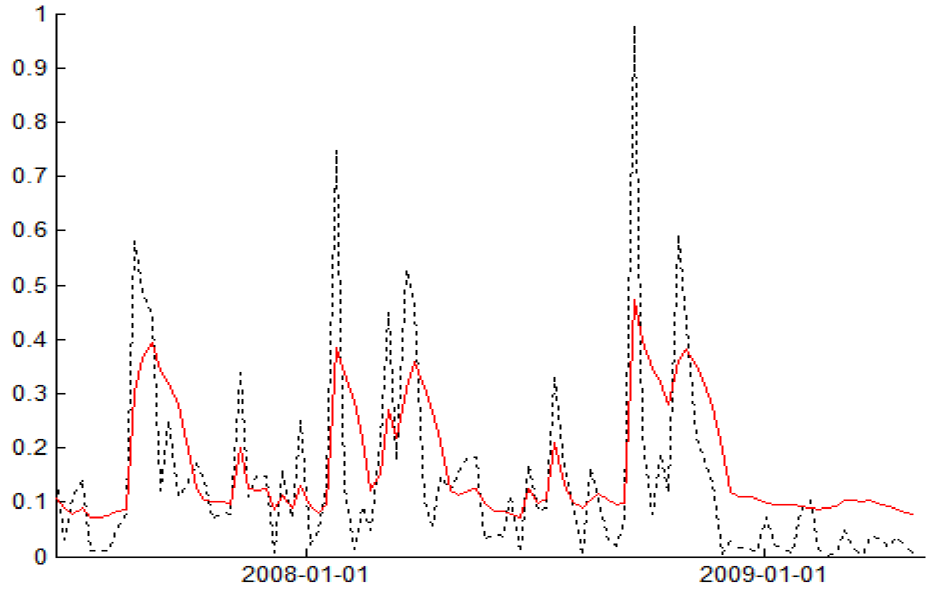
(d) AGARCHX-T model



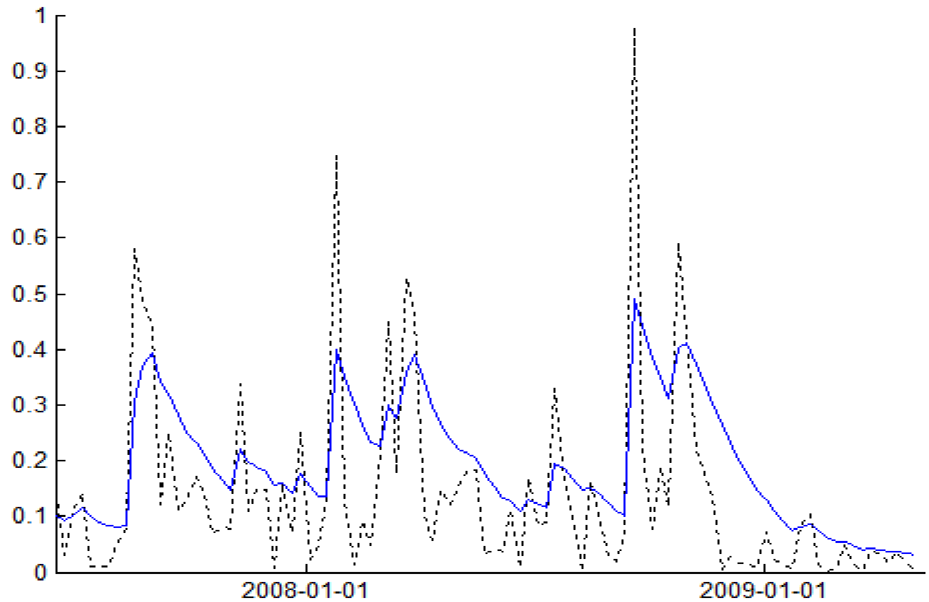
Note: The dotted line in (a) and (b) is the true volatility proxied by $|r_t - r_{t-1}|$. The dotted line in (c) and (d) is the true volatility proxied by $\hat{\sigma}^2_{GK}$. The line marked in bold is the volatility estimates $\hat{\sigma}_t$.

Figure 5 Plots of Out-of-Sample Volatility Forecasts for the U.S. Short Rates over the Sample Period 2007/06/15 - 2009/05/08

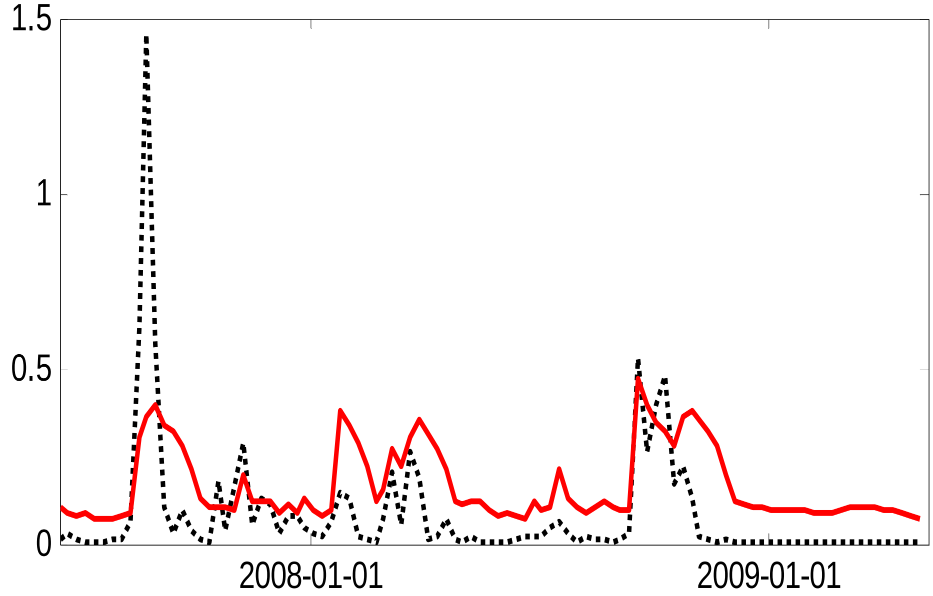
(a) Semiparametric approach



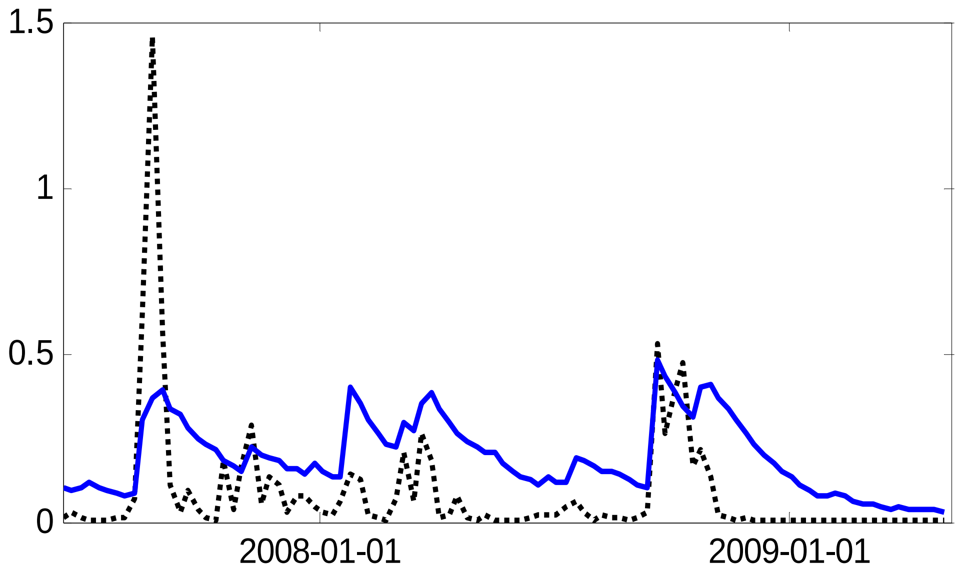
(b) AGARCHX-T model



(c) Semiparametric approach

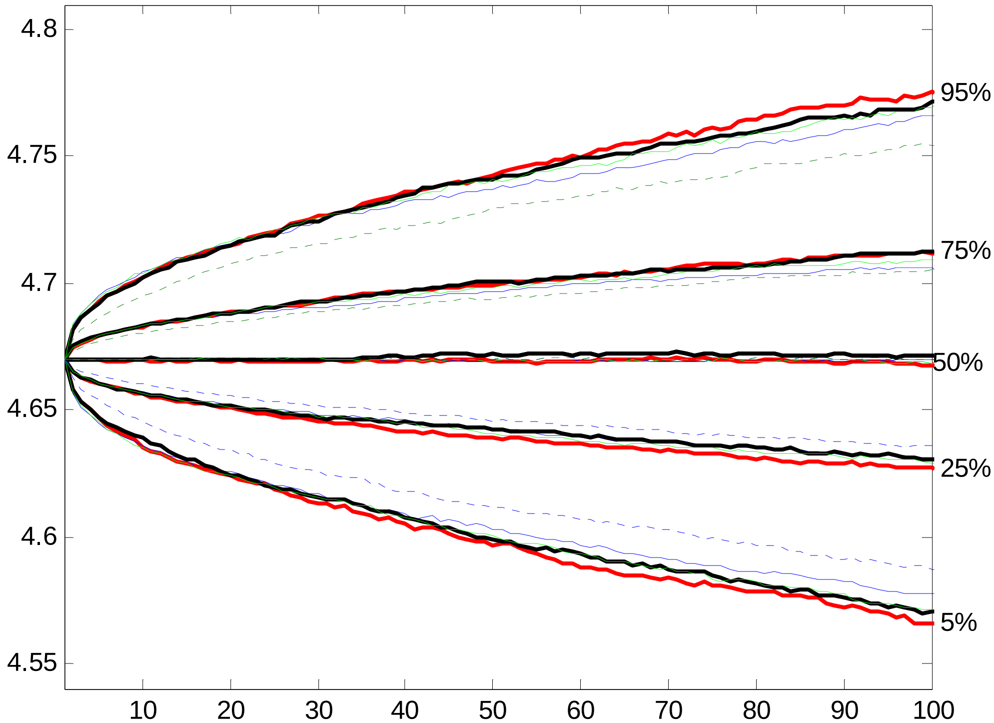


(d) AGARCHX-T model



Note: The dotted line in (a) and (b) is the true volatility proxied by $| r_t - r_{t-1} |$. The dotted line in (c) and (d) is the true volatility proxied by $\hat{\sigma}^2_{GK}$. The line marked in bold is the volatility estimate $\hat{\sigma}_t$.

Figure 6 **Confidence Intervals for Simulated Interest Rates from BHK and Semiparametric Models**



Note: The solid lines (reading from the outermost to the innermost lines) are for the AGARCHX-T, GARCHX-T, AGARCXH-N and GARCHX-N models. The dotted lines are for the semiparametric model.

Table 1 Estimates of Mean Squared and Mean Absolute Volatility Estimation Error for Simulated Data

(a) DGP – linear drift

	Normal				Student's t				Skewed t			
Models	MSE	Std. err.	MAE	Std. err.	MSE	Std. err.	MAE	Std. err.	MSE	Std. err.	MAE	Std. err.
Estimated with a linear drift function												
GARCH	0.3262	0.1210	0.3794	0.0638	0.3558	0.1357	0.4044	0.0716	0.3229	0.1193	0.3772	0.0559
Iteration 1	0.2316	0.1065	0.3561	0.0664	0.2310	0.0968	0.3561	0.0638	0.2308	0.0962	0.3558	0.0635
Iteration 2	0.2036	0.0861	0.3336	0.0617	0.2028	0.0944	0.3337	0.0629	0.2027	0.0951	0.3333	0.0631
Iteration 3	0.2065	0.0886	0.3359	0.0598	0.2030	0.0839	0.3336	0.0595	0.2033	0.0847	0.3336	0.0595
Iteration 4	0.2126	0.0961	0.3411	0.0627	0.2038	0.0855	0.3349	0.0607	0.2057	0.0845	0.3370	0.0593
Iteration 5	0.2085	0.0855	0.3402	0.0603	0.2057	0.0876	0.3371	0.0597	0.2062	0.0881	0.3379	0.0602
Iteration 6	0.2095	0.0846	0.3408	0.0605	0.2080	0.0864	0.3389	0.0607	0.2083	0.0876	0.3392	0.0624
Iteration 7	0.2155	0.0934	0.3445	0.0628	0.2094	0.0864	0.3398	0.0602	0.2098	0.0868	0.3399	0.0604
Iteration 8	0.2137	0.0906	0.3426	0.0623	0.2117	0.0916	0.3404	0.0620	0.2119	0.0918	0.3401	0.0619
Final smoothing	0.2158	0.0929	0.3442	0.0617	0.2123	0.0909	0.3415	0.0611	0.2125	0.0911	0.3417	0.0614
AGARCHX	0.2738	0.1193	0.3674	0.0693	0.2650	0.1011	0.3631	0.0652	0.2705	0.1052	0.3640	0.0511
GARCHX	0.3194	0.1234	0.3789	0.0711	0.2945	0.1217	0.3704	0.0686	0.3116	0.1143	0.3755	0.0532
Estimated with a nonlinear drift function												
GARCH	0.3286	0.1234	0.3785	0.0650	0.3552	0.1352	0.4025	0.0717	0.3224	0.1324	0.3826	0.0600
Iteration 1	0.2310	0.1071	0.3552	0.0677	0.2306	0.0972	0.3551	0.0644	0.2378	0.0984	0.3638	0.0661
Iteration 2	0.2038	0.0871	0.3331	0.0623	0.2039	0.0951	0.3343	0.0644	0.2130	0.0961	0.3436	0.0642
Iteration 3	0.2081	0.0897	0.3368	0.0611	0.2031	0.0835	0.3335	0.0597	0.2154	0.0888	0.3454	0.0626
Iteration 4	0.2147	0.0996	0.3413	0.0639	0.2053	0.0856	0.3362	0.0611	0.2143	0.0860	0.3460	0.0613
Iteration 5	0.2082	0.0857	0.3397	0.0613	0.2069	0.0892	0.3374	0.0604	0.2176	0.0918	0.3484	0.0636
Iteration 6	0.2117	0.0852	0.3423	0.0618	0.2098	0.0860	0.3397	0.0605	0.2192	0.0884	0.3499	0.0628
Iteration 7	0.2143	0.0907	0.3438	0.0632	0.2095	0.0864	0.3392	0.0599	0.2211	0.0901	0.3502	0.0629
Iteration 8	0.2173	0.0958	0.3450	0.0641	0.2135	0.0921	0.3414	0.0621	0.2233	0.0942	0.3512	0.0644
Final smoothing	0.2167	0.0927	0.3449	0.0630	0.2134	0.0910	0.3418	0.0613	0.2239	0.0940	0.3522	0.0639
AGARCHX	0.2846	0.1237	0.3689	0.0714	0.2739	0.1223	0.3672	0.0692	0.2814	0.1248	0.3681	0.0663
GARCHX	0.3205	0.1260	0.3813	0.0725	0.3131	0.1255	0.3811	0.0710	0.3207	0.1269	0.3859	0.0701

(b) DGP – Nonlinear drift

	Normal				Student's t				Skewed t			
Models	MSE	Std. err.	MAE	Std. err.	MSE	Std. err.	MAE	Std. err.	MSE	Std. err.	MAE	Std. err.
Estimated with a linear drift function												
GARCH	0.1094	0.0708	0.1893	0.0741	0.1077	0.0720	0.1877	0.0670	0.1042	0.0714	0.1907	0.0667
Iteration 1	0.0808	0.0651	0.1791	0.0501	0.0809	0.0649	0.1791	0.0500	0.0805	0.0637	0.1788	0.0497
Iteration 2	0.0704	0.0532	0.1669	0.0461	0.0685	0.0524	0.1651	0.0449	0.0692	0.0529	0.1649	0.0443
Iteration 3	0.0756	0.0568	0.1747	0.0498	0.0780	0.0758	0.1739	0.0522	0.0764	0.0722	0.1735	0.0509
Iteration 4	0.0710	0.0517	0.1693	0.0431	0.0729	0.0581	0.1703	0.0415	0.0716	0.0550	0.1695	0.0439
Iteration 5	0.0772	0.0605	0.1740	0.0473	0.0739	0.0558	0.1725	0.0460	0.0770	0.0596	0.1746	0.0523
Iteration 6	0.0782	0.0680	0.1748	0.0533	0.0700	0.0521	0.1684	0.0427	0.0721	0.0569	0.1691	0.0448
Iteration 7	0.0755	0.0568	0.1746	0.0467	0.0756	0.0614	0.1724	0.0466	0.0750	0.0583	0.1721	0.0458
Iteration 8	0.0782	0.0623	0.1750	0.0471	0.0754	0.0577	0.1752	0.0486	0.0742	0.0577	0.1732	0.0467
Final smoothing	0.0772	0.0603	0.1750	0.0472	0.0756	0.0591	0.1734	0.0464	0.0753	0.0595	0.1725	0.0462
AGARCHX	0.0967	0.0695	0.1851	0.0652	0.0930	0.0645	0.1825	0.0607	0.0944	0.0673	0.1838	0.0625
GARCHX	0.0991	0.0705	0.1872	0.0643	0.0956	0.0671	0.1860	0.0618	0.0968	0.0680	0.1867	0.0632
Estimated with a nonlinear drift function												
GARCH	0.0998	0.1057	0.1865	0.0724	0.0843	0.0646	0.1823	0.0647	0.0963	0.0693	0.1909	0.0653
Iteration 1	0.0917	0.0961	0.1816	0.0591	0.0846	0.0782	0.1786	0.0518	0.0830	0.0564	0.1843	0.0502
Iteration 2	0.0856	0.0948	0.1727	0.0602	0.0688	0.0519	0.1630	0.0422	0.0719	0.0572	0.1686	0.0480
Iteration 3	0.0888	0.0970	0.1772	0.0591	0.0754	0.0555	0.1716	0.0478	0.0726	0.0528	0.1721	0.0469
Iteration 4	0.0852	0.0907	0.1751	0.0559	0.0709	0.0501	0.1682	0.0432	0.0751	0.0579	0.1739	0.0494
Iteration 5	0.0844	0.0899	0.1738	0.0527	0.0779	0.0597	0.1733	0.0481	0.0798	0.0628	0.1778	0.0499
Iteration 6	0.0841	0.0907	0.1760	0.0551	0.0740	0.0538	0.1721	0.0462	0.0777	0.0603	0.1764	0.0490
Iteration 7	0.0839	0.0884	0.1766	0.0543	0.0737	0.0557	0.1713	0.0454	0.0768	0.0597	0.1751	0.0476
Iteration 8	0.0869	0.0906	0.1783	0.0550	0.0747	0.0542	0.1732	0.0458	0.0789	0.0532	0.1769	0.0492
Final smoothing	0.0862	0.0907	0.1766	0.0535	0.0747	0.0550	0.1731	0.0463	0.0790	0.0551	0.1770	0.0493
AGARCHX	0.0945	0.1022	0.1846	0.0619	0.0806	0.0612	0.1804	0.0521	0.0901	0.0635	0.1825	0.0534
GARCHX	0.0977	0.1025	0.1852	0.0637	0.0838	0.0637	0.1817	0.0580	0.0942	0.0668	0.1843	0.0599

Note: The results are for simulated data from the DGP in equation (14) for the linear drift and equation (15) for the nonlinear drift with the conditional variance following equation (16). The sample size is 1000 and the number of replications is 50. The rows labeled as iterations and final smoothing are the results for the semiparametric approach.

Table 2 Summary Statistics for the U.S. Short Rates

Variable	Mean	Standard deviation	Skewness	Kurtosis	JB test	Q(10)
r_t	5.8252	3.0448	0.7665	1.0476	271.82 [0.00]	18017.36 [0.00]
Δr_t	-0.0028	0.2319	-0.6466	17.9637	25557.59 [0.00]	185.53 [0.00]
Variable	ARCH(10)	τ_μ	τ_μ^{GLS}	τ_β	τ_β (BW)	
r_t	35.5992 [0.00]	-1.4902	-1.5440	-2.4301	-2.5147	

Note: The JB test represents the Jarque-Bera test of normality. Q(10) is the Ljung-Box test of serial correlation of order 10. ARCH(10) is the test for ARCH effect up to order 10 for the resulting residual of an AR(12) regression on short rate. τ_μ and τ_μ^{GLS} are the ADF and the GLS-based Dickey-Fuller test statistics and their 5% critical values are – 2.8629 and –1.95, respectively. τ_β and τ_β (BW) are the test statistics for Seo (1999) test with the latter using the Bollerslev and Wooldridge (1992) robust standard errors. The simulated critical values of τ_β and τ_β (BW) are –1.9073 and –1.8891 at the 5% significance level.

Table 3 Short-Term Interest Rate Model Estimates (1973/02/09 - 2007/06/08)

(a) Linear drift specification

	Normal			Student's t			Hansen's Skewed t		
Models	GARCH	GARCHX	AGARCHX	GARCH	GARCHX	AGARCHX	GARCH	GARCHX	AGARCHX
μ	0.0042	0.0039	0.0033	0.0037	0.0018	0.0014	0.0026	0.0041	0.0015
	(0.0063)	(0.0030)	(0.0031)	(0.0039)	(0.0024)	(0.0026)	(0.0019)	(0.0058)	(0.0045)
λ	-0.0014	-0.0002	-0.0004**	-0.0002**	-0.0003**	-0.0003**	-0.0007**	-0.0004	-0.0006
	(0.0016)	(0.0008)	(0.0002)	(0.0001)	(0.0001)	(0.0001)	(0.0003)	(0.0008)	(0.0008)
α_0	0.0001	0.0001	0.0003*	0.0000	0.00002	0.00002	0.0002	0.0003*	0.0000
	(0.0003)	(0.0002)	(0.0001)	(0.00001)	(0.00002)	(0.00004)	(0.0005)	(0.0001)	(0.0001)
α_1	0.1055*	0.3031*	0.2157*	0.1181*	0.2415*	0.2643*	0.1825*	0.2016*	0.2389*
	(0.0252)	(0.0695)	(0.0757)	(0.0258)	(0.0455)	(0.0478)	(0.0731)	(0.0680)	(0.0412)
α_2	0.8915*	0.6719*	0.6643*	0.8815*	0.7110*	0.7054*	0.8107*	0.7589*	0.7182*
	(0.0252)	(0.1069)	(0.1857)	(0.0258)	(0.0575)	(0.0518)	(0.0359)	(0.1205)	(0.1033)
α_3	-	-	0.0337*	-	-	0.0082*	-	-	0.0028*
			(0.0137)			(0.0011)			(0.0010)
b	-	0.0076*	0.0006*	-	0.0010*	0.0005*	-	0.0012*	0.0003*
		(0.0014)	(0.0001)		(0.0003)	(0.0001)		(0.0004)	(0.0001)
δ	-	0.4417*	0.5231*	-	2.6622*	3.14407*	-	2.3817*	2.9637*
		(0.0861)	(0.1494)		(0.3310)	(0.7811)		(0.4930)	(0.5182)
v	-	-	-	5.2401*	4.4148*	4.4212*	4.9613*	4.1945*	4.3182*
				(0.4109)	(0.4034)	(1.5303)	(0.3153)	(0.5738)	(0.5016)
η							-0.0005	-0.0001	-0.0002
							(0.0015)	(0.0021)	(0.0014)
LL	2810.19	2868.07	2871.29	2841.77	2912.65	2946.31	2135.35	2139.57	2146.18
$Q(\varepsilon_t / \sigma_t)$	194.8024	168.6053	170.8604	189.8495	170.4310	172.7072	153.6514	161.4817	169.2258
	[0.0000]	[0.0000]	[0.0000]	[0.0000]	[0.0000]	[0.0000]	[0.0000]	[0.0000]	[0.0000]
$Q(\varepsilon_t^2 / \sigma_t^2)$	22.1335	11.0381	11.5038	20.394	11.7280	12.0020	21.3148	10.3564	10.0106
	[0.0361]	[0.5256]	[0.4863]	[0.0599]	[0.4614]	[0.4455]	[0.0459]	[0.5848]	[0.6151]

(b) Nonlinear drift specification

	Normal			Student's t			Hansen's Skewed t		
Models	GARCH	GARCHX	AGARCHX	GARCH	GARCHX	AGARCHX	GARCH	GARCHX	AGARCHX
μ	0.0513*** (0.0269)	0.036 (0.0247)	0.0356 (0.0236)	0.0158 (0.0247)	0.0034 (0.0031)	0.0018 (0.0024)	0.0103 (0.0256)	0.0023** (0.0011)	0.0214 (0.0668)
λ_1	-0.0009 (0.0007)	0.0047** (0.0017)	0.0044** (0.0021)	-0.0019 (0.0026)	0.0011 (0.0007)	0.0016* (0.0004)	0.0015 (0.0037)	0.0013 (0.0035)	0.0006 (0.0014)
λ_2	-0.0003 (0.0005)	-0.0000 (0.0005)	-0.0000 (0.0005)	-0.0000 (0.0006)	-0.0001 (0.0002)	-0.0002 (0.0006)	0.0007 (0.0005)	-0.0000 (0.0001)	-0.0004 (0.0003)
λ_3	-0.0004 (0.0291)	-0.0002 (0.0181)	-0.0002 (0.0195)	-0.0001 (0.0171)	-0.0002 (0.0042)	-0.0001 (0.0035)	-0.0002 (0.0725)	-0.0001 (0.0395)	-0.0002 (0.0497)
α_0	0.0001 (0.0003)	0.0004** (0.0002)	0.0005* (0.0002)	0.00014** (0.00006)	0.0003* (0.0001)	0.0002** (0.0001)	0.0002** (0.0001)	0.0003* (0.0001)	0.0001* (0.0000)
α_1	0.1645* (0.0082)	0.3180* (0.0514)	0.2157* (0.0757)	0.1657* (0.0337)	0.2643* (0.0478)	0.2583* (0.0507)	0.1367* (0.0413)	0.1714* (0.0385)	0.2010* (0.0294)
α_2	0.8255* (0.1088)	0.6619* (0.1311)	0.6726* (0.1342)	0.8213* (0.0415)	0.7054* (0.0518)	0.7217* (0.0507)	0.8415* (0.2106)	0.8107* (0.1359)	0.7718* (0.1547)
α_3	-	-	0.0298* (0.0107)	-	-	0.0147* (0.0035)	-	-	0.0108* (0.0030)
b	-	0.0593* (0.0065)	0.0009* (0.0001)	-	0.0006* (0.0001)	0.0069* (0.0012)	-	0.0018** (0.0009)	0.0035* (0.0009)
δ	-	0.0540* (0.0059)	3.236* (0.3405)	-	4.5966* (0.743)	0.3979* (0.0671)	-	4.2613* (0.9180)	0.2859* (0.0101)
v	-	-	-	4.9327* (0.3879)	4.4043* (0.3871)	4.3314* (0.4398)	4.2138* (0.1310)	4.1017* (0.6133)	4.0981* (0.5819)
η	-	-	-	-	-	-	-0.0008 (0.0561)	-0.0003 (0.0318)	-0.0005 (0.1036)
LL	2443.13	2474.76	2491.27	2334.80	2355.82	2369.91	1753.42	1814.79	1830.25
$Q(\varepsilon_t / \sigma_t)$	165.1314 [0.0000]	128.3772 [0.0000]	131.0561 [0.0000]	141.2809 [0.0000]	137.1750 [0.0000]	148.9106 [0.0000]	123.5912 [0.0000]	146.2058 [0.0000]	151.8333 [0.0000]
$Q(\varepsilon_t^2 / \sigma_t^2)$	21.0615 [0.0495]	17.9813 [0.1163]	18.0210 [0.1151]	19.5918 [0.0752]	15.3428 [0.2233]	14.7311 [0.2565]	20.1113 [0.0650]	12.4285 [0.4119]	12.0091 [0.4449]

Note: The GARCHX (AGARCHX) model refers to the symmetric (asymmetric) GARCH model with additive levels effects given by equation 4 (5). LL denotes the log-likelihood value, $Q(\varepsilon_t / \sigma_t)$ and $Q(\varepsilon_t^2 / \sigma_t^2)$ are the Ljung-Box test statistics for serial correlation in the standardised and squared standardised residuals up to order twelve, respectively. *, ** and *** denote significance at 1%, 5% and 10% significance levels.

Table 4 The Goodness-of-Fit of In-Sample Volatility Estimates of Parametric and Semiparametric Models of U.S. Short Rates over the Period 1973/02/09 - 2007/06/08

(a) Volatility Benchmark $|r_t - r_{t-1}|$

	Normal				Student's t			
Linear drift	MSE	MAE	AIC	R^2_{vol}	MSE	MAE	AIC	R^2_{vol}
GARCH	0.0441	0.1632	-465.2797	0.3539	0.0444	0.1645	-451.8839	0.3564
Iteration 1	0.0410	0.1599	-584.3063	0.3982	0.0411	0.1601	-583.1900	0.3989
Iteration 2	0.0413	0.1595	-572.9455	0.3941	0.0415	0.1601	-564.1346	0.3937
Iteration 3	0.0409	0.1583	-589.5497	0.3929	0.0410	0.1584	-586.7339	0.3926
Iteration 4	0.0408	0.1578	-593.7955	0.3913	0.0408	0.1579	-592.4436	0.3910
Iteration 5	0.0408	0.1576	-593.4775	0.3901	0.0409	0.1577	-589.1935	0.3890
Iteration 6	0.0407	0.1571	-598.9764	0.3903	0.0407	0.1572	-596.3312	0.3893
Iteration 7	0.0407	0.1571	-599.0108	0.3906	0.0408	0.1573	-593.5321	0.3892
Iteration 8	0.0406	0.1567	-603.5222	0.3909	0.0406	0.1568	-601.1995	0.3901
Final smoothed	0.0406	0.1569	-601.8026	0.3908	0.0407	0.1572	-601.9206	0.3899
AGARCHX	0.0433	0.1609	-468.8008	0.3445	0.0421	0.1584	-498.5179	0.3533
GARCHX	0.0439	0.1617	-459.0652	0.3431	0.0428	0.1596	-486.4507	0.3526

Nonlinear drift	MSE	MAE	AIC	R^2_{vol}	MSE	MAE	AIC	R^2_{vol}
GARCH	0.0441	0.1633	-463.5183	0.3534	0.0444	0.1645	-453.0660	0.3567
Iteration 1	0.0411	0.1602	-581.6097	0.3981	0.0411	0.1601	-583.1312	0.3989
Iteration 2	0.0414	0.1597	-571.1824	0.3943	0.0415	0.1601	-563.9604	0.3942
Iteration 3	0.0409	0.1584	-588.7446	0.3932	0.0410	0.1584	-586.7507	0.3928
Iteration 4	0.0408	0.1580	-593.2638	0.3919	0.0408	0.1579	-592.6804	0.3911
Iteration 5	0.0409	0.1578	-590.8762	0.3901	0.0409	0.1577	-589.8344	0.3893
Iteration 6	0.0407	0.1573	-597.7895	0.3905	0.0408	0.1574	-594.9734	0.3897
Iteration 7	0.0407	0.1573	-597.6123	0.3905	0.0408	0.1573	-593.8119	0.3896
Iteration 8	0.0406	0.1570	-601.9312	0.3914	0.0406	0.1569	-600.9026	0.3903
Final smoothed	0.0407	0.1572	-600.0607	0.3907	0.0407	0.1572	-600.0972	0.3901

(b) Volatility Benchmark $\hat{\sigma}^2_{GK}$

	Normal				Student's t			
Linear drift	MSE	MAE	AIC	R^2_{vol}	MSE	MAE	AIC	R^2_{vol}
GARCH	0.0263	0.1366	-1325.2800	0.4139	0.0270	0.1387	-1281.6440	0.4128
Iteration 1	0.0249	0.1340	-1417.3380	0.4309	0.0251	0.1344	-1409.1910	0.4307
Iteration 2	0.0245	0.1321	-1448.4510	0.4322	0.0248	0.1328	-1428.1230	0.4312
Iteration 3	0.0242	0.1311	-1464.1010	0.4284	0.0244	0.1313	-1455.2030	0.4270
Iteration 4	0.0243	0.1308	-1462.5520	0.4259	0.0244	0.1311	-1452.1460	0.4245
Iteration 5	0.0243	0.1307	-1462.1710	0.4264	0.0244	0.1308	-1454.1570	0.4246
Iteration 6	0.0241	0.1303	-1470.9490	0.4251	0.0243	0.1304	-1462.5230	0.4223
Iteration 7	0.0242	0.1303	-1469.1290	0.4257	0.0243	0.1306	-1459.5990	0.4241
Iteration 8	0.0241	0.1299	-1475.9890	0.4262	0.0242	0.1302	-1466.1370	0.4242
Final smoothed	0.0241	0.1301	-1474.4380	0.4263	0.0242	0.1304	-1464.8260	0.4240
AGARCHX	0.0268	0.1340	-1293.6400	0.4189	0.0261	0.1335	-1334.6530	0.4219
GARCHX	0.0266	0.1342	-1302.4110	0.4167	0.0258	0.1327	-1352.3520	0.4225

Nonlinear drift	MSE	MAE	AIC	R^2_{vol}	MSE	MAE	AIC	R^2_{vol}
GARCH	0.0266	0.1369	-1310.7710	0.4129	0.0270	0.1387	-1283.1120	0.4129
Iteration 1	0.0250	0.1343	-1411.8520	0.4313	0.0250	0.1344	-1409.5440	0.4307
Iteration 2	0.0245	0.1323	-1444.5760	0.4325	0.0248	0.1329	-1427.3420	0.4320
Iteration 3	0.0243	0.1313	-1460.7870	0.4287	0.0244	0.1314	-1455.9540	0.4275
Iteration 4	0.0243	0.1310	-1459.0720	0.4265	0.0244	0.1311	-1453.4080	0.4249
Iteration 5	0.0243	0.1309	-1458.5770	0.4270	0.0244	0.1308	-1455.7790	0.4251
Iteration 6	0.0242	0.1305	-1469.1250	0.4258	0.0243	0.1305	-1461.6970	0.4234
Iteration 7	0.0242	0.1304	-1468.5080	0.4263	0.0243	0.1307	-1459.1260	0.4244
Iteration 8	0.0242	0.1303	-1469.7660	0.4268	0.0242	0.1303	-1466.3600	0.4247
Final smoothed	0.0241	0.1303	-1470.8810	0.4257	0.0242	0.1304	-1464.5850	0.4240

Note: The rows labeled as iterations and final smoothed are the results for the sem-iparametric approach. The prefix 'A' denotes asymmetric while the suffix 'X' denotes levels effects.

Table 5 The Out-of-Sample Volatility Forecast Performance of Parametric and Semiparametric Models of U.S. Short Rates over the Period 2007/06/15 - 2009/05/08

(a) Volatility Benchmark $\left|r_t - r_{t-1}\right|$

	GARCH	Final Smoothed	AGARCHX-T	GARCHX-T	AGARCHX-N	GARCHX-N
MSE	0.0349	0.0318	0.0339	0.0344	0.03499	0.0352
MAE	0.1431	0.1225	0.1295	0.1319	0.1346	0.1359

(b) Volatility Benchmark $\hat{\sigma}^2_{GK}$

	GARCH	Final Smoothed	AGARCHX-T	GARCHX-T	AGARCHX-N	GARCHX-N
MSE	0.0498	0.0395	0.0458	0.0466	0.0477	0.0486
MAE	0.1744	0.1402	0.1581	0.1643	0.1666	0.1678

Note: See note to Table 4. ‘T’ and ‘N’ denote Student’s t and normal distributions, respectively.