Discussion Paper #2006-8

Is Nonlinear Drift Implied by the Short-End of the Term Structure?

by

Hideyuki Takamizawa

November 2006

Is Nonlinear Drift Implied by the Short-End of the Term Structure?

Hideyuki Takamizawa [*]

November 22, 2006

Abstract

Nonlinear drift models of the short-rate are estimated using data on the short-end of the term structure, where the cross-sectional relation is obtained by an analytical approximation. We find that (i) nonlinear physical drift is not implied unless it is strongly affected by cross-sectional dimensions of the data; (ii) nonlinear risk-neutral drift that allows for fast mean-reversion for high rates is desirable to explain and predict observed patterns of yield spreads; and (iii) for higher-frequency data from which transitory shocks are removed, (ii) still remains valid although the nonlinearity is somewhat reduced.

Key Words: Short-Rate; Nonlinear Drift; Term Structure; Linear Approximation.

[*] Address correspondence to: Graduate School of Economics, Hitotsubashi University, Kunitachi Tokyo 186-8601, JAPAN. Phone & Fax: +81–42–580–8524. E-mail: htakamiz@econ.hit-u.ac.jp

1 Introduction

The drift of the instantaneous risk-free rate, the short-rate, has a crucial role both in capturing the time-series behavior and in pricing the cross-section of interest-rate claims. This has drawn attention to the shape of the drift. In particular, since Aït-Sahalia (1996), the nonlinearity in the drift has been intensively discussed. Nonlinear drift, having a more flexible form than linear drift, can produce different speeds of mean-reversion at different levels of the short-rate. Specifically, fast mean-reversion is generated at very high and low levels, but little mean-reversion occurs at middle levels. From an economic perspective, nonlinear drift seems to provide a rationale for the puzzling behavior of interest rates: in spite of near-unit root behavior, they do not diverge. It can further be rooted in the behavior of central banks, which adjust interest-rate levels, depending on economic and market conditions, in a certain range: see Aït-Sahalia (1996, p.407) for discussion regarding economic aspects of nonlinear drift.

In the estimation, Aït-Sahalia (1996) points out that parametric models adequately matching a nonparametrically estimated marginal density have nonlinear drift. Jiang (1997) also utilizes the relation between the marginal density and drift-diffusion functions to obtain a nonparametric estimator of the drift, which exhibits nonlinearity. Stanton (1997) finds that while nonparametric estimators of the drift are nearly zero for most of the observed range of U.S. interest rates, they decrease rapidly at extremely high levels. Conley et al. (1997) document that in controlling the value of a parameter of constant elasticity of variance, nonlinear drift more adequately matches data than does linear drift. Ang and Bekaert (2002) show that nonlinear drift is naturally implied when regime-switching models with state-dependent transition probabilities are estimated.

Conversely, Pritsker (1998) demonstrates that the nonparametric-based test proposed by Aït-Sahalia (1996) too often rejects the null of linear drift, due to the extremely slow rate of convergence of the test statistic given the high persistence of interest rates. Chapman and Pearson (2000) perform Monte Carlo simulations, in which both parametric and nonparametric estimators of the drift are obtained using data generated from a linear drift model. They show that these estimators tend to be negatively (positively) biased for very high (low) rates, leading to spurious nonlinearities. [1] Li et al. (2004) report that when the condition that the short-rate stays in a predetermined range is incorporated into the estimation of the drift, a linear drift model is not rejected. Recent studies using simulation-based techniques have reported that there is not sufficient evidence in favor of nonlinear drift. Durham (2003), employing the simulated maximum likelihood method, points out that terms beyond a constant in the drift do little to improve the fit. Jones (2003), employing the Bayesian MCMC approach, finds that the nonlinearity is not only an outcome of particular distributional assumptions reflected in chosen priors, but also an outcome of estimating misspecified models from high-frequency data.

The lack of consensus reflects fundamental difficulties in the estimation of the drift, which requires data over a long period of time. The high persistence of interest rates reinforces these difficulties. Besides, when the nonlinearity of the drift is considered, the problem becomes even more serious. We normally need to show that the speed of mean-reversion at the high and low levels differs from that in the middle. However, observations at these extreme levels are usually scarce. It is therefore unavoidable that the shapes of the drift at these extreme levels are identified with much less precision. The problem is not easy to solve even though time-series observations over several decades are available.

In this paper, we adopt an alternative approach to estimating the drift. Instead of relying solely on time-series data, we utilize multiple series of data on the term structure of interest rates. By further specifying the price of risk, the term structure can be derived. Part of the information on the short-rate dynamics is then reflected in the cross-section of discount-bond yields. The drift needs to be consistent with both time-series and cross-sectional dimensions of the data, which we expect leads to more precise estimators of the drift.

The estimation, however, is time-consuming. We need both to solve a valuation equation relating the short-rate to yields in the cross-section and to optimize an objective function with respect to model parameters for estimation. Since we consider short-rate models with nonlinear drift, closed-form models of the term structure cannot generally be obtained, which is a major obstacle to using term structure data. To overcome this difficulty, we utilize an approximation proposed by Takamizawa and Shoji (2003): a local linear approximation is applied to the short-rate process with nonlinear drift and diffusion functions, and the partial differential equation valuing a discount bond can be solved analytically. Once the yield function is obtained in closed form, the estimation is carried out by the maximum likelihood (ML) method, in which the models are fitted to both time-series and cross-sectional dimensions of the data.

Since our primary interest is in the drift of the short-rate process, we exclusively focus on the stochastic behavior of the short-rate. Therefore, in constructing the term structure, we also assume a single factor. Accordingly, we fit the models to data on the *short-end* of the term structure, short-yield data, where the assumption is not overly restrictive, even though multiple factors are necessary for describing the entire term structure.

We emphasize that the short-end of the term structure is indeed informative. Figure 1 graphs spreads of the three- and six-month Eurodollar rates over the one-month Eurodollar rate against the level of the one-month rate on a weekly basis. While there seems no clear pattern for low to middle rates, large negative spreads, particularly those of the six-month rate, can be seen for high rates. This may indicate that the drift in the risk-neutral measure, the risk-neutral drift, is nonlinear. More precisely, suppose the risk-neutral drift decreases rapidly as the short-rate increases. Then, in the risk-neutral world, the short-rate at high levels will be more likely to decrease, which reduces the rate of increase in $\int_t^T r_u du$ and hence the rate of decrease in $\exp(-\int_t^T r_u du)$. Since discount-bond prices are given by $E_t^Q[\exp(-\int_t^T r_u du)]$, where the conditional expectation E_t^Q is taken with respect to the risk-neutral measure, it follows that the prices are less discounted on average due to fast mean-reversion. The yields to maturity are then low relative to the short-rate, resulting in large negative spreads. A desirable specification of the risk-neutral drift is therefore implied by short-yield data.

A desirable specification of the drift in the physical measure, the physical drift, is also possibly implied. Through the absence of arbitrage, the physical drift is determined by the risk-neutral drift and the risk premium. The risk-neutral drift is likely to be restricted by term structure data, as explained above. The risk-premium function is also restricted by the absence of arbitrage: see, for example, Cox, Ingersoll, and Ross (CIR) (1985, p.398). Then, the resulting physical drift may not be determined arbitrarily, either. While this perspective cannot be achieved using time-series data alone, it can using term structure data.

Our findings regarding the nonlinearity in both the physical and risk-neutral drifts are as follows. First, nonlinear physical drift is implied only in cases where nonlinear terms are strongly affected by cross-sectional dimensions of the data. These cases, however, are rather restrictive. In general, the goodness-of-fit to time-series dimensions of the data is not much reduced without nonlinear terms. Second, nonlinear risk-neutral drift is more desirable. Due to higher-order terms, both large negative spreads for high rates and very small spreads for low rates are consistently explained. Third, nonlinear risk-neutral drift is still evident for higher-frequency data from which transitory shocks are removed, although the nonlinearity is somewhat reduced.

The rest of the paper is organized as follows. Section 2 specifies parametric models of the short-rate with nonlinear drift and constant elasticity of variance, and derives an analytical approximation of the term structure. Section 3 explains the estimation framework. Section 4 reports empirical results for weekly data. Section 5 examines whether nonlinear terms in both the physical and risk-neutral drifts contribute to out-of-sample prediction. Section 6 examines using daily data whether nonlinear risk-neutral drift is still preferable after controlling transitory shocks probably contained in the daily-observed short-rate. Section 7 provides concluding remarks.

2 Models

Following Aït-Sahalia (1996, 1999) and Conley et al. (1997), we consider the following parametric model of the short-rate process with nonlinear drift and constant elasticity of variance/volatility (CEV), which is standard in the literature when it comes to nonlinear drift models of the short-rate:

$$dr_t = (\alpha_{-1}/r_t + \alpha_0 + \alpha_1 r_t + \alpha_2 r_t^2)dt + \sigma r_t^{\gamma} dW_t \ , \tag{1}$$

where W_t is Brownian Motion in the physical measure.

Next, we specify the price of risk, $\lambda(r)$, consistently with non-arbitrage. We consider as a non-arbitrage condition boundedness of $\lambda(r)$ on $\{r : \sigma(r) = 0\}$, where $\sigma(r)$ stands for the diffusion term of the short-rate process. Accordingly, the risk premium given by $\lambda(r)\sigma(r)$ is zero if $\sigma(r) = 0$. A similar condition is adopted by, for example, Stanton (1997) and Jiang (1998). To keep models simple, we also assume that both the physical and risk-neutral drifts, denoted, respectively, as $\mu(r)$ and $\mu^Q(r)$, are of the same functional form (this assumption is later withdrawn). This leads to an additional restriction on $\lambda(r)$ through the following identity:

$$\mu^Q(r) = \mu(r) - \lambda(r)\sigma(r) \ . \tag{2}$$

Then, possible specifications are [2]

$$\lambda(r_t) = \lambda_2 r_t^{2-\gamma} \qquad (1 < \gamma \leq 2) \ , \tag{3}$$

$$\lambda(r_t) = \lambda_1 r_t^{1-\gamma} + \lambda_2 r_t^{2-\gamma} \qquad (0 < \gamma \leq 1) \ . \tag{4}$$

Consequently, the risk-neutral process of the short-rate is modeled by the following stochastic differential equation (SDE):

$$dr_t = (\alpha_{-1}/r_t + \alpha_0 + \alpha_1 r_t + \beta_2 r_t^2)dt + \sigma r_t^\gamma dW_t^Q \qquad (1 < \gamma \leq 2)\ , \tag{5}$$

$$dr_t = (\alpha_{-1}/r_t + \alpha_0 + \beta_1 r_t + \beta_2 r_t^2)dt + \sigma r^\gamma dW_t^Q \qquad (0 < \gamma \leq 1)\ , \tag{6}$$

where W_t^Q is Brownian Motion in the risk-neutral measure, and where $\beta_i \equiv \alpha_i - \sigma\lambda_i$ $(i = 1, 2)$. In the empirical analysis, we report estimation results of β_i, instead of λ_i, as the shapes of both $\mu(r)$ and $\mu^Q(r)$ are of primary interest.

Many previous studies, using long time-series data on U.S. interest rates, report that the estimate of the CEV parameter, γ, is more than one but less than two: see Chan, Karolyi, Longstaff, and Sanders (CKLS) (1992) among others. Therefore, when we treat γ as a free parameter, we conservatively adopt the model given by (1) and (3) (or (5)). On the other hand, when we wish to increase the number of parameters in $\lambda(r)$, we may have to restrict the value of γ. The alternative model given by (1) and (4) (or (6)) allows us to examine whether estimated shapes of $\mu(r)$ and $\mu^Q(r)$ differ between the two models. Of particular interest is whether the difference in the number of shared parameters (three for the former and two for the latter) has a nontrivial impact on the shape of $\mu(r)$.

Next, we derive the price of a default-free discount bond. Let $P(r_t, t, T)$ denote the price at time t with maturity time T. Then, by the standard non-arbitrage argument, it is the solution to the following partial differential equation (PDE):

$$\frac{1}{2}(\sigma r_t^\gamma)^2 \frac{\partial^2 P}{\partial r_t^2}(r_t, t, T) + \mu^Q(r_t)\frac{\partial P}{\partial r_t}(r_t, t, T) + \frac{\partial P}{\partial t}(r_t, t, T) - r_t P(r_t, t, T) = 0\ , \tag{7}$$

with the boundary condition, $P(r_T, T, T) = 1$. Since the PDE cannot generally be solved in closed form, we employ an analytical approximation for $P(r_t, t, T)$ proposed by Takamizawa and Shoji (2003) to keep the computational burden manageable. The approximate solution, denoted as $\tilde{P}(r_t, \tau)$ with $\tau = T - t$, is derived as

$$\tilde{P}(r_t, \tau) = \exp\{-A(\tau; r_t) - B(\tau; r_t)r_t\}\ , \tag{8}$$

where $A(\tau; r_t)$ and $B(\tau; r_t)$ are given in Appendix. The yield to maturity of a discount bond is then given by $\tilde{Y}(r_t, \tau) = \frac{1}{\tau}\{A(\tau; r_t) + B(\tau; r_t)r_t\}$.

The accuracy of the approximation decreases with maturity length, τ. However, we have verified that the approximation does not cause serious estimation problems, as long as standard nonlinear drift models are estimated using data on U.S. interest rates with short maturities: the details of the accuracy analysis are available upon request.

3 Estimation Framework

3.1 Data

We use weekly data on the one-, three-, and six-month Eurodollar deposit rates, which are available at H-15 Federal Reserve Statistical Release (Selected Interest Rate Series). Although weekly data (on a Friday basis) is downloadable, we construct it from daily data by picking up every set of Wednesday observations. If it is missing on a particular week, we choose it from another day of the week in order of Tuesday, Thursday, Friday, and Monday. Exceptionally, we replace two sets of Wednesday observations that are possibly outliers with other sets within the same week: 12/24/1980 is replaced with 12/23 (Tue.), and 12/26/1990 with 12/28 (Fri.). We use for estimation data from January 6, 1971 to December 29, 1999 (1513 obs.), and for out-of-sample prediction data from January 5, 2000 to December 28, 2005 (313 obs.). We also use an alternative dataset consisting of U.S. Treasury bill yields with maturities of three, six, and twelve months, covering the period from July 15, 1959 to August 22, 2001. The estimation results are broadly similar to those for the Eurodollar data. Hence, they are not reported in this paper, but are available upon request.

3.2 Objective function for estimation

We employ the ML method to estimate the models, where they are fitted to both time-series and cross-sectional dimensions of the data. [3] The density function at a given point in time is decomposed into two parts, denoted as f_T and f_C: f_T is the transition density for the time-series behavior of the short-rate, whereas f_C is the density for measurement errors added to term structure models.

For the short-rate model given by (1), no analytical expression is known for f_T. To compute it, therefore, we employ the method proposed by Aït-Sahalia (1999, 2002). The method provides an analytical approximation of f_T, expressed as the multiplication of the normal density and correction terms given by power series in an observation interval, Δ. We truncate the series at the first order, since, according to Aït-Sahalia (1999), sufficient accuracy is achieved by the first-order approximation for regular financial data.

f_C, on the other hand, is based on the following regression model:

$$\begin{pmatrix} y_{3M,t} \\ y_{6M,t} \end{pmatrix} = \begin{pmatrix} \tilde{Y}(r_t, 0.25) \\ \tilde{Y}(r_t, 0.50) \end{pmatrix} + \begin{pmatrix} v_1 & 0 \\ v_2\rho & v_2\sqrt{1-\rho^2} \end{pmatrix} \begin{pmatrix} u_{1,t} \\ u_{2,t} \end{pmatrix}, \tag{9}$$

where

$$u_{i,t} = \phi_i u_{i,t-\Delta} + \sqrt{1-\phi_i^2} z_{i,t} \ , \quad z_{i,t} \sim i.i.d.\, N(0,1) \quad (i = 1,2) \ , \tag{10}$$

and where $z_{1,t}$ and $z_{2,t}$ are independent from r_t. We note that each unconditional variance of u_i is set to one given $|\phi_i| < 1$. The magnitude of the measurement errors is then captured by (v_1, v_2). By incorporating the first-order autocorrelations as well as the contemporaneous correlation, distributional features of the measurement errors become more realistic. In fact, the log-likelihood value increases dramatically by assuming the autocorrelations. This does not necessarily mean, however, that the descriptive power of term structure models is improved. Nevertheless, we adopt this assumption to prevent both $\mu(r)$ and $\mu^Q(r)$ from being excessively affected by cross-sectional dimensions. In other words, we wish to avoid a situation where nonlinear terms, which do not originally exist, would become significant due to poor assumptions for the measurement error distribution.

Let θ_T and θ_C be parameter vectors on which f_T and f_C depend: $\theta_T = ((\alpha_j)_{j=-1}^2, \sigma, \gamma)$, and $\theta_C = ((\alpha_j)_{j=-1}^J, (\beta_j)_{j=J+1}^2, \sigma, \gamma, \rho, (v_j, \phi_j)_{j=1}^2)$, where $J = 0$ for $0 < \gamma \leq 1$ and $J = 1$ for $1 < \gamma \leq 2$. Then, an objective function for estimation is

$$\sum_t \{\ln f_T(r_t; r_{t-\Delta}, \theta_T) + \ln f_C(y_t; r_t, \theta_C)\} \ , \tag{11}$$

where $y_t = (y_{3M,t},\, y_{6M,t})$. We note that in estimating all the parameters, both ϕ_1 and ϕ_2 approach one, and that the structural parameters of interest take on unreasonable values. To avoid this problem, we fix $(\phi_1,\, \phi_2)$. To obtain the reasonable values, we first estimate two representative models with $(\phi_1,\, \phi_2)$ set to zero, and then estimate $(\phi_1,\, \phi_2)$ from the residual series of each model. The two representative models, which we later label (GF *) and (G1 ****), are distinguished mainly by the value of γ: it is a free parameter for the former while it is fixed at one for the latter. The resulting estimates for (GF *) and (G1 ****) are (0.851, 0.819) and (0.853, 0.824), respectively. The former (latter) values are repeatedly used for estimating other models with γ free (fixed at one).

4 Empirical Results for Weekly Data

4.1 Estimation results for models with γ free

We begin with estimation of the short-rate model given by (1) using time-series data on the one-month rate alone. That is, only the first component of (11), $\sum_t \ln f_T(r_t; r_{t-\Delta}, \theta_T)$, is maximized over θ_T. The column of Table 1 labeled (Time) presents the results. First, the estimate of γ is 1.31, which is significantly different from one. Hence, as long as we estimate parsimonious models using long-term data on U.S. interest rates, the conservative specification of $\lambda(r)$ given by (3) may be acceptable. Second, neither α_{-1} nor α_2 is estimated precisely. The result agrees with that of Aït-Sahalia (1999, Table VI), whose ML technique we follow. While the data used in his study (monthly data on the Fed funds rate over 1963/01–1998/12) are different from ours, the estimates (standard errors) of $\alpha_{-1} \times 10^2$ and α_2 are similar: 0.069 (0.2) and -4.059 (6.4), respectively. These results indicate the difficulty of the precise estimation of the nonlinear terms using time-series data alone.

Given the imprecise estimates, the null hypothesis of $\alpha_{-1} = \alpha_2 = 0$ is not rejected at any conventional significance level: the likelihood-ratio statistic (p-value, d.f.) is 0.67 (0.72, 2). Furthermore, the null of zero drift is not rejected, either: the likelihood-ratio statistic (p-value, d.f.) is 5.36 (0.25, 4). That is, the precise estimation of the drift itself is difficult in our data.

We now estimate using short-yield data the model given by (1) and (5), which we label (GF *) (the model with γ free). The asterisk denotes that α_{-1} is a free parameter. When we restrict $\alpha_{-1} = 0$, it is replaced with 0. The column of Table 1 labeled (GF *) presents the results. First, the estimate of α_2 is -7.09 and now significant, the reason for which is explained below. Second, the estimate of α_{-1} is still insignificant even after introducing cross-sectional constraints. Third, the estimates of the volatility parameters, (σ, γ), are $(0.72, 1.30)$, which are little changed from those using time-series data alone, $(0.73, 1.31)$. The result indicates that time-series dimensions of the data almost exclusively determine the values of the volatility parameters. Also, this is consistent with the fact that the volatility is invariant to changes of probability measures. Fourth, the estimate of β_2 is -4.04 and significant, which is in line with our expectations as mentioned in Introduction. A positive term premium is implied, as the resulting estimate of $\lambda_2 = (\alpha_2 - \beta_2)/\sigma$ is negative. This, in fact, is the key to the significant estimate of α_2. Specifically, although α_2 is the only unshared parameter in $\mu(r)$, and hence supposed to be adjusted to time-series dimensions of the data, it is also constrained by cross-sectional dimensions through the term premium: for the term premium to be positive $\alpha_2 < \beta_2$ is required given $\sigma > 0$. In addition, by taking account of $\beta_2 < 0$, which is supported by short-yield data, the constraint on α_2 is actually $\alpha_2 < \beta_2 < 0$.

Next, we test for $\alpha_{-1} = 0$. The likelihood-ratio statistic is 0.07, and the null is not rejected at the 5% significance level. Strong evidence in favor of fast mean-reversion at low levels cannot therefore be obtained in either the physical or the risk-neutral measure. Parameter estimates for (GF 0) are also presented in Table 1. The estimates of α_2 and β_2 are both significant, and similar in magnitude to those for (GF *). Further constraints on nonlinear terms are therefore not supported by short-yield data. In particular, the rejection of $\beta_2 = 0$ is extremely strong. In actually estimating the model with $\alpha_{-1} = \beta_2 = 0$, the slope of $\mu^Q(r)$ is negative, indicating that the short-rate in the risk-neutral measure mean-reverts for both high and low rates at a constant speed. However, this behavior is not consistent with either large negative spreads for high rates or small spreads for low rates. Due to the quadratic term in $\mu^Q(r)$, interest rates at both extreme levels can adequately be explained. In other words, it indeed gives an additional degree of freedom in fitting models to short-yield data.

In Panels (a) and (b) of Figure 2, $\mu(r)$ and $\mu^Q(r)$ are plotted against the level of r. The shape of $\mu(r)$ estimated using time-series data alone appears to differ from that estimated using short-yield data. However, the difference may actually be minor, as the estimates using time-series data alone are imprecise. If this is the case, we could say that a desirable shape of $\mu(r)$ is identified from short-yield data. We then test for whether the identified drift is restrictive for time-series data alone. Specifically, the null hypothesis is $(\alpha_{-1}, \alpha_0, \alpha_1, \alpha_2) = (0.000,\ 0.015,\ 0.358,\ -6.830)$ taken from (GF 0). Then, the rest of the parameters, (σ, γ), are estimated using time-series data to compute the likelihood value. The likelihood-ratio statistic (p-value, d.f.) is 10.92 (0.03, 4). Identifying the shape of $\mu(r)$ by introducing cross-sectional constraints is therefore not costless for the GF-type models. We further test in the next section for whether the identified drift is useful for out-of-sample prediction. Between (GF *) and (GF 0), there is little difference in the shapes of $\mu(r)$ and $\mu^Q(r)$. A small deviation observed for $r < 0.03$ may be spurious, taking account of the fact that the in-sample minimum of r is 0.03.

4.2 Estimation results for models with $\gamma = 1$

When we increase terms in $\lambda(r)$ as given in (4), we actually have to restrict the value of γ as long as the parsimony assumption is maintained that both $\mu(r)$ and $\mu^Q(r)$ are of the same functional form. Based on the actual estimate exceeding one, we set $\gamma = 1$. We note, however, that the constraint is strongly rejected, as we see below.

We also begin with estimation of the short-rate model with $\gamma = 1$ using time-series data alone. The column of Table 2 labeled (Time) presents the results. First, the null hypothesis of $\gamma = 1$ is strongly rejected: the likelihood-ratio statistic is 53. Second, none of the estimates are significant. Again, the null of zero drift is not rejected at any conventional significance level: the likelihood-ratio statistic (p-value, d.f.) is 5.66 (0.23, 4).

We next estimate the model given by (1) and (6), which we label (G1 ****) (the model with γ fixed at one). The asterisks denote that the parameters in order of $(\alpha_{-1}, \alpha_2, \lambda_1, \lambda_2)$ are free. When one or more of them are set to zero, the corresponding asterisks are replaced with 0s. The column of Table 2 labeled (G1 ****) presents the results. First, neither α_{-1} nor α_2 is estimated precisely, as is the case for time-series data alone. The insignificant estimate of α_2 contrasts sharply with the previous result for (GF *). This is because for (G1 ****) neither the sign nor the magnitude of α_2 is crucial for explaining the term premium. More specifically, for a positive term premium, $\lambda(r) = \lambda_1 + \lambda_2 r < 0$ is required given $\sigma(r) > 0$. However, the inequality possibly holds even for $\lambda_2 > 0$ $(\alpha_2 > \beta_2)$, when λ_1 is sufficiently negative. Indeed, the estimates of α_1 and β_1 are 0.37 and 1.08, respectively, leading to $\lambda_1 = (\alpha_1 - \beta_1)/\sigma < 0$. Moreover, in computing $\lambda(r)$, it is negative until the short-rate is slightly below 0.14, and then becomes positive. We notice in Figure 1 that the proportion of negative spreads increases for $r > 0.14$. Hence, $\lambda(r) > 0$, or equivalently a negative term premium, for high rates is more consistent with the actual data.

We next test for $\alpha_{-1} = \alpha_2 = 0$, the linearity in $\mu(r)$. The likelihood-ratio statistic (p-value, d.f.) is 5.17 (0.08, 2), and the null is not rejected at the 5% significance level. The estimation results for the null model, (G1 00**), are also presented in Table 2. The estimates of α_0 and α_1 are now significant, implying that the short-rate mean-reverts in the physical measure at a constant speed. [4] We also test for whether the identified drift is restrictive for time-series data alone. Similarly, the null hypothesis is $(\alpha_{-1}, \alpha_0, \alpha_1, \alpha_2) =$ (0.000, 0.025, -0.345, 0.000) taken from (G1 00**), and the rest of the parameter, σ, is estimated using time-series data. The likelihood-ratio statistic is 0.30, and the null is not rejected at any conventional significance level. Identifying the shape of $\mu(r)$ by introducing cross-sectional constraints does not therefore entail cost for the G1-type models. This can also be confirmed in Figure 2(c). The shape of $\mu(r)$ estimated using time-series data indeed resembles that for (G1 00**).

The estimates of β_2, on the other hand, are significant for both (G1 ****) and (G1

00**): -6.05 and -3.37, respectively. While they appear to differ, the difference does not yet arise in a practical sense for the reasonable range of r. Looking at Figure 2(d), both shapes almost coincide, except at the very low level of r where the discrepancy appears to be exaggerated for the following reasons: the estimate of α_{-1} for (G1 ****) is imprecise and there is no observation of r below 0.03 in the in-sample data. Furthermore, comparing the shape of $\mu^Q(r)$ for (G1 00**) to that for (GF 0), *i.e.*, the dotted lines between Panels (b) and (d), we see little difference. Nonlinear risk-neutral drift indeed seems robust.

(G1 ****) allows us to test for whether a single term in $\lambda(r)$ is sufficient for explaining short-yield data. If this is the case, the specification of $\lambda(r)$ for (GF *) turns out to be not so restrictive. Then, the resulting nonlinear physical drift, though it is more or less restrictive for time-series data alone, can be justified in terms of a better fit to short-yield data.

First, we test for $\lambda_1 = 0$, which is equivalent to $\alpha_1 = \beta_1$. The model labeled (G1**0*) is also a special case of (GF *) where $\gamma = 1$ is placed. The result of the likelihood-ratio test is that (G1 **0*) is strongly rejected against (G1 ****): the test statistic is 15.7. Of particular note is the significantly negative estimate of α_2, -7.67, shown in Table 2. Nonlinear physical drift is thus implied, however, as an outcome of the constraint that is hardly supported by the data. Next, we test for $\lambda_2 = 0$, which is equivalent to $\alpha_2 = \beta_2$. Hence, we at the same time test for whether the speed of mean-reversion for high rates is identical in both measures. The test statistic (p-value, d.f.) is 5.97 (0.015, 1). The null is therefore rejected at the 5% significance level but not at the 1% level. The estimation results for (G1 ***0) are also presented in Table 2. The estimate of α_2 is -5.50 and significant, which is in line with our expectations, as $\beta_2 < 0$ is consistently supported by the data. Nonlinear physical drift is thus implied with the constraint not as strong as that of $\lambda_1 = 0$.

The reason for the difference in the test results can be explained as follows. By setting $\lambda_1 = 0$ ($\alpha_1 = \beta_1$), α_2 is the only unshared parameter in $\mu(r)$, and hence supposed to be adjusted to time-series dimensions of the data. However, α_2 is constrained as $\alpha_2 < \beta_2 < 0$ due to a positive term premium implied by the data. The lack of flexibility in α_2 results in the strong rejection. On the other hand, α_1 is the analogous parameter in $\mu(r)$ by setting $\lambda_2 = 0$ ($\alpha_2 = \beta_2$). The extent to which α_1 is constrained by cross-sectional dimensions is weaker: $\alpha_1 < \beta_1$. The strong rejection is therefore avoided.

Apart from the inverse term, the statistical significance of which cannot be obtained in our data, the nonlinearity in $\mu(r)$ depends on the significance of the quadratic term. At present, the conditions under which α_2 becomes significant can be summarized as follows:

[C1] α_2 is linked to the coefficient of the lower-order term in $\lambda(r)$, and

[C2] $\alpha_2 = \beta_2$.

[C1] holds for (GF *) and (G1 **0*), while [C2] holds for (G1 ***0). In both cases, the quadratic term in $\mu(r)$ is strongly linked to cross-sectional dimensions of the data. As we have seen, however, these links are more or less restrictive. Conversely, if these links are removed, the significance of α_2 depends largely on time-series dimensions of the data. If the significance were indeed obtained from time-series data alone, it would become more pronounced for short-yield data. Otherwise, the significance cannot be expected even though cross-sectional constraints are introduced, as is the case for (G1 ****)

4.3 Estimation results for more general models

We further examine whether α_2 is indeed significant under [C1] or [C2], using more general models where the parsimony assumption that $\mu(r)$ and $\mu^Q(r)$ are of the same functional form is withdrawn. First, we add a linear term to (3) while treating γ as a free parameter, *i.e.*, $\lambda(r) = \lambda_2 r^{2-\gamma} + \lambda_3 r$. Then, the SDE for the risk-neutral process of the short-rate is

$$dr_t = (\alpha_{-1}/r_t + \alpha_0 + \alpha_1 r_t + \beta_2 r_t^2 + \beta_3 r_t^{\gamma+1})dt + \sigma r_t^\gamma dW_t^Q \ , \tag{12}$$

where $\beta_2 = \alpha_2 - \sigma\lambda_2$ and $\beta_3 = -\sigma\lambda_3$. [C1] holds for this model. The estimate (standard error) of α_2 is 11.64 (6.94), and not significant. Thus, [C1] does not seem to be decisive in general. The insignificant estimate is actually preferable in this case, as the positive value of α_2 leads to the possibility that the short-rate in the physical measure diverges. The result of $\alpha_2 > 0$ is due to cross-sectional constraints. Specifically, although α_2 is still constrained by a positive term premium, the constraint is no longer $\alpha_2 < \beta_2 < 0$, as previously implied for (GF *) and (G1 **0*). But rather, it is $\alpha_2 < \beta_2$. This is because $\mu^Q(r)$ has the term of higher than quadratic order, so fast mean-reversion at high levels in the risk-neutral measure is possible without $\beta_2 < 0$. Indeed, the estimates of β_2 and β_3 are 39.41 and -52.50, respectively. Both positive term premium and fast mean-reversion are thus implied, however, at the cost of the unfavorable shape of $\mu(r)$.

By placing $\lambda_2 = 0$ on the above model, [C2] holds. The estimate (standard error) of α_2 is -8.30 (3.00). Nonlinear physical drift is thus implied under [C2]. The estimate of β_3 is 4.99, which is consistent with a positive term premium. However, $\mu^Q(r)$ is now increasing for high rates, leading to the possibility that the short-rate in the risk-neutral measure diverges. Not surprisingly, the null of $\lambda_2 = 0$ is strongly rejected, indicating that the significance of α_2 is an outcome of the undesirable constraint.

Conversely, we remove [C1] and [C2]. We add a constant term to (3) while treating γ as a free parameter, *i.e.*, $\lambda(r) = \lambda_1 + \lambda_2 r^{2-\gamma}$. Then, the SDE for the risk-neutral process of the short-rate is

$$dr_t = (\alpha_{-1}/r_t + \alpha_0 + \alpha_1 r_t + \beta_1 r_t^{\gamma} + \beta_2 r_t^2)dt + \sigma r_t^{\gamma} dW_t^Q \ , \tag{13}$$

where $\beta_1 = -\sigma\lambda_1$ and $\beta_2 = \alpha_2 - \sigma\lambda_2$. We note that α_2 is no longer linked to λ_1, the coefficient of the lower-order term in $\lambda(r)$. We label the model (GF-GEN ****) (the γ-free general model). The asterisks are defined analogously to those for (G1 ****). Table 3 presents the estimation results, which are indeed similar to those for (G1 ****). First, the estimate of α_2 is 2.23 and insignificant. Second, the null hypothesis of $\alpha_{-1} = \alpha_2 = 0$ is not rejected at any conventional significance level. Third, the null of $\lambda_1 = 0$ is strongly rejected. It is noted that the null model, (GF-GEN **0*), is equivalent to (GF *). Then, nonlinear physical drift implied for (GF *), which satisfies [C1], is also an outcome of the undesirable constraint. Fourth, the null of $\lambda_2 = 0$ is also rejected, but not as strongly as that of $\lambda_1 = 0$: the likelihood-ratio statistic (p-value, d.f.) is 6.96 (0.008, 1). The estimation results for the null model, (GF-GEN ***0), are presented in Table 3. The estimate of α_2 is -4.66 and significant. Nonlinear physical drift is thus implied under [C2], however, as an outcome of the undesirable constraint.

As we have seen, [C1] and [C2] are easily removed by simply adding a constant term to (3). Then, with more flexible specifications of $\lambda(r)$, models that are free from [C1] and [C2] can also easily be created. While nonlinear physical drift is implied for some parsimonious (restricted) models, it does not seem compulsory in general.

4.4 Sub-sample analysis

We check the robustness of the results using sub-sample data for the following three sub-periods: (i) 01/1971–12/1989 (991 obs.), (ii) 01/1990–12/1999 (522 obs.), and (iii) 06/1973–02/1995 (1134 obs.). The choice of period (i) is based on CKLS (1992), in which the sample period is chosen by the end of 1989. Period (ii) can be referred to as the post-CKLS period. Period (iii) is taken from Aït-Sahalia (1996). Due to space limitations, the estimation results are briefly reported, but the details are available upon request.

For period (i), both $\mu(r)$ and $\mu^Q(r)$ are nonlinear due to a positive and significant estimate of α_{-1}, indicating that the short-rate mean-reverts for low rates in both measures. The result is not surprising, as interest rates observed in this period do not remain at low levels, which can naturally be interpreted as a consequence of mean-reversion. As for the quadratic terms, on the other hand, similar results are obtained: the estimates of β_2 are consistently negative and significant, whereas the estimates of α_2 are significant for parsimonious models satisfying [C1] or [C2]. For period (ii), neither $\mu(r)$ nor $\mu^Q(r)$ is estimated to be nonlinear. The result is not surprising, either, as both the short-rate and spread are in the middle range, which can reasonably be explained without nonlinear terms that generate faster mean-reversion at the extreme levels of the short-rate. For period (iii), the results are very similar to those using the full-sample data. Overall, while the extent to which the nonlinearities are implied depends on sample periods, the results of the sub-sample analysis are basically in line with our claims. In particular, although the estimate of β_2 is not significant for period (ii), this is not contradictory to the claim that the quadratic term in $\mu^Q(r)$ provides an additional degree of freedom in explaining interest rates at the extreme levels.

5 Out-of-Sample Prediction

We examine whether nonlinear drifts contribute to the prediction of future interest-rate levels using out-of-sample data from January 5, 2000 to December 28, 2005 (313 obs.). This period includes both easing and tightening cycles, so reliable results concerning the predictive power of models are expected. Figure 3(a) graphs the time-series of the one-month rate, a proxy for the short-rate, and the spread between the one- and six-month rates in the out-of-sample period. We see clearly that there are periods in which the level of the short-rate is rapidly changing and those in which it is relatively stable. Of particular note is that during the entire 2001 the short-rate continues to decrease to the level of below 0.02, which is lower than the in-sample minimum, 0.03. In the subsequent period, it gradually decreases further to around 0.01, and remains at this extremely low level. We then expect that models with mean-reversion will forecast poorly. The short-rate finally starts to rise in June 2004, and the rapid rise continues during the rest of the out-of-sample period.

The levels of the spread, on the other hand, are relatively stable compared with those of the short-rate. Large negative spreads are observed around early 2001 when the short-rate starts to decrease. The spread fluctuates around zero in the subsequent period when the short-rate further decreases and remains at very low levels. Just before the short-rate starts to rise, the level of the spread rises, and then becomes stable over the period of rapid rise in the short-rate. We then expect that models with nonlinear risk-neutral drift, which allows for keeping spreads small, will forecast accurately.

5.1 Comparison criteria

We consider the following h-step ahead model prediction errors:

$$
\begin{aligned}
e_{1,t+h\Delta} &= r_{t+h\Delta} - E_t[r_{t+h\Delta}] \,, && (14)\\
e_{2,t+h\Delta} &= r^2_{t+h\Delta} - E_t[r^2_{t+h\Delta}] \,, && (15)\\
e_{3,\tau,t+h\Delta} &= y_{\tau,t+h\Delta} - E_t[Y(r_{t+h\Delta},\tau)] \,, && (16)\\
e_{4,\tau,t+h\Delta} &= y^2_{\tau,t+h\Delta} - E_t[Y(r_{t+h\Delta},\tau)^2] \,, && (17)
\end{aligned}
$$

where the conditional expectation is taken with respect to the physical measure for each model. First, e_1 and e_2 are prediction errors for the level and squared level of the short-rate. The reason for examining e_2 is to pay attention also to the impact of volatility specification on the prediction. Second, e_3 and e_4 are prediction errors for the level and squared level of a τ-maturity yield. The total predictive power of a model can be measured by the magnitude of e_3 and e_4. When it is small, it follows that predictions of both the short-rate and yields in the cross-section are precise. On the other hand, when we wish to focus solely on the descriptive power of term structure models, we examine

$$
e_{5,\tau,t} = y_{\tau,t} - Y(r_t,\tau) \,, \tag{18}
$$

where r_t is taken from out-of-sample data, not a predicted value of a model.

When analytical expressions of the conditional moments are unavailable, we also employ the method proposed in Aït-Sahalia (2002, eq.(4.3)), maintaining the consistency in the evaluation of the density and moments. Here, we truncate the series at the second order, because we consider forecasting periods of up to six months, much longer than a week in computing the transition density function. We further note that the approximation, $\tilde{Y}(r_{t+h\Delta,\tau})$, is used instead of $Y(r_{t+h\Delta,\tau})$ when the latter is unavailable in closed form. This seems justified given the accuracy of the approximation for short maturities.

To measure the magnitude of these prediction errors, we consider the following criteria: mean absolute error (MAE), mean absolute percentage error (MAPE), root mean squared error (RMSE), and root mean squared percentage error (RMSPE). Since the performance rankings based on the MAE (MAPE) criterion generally agree with those based on the RMSE (RMSPE) criterion, we report the results based on the MAE and MAPE criteria. The MAE and MAPE are computed as

$$\text{MAE}: \ \frac{1}{N-h}\sum_t |e_{i,\tau,t+h\Delta}| \ , \qquad \text{MAPE}: \ \frac{1}{N-h}\sum_t \left|\frac{e_{i,\tau,t+h\Delta}}{x_{i,\tau,t+h\Delta}}\right| \ , \tag{19}$$

where $i = 1, ..., 5$ and $\tau = (0, 0.25, 0.5)$ ($\tau = 0$ corresponds to the short-rate), and where $(x_1, x_2, x_3, x_4, x_5) = (r, r^2, y_\tau, y_\tau^2, y_\tau)$.

5.2 Competing models

Among the nonlinear drift models estimated from weekly data, we choose (GF 0) and (G1 00**): both $\mu(r)$ and $\mu^Q(r)$ are nonlinear for the former and only $\mu^Q(r)$ is nonlinear for the latter. We also consider as benchmarks simpler models that are arbitrage-free. Specifically, we adopt the models proposed by Vasicek (1977), and CIR (1985):

$$\text{(Vasicek)} \qquad dr_t = (\alpha_0 + \alpha_1 r_t)dt + \sigma dW_t \qquad \text{and} \quad \lambda(r_t) = \lambda_0 + \lambda_1 r_t \ , \tag{20}$$

$$\text{(CIR)} \qquad dr_t = (\alpha_0 + \alpha_1 r_t)dt + \sigma\sqrt{r_t}\, dW_t \qquad \text{and} \quad \lambda(r_t) = \lambda_1\sqrt{r_t} \ . \tag{21}$$

Parameters of these models, which are estimated by the same procedure as outlined in Section 3.2, are as follows:

$$\begin{aligned} &\text{(Vasicek)} \quad (\alpha_0,\, \alpha_1,\, \sigma,\, \lambda_0,\, \lambda_1) &&= (0.034,\, -0.448,\, 0.030,\, -0.868,\, 3.069)\ , \\ &\text{(CIR)} \qquad (\alpha_0,\, \alpha_1,\, \sigma,\, \lambda_1) &&= (0.057,\, -0.761,\, 0.096,\, -2.605)\ . \end{aligned}$$

We also consider models in which the short-rate in the physical measure follows a martingale, *i.e.*, $\mu(r) = 0$. Such models are worth considering, as the null of $\mu(r) = 0$ is not rejected at conventional significance levels using time-series data alone. Besides, taking the short-rate behavior observed in the out-of-sample period into consideration, zero physical drift (ZPD) may actually be appropriate for prediction. [5] The ZPD models are obtained as special cases of (Vasicek) and (G1 00**), which are labeled as (Vas. ZPD) and (G1 ZPD) for brevity, respectively. [6] For (Vas. ZPD), we further set $\lambda_1 = 0$ to distinguish it from (Vasicek) also with respect to whether or not the short-rate in the risk-neutral measure mean-reverts. [7] Parameter estimates for (Vas. ZPD) and (G1 ZPD) are as follows:

$$\begin{aligned} &\text{(Vas. ZPD)} \quad (\sigma,\, \lambda_0) &&= (0.030,\, -0.562)\ , \\ &\text{(G1 ZPD)} \quad (\beta_1,\, \beta_2,\, \sigma) &&= (0.723,\, -5.666,\, 0.324)\ . \end{aligned}$$

5.3 Comparison results of the forecasting performance

We consider $h = 1, 4, 13$, and 26, *i.e.*, the one-week, four-week, three-month, and six-month forecasting periods. Tables 4 and 5 present MAEs ($\times 10^4$) and MAPEs ($\times 10^2$), respectively. In each row, the smallest and second smallest numbers are expressed in bold and italic. To ease interpretation of the results, Panels (b) through (f) of Figure 3 graph the time-series of some prediction errors, $e_{1,t+h\Delta}$, $e_{2,t+h\Delta}$, $e_{3,\tau,t+h\Delta}$, $e_{4,\tau,t+h\Delta}$, and $e_{5,\tau,t}$ for $\tau = 0.5$ and $h = 26$, from (Vasicek), (Vas. ZPD), and (G1 ZPD).

First, the results of the forecasting performance on the short-rate level are presented in Panel A of Tables 4 and 5. For all forecasting periods, the ZPD models, (Vas. ZPD) and (G1 ZPD), outperform the other models with mean-reversion. The result is in fact not surprising. The mean-reverting models predict that the short-rate will rise when it is currently at around the in-sample minimum, 0.03. In reality, however, the short-rate changes in the opposite direction and remains at the extremely low level, as shown in Figure 3(a). As a result, these models forecast poorly compared to the ZPD models that do not predict increase (nor decrease) in the short-rate. Indeed, looking at Figure 3(b), the gap in the forecasting performance between (G1 ZPD) and (Vasicek) is most pronounced when the short-rate is extremely low. Conversely, limiting our attention to the period of rapid rise in the short-rate, the mean-reverting models have a superior forecasting performance. If this period further continues, the gap in the forecasting performance between the ZPD and mean-reverting models will narrow. Otherwise, the gap will remain, as both models perform equally well when the short-rate fluctuates around the long-term mean.

Taking into consideration the forecasting performance on the squared short-rate level, (G1 ZPD) is preferable to (Vas. ZPD). Panels B show that (G1 ZPD) outperforms (Vas. ZPD) for all forecasting periods. In particular, the superior performance is more pronounced in the MAPE criteria, as the magnitude of e_2 is much smaller when the (squared) short-rate level is extremely low, as shown in Figure 3(c). Moreover, (GF 0) and (G1 00**) also outperform (Vas. ZPD) for all forecasting periods in both criteria, except for one case where (GF 0) is outperformed in the MAE criteria for $h = 26$. These results indicate that level-dependent volatility is desirable also for the out-of-sample prediction. This, of course, holds true as long as the prediction of the short-rate level is moderately accurate. Despite level-dependent volatility, the forecasting performance of (CIR) is also poor, owing to the poor predictive power for the short-rate level.

The results of the forecasting performance on the level of the three- and six-month rates are presented in Panels C and D. (G1 ZPD) exhibits the best performance for all forecasting periods. (Vas. ZPD) ranks second. Predictions for (GF 0) and (G1 00**) are similar, and more accurate than those for (Vasicek) and (CIR). Looking at Figure 3(d), the superior performance of the ZPD models over (Vasicek) is also pronounced after the short-rate falls below the in-sample minimum.

The results of the forecasting performance on the squared level of the three- and six-month rates are presented in Panels E and F. (G1 ZPD) also ranks first, except $e_{4,0.5}$ for $h = 1$ in the MAE criterion. The superior performance is more pronounced in the MAPE criterion for the same reason as for e_2, which is also confirmed in Figure 3(e). (Vas. ZPD) generally ranks second. In the MAPE criterion, (GF 0) outperforms (Vas. ZPD) in some cases: $e_{4,0.25}$ for $h = 4, 13$. This is attributable to the superior forecasting performance of (GF 0) on the squared short-rate level for $h = 4, 13$, shown in Panel B.

The descriptive power of term structure models also has a crucial role in the prediction of yields. Looking at the magnitude of e_5 shown in Panels G, (G1 ZPD) exhibits the best performance and (Vas. ZPD) follows. The reason for the superior performance is attributable to little mean-reversion at low levels in the risk-neutral measure, which allows for capturing very small spreads for low rates actually observed throughout the out-of-sample period. In fact, the smaller the value of $\mu^Q(r)$ at $r = 0.01$, the smaller the MAE of e_5. For example, both values are smallest for (G1 ZPD) and largest for (Vasicek). We also confirm in Figure 3(f) the superior performance of (G1 ZPD) over (Vasicek), which is pronounced in the period when the short-rate is extremely low and the spread is around zero. Of particular note is that (G1 ZPD) still outperforms (Vasicek), even in the period of rapid rise in the short-rate when the latter outperforms the former in the prediction of the short-rate level. It then follows that a desirable shape of $\mu^Q(r)$ is robust throughout the in- and out-of-sample periods, in contrast with a desirable shape of $\mu(r)$ which seems to vary in different sub-periods.

In view of the out-of-sample prediction of both the short-rate and yields in the cross-section, the role of nonlinear risk-neutral drift is reconfirmed as decisive. On the other hand, zero physical drift seems more appropriate than mean-reverting drift when a relatively long out-of-sample period is chosen in our data. This suggests that caution is required in identifying the shape of $\mu(r)$ by introducing cross-sectional constraints.

6 Estimation Using Daily Data

To further examine the robustness of the previous findings, we estimate nonlinear drift models of the short-rate using daily data. One notable feature of higher-frequency data is that while on the one hand observations remain at a certain level for a while, they can change instantly and drastically on the other. Behind this behavior, there seems to exist a transitory component. If so, the nonlinearity may be exaggerated when the drift is estimated directly from daily observations, as pointed out by Jones (2003). In his study, a nonlinear drift model is estimated after controlling transitory shocks: the observed short-rate can deviate from, but mean-reverts to, an unobserved *true* rate that is free from transitory shocks, while the true rate is also a stochastic process with possibly nonlinear drift.

We also model transitory shocks explicitly, but differently from Jones (2003), as explained below. Of particular interest here is whether nonlinear risk-neutral drift is still preferable. As we have seen in previous sections using weekly data, the quadratic term in the risk-neutral drift provides an additional degree of freedom in explaining interest rates at the extreme levels. However, these rates are more likely affected by transitory shocks than are middle rates. Then, explicitly modeling the transitory component may be more effective than considering nonlinear risk-neutral drift.

The source of daily data is the same as that of weekly data. The sample period for estimation is from January 4, 1971 to December 30, 2005. Aside from the reason that we do not perform the out-of-sample prediction using daily data, we are interested in how parameter estimates, especially that of the inverse term in the drifts, are affected by including recent observations that exhibit mean-reverting behavior at low levels, as shown in Figure 3(a). If one or more of $(r_t,\ y_{3M,t},\ y_{6M,t})$ are missing at time t, we discard the whole time-t observations. We further exclude as outliers the observations on 12/30/1971 and 12/31/1971. Special treatments are not made for weekends, holidays, and missing observations. The time interval between successive observations is assumed to be constant and set to 1/252.

6.1 Model for daily data

We assume that the short-rate r_t is observed with noise, which is modeled by simply adding a noise process ϵ_t to the true, or smoothed, short-rate process r_t^*:

$$r_t = r_t^* + \epsilon_t \ . \tag{22}$$

We naturally treat $\{r_t^*, \epsilon_t\}$ as latent processes. The physical process of r_t^* is modeled by the SDE of the same form as before:

$$dr_t^* = (\alpha_{-1}/r_t^* + \alpha_0 + \alpha_1 r_t^* + \alpha_2 {r_t^*}^2)dt + \sigma {r_t^*}^\gamma dW_{1,\,t} \ , \tag{23}$$

where $W_{1,\,t}$ is Brownian motion in the physical measure. The physical process of ϵ_t is modeled as simply as possible, because its role is merely to separate transitory shocks from a daily-observed proxy for r_t. That is,

$$d\epsilon_t = \kappa\, \epsilon_t\, dt + \xi\, dW_{2,\,t} \ , \tag{24}$$

where $W_{2,\,t}$ is Brownian motion in the physical measure. We assume that $W_{1,\,t}$ and $W_{2,\,t}$ are mutually independent, for r_t^* and ϵ_t to be mutually independent. Also, we set the unconditional mean of ϵ_t to zero given $\kappa < 0$. The price of risk attributed to uncertain variation in r_t^* is also given by (3) or (4), where r_t is replaced by r_t^*. On the other hand, we assume that investors do not require the risk premium attributed to uncertain variation in ϵ_t. That is, the price of risk for ϵ_t is assumed to be zero.

It is more natural to consider that the yields in the cross-section are also affected by transitory shocks at a daily frequency. This impact, however, may decrease with increase in the maturity, as the one-month rate is the most volatile and the six-month rate is the least in our data. To incorporate this feature, we model the price of a discount bond as

$$\begin{aligned} P(r_t^*, \epsilon_t, t, T) &= E_t^Q\left[\exp\left\{-\int_t^T r_u du\right\}\right] \\ &= E_t^Q\left[\exp\left\{-\int_t^T (r_u^* + \epsilon_u) du\right\}\right] \\ &= E_t^Q\left[\exp\left\{-\int_t^T r_u^* du\right\}\right] E_t^Q\left[\exp\left\{-\int_t^T \epsilon_u du\right\}\right] \ , \end{aligned} \tag{25}$$

where E_t^Q stands for the conditional expectation with respect to the risk-neutral measure, and where the last equality follows from the assumption that r_t^* and ϵ_t are mutually independent. In contrast with one-factor models, the noise process systematically enters into the stochastic discount factor, which leads to the desired property, as explained below.

The first expectation on the right-hand side of (25) is further developed to the analytical approximation for estimation. The second expectation has a closed-form expression as the Vasicek model. Then, the approximation of a τ-maturity yield is given by

$$\tilde{Y}(r_t^*, \epsilon_t, \tau) = \frac{1}{\tau}\left\{A(\tau; r_t^*) + B(\tau; r_t^*) r_t^* + C(\tau) + D(\tau)\epsilon_t\right\} \ , \tag{26}$$

where $A(\tau; r_t^*)$ and $B(\tau; r_t^*)$ are given in Appendix, and $C(\tau)$ and $D(\tau)$ are given by

$$D(\tau) = -\frac{1 - e^{\kappa\tau}}{\kappa}, \tag{27}$$

$$C(\tau) = -\left(\frac{\xi}{2\kappa}\right)^2 \left\{\kappa D(\tau)^2 - 2D(\tau) + 2\tau\right\}. \tag{28}$$

We note that given $\kappa < 0$, $D(\tau)/\tau$ approaches zero as τ increases. Therefore, the longer the maturity, the less the yield is affected by ϵ_t. The constant term, $C(\tau)/\tau$, remains, however.

6.2 Inversion and estimation method

We recover (r_t^*, ϵ_t) from observed variables through the two equations. One is (22) with the one-month rate used as a proxy for r_t, and the other is (26) with the left-hand side replaced with an observed yield. That is, we assume that either the three-month or the six-month rate is explained exactly by the model at any point in time. For simplicity, we denote the yield used for inversion as y_t. Since (26) is nonlinear in r_t^*, we have to numerically solve the two equations for (r_t^*, ϵ_t), but the convergence is achieved with a few iterations.

We also employ the ML method. The density for a time-series part is based on (23) and (24). The density for a cross-sectional part is based on the following regression model: $\hat{y}_t = \tilde{Y}(r_t^*, \epsilon_t, \hat{\tau}) + v\, u_t$, where $\hat{y}_t$ stands for the yield not used for inversion, and where $u_t = \phi\, u_{t-\Delta} + \sqrt{1-\phi^2} z_t$, $z_t \sim i.i.d.\ N(0, 1)$. The unconditional variance of u_t is set to one given $|\phi| < 1$, for v to represent the magnitude of the measurement error. [8]

The density function at time t conditioned on time $t - \Delta$ is expressed as

$$\begin{aligned} f(r_t, y_t, \hat{y}_t \,|\, r_{t-\Delta}, y_{t-\Delta}) &= f(r_t^*, \epsilon_t, \hat{y}_t \,|\, r_{t-\Delta}^*, \epsilon_{t-\Delta}) \begin{vmatrix} \dfrac{dr_t}{dr_t^*} & \dfrac{dy_t}{dr_t^*} \\ \dfrac{dr_t}{d\epsilon_t} & \dfrac{dy_t}{d\epsilon_t} \end{vmatrix}^{-1} \\ &= f_T(r_t^*, \epsilon_t \,|\, r_{t-\Delta}^*, \epsilon_{t-\Delta}) \left|\frac{dy_t}{d\epsilon_t} - \frac{dy_t}{dr_t^*}\right|^{-1} f_C(\hat{y}_t \,|\, r_t^*, \epsilon_t) \\ &= f_{T,r^*}(r_t^* \,|\, r_{t-\Delta}^*) f_{T,\epsilon}(\epsilon_t \,|\, \epsilon_{t-\Delta}) \left|\frac{dy_t}{d\epsilon_t} - \frac{dy_t}{dr_t^*}\right|^{-1} f_C(u_t \,|\, r_t^*, \epsilon_t)|v|^{-1}. \end{aligned} \tag{29}$$

The second equality follows from the decomposition into the time-series (marginal) and cross-sectional (conditional) components and from the fact that $\hat{y}_t$ is explained by (r_t^*, ϵ_t), which are Markovian. In addition, $dr_t/dr_t^* = dr_t/d\epsilon_t = 1$ holds from (22). The last equality follows from the assumption that r_t^* and ϵ_t are mutually independent.

The transition density f_{T,r^*} is computed after discretizing (23) by the Euler method. At a daily frequency, the crude first-order approximation is reported to work well: see, for example, Jones (2003, Appendix C). As for $f_{T,\epsilon}$ and f_C, the closed-form expressions are available. Finally, $dy_t/d\epsilon_t$ and dy_t/dr_t^* in the Jacobian have been evaluated and stored in the iteration procedure for recovering (r_t^*, ϵ_t).

6.3 Empirical results for daily data

To focus on the extent to which the nonlinearities will change for the smoothed short-rate, we report the estimation results for the model with γ free, where both the physical and risk-neutral drifts exhibit marked nonlinearities for weekly data. For notational convenience, we label the model with the three-month (six-month) rate used for inversion (GF-TC 3) ((GF-TC 6)) (the model with γ free and with the transitory component). For comparison, we also estimate (GF *). That is, we directly estimate the behavior of the daily-observed short-rate using the model adopted for weekly data.

Table 6 presents the results. The behavior of r_t is more mean-reverting for daily data than for weekly data. The estimates of α_2 and β_2 for (GF *) are -9.43 and -6.65, which are larger in absolute value than those using weekly data, -7.09 and -4.04. In addition, the estimate of γ is 1.51, which is significantly increased from 1.30. That is, the short-rate mean-reverts more rapidly for high rates and the volatility is more elastic to changes in r_t. The same can be said for the case in which weekly data over the same period, 1971–2005, are used for estimating (GF *): the estimates of α_2, β_2, and γ are -8.36, -5.28, and 1.28, respectively. These results are in line with the aforementioned properties of higher-frequency data, and hence suggest the presence of the transitory component, which is supposed to dissipate at a weekly frequency. This is indeed the case, as we see below.

Another important result is the positive and significant estimate of α_{-1}, 0.00011. We note that using weekly data over the same period, the estimate of α_{-1} is -0.00006 and insignificant. Figure 4 graphs the drift functions estimated from daily data. They, however, do not imply fast mean-reversion at the sample minimum, 0.0094. The result is actually preferable in both statistical and economic senses: the observed patterns of the spread can be explained, yet the short-rate does not reach zero in finite time.

When the noise process is explicitly modeled and the three-month rate is used for inversion, both the speed of mean-reversion for high rates and the elasticity of volatility return to the level implied by weekly data. First, the estimates of α_2 and β_2 are -6.40 and -3.53, which are slightly decreased from those using weekly data, shown in Table 1. We see clearly in Figure 4 that compared with the case in which the model is fitted directly to r_t, the rate of decrease in both the physical and risk-neutral drifts at high levels of r^* is moderate. At the very low level, both drifts for (GF-TC 3) become negative due to the negative estimate of α_{-1}. The negative estimate of α_{-1} is undesirable, as the smoothed short-rate possibly goes through the lower boundary of zero. We note, however, that the negative estimate, with various precisions, is also reported by previous studies: Durham (2003, Table 7) when bond yield data are used, and Jones (2003, Tables 1 and 4) when the Jeffreys prior is used. At least, evidence in favor of fast mean-reversion at low levels cannot be obtained for r_t^*, either. Second, the estimates of σ and γ are 0.71 and 1.33, which are also close to those using weekly data. Thus, the volatility of changes in r_t^* is similar to, but actually slightly lower than, that of changes in r_t observed at a weekly frequency.

The behavior of the noise process, ϵ_t, on the other hand, is indeed transitory. The estimate of κ, which measures the speed of mean-reversion, is -83.25. The mean half-life is about 0.0083, or equivalently 2.10 days by multiplying 252. At a weekly frequency, therefore, the impact of ϵ_t on r_t is minor. The mean half-life is nonetheless longer than that estimated by Jones (2003) using the seven-day Eurodollar rate, the reason for which is explained as follows. We use for estimation the longer-term rates and model the noise process such that it can affect these rates. The noise process does not therefore die out too quickly.

When the six-month rate is used for inversion, the overall picture of the results changes from that for (GF-TC 3). First, although the estimates of α_2 and β_2 are both significant, their absolute magnitude becomes smaller: -2.82 and -0.54. In particular, the reduction in the absolute value of β is substantial, and the risk-neutral drift does not become negative even at the sample maximum of r^*, 0.1981, as shown in Figure 4(b). At the very low level of r^*, on the other hand, we notice that both $\mu(r^*)$ and $\mu^Q(r^*)$ for (GF-TC 6) become more negative. This problem may not be as serious as it seems, however, as r^* implied by (GF-TC 6) does not become too small: the sample minimum is 0.0135, which is larger than that implied by (GF-TC 3), 0.0103. Second, while the estimate of γ is little changed, that of σ is significantly decreased, suggesting that the smoothing is excessive. Third, and related to the oversmoothing of r_t, the persistence of ϵ_t is significantly increased. The estimate of κ is -23.11, indicating that the mean half-life is increased to 0.030, or equivalently 7.56 days.

In Figure 5, the time-series of $\{r_t^*, \epsilon_t\}$ are plotted for both (GF-TC 3) and (GF-TC 6).

We note that $r_t = r_t^* + \epsilon_t$ holds by construction, and hence that the time-series of the one-month rate is recovered for the entire period. Although it is difficult to visually observe differences between the two figures, the smoothed short-rate for (GF-TC 6) appears to be less volatile, particularly at high levels of r^*. To compensate for this smaller variation, the noise process appears to be more volatile.

Behind the difference in the persistence of ϵ_t is the difference in statistical properties between the three- and six-month rates. Specifically, the six-month rate behaves more differently from the one-month rate than does the three-month rate, as partly reflected in Figure 1. In addition, we have observed in Tables 1–3 that the estimates of v_2 are always larger than those of v_1. These indicate that the six-month rate requires larger measurement errors when explained by the one-month rate. Nevertheless, when a model is fitted exactly to the six-month rate, the noise process behaves more like an alternative factor for explaining the cross-sectional relation, rather than like transitory shocks affecting the time-series variation in r_t. Thus, the noise process does not die out quickly to be an effective explanatory variable for the six-month rate. The difference in the behavior of r_t^* can be explained in line with this argument. In particular, since the noise process for (GF-TC 6) more effectively helps explain the cross-sectional relation, the significance of β_2 is reduced more pronouncedly.

In terms of separating transitory shocks from the short-rate observed at a daily frequency, using the three-month rate for inversion seems more appropriate, as the resulting behavior of the smoothed short-rate is not much changed from that of the short-rate observed at a weekly frequency.

7 Concluding Remarks

We have estimated the behavior of the short-rate using data on the one-, three-, and six-month Eurodollar deposit rates with the method of maximum likelihood. The cross-sectional relation between the short-rate and yields was obtained by an analytical approximation, which is accurate for short maturities.

We found that nonlinear physical drift is implied when nonlinear terms are strongly linked to cross-sectional dimensions of the data. Specifically, the quadratic term becomes significant to be consistent with both the time-series behavior of the short-rate and a positive term premium in the cross-section of yields. The links, however, are more or less restrictive. Besides, they can be easily removed. Without the links, the nonlinearity is difficult to identify with precision. In view of the out-of-sample prediction, zero physical drift is actually preferable. Although zero physical drift may be better recognized as a part of nonlinear drift that produces strong mean-reversion outside the historically observed range of the short-rate, it is difficult to find strong evidence supporting for nonlinear physical drift from a statistical perspective. Nonlinear risk-neutral drift, on the other hand, is strongly supported by the data. In particular, the quadratic term is the key to explaining both large spreads for high rates and small spreads for low rates, which linear drift models cannot adequately explain. Evidence supporting for nonlinear risk-neutral drift is also confirmed by daily data, where transitory shocks are removed, although the nonlinearity is somewhat reduced.

While this study presented an attempt to estimate nonlinear drift models of the short-rate using data on the short-end of the term structure, there are a number of limitations, which indicate the necessity of further studies. First, we limit the maturity spectrum to the short-end. Of more fundamental interest is whether nonlinear drift is implied by the *entire* term structure. For the analysis, multi-factor models are preferable, which allow for more interesting (but certainly more complicated) tests for the (non)linearity in the drift of the short-rate process together with the (non)linearity in the drifts of other factor processes. Second, we limit the volatility specification by assuming the constant elasticity of volatility. More realistic specifications of the volatility, such that it is driven by another Brownian motion, are expected to significantly improve the time-series fit. Whether or not they can contribute to a better fit to the cross-section of yields is of interest. Furthermore, the volatility behavior implicit in term structure data may be worth exploring. Third, we limit the number of regimes to one. In other words, model parameters are fixed throughout the sample period. It is more realistic to consider different regimes of interest rates, specifically high and low volatility regimes: see, for example, Gray (1996), Ang and Bekaert (2002), and Bansal and Zhou (2002). When cross-sectional data are also included, the length of the data is not very crucial for efficient estimation, which alleviates the difficulties in regime-switching models.

Appendix: Approximation of the Nonlinear Term Structure

We briefly explain how to derive an approximation of the nonlinear term structure proposed by Takamizawa and Shoji (2003). First, we approximate the short-rate process in the risk-neutral measure. By applying the Ito formula to $\mu^Q(r_u, u)$ where $u \in [t, t+\tau]$ for $\tau > 0$, we have

$$\mu^Q(r_u, u) = \mu^Q(r_t, t) + \int_t^u \frac{\partial \mu^Q}{\partial r}(r_s, s)dr_s + \int_t^u \left\{ \frac{1}{2}\frac{\partial^2 \mu^Q}{\partial r^2}(r_s, s)\sigma^2(r_s, s) + \frac{\partial \mu^Q}{\partial s}(r_s, s) \right\} ds \ . \tag{30}$$

Fixing integrands at time t leads to

$$\tilde{\mu}^Q(r_u, u) = \mu^Q(r_t, t) + \frac{\partial \mu^Q}{\partial r}(r_t, t)(r_u - r_t) + \left\{ \frac{1}{2}\frac{\partial^2 \mu^Q}{\partial r^2}(r_t, t)\sigma^2(r_t, t) + \frac{\partial \mu^Q}{\partial t}(r_t, t) \right\} (u - t) \ , \tag{31}$$

where $\tilde{\mu}^Q$ on the left-hand side clarifies the approximation. Collecting terms provides

$$\tilde{\mu}^Q(r_u, u) = a_2(t)r_u + a_1(t)u + a_0(t) \ , \tag{32}$$

where

$$a_2(t) = \frac{\partial \mu^Q}{\partial r}(r_t, t) \ , \tag{33}$$

$$a_1(t) = \frac{1}{2}\frac{\partial^2 \mu^Q}{\partial r^2}(r_t, t)\sigma^2(r_t, t) + \frac{\partial \mu^Q}{\partial t}(r_t, t) \ , \tag{34}$$

$$a_0(t) = \mu^Q(r_t, t) - a_2(t)r_t - a_1(t)t \ . \tag{35}$$

Similarly, we approximate $\sigma^2(r_u, u)$ as $\tilde{\sigma}^2(r_u, u) = b_2(t)r_u + b_1(t)u + b_0(t)$, where $b_i(t)$ is provided analogously to $a_i(t)$ with μ^Q replaced by σ^2.

Let $\{\tilde{r}_u : u \in [t, t+\tau]\}$ be an approximate process of the short-rate having $\tilde{\mu}^Q$ and $\tilde{\sigma}^2$. By construction, $\tilde{r}_t = r_t$ holds. The SDE is then

$$d\tilde{r}_u = \{a_2(t)\tilde{r}_u + a_1(t)u + a_0(t)\} du + \sqrt{b_2(t)\tilde{r}_u + b_1(t)u + b_0(t)}\, dW_u^Q \ . \tag{36}$$

Under the approximate process, the price of a discount bond, $\tilde{P}(\tilde{r_u}, u, t+\tau)$, is the solution to the following PDE (the coefficients are abbreviated as a_i and b_i $(i = 0, 1, 2)$):

$$\frac{1}{2}\{b_2\tilde{r}_u + b_1u + b_0\} \frac{\partial^2 \tilde{P}}{\partial \tilde{r}_u^2}(\tilde{r}_u, u, t+\tau) + \{a_2\tilde{r}_u + a_1u + a_0\} \frac{\partial \tilde{P}}{\partial \tilde{r}_u}(\tilde{r}_u, u, t+\tau) + \frac{\partial \tilde{P}}{\partial u}(\tilde{r}_u, u, t+\tau) - \tilde{r}_u\tilde{P}(\tilde{r}_u, u, t+\tau) = 0 \ , \tag{37}$$

with the boundary condition $\tilde{P}(\tilde{r}_{t+\tau}, t+\tau, t+\tau) = 1$. Since all the coefficients of the partial derivatives become linear in $\tilde{r}_u$, the closed-form solution can be derived as for affine term structure models. Specifically, we assume the solution to be $\tilde{P}(\tilde{r}_u, u, t+\tau) = \exp\{-A(u,t+\tau) - B(u,t+\tau)\tilde{r}_u\}$. By appropriately differentiating the above expression and then substituting the resulting derivatives into the above PDE, we obtain ordinary differential equations for $A(u,t+\tau)$ and $B(u,t+\tau)$ with boundary conditions $A(t+\tau,t+\tau) = B(t+\tau,t+\tau) = 0$. Through somewhat tedious calculation, the solution is

$$\tilde{P}(r_t,\tau) = \exp\{-A(\tau;r_t) - B(\tau;r_t)r_t\}\,, \tag{38}$$

where

$$\begin{aligned} B(\tau;r_t) &= \frac{2(e^{\psi\tau}-1)}{g(\tau)}\,, &(39)\\ A(\tau;r_t) &= \frac{2}{b_2^2}(b_2\mu^Q(r,t) - a_2\sigma^2(r,t) + b_1)(\ln g(\tau) - \ln 2\psi - \frac{k_1\tau}{2}) - \frac{2k_3\tau}{b_2^2}\ln 2\psi \\ &\quad + \frac{1}{b_2}(\sigma^2(r,t) - b_2 r)(B(\tau)-\tau) - \frac{\tau^2}{2b_2^2}(b_1b_2 + k_1k_3) + \frac{2k_3}{b_2^2}\int_0^\tau \ln g(u)du\,, &(40)\end{aligned}$$

and where $\psi = (2b_2 + a_2^2)^{0.5}$, $g(\tau) = k_1e^{\psi\tau} + k_2$, $k_1 = \psi - a_2$, $k_2 = \psi + a_2$, and $k_3 = b_2a_1 - b_1a_2$.

Endnote

1 It is noted that caution is required in interpreting their results. Regardless of estimation techniques involved, it is originally unrealistic to obtain linear drift estimators from the artificial data. In the parametric case, for example, the linear drift is obtained only when the coefficients of nonlinear terms, such as α_{-1} and α_2 of the inverse and quadratic terms, are exactly zero, which, however, is unlikely in the numerical experiments. Nevertheless, when estimated drift functions are plotted against the level of the short-rate, nonlinear shapes are exaggerated. This is true even though α_{-1} is almost negligible but non-zero (say, 10^{-8}) when the horizontal axis begins at sufficiently near zero.

2 More than two terms in $\lambda(r)$ seem unnecessary for explaining short-yield data used here. We also specify $\lambda(r) = \lambda_1 r^{1-\gamma} + \lambda_2 r^{2-\gamma} + \lambda_3 r^{3-\gamma}$. We estimate models with and without the last term in $\lambda(r)$ while setting $\gamma = 1$, and then test for $\lambda_3 = 0$ based on the likelihood ratio. The null is not rejected at any conventional significance level.

3 This estimation framework is similar to one of those commonly used for affine term structure models: see, for example, Chen and Scott (1993), Pearson and Sun (1994), and Duffee (2002).

4 We note that for the G1-type models $\mu(r)$ is not necessarily estimated precisely. In the previous version of this paper, we estimated the models using data over the period 1971–2003. The result is that the null of $\mu(r) = 0$ is not rejected at conventional significance levels.

5 It may be more appropriate to state that zero physical drift is effective for the historically observed range of the short-rate. Outside this range, (strong) mean-reversion may be more desirable from an economic perspective, as mentioned in Introduction.

6 We do not consider the ZPD models based on (CIR) or (GF 0), as the assumption of ZPD is too restrictive. For (CIR), the assumption leads to $\mu^Q(r) = -\lambda\sigma r$, which is increasing for $\lambda < 0$ implied by a positive term premium. For (GF 0), only the quadratic term is left in $\mu^Q(r)$.

7 We do not consider a model in which the short-rate in the risk-neutral measure follows a martingale, *i.e.*, $dr_t = \sigma dW_t^Q$, as we can hardly recognize it as a basis for a term structure model. Since σ is determined exclusively by time-series dimensions of the data, as is noted earlier, the model has virtually no descriptive power for the cross-section of yields.

8 We also pin down the autocorrelation parameter, ϕ, by the same procedure as outlined in Section 3.2. When the three-month (six-month) rate is exactly fitted, the estimate of ϕ is 0.926 (0.858).

Reference

Aït-Sahalia, Y. (1996), "Testing continuous-time models of the spot interest rate," *Review of Financial Studies*, 9, 385-426.

Aït-Sahalia, Y. (1999), "Transition densities for interest rate and other nonlinear diffusions," *Journal of Finance*, 54, 1361-1395.

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

Ang, A., and Bekaert, G. (2002), "Short rate nonlinearities and regime switches," *Journal of Economic Dynamics and Control*, 26, 1243-1274.

Bansal R., and Zhou, H. (2002), "Term structure of interest rates with regime shifts," *Journal of Finance*, 57, 1997-2043.

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.

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

Chen, R., and Scott, L. (1993), "Maximum likelihood estimation for a multifactor equilibrium model of the term structure of interest rates," *Journal of Fixed Income*, 3, 14-31.

Conley, T. G., Hansen, L. P., Luttmer, E. G. J., and Scheinkman, J. A. (1997), "Short-term interest rates as subordinated diffusions," *Review of Financial Studies*, 10, 525-577.

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

Duffee, G. R. (2002), "Term premia and interest rate forecasts in affine models," *Journal of Finance*, 57, 405-443.

Durham, G. B. (2003), "Likelihood-based specification analysis of continuous-time models of the short-term interest rate," *Journal of Financial Economics*, 70, 463-487.

Gray, S. F. (1996), "Modeling the conditional distribution of interest rates as a regime-switching process," *Journal of Financial Economics*, 42, 27-62.

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.

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

Li, M., Pearson, N. D., and Poteshman, A. M. (2004), "Conditional estimation of diffusion processes," *Journal of Financial Economics*, 74, 31-66.

Pearson, N. D., and Sun, T. (1994), "Exploiting the conditional density in estimating the term structure: An application to the Cox, Ingersoll, and Ross model," *Journal of Finance*, 49, 1279-1304.

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

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

Takamizawa, H., and Shoji, I. (2003), "Modeling the term structure of interest rates with general short-rate models," *Finance & Stochastics*, 7, 323-335.

Vasicek, O. A. (1977), "An equilibrium characterization of the term structure," *Journal of Financial Economics*, 5, 177-188.

	(Time)		(GF *)		(GF 0)	
α_{-1} $\times 10^2$	0.056	(0.191)	0.019	(0.061)	0.000	
α_0	-0.027	(0.107)	0.008	(0.026)	0.015	(0.008)
α_1	0.709	(1.777)	0.439	(0.336)	0.358	(0.184)
α_2	-5.270	(8.818)	-7.085	(1.558)	-6.830	(1.350)
β_2			-4.035	(1.321)	-3.782	(0.954)
σ	0.734	(0.190)	0.720	(0.184)	0.723	(0.180)
γ	1.311	(0.103)	1.301	(0.101)	1.303	(0.099)
v_1 $\times 10^2$			0.408	(0.020)	0.408	(0.020)
v_2 $\times 10^2$			0.450	(0.018)	0.450	(0.018)
$\sum \ln f_T$	6607.7		6601.99		6602.25	
$\sum \ln f_C$			15105.81		15105.51	
LogL	6607.7		21707.80		21707.76	

Table 1: Parameter estimates (standard errors) are presented. The SDE for the short-rate process in the physical measure is given by

$$dr_t = (\alpha_{-1}/r_t + \alpha_0 + \alpha_1 r_t + \alpha_2 r_t^2)dt + \sigma r_t^{\gamma} dW_t \ ,$$

with the risk-neutral drift given as $\alpha_{-1}/r_t + \alpha_0 + \alpha_1 r_t + \beta_2 r_t^2$. The data consist of weekly observations for the one-, three-, and six-month Eurodollar deposit rates covering the period from January 6, 1971 to December 29, 1999 (1513 obs.).

	(Time)		(G1 ****)		(G1 00**)		(G1 **0*)		(G1 ***0)	
$\alpha_{-1} \times 10^2$	0.054	(0.256)	0.232	(0.163)	0.000		0.115	(0.163)	0.163	(0.160)
α_0	−0.010	(0.142)	−0.058	(0.061)	0.025	(0.007)	−0.024	(0.064)	−0.038	(0.062)
α_1	0.251	(2.313)	0.369	(0.652)	−0.345	(0.131)	0.766	(0.745)	0.596	(0.708)
α_2	−2.591	(11.169)	−0.940	(3.112)	0.000		−7.672	(2.685)	−5.503	(2.434)
β_1			1.077	(0.699)	0.202	(0.172)	0.766		0.896	(0.717)
β_2			−6.048	(2.377)	−3.373	(0.937)	−5.114	(2.530)	−5.503	
σ	0.325	(0.012)	0.326	(0.012)	0.325	(0.012)	0.328	(0.012)	0.327	(0.012)
$v_1 \times 10^2$			0.409	(0.020)	0.410	(0.020)	0.410	(0.020)	0.409	(0.020)
$v_2 \times 10^2$			0.450	(0.018)	0.451	(0.018)	0.451	(0.018)	0.451	(0.018)
$\sum \ln f_T$	6581.4		6578.8		6581.3		6575.0		6577.9	
$\sum \ln f_C$			15117.0		15112.0		15113.0		15115.0	
LogL	6581.4		21695.8		21693.3		21688.0		21692.9	

Table 2: Parameter estimates (standard errors) are presented. The SDE for the short-rate process in the physical measure is given by $dr_t = (\alpha_{-1}/r_t + \alpha_0 + \alpha_1 r_t + \alpha_2 r_t^2)dt + \sigma r_t dW_t$, with the risk-neutral drift given as $\alpha_{-1}/r_t + \alpha_0 + \beta_1 r_t + \beta_2 r_t^2$. The data consist of weekly observations for the one-, three-, and six-month Eurodollar deposit rates covering the period from January 6, 1971 to December 29, 1999 (1513 obs.).

	(GF-GEN ****)		(GF-GEN 00**)		(GF-GEN ***0)	
α_{-1} $\times 10^2$	0.062	(0.160)	0.000		0.033	(0.155)
α_0	0.009	(0.063)	0.027	(0.006)	0.008	(0.059)
α_1	-0.438	(0.784)	-0.384	(0.120)	0.158	(0.652)
α_2	2.230	(3.916)	0.000		-4.663	(2.072)
β_1	2.039	(1.095)	1.498	(0.638)	0.640	(0.232)
β_2	-6.048	(2.073)	-5.004	(1.552)	-4.663	
σ	0.715	(0.199)	0.721	(0.175)	0.704	(0.189)
γ	1.301	(0.111)	1.304	(0.097)	1.294	(0.107)
v_1 $\times 10^2$	0.407	(0.020)	0.408	(0.020)	0.408	(0.020)
v_2 $\times 10^2$	0.449	(0.018)	0.449	(0.018)	0.450	(0.018)
$\sum \ln f_T$	6605.4		6606.5		6604.1	
$\sum \ln f_C$	15109.5		15107.6		15107.3	
LogL	21714.9		21714.1		21711.4	

Table 3: Parameter estimates (standard errors) are presented. The SDE for the short-rate process in the physical measure is given by

$$dr_t = (\alpha_{-1}/r_t + \alpha_0 + \alpha_1 r_t + \alpha_2 r_t^2)dt + \sigma r_t^\gamma dW_t \ ,$$

with the risk-neutral drift given as $\alpha_{-1}/r_t + \alpha_0 + \alpha_1 r_t + \beta_1 r_t^\gamma + \beta_2 r_t^2$. The data consist of weekly observations for the one-, three-, and six-month Eurodollar deposit rates covering the period from January 6, 1971 to December 29, 1999 (1513 obs.).

	(Vasicek)	(CIR)	(GF 0)	(G1 00**)	(Vas. ZPD)	(G1 ZPD)
Panel A: Prediction errors for the short-rate: e_1						
$h = 1$	5.84	7.98	5.30	*5.25*	**4.18**	**4.18**
$h = 4$	20.42	29.41	18.18	*18.06*	**13.78**	**13.78**
$h = 13$	62.10	90.21	55.50	*54.47*	**41.21**	**41.21**
$h = 26$	118.89	169.14	109.97	*104.77*	**79.71**	**79.71**
Panel B: Prediction errors for the squared short-rate: e_2						
$h = 1$	0.48	0.47	0.36	*0.34*	0.37	**0.31**
$h = 4$	1.72	1.72	1.20	*1.14*	1.25	**1.00**
$h = 13$	5.41	5.72	3.73	*3.52*	3.85	**3.03**
$h = 26$	10.68	11.85	7.64	*7.05*	7.50	**5.80**
Panel C: Prediction errors for the three-month rate: $e_{3,0.25}$						
$h = 1$	51.17	51.66	24.97	31.76	*17.89*	**15.77**
$h = 4$	64.21	72.05	37.08	42.49	*26.29*	**24.26**
$h = 13$	103.09	129.73	73.27	76.25	*54.30*	**52.90**
$h = 26$	157.35	205.39	127.77	124.21	*92.62*	**92.14**
Panel D: Prediction errors for the six-month rate: $e_{3,0.50}$						
$h = 1$	89.66	88.40	42.93	56.33	*30.66*	**27.87**
$h = 4$	102.10	107.81	54.85	66.77	*38.80*	**35.66**
$h = 13$	138.93	162.59	90.13	98.36	*66.76*	**64.64**
$h = 26$	191.27	235.53	144.36	144.33	*104.29*	**103.87**
Panel E: Prediction errors for the squared three-month rate: $e_{4,0.25}$						
$h = 1$	2.96	2.87	1.59	1.91	*1.23*	**1.19**
$h = 4$	4.23	4.26	2.46	2.64	*2.10*	**1.83**
$h = 13$	7.97	8.53	5.07	5.02	*4.77*	**3.99**
$h = 26$	13.41	15.05	9.25	8.65	*8.48*	**6.86**
Panel F: Prediction errors for the squared six-month rate: $e_{4,0.50}$						
$h = 1$	5.54	5.35	2.86	3.58	**2.07**	*2.17*
$h = 4$	6.82	6.83	3.77	4.35	*2.96*	**2.77**
$h = 13$	10.61	11.35	6.49	6.74	*5.69*	**4.99**
$h = 26$	16.18	18.15	10.92	10.52	*9.50*	**7.98**
Panel G: Errors for yields given actual short-rate data						
$e_{5,0.25}$	46.92	44.82	20.43	27.56	*16.05*	**13.59**
$e_{5,0.50}$	85.66	81.95	38.44	52.26	*28.84*	**25.86**

Table 4: Mean absolute errors (MAE, $\times 10^4$) for the out-of-sample prediction are presented. In each row, the smallest and second smallest numbers are expressed in bold and italic. e_1 and e_2 are prediction errors for the level and squared level of the short-rate. $e_{3,\tau}$ and $e_{4,\tau}$ are prediction errors for the level and squared level of a τ-maturity yield. $e_{5,\tau}$ is the difference between observed and theoretical τ-maturity yields, given data on the short-rate. The out-of-sample period is from January 5, 2000 to December 28, 2005 (313 obs.).

	(Vasicek)	(CIR)	(GF 0)	(G1 00**)	(Vas. ZPD)	(G1 ZPD)
Panel A: Prediction errors for the short-rate: e_1						
$h = 1$	2.92	4.36	*2.41*	2.49	**1.61**	**1.61**
$h = 4$	10.84	16.87	*8.75*	9.05	**5.22**	**5.22**
$h = 13$	34.77	53.41	*28.73*	28.84	**15.64**	**15.64**
$h = 26$	70.17	103.19	61.83	*58.92*	**31.83**	**31.83**
Panel B: Prediction errors for the squared short-rate: e_2						
$h = 1$	11.43	9.89	*5.01*	5.23	7.12	**3.31**
$h = 4$	46.23	42.31	*19.31*	20.19	27.92	**10.94**
$h = 13$	160.98	166.47	*73.40*	75.03	95.75	**35.19**
$h = 26$	353.07	400.41	190.58	*186.12*	211.53	**79.89**
Panel C: Prediction errors for the three-month rate: $e_{3,0.25}$						
$h = 1$	28.93	29.28	12.29	16.57	*8.99*	**6.39**
$h = 4$	36.34	41.09	18.44	22.57	*12.07*	**9.70**
$h = 13$	59.07	75.42	38.17	41.64	*22.79*	**20.82**
$h = 26$	92.69	122.20	71.03	70.71	*39.09*	**37.60**
Panel D: Prediction errors for the six-month rate: $e_{3,0.50}$						
$h = 1$	49.33	48.71	20.88	28.81	*15.51*	**11.13**
$h = 4$	56.30	59.65	26.90	34.65	*18.44*	**14.31**
$h = 13$	77.52	91.35	46.03	52.79	*28.98*	**25.58**
$h = 26$	108.83	134.53	77.67	80.26	*44.76*	**42.20**
Panel E: Prediction errors for the squared three-month rate: $e_{4,0.25}$						
$h = 1$	76.08	72.97	27.10	37.65	*24.29*	**13.53**
$h = 4$	113.31	112.44	*43.13*	54.54	44.77	**21.22**
$h = 13$	234.02	252.22	*103.41*	116.53	112.70	**48.80**
$h = 26$	431.21	496.93	229.70	236.19	*228.24*	**97.81**
Panel F: Prediction errors for the squared six-month rate: $e_{4,0.50}$						
$h = 1$	142.41	136.93	49.10	71.24	*39.91*	**24.19**
$h = 4$	180.08	180.01	66.63	90.14	*59.70*	**32.14**
$h = 13$	300.41	325.40	130.76	156.50	*125.62*	**61.47**
$h = 26$	492.46	566.76	259.62	278.66	*236.68*	**112.07**
Panel G: Errors for yields given actual short-rate data						
$e_{5,0.25}$	26.51	25.33	10.02	14.25	*8.33*	**5.45**
$e_{5,0.50}$	47.05	45.05	18.64	26.54	*14.83*	**10.21**

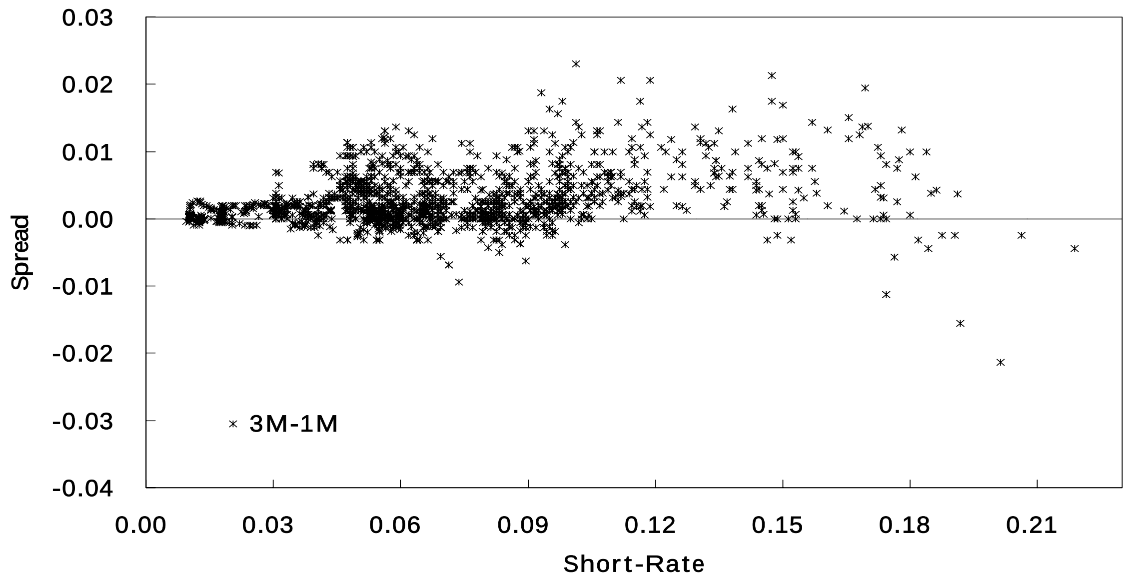
Table 5: Mean absolute percentage errors (MAPE, $\times 10^2$) for the out-of-sample prediction are presented. In each row, the smallest and second smallest numbers are expressed in bold and italic. e_1 and e_2 are prediction errors for the level and squared level of the short-rate. $e_{3,\tau}$ and $e_{4,\tau}$ are prediction errors for the level and squared level of a τ-maturity yield. $e_{5,\tau}$ is the difference between observed and theoretical τ-maturity yields, given data on the short-rate. The out-of-sample period is from January 5, 2000 to December 28, 2005 (313 obs.).

	(GF *)		(GF-TC 3)		(GF-TC 6)	
$\alpha_{-1} \times 10^2$	0.011	(0.002)	−0.026	(0.002)	−0.047	(0.008)
α_0	−0.020	(0.002)	0.014	(0.001)	0.024	(0.006)
α_1	1.006	(0.042)	0.395	(0.0004)	0.021	(0.074)
α_2	−9.431	(0.347)	−6.401	(0.716)	−2.815	(0.606)
β_2	−6.651	(0.218)	−3.525	(0.091)	−0.540	(0.152)
σ	0.919	(0.034)	0.713	(0.029)	0.595	(0.053)
γ	1.512	(0.013)	1.334	(0.014)	1.331	(0.031)
κ			−83.25	(2.463)	−23.11	(0.787)
ξ			0.046	(0.0004)	0.048	(0.0004)
$v \times 10^2$			0.252	(0.002)	0.181	(0.001)
$v_1 \times 10^2$	0.321	(0.002)				
$v_2 \times 10^2$	0.356	(0.003)				
LogL	134029		138811		138747	

Table 6: Parameter estimates (standard errors) are presented. (GF *) is a model adopted for weekly data. In (GF-TC), the observed short-rate is modeled as $r_t = r_t^* + \epsilon_t$. The SDEs for r_t^* and ϵ_t in the physical measure are given by

$$dr_t^* = (\alpha_{-1}/r_t^* + \alpha_0 + \alpha_1 r_t^* + \alpha_2 {r_t^*}^2)dt + \sigma {r_t^*}^\gamma dW_{1,t} \ , \quad \text{and} \quad d\epsilon_t = \kappa\, \epsilon_t\, dt + \xi\, dW_{2,t} \ ,$$

where $W_{1,t}$ and $W_{2,t}$ are mutually independent Brownian motions. The risk-neutral drift of r_t^* is given as $\alpha_{-1}/r_t^* + \alpha_0 + \alpha_1 r_t^* + \beta_2 {r_t^*}^2$. The latent processes, $\{r_t^*, \epsilon_t\}$, are recovered from the one-month rate and either the three-month rate (GF-TC 3) or the six-month rate (GF-TC 6). The daily data covers the period from January 4, 1971 to December 30, 2005 (8883 obs.).



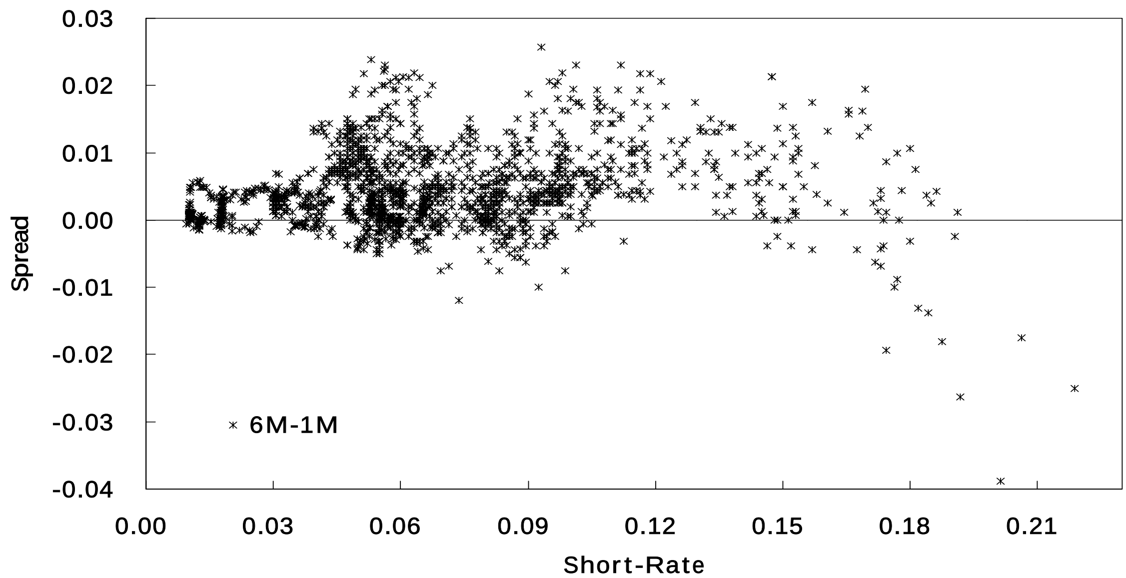


Figure 1: Spreads of the three-month (above) and six-month (below) Eurodollar deposit rates over the one-month rate are plotted against the level of the one-month rate. Weekly data (on a Wednesday basis) covers the period from January 6, 1971 to December 28, 2005 (1825 obs.).

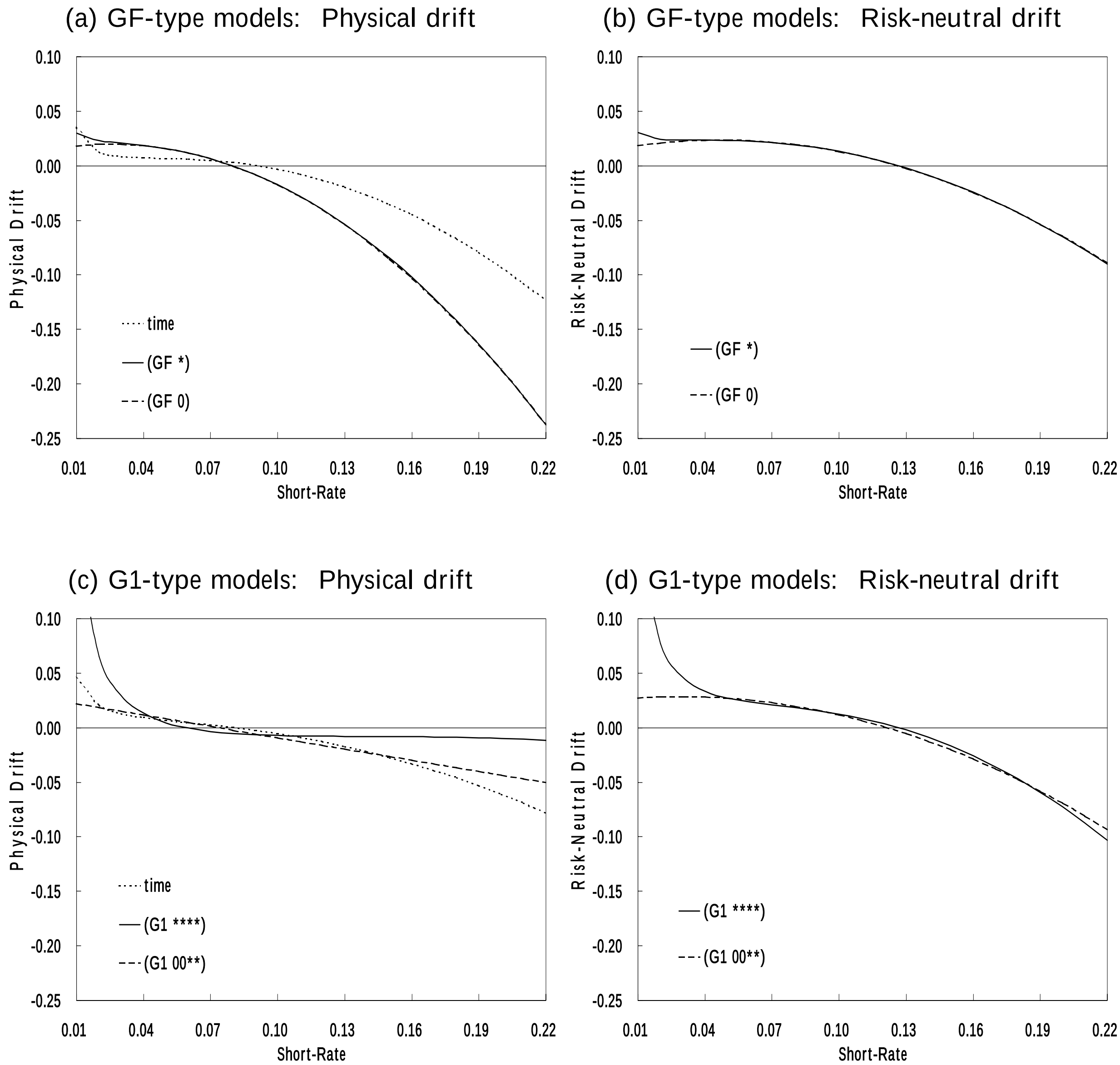


Figure 2: The physical and risk-neutral drift functions are plotted for the GF-type models (upper panels) and the G1-type models (lower panels). The dotted lines labeled “time” correspond to the physical drift estimated using time-series data alone.

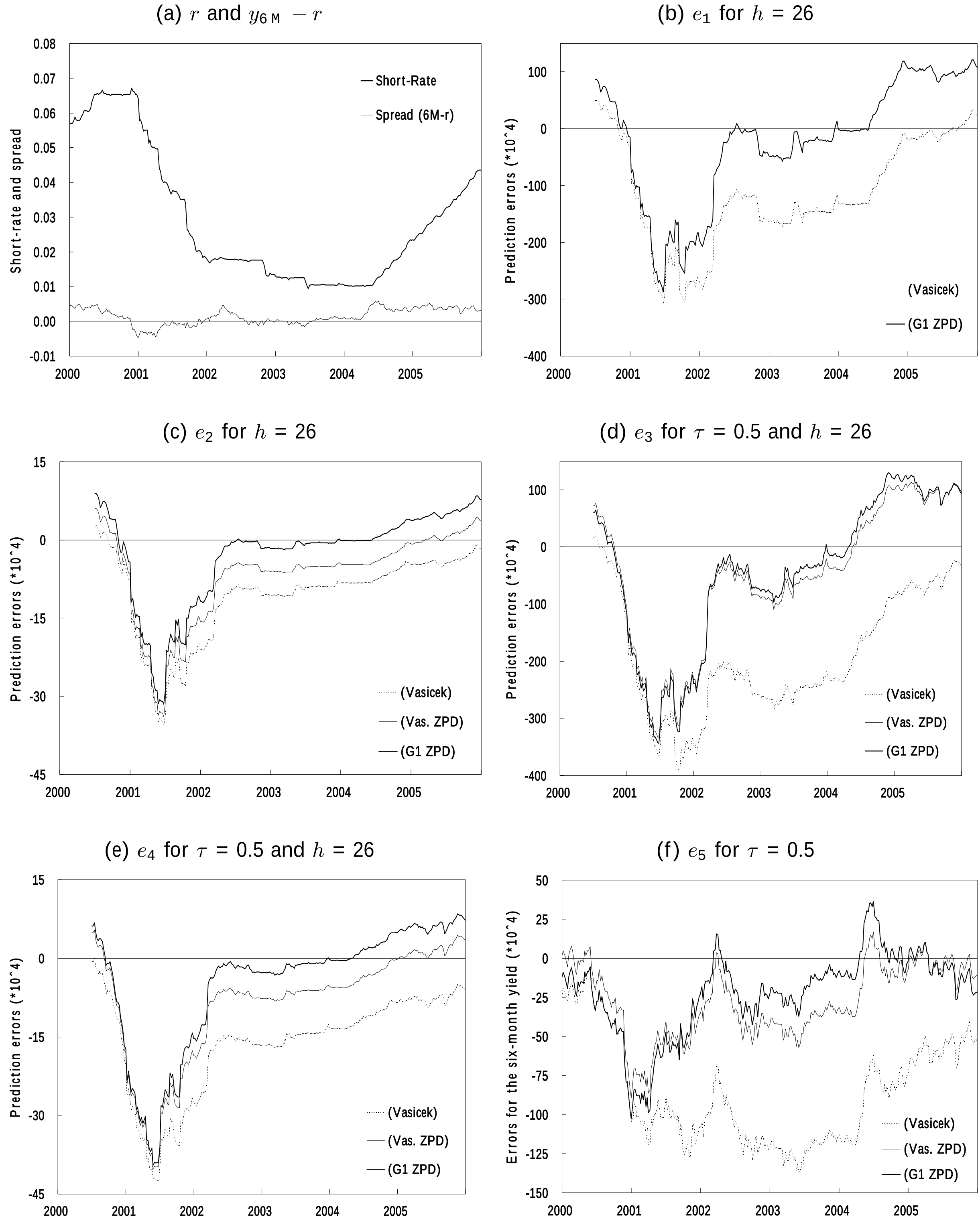


Figure 3: Panel (a) plots the time-series of the one-month rate, a proxy for r, and the spread between the one- and six-month rates. Panels (b) through (f) plot the time-series of some prediction errors from (Vasicek), (Vas. ZPD), and (G1 ZPD) for the 26-week prediction period (h = 26). The selected errors are those for (b) the short-rate, (c) the squared short-rate, (d) the six-month-rate, (e) the squared six-month rate, and (f) the six-month rate given short-rate data.

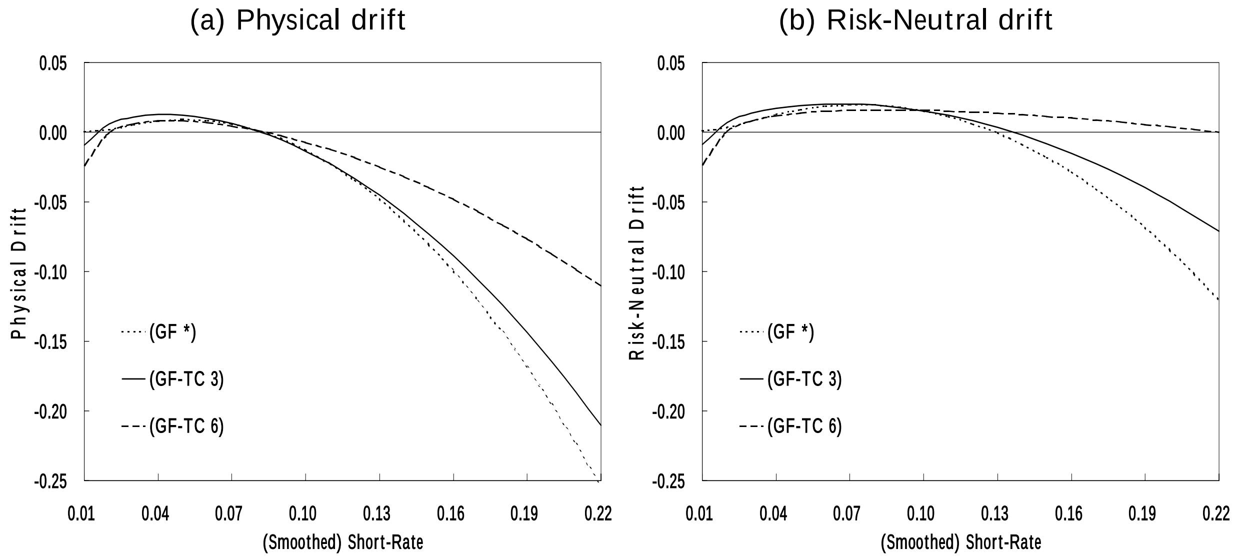
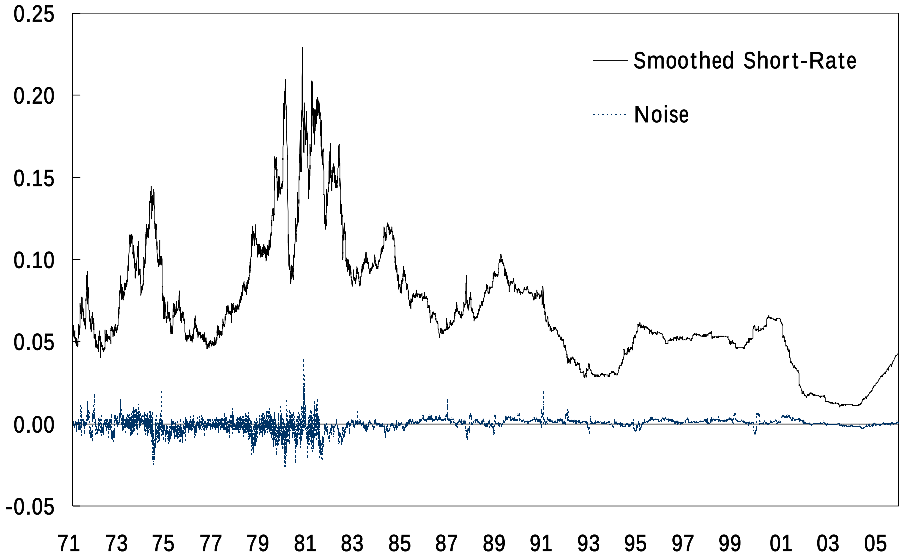


Figure 4: The physical and risk-neutral drift functions are plotted for (GF *), (GF-TC 3), and (GF-TC 6). (GF *) is a model adopted for weekly data. (GF-TC 3) and (GF-TC 6) accommodate a transitory component: the former (latter) uses for inversion the three-month (six-month) rate.

(a) (GF-TC 3)



(b) (GF-TC 6)

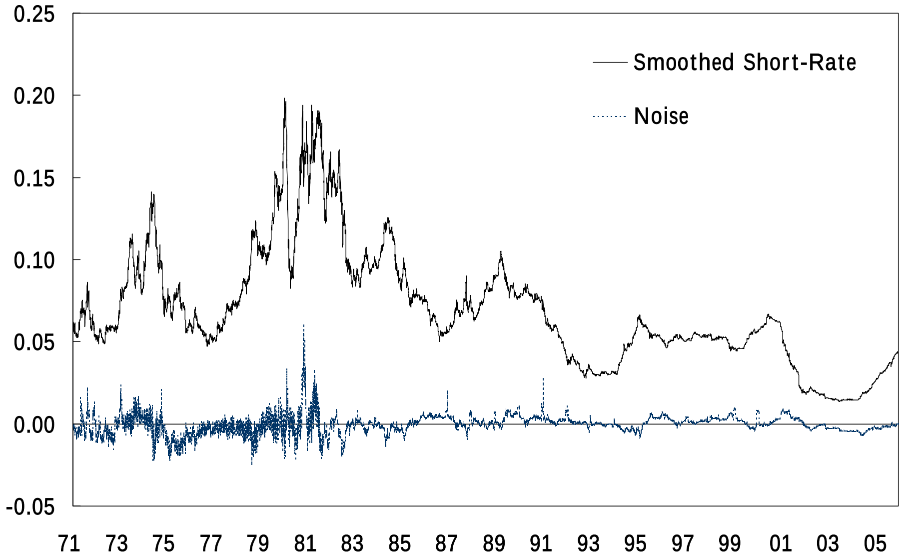


Figure 5: The time-series of the smoothed short-rate and the noise process are plotted for (GF-TC3) and (GF-TC6): the former (latter) uses the three-month (six-month) rate for extracting the latent processes. The sum of the two processes provides the time-series of the observed one-month rate by construction. The daily data cover the period from January 4, 1971 to December 30, 2005 (8883 obs.)