

A Term Structure Model of Interest Rates with Quadratic Volatility

Hideyuki Takamizawa *

Graduate School of Commerce and Management,

Hitotsubashi University

Kunitachi Tokyo 186-8601, JAPAN

E-mail: Hi.Takamizawa@r.hit-u.ac.jp

This draft: December 9, 2017

Abstract

This study proposes a no-arbitrage term structure model that can capture the volatility of interest rates without sacrificing the goodness-of-fit to the cross-section and predictive ability about the level of interest rates. The key feature of the model is the covariance matrix of changes in factors, which is specified as quadratic functions of factors. The quadratic specification can capture intense volatility even with spanned factors, which is not the case for the affine specification. Furthermore, since the quadratic specification guarantees the positive definiteness of the covariance matrix without restricting the sign of factors, it allows for a flexible specification of the physical drift as does the Gaussian term structure model, contributing also to accurate level prediction.

Keywords: Term structure, Interest rate, Volatility, Affine model, Prediction.

JEL codes: C58, E43, G12, G17.

*I am grateful to two anonymous referees for comments, which enabled me to improve the paper. Also, the support from JSPS KAKENHI Grant (15K03538), Tokio Marine Kagami Memorial Foundation Research Grant, and JCER Research Grant are gratefully acknowledged. All errors are my own.

1 Introduction

No-arbitrage affine term structure models are known to have difficulties in that stochastic volatility factors, when extracted from the cross-section of interest rates, do not behave similarly with standard volatility measures. In particular, it is often the case that the volatility implied by the models fluctuates too little or correlates negatively with standard measures; see, for example, Collin-Dufresne, Goldstein, and Jones (CDGJ) (2009), and Jacobs and Karoui (2009). The purpose of this study is to propose non-affine models that can capture interest-rate volatility without sacrificing the goodness-of-fit to the cross-section and predictive ability about the level of interest rates.

There are mainly two sources of the problem of volatility implied by the affine models. The first is specification: the covariance matrix of changes in factors is specified as a linear function of factors. The second is estimation: some interest rates are assumed to be measured without error to extract factors from the cross-section. By combining these two sources, it is implied that a conditional variance of an interest rate, for any horizon and maturity, can be described by a linear combination of interest rates. This implication, however, is statistically rejected by Andersen and Benzoni (2010).

To improve the fit to the volatility, previous studies modify the affine models in terms of specification, or estimation, or both. Collin-Dufresne and Goldstein (2002) propose an affine term structure model with unspanned stochastic volatility factors that do not appear in the cross-section of interest rates and hence are free to fit the time-series of interest rates or the cross-section of bond derivatives. They show that originally spanned stochastic volatility factors in affine models can be made unspanned by imposing certain parameter constraints. The constraints, however, are usually at the cost of reducing the cross-sectional fit, leading to statistical rejection in favor of unconstrained affine models; see, for example, Bikbov and Chernov (2009), and Thompson (2008). Thompson (2008) proposes an estimation method that prevents affine models with stochastic volatility factors from overfitting the cross-section. This method, however, is not a fundamental solution of the trade-off between fitting the time-series and cross-section as it reduces the latter.

Our approach to resolving the trade-off is to change the specification of volatility while maintaining the assumption that some interest rates are measured without error for the cross-sectional fit. Hence, all factors in the proposed models are spanned by interest rates, which means that none of the factors may behave like volatility. The question then is how to drive the volatility by non-volatility factors. Our answer is to use quadratic functions of factors as a driving force of

volatility. An intuition behind this specification is an ARCH model, where the volatility is driven by the squared change of an interest rate (after subtracting the conditional mean) between time t and time $t + \Delta$. This change may be well captured by squaring and appropriately combining many interest rates at time t . In fact, using time-series models of interest rates, Takamizawa (2015) shows evidence in favor of a quadratic specification of volatility. However, he does not show that this evidence can be extended to no-arbitrage models, in which the factors are treated as latent variables and hence should be estimated simultaneously.

In addition to the potential of resolving the trade-off between fitting the time-series and cross-section of interest rates, the quadratic specification might resolve another trade-off between predicting the level and volatility of interest rates, a trade-off within time-series properties. The affine models also have this trade-off. Intuitively, as Duffee (2002) demonstrates, the Gaussian term structure model, which is an affine model with constant volatility, is good at predicting the level of interest rates. However, it cannot then explain time-varying volatility. The second trade-off stems from the conditions to keep stochastic volatility factors nonnegative. The so-called admissibility conditions include the following: (i) the instantaneous correlations between volatility factors are zero; (ii) the constant term in the drift of change in a volatility factor is positive; (iii) the drift of change in a volatility factor does not depend on non-volatility factors that can change signs; and (iv) the drift of change in a volatility factor can depend on the other volatility factors but only positively. The more the number of volatility factors, the more binding are the admissibility conditions. The Gaussian model is the only affine model that is free of the admissibility conditions. Similar to the Gaussian model, the proposed models do not need the admissibility conditions because the quadratic specification guarantees the positive definiteness of the covariance matrix. In this manner, the proposed models can inherit the strength of the Gaussian model in terms of predicting the level of interest rates while overcoming its weakness of constant volatility.

An obvious weakness of the proposed models, on the other hand, is that there is no closed-form solution for no-arbitrage bond prices and yields. This study overcomes this weakness by relying on an approximation method proposed by Takamizawa and Shoji (2009), which approximates conditional moments of multi-dimensional diffusion processes as the solution to ordinary differential equations. Since bond prices are computed as conditional expectation of a stochastic discount factor, it is straightforward to apply this method. The accuracy of the approximation is maintained for the proposed models with reasonable parameter values.

Empirical analysis shows that the proposed models indeed predict the volatility better than the

affine models while having a similar descriptive power for the cross-section and predictive power for the level. The affine and proposed models are both estimated by the quasi-maximum likelihood method with the assumption that some interest rates are measured without error to maintain the descriptive power for the cross-section. The proposed models attain larger likelihood values than the affine models. The larger values are attributed to improving not the cross-sectional fit but the time-series fit. The improvement in the time-series fit leads to a higher predictive accuracy for the volatility of interest rates, especially with long maturities. However, it does not raise the predictive accuracy for the level of interest rates.

The rest of the manuscript is structured as follows. Section 2 introduces first the affine models and then the proposed models. Section 3 explains estimation and prediction methods. Sections 4 through 6 report empirical results of estimation, volatility prediction, and level prediction, respectively. Section 7 concludes. Technical arguments on the approximation of no-arbitrage bond prices and its accuracy are collected in Appendices.

2 Models

The purpose of this study is to propose no-arbitrage term structure models that can capture the volatility without sacrificing the goodness-of-fit to the cross-section of interest rates. To achieve this purpose, we change the covariance matrix of changes in factors from affine to quadratic specifications. Meanwhile, we do not change the drift vectors in both the risk-neutral and physical probability measures from the affine specification. This is aimed at controlling for the drift vectors and attributing the improvement, if any, to the covariance matrix.

The affine models are first introduced. They can be classified by the number of factors driving the volatility. Using the notation of affine models proposed by Dai and Singleton (2000), let $A_m(n)$ denote an affine term structure model that has n factors in total, among which m factors drive the volatility. This study uses $n = 3$ following Litterman and Scheinkman (1991) and the subsequent studies, and then considers $m = 1, 2, 3$. For each m , the affine covariance matrix is replaced by a quadratic counterpart. The same quadratic specification is used for all m . However, the effect of the quadratic specification will differ in m because the severity of the admissibility conditions imposed on the original affine models differs in m .

Section 2.1 presents $A_m(3)$ ($m = 1, 2, 3$) and Section 2.2 introduces quadratic specifications into each of the affine models. Section 2.3 provides a brief explanation of how to compute no-arbitrage

bond prices by the proposed models.

2.1 Affine models

Let X_t be a three-dimensional state vector. For the affine models, the risk-neutral and physical distributions of instantaneous change in X_t can be expressed generally as

$$\text{(Risk-neutral)} \quad dX_t \sim N[(K_0^Q + K_1^Q X_t) dt, \Sigma_t dt], \quad (1)$$

$$\text{(Physical)} \quad dX_t \sim N[(K_0 + K_1 X_t) dt, \Sigma_t dt]. \quad (2)$$

The specification of Σ_t differs in m . Furthermore, the specification of K_i^Q ($i = 0, 1$) depends on the rotation of factors as well as m . This study adopts the rotation proposed by CDGJ (2008). Finally, the specification of K_i ($i = 0, 1$) depends on the specification of market price of risks in addition to m . This study adopts the so-called essentially affine specification proposed by Duffee (2002).

We next parameterize the vectors and matrices in (1) and (2) for each m .

2.1.1 $A_1(3)$

A state vector X_t consists of $X_t = (r_t \ \mu_t \ x_{3,t})'$, where r_t is the instantaneous risk-free rate, μ_t is the risk-neutral, conditional mean of change in r_t , and $x_{3,t}$ represents a volatility factor. Then, the parameters in (1) and (2) are given by

$$K_0^Q = \begin{pmatrix} 0 \\ k_{\mu,0}^Q \\ k_{3,0}^Q \end{pmatrix}, \quad K_1^Q = \begin{pmatrix} 0 & 1 & 0 \\ k_{\mu,1}^Q & k_{\mu,2}^Q & k_{\mu,3}^Q \\ 0 & 0 & k_{3,3}^Q \end{pmatrix}, \quad K_0 = \begin{pmatrix} k_{r,0} \\ k_{\mu,0} \\ k_{3,0}^Q \end{pmatrix}, \quad K_1 = \begin{pmatrix} k_{r,1} & k_{r,2} & k_{r,3} \\ k_{\mu,1} & k_{\mu,2} & k_{\mu,3} \\ 0 & 0 & k_{3,3} \end{pmatrix}, \quad (3)$$

$$\Sigma_t = \begin{pmatrix} s_{rr,0} & s_{r\mu,0} & 0 \\ s_{r\mu,0} & s_{\mu\mu,0} & 0 \\ 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} s_{rr,3} & s_{r\mu,3} & s_{r3,3} \\ s_{r\mu,3} & s_{\mu\mu,3} & s_{\mu3,3} \\ s_{r3,3} & s_{\mu3,3} & s_{33,3} \end{pmatrix} x_{3,t}, \quad (4)$$

where the first and second matrices on the right-hand side (RHS) of (4) are nonnegative and positive definite, respectively.

Since $x_{3,t}$ must be nonnegative in both the risk-neutral and physical probability measures, the drift of change in $x_{3,t}$ should not depend on r_t nor μ_t ; the (3,1) and (3,2) elements of K_1^Q and K_1 are set to zero. Also of note is that by the essentially affine specification of the market price

of risks, the first and second rows of K_i ($= 0, 1$) can be different from those of K_i^Q . This is not the case for $x_{3,t}$: the constant term $k_{3,0}^Q$, which is positive as one of the admissibility conditions, is shared between the two probability measures. Finally, for identification of $x_{3,t}$, $k_{\mu,3}^Q = 1$ is placed. By this normalization, all factors have a similar level, which eases comparison of their dynamics. It is, however, noted that an alternative normalization such as $s_{33} = 1$ does not change the model's fit.

2.1.2 $A_2(3)$

A state vector X_t consists of $X_t = (r_t \ x_{2,t} \ x_{3,t})'$, where r_t is the instantaneous risk-free rate, and $x_{i,t}$ ($i = 2, 3$) represents volatility factors. Then,

$$K_0^Q = \begin{pmatrix} k_{r,0}^Q \\ k_{2,0}^Q \\ k_{3,0}^Q \end{pmatrix}, K_1^Q = \begin{pmatrix} k_{r,1}^Q & k_{r,2}^Q & k_{r,3}^Q \\ 0 & k_{2,2}^Q & k_{2,3}^Q \\ 0 & k_{3,2}^Q & k_{3,3}^Q \end{pmatrix}, K_0 = \begin{pmatrix} k_{r,0} \\ k_{2,0}^Q \\ k_{3,0}^Q \end{pmatrix}, K_1 = \begin{pmatrix} k_{r,1} & k_{r,2} & k_{r,3} \\ 0 & k_{2,2} & k_{2,3}^Q \\ 0 & k_{3,2}^Q & k_{3,3} \end{pmatrix}, \quad (5)$$

$$\Sigma_t = \begin{pmatrix} s_{rr,0} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} s_{rr,2} & s_{r2,2} & 0 \\ s_{r2,2} & s_{22,2} & 0 \\ 0 & 0 & 0 \end{pmatrix} x_{2,t} + \begin{pmatrix} s_{rr,3} & 0 & s_{r3,3} \\ 0 & 0 & 0 \\ s_{r3,3} & 0 & s_{33,3} \end{pmatrix} x_{3,t}, \quad (6)$$

where all matrices on the RHS of (6) are nonnegative definite.

By the essentially affine specification of the market price of risks, all parameters in the drift of change in r_t can be different between the two probability measures. Meanwhile, the physical drift for the volatility factors is restricted in a way where only the diagonal elements of K_1 can be different from those of K_1^Q . To prevent $x_{i,t}$ ($i = 2, 3$) from being negative in both probability measures, the following constraints are imposed on the drift parameters: $k_{2,0}^Q > 0$ and $k_{3,0}^Q > 0$; the (2,1) and (3,1) elements of K_1^Q (and hence K_1) are zero; and $k_{2,3}^Q \geq 0$ and $k_{3,2}^Q \geq 0$. By these constraints together with the specification of Σ_t , the correlation between $x_{2,t}$ and $x_{3,t}$ cannot take a negative value. For identification of $x_{i,t}$ ($i = 2, 3$), $k_{r,2}^Q = -k_{r,3}^Q = 1$ is placed. An alternative normalization such as $s_{22,2} = s_{33,3} = 1$ does not change the model's fit: in this case, the estimate of $k_{r,2}^Q$ is positive while that of $k_{r,3}^Q$ is negative. This is why $k_{r,3}^Q = -1$ is placed instead of $k_{r,3}^Q = 1$.

2.1.3 $A_3(3)$

All components of a state vector X_t are volatility factors, $X_t = (x_{1,t} \ x_{2,t} \ x_{3,t})'$. The instantaneous risk-free rate r_t is a linear combination of the elements of X_t as

$$r_t = \delta_0 + \delta_1 x_{1,t} + \delta_2 x_{2,t} + \delta_3 x_{3,t}. \quad (7)$$

The parameters in (1) and (2) are given by

$$K_0^Q = \begin{pmatrix} k_{1,0}^Q \\ k_{2,0}^Q \\ k_{3,0}^Q \end{pmatrix}, \quad K_1^Q = \begin{pmatrix} k_{1,1}^Q & k_{1,2}^Q & k_{1,3}^Q \\ k_{2,1}^Q & k_{2,2}^Q & k_{2,3}^Q \\ k_{3,1}^Q & k_{3,2}^Q & k_{3,3}^Q \end{pmatrix}, \quad K_0 = \begin{pmatrix} k_{1,0}^Q \\ k_{2,0}^Q \\ k_{3,0}^Q \end{pmatrix}, \quad K_1 = \begin{pmatrix} k_{1,1} & k_{1,2} & k_{1,3} \\ k_{2,1} & k_{2,2} & k_{2,3} \\ k_{3,1} & k_{3,2} & k_{3,3} \end{pmatrix}, \quad (8)$$

$$\Sigma_t = \begin{pmatrix} s_{11}x_{1,t} & 0 & 0 \\ 0 & s_{22}x_{2,t} & 0 \\ 0 & 0 & s_{33}x_{3,t} \end{pmatrix}, \quad (9)$$

where $s_{ii} > 0$ ($i = 1, 2, 3$).

By the essentially affine specification of the market price of risks, only the diagonal elements of K_1 are different from those of K_1^Q , and the remaining parameters are shared between the two probability measures. For $x_{i,t}$ ($i = 1, 2, 3$) to be nonnegative, it is required that all elements of K_0^Q (and hence K_0) be positive and that the off-diagonal elements of K_1^Q (and hence K_1) be nonnegative. These constraints, together with the specification of Σ_t , exclude the possibility that the correlations between any pair of factors are negative.

For the identification of $x_{i,t}$ ($i = 1, 2, 3$), $\delta_1 = -\delta_2 = \delta_3 = 1$ is placed in (7). An alternative normalization such as $s_{11} = s_{22} = s_{33} = 1$ does not change the model's fit: in this case, the estimates δ_1 and δ_3 are positive whereas that of δ_2 is negative, leading to the above normalization based on δ_i ($i = 1, 2, 3$). Additionally, $\delta_0 = -1$ is placed for two reasons. First, it is difficult to obtain a precise estimate of δ_0 . Second, a sufficiently negative value (relative to the level of factors) gives volatility factors a wide margin for taking positive values.

As seen above, more restrictions are placed by increasing m . The literature prefers $m = 1$, which strikes the balance between time-varying volatilities and flexible correlations. In the next subsection, we replace the affine covariance matrix in each of the affine models with a quadratic counterpart, aiming at pursuing both time-varying volatilities and flexible correlations. Although the specification of the quadratic covariance matrix is the same for all m , its effect can be different because of the different restrictions in the original affine models.

2.2 Proposed models

We specify Σ_t in (1) and (2) as quadratic functions of X_t . The remaining components of the distributions are the same as those for the original affine models, which allows us to highlight the effect of a quadratic Σ_t . To guarantee the positive definiteness of Σ_t , it is convenient to first decompose Σ_t and then specify the components. We consider two decompositions.

2.2.1 LSD model

LSD is an abbreviation of Linear Standard Deviation. The first decomposition of Σ_t follows Bollerslev (1990) and Engle (2002):

$$\Sigma_t = H_t R H_t, \quad (10)$$

where H_t is a diagonal matrix and R is a constant correlation matrix.¹ Then, the i -th diagonal element of H_t , denoted as $h_i(X_t)$, is specified by a linear function of X_t :

$$h_i(X_t) = \beta_0^i + \beta^i X_t \quad (i = 1, 2, 3). \quad (11)$$

A model having (1)–(2) and (10)–(11) is labeled as $A_m(3)$ -LSD ($m = 1, 2, 3$).

2.2.2 QEV model

QEV is an abbreviation of Quadratic EigenValues. The second decomposition of Σ_t is the eigenvalue decomposition:

$$\Sigma_t = P L_t P', \quad (12)$$

where L_t is a diagonal matrix consisting of the eigenvalues, and P is an orthonormal matrix consisting of the corresponding eigenvectors, which are fixed as in the previous studies; Fan et al. (2003), Han (2007), Jarrow et al. (2007), Longstaff et al. (2001), and Pérignon and Villa (2006).

P can be expressed as

$$P = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \varphi_3^P & -\sin \varphi_3^P \\ 0 & \sin \varphi_3^P & \cos \varphi_3^P \end{pmatrix} \begin{pmatrix} \cos \varphi_2^P & 0 & -\sin \varphi_2^P \\ 0 & 1 & 0 \\ \sin \varphi_2^P & 0 & \cos \varphi_2^P \end{pmatrix} \begin{pmatrix} \cos \varphi_1^P & -\sin \varphi_1^P & 0 \\ \sin \varphi_1^P & \cos \varphi_1^P & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad (13)$$

¹In the previous version of this manuscript, a state-dependent correlation matrix is also considered. It offers little improvement, however, over a constant correlation matrix because none of the parameters related to the state-dependence are statistically significant.

where the parameters to be estimated are $\sin \varphi_i^P$ ($i = 1, 2, 3$). For identification, $\varphi_i^P \in [-\pi/2, \pi/2]$ is imposed so that $\cos \varphi_i^P = \sqrt{1 - \sin^2 \varphi_i^P}$.

For Σ_t to be positive definite, all eigenvalues in L_t must be positive. Then, the i -th diagonal element of L_t , denoted as $l_i(X_t)$, is specified by a quadratic function of X_t :

$$l_i(X_t) = c_0^i + X_t' \Gamma^i X_t \quad (i = 1, 2, 3), \quad (14)$$

where $c_0^i > 0$ and Γ^i is a nonnegative definite matrix. Similar to Σ_t , Γ^i is parameterized based on the spectral decomposition:

$$\Gamma^i = Q^i M^i Q^{i'} \quad (i = 1, 2, 3), \quad (15)$$

where

$$M^i = \begin{pmatrix} m_1^i & 0 & 0 \\ 0 & m_2^i & 0 \\ 0 & 0 & m_3^i \end{pmatrix}, \quad \text{with } 0 \leq m_1^i \leq m_2^i \leq m_3^i, \quad (16)$$

and

$$Q^i = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \varphi_3^i & -\sin \varphi_3^i \\ 0 & \sin \varphi_3^i & \cos \varphi_3^i \end{pmatrix} \begin{pmatrix} \cos \varphi_2^i & 0 & -\sin \varphi_2^i \\ 0 & 1 & 0 \\ \sin \varphi_2^i & 0 & \cos \varphi_2^i \end{pmatrix} \begin{pmatrix} \cos \varphi_1^i & -\sin \varphi_1^i & 0 \\ \sin \varphi_1^i & \cos \varphi_1^i & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad (17)$$

with $\varphi_j^i \in [-\pi/2, \pi/2]$. It is noted that $\sin \varphi_j^i$ cannot be identified for some m_j^i . For instance, when $m_j^i = 0$ for all j , $\sin \varphi_j^i$ cannot be identified for all j . In such cases, $\sin \varphi_j^i = 0$ is placed. Furthermore, when the value of a parameter with sign constraint reaches the boundary in the estimation, it is fixed in the following and then the remaining parameters are re-estimated: $c_0^i = 10^{-8}$ (as $c_0^i > 0$); $m_j^i = 0$ (as $m_j^i \geq 0$).

A model having (1)–(2) and (12)–(17) is labeled as $A_m(3)$ -QEV ($m = 1, 2, 3$).

2.2.3 Difference between the LSD and QEV models

Though Σ_t is quadratic in X_t for both models, the instantaneous variance has a different form. For the LSD model, the instantaneous variance of the i -th factor is given by $h_i^2(X_t)$, that is, a squared linear combination of X_t . Hence, it is easy to see which factor is more influential for driving the variance. For the QEV model, it is given by $p_{i1}^2 l_1(X_t) + p_{i2}^2 l_2(X_t) + p_{i3}^2 l_3(X_t)$, where p_{ij} is the (i, j) element of P . Then, this linear combination of quadratic functions may be more appropriate for capturing complicated behavior of volatility that is intense in some periods and calm in the other. It is, however, difficult to distinguish influential factors for the variance in the QEV model.

2.3 An approximation of no-arbitrage bond prices and yields

Let $P(X_t, \tau)$ be the price at time t of a zero-coupon bond with τ years to maturity. Then, by the absence of arbitrage, it is given by

$$P(X_t, \tau) = E_t^Q \left[\exp \left\{ - \int_t^{t+\tau} r_u du \right\} \right], \quad (18)$$

where $E_t^Q[\cdot]$ stands for the conditional expectation under the risk-neutral probability measure. The yield to maturity of a τ -year zero-coupon bond is given by $Y(X_t, \tau) = -\frac{1}{\tau} \ln P(X_t, \tau)$.

Since the proposed models are non-affine, there is no closed-form for $P(X_t, \tau)$. To make it possible to estimate the models with no-arbitrage conditions, $P(X_t, \tau)$ is approximated using a method proposed by Takamizawa and Shoji (2009). The method approximates a vector of conditional moments of multi-dimensional diffusion processes as the solution to a system of ordinary differential equations. Since the zero-coupon bond price is given as the conditional expectation, the method can be directly applied. The outline of the method is provided in Appendix A and the accuracy of the approximation is examined in Appendix B. In brief, the accuracy is maintained at least for maturities of up to ten years when reasonable values of parameter and state vectors are provided.

3 Estimation and prediction method

3.1 Dataset

Weekly (Wednesday) data on the U.S. dollar LIBOR and swap rates are used, which cover the period from January 4, 1991 to May 27, 2009 with 961 observations in total. The data are divided into in-sample for estimation, ending in April 9, 2003 with 641 observations, and out-of-sample for prediction. This partition is intended to incorporate information on the lowest range of interest rates into model estimation as well as to reserve sufficient out-of-sample observations.

The LIBOR rates with maturities of 6 and 12 months, and swap rates with maturities of 2, 3, 4, 5, 7, and 10 years are used to obtain zero-coupon bond yields on a continuously compounded basis by a bootstrap method with linear interpolation to discount functions. The maturities of the zero yields used for the analysis are 0.5, 1, 2, 3, 5, and 10 years. Following the previous studies (e.g., CDGJ, 2008 and Duffee, 2002), this study assumes that the yields with maturities of 0.5, 2, and 10 years are measured without error to extract latent factors and that the rest of the yields are measured with error. While there may be a concern that this choice is arbitrary, it is actually

convenient. By selecting yields that are measured without error from each of the short, medium, and long segments of the yield curve and combining them appropriately, proxies for the level, slope, and curvature of the yield curve can also be measured without error.

3.2 Realized volatility measure

A realized measure of the one-week ahead variance of a τ -year yield is computed using daily data as

$$RV_{t,t+\Delta,\tau} = \sum_{j=1}^{m_{t+\Delta}} \left(y_{t+\frac{\Delta}{m_{t+\Delta}}j,\tau} - y_{t+\frac{\Delta}{m_{t+\Delta}}(j-1),\tau} \right)^2 \quad (\tau = 0.5, 1, 2, 3, 5, 10), \quad (19)$$

where Δ is a week interval, set at $1/52$, and m_t stands for the number of observations during a week ending at time t (usually $m_t = 5$). The realized measure is generated every Wednesday. The total number of observations is 960, among which the first 640 observations belong to the in-sample.

A realized measure of the h -week ahead variance of a τ -year yield is computed as

$$RV_{t,t+h\Delta,\tau} = \sum_{j=1}^h RV_{t+(j-1)\Delta,t+j\Delta,\tau}. \quad (20)$$

The annualized variance is obtained by dividing $RV_{t,t+h\Delta,\tau}$ by $h\Delta$. $RV_{t,t+h\Delta,\tau}$ is also generated weekly, which means that some overlapping daily observations are used in successive observations of $RV_{t,t+h\Delta,\tau}$ with $h > 1$. This study considers $h = 4, 13, 26$.

3.3 Model estimation

The models are estimated by the quasi maximum likelihood method. To clarify the parameters to estimate, let $\Theta = \Theta_P \cup \Theta_Q$ be the parameter vector of a term structure model, where Θ_P consists of the parameters in K_0 and K_1 as well as those in Σ_t and Θ_Q consists of the parameters in K_0^Q and K_1^Q as well as those in Σ_t .

Let $y_{t,\tau}$ be a yield of a zero-coupon bond observed at time t with τ years to maturity, and $Y_t^p = (y_{t,0.5} \ y_{t,2} \ y_{t,10})'$ be a vector of yields measured without error. The corresponding vector of model-implied yields is denoted as

$$\Upsilon(X_t; \Theta_Q) = (Y(X_t, 0.5; \Theta_Q) \ Y(X_t, 2; \Theta_Q) \ Y(X_t, 10; \Theta_Q))'. \quad (21)$$

Then, X_t is extracted by solving $Y_t^p = \Upsilon(X_t; \Theta_Q)$ for X_t . Since for the proposed models, $Y(X_t, \tau; \Theta_Q)$ has no closed-form, this extraction is performed numerically. However, only a few

iterations are sufficient if a good initial value of X_t is given. It is actually the value of X_t implied by the original affine model.

The rest of the yields, denoted as $Y_t^e = (y_{t,1} \ y_{t,3} \ y_{t,5})'$, are measured with error, denoted as $U_t = (u_{t,1} \ u_{t,3} \ u_{t,5})'$. It is assumed to be independent of X_s for any s and follow

$$U_t \sim i.i.d.N(\mathbf{0}, \varsigma^2 \mathbf{I}) . \quad (22)$$

The reason for assuming such a simple distribution is to let the models explain various features of the data as much as possible.

The joint density function at time t conditioned on time $t - \Delta$ can be written and developed as follows:

$$\begin{aligned} f(Y_t^p, Y_t^e | Y_{t-\Delta}^p; \Theta, \varsigma^2) &= f(X_t, U_t | X_{t-\Delta}; \Theta, \varsigma^2) \left| \frac{d\Upsilon(X_t; \Theta_Q)}{dX_t'} \right|^{-1} \\ &= f_T(X_t | X_{t-\Delta}; \Theta_P) f_C(U_t | X_t; \Theta_Q, \varsigma^2) \left| \frac{d\Upsilon(X_t; \Theta_Q)}{dX_t'} \right|^{-1} . \end{aligned} \quad (23)$$

The first equality is from changes of variables from Y_t^p to X_t , by which the Jacobian term arises, and from Y_t^e to U_t . The second equality is from the decomposition of the joint density into the marginal (f_T) and conditional (f_C) components with X_t being Markovian.

For the LSD and QEV models, the transition density, f_T , has no closed-form with finite Δ . It is then approximated by the multivariate normal density function, which seems to be justified by a relatively short interval, $\Delta = 1/52$. The conditional first and second moments to be substituted are computed with the same method used for pricing bonds. It is noted that these moments can be computed exactly because the drift vector is linear in X_t and the covariance matrix is at most quadratic in X_t . The Jacobian term is computed in the process of extracting X_t . On the other hand, f_C is the multivariate normal density function from (22). The objective function for estimating Θ is then

$$\sum_t \left\{ \ln f_T(X_t | X_{t-\Delta}; \Theta_P) + \ln f_C(U_t | X_t; \Theta_Q, \varsigma^2) - \ln \left| \frac{d\Upsilon(X_t; \Theta_Q)}{dX_t'} \right| \right\} . \quad (24)$$

Once Θ is estimated using the in-sample data, it is held fixed throughout the out-of-sample period. That is, Θ is not re-estimated each time the out-of-sample prediction is made. This approach does not seem to be a serious concern because the volatility can be captured without changing the structural parameters, as will be seen in Section 5.1.

3.4 Volatility prediction

The models predict the h -week-ahead annualized standard deviation of a τ -year yield. Let $pred_{t,t+h\Delta,\tau}$ denote a model forecast defined as follows:

$$pred_{t,t+h\Delta,\tau} = \sqrt{\frac{\text{var}_t[Y(X_{t+h\Delta}, \tau)]}{h\Delta}} \quad (h = 4, 13, 26 ; \tau = 0.5, 1, 2, 3, 5, 10) , \quad (25)$$

where $\text{var}_t[\cdot]$ stands for a model-implied conditional variance computed under the physical probability measure. It is noted that in computing $pred_{t,t+h\Delta,\tau}$, the standard deviation of measurement errors, ς , introduced for estimation is omitted as it is a small constant.

Since for the proposed models, there is no closed-form of $\text{var}_t[Y(X_{t+h\Delta}, \tau)]$, it is computed using the Monte Carlo method. Let $\{X_s^{(i)}\}_t^{t+h\Delta}$ denote the i -th path of the state vector, starting from X_t (known initial value) and ending in $X_{t+h\Delta}^{(i)}$ (simulated value). It is generated from a discretized version of (2), where dt is simply replaced by $\Delta/10$, the frequency corresponding roughly to two observations per day. Once $X_{t+h\Delta}^{(i)}$ is generated, it is substituted into the approximate function, $Y(X_{t+h\Delta}^{(i)}, \tau)$. This computation is repeated $n = 10,000$ times with antithetic variates. A simulated value of conditional variance of a τ -year yield is given by $\frac{1}{n} \sum_{i=1}^n Y(X_{t+h\Delta}^{(i)}, \tau)^2 - \{\frac{1}{n} \sum_{i=1}^n Y(X_{t+h\Delta}^{(i)}, \tau)\}^2$.

Let $rsd_{t,t+h\Delta,\tau} = \sqrt{\frac{RV_{t,t+h\Delta,\tau}}{h\Delta}}$ be an annualized, h -week standard deviation realized at time $t + h\Delta$. This study considers two criteria for evaluating the forecasting performance. The first is the root mean squared error (RMSE), in which the prediction error $e_{t+h\Delta,\tau}$ is computed as $e_{t+h\Delta,\tau} = rsd_{t,t+h\Delta,\tau} - pred_{t,t+h\Delta,\tau}$. The second is the correlation between the realized and predicted values, $corr(rsd_{t,t+h\Delta,\tau}, pred_{t,t+h\Delta,\tau})$.

4 Results of model estimation

Before estimation, we restrict the parameters of the physical drift, which applies to both the affine and proposed models as they have the same specification of the drift. Specifically, only the diagonal elements of K_1 are estimated as free parameters, and the rest of the parameters in the physical drift are the same as those in the risk-neutral drift. It is noted that this is not an additional restriction for the $A_3(3)$ model. The restriction is intended to match the condition on estimating the physical drift among the models. In Section 6, it is removed from the $A_1(3)$ and $A_1(3)$ -LSD/QED models to examine an additional benefit of a more flexible specification of the physical drift for predicting the level of interest rates.

To keep the models simple, insignificant parameters in the previous round of estimation are set to zero and the remaining parameters are re-estimated. Similarly, when a parameter with sign constraint hits the boundary, it is fixed at the boundary value and the remaining parameters are re-estimated. The estimation results are reported for each m .

4.1 Results for the $A_1(3)$ and $A_1(3)$ -LSD/QEV models

Tables 1–3 present parameter estimates (standard errors) for the $A_1(3)$ and $A_1(3)$ -LSD/QEV models. Standard errors are computed by the outer product of gradients of the log-likelihood function. First of all, the value of the maximum log-likelihood, LogL, for the LSD and QEV models is larger by 70 and 112, respectively, than that for the $A_1(3)$ model, indicating that the overall fit is improved by the quadratic specification of volatility. But the standard deviation of measurement errors ς is estimated to be between 6.13 and 6.15 basis points (bps). These results together imply that the improvement by the quadratic specification is attributed not to the cross-sectional fit but to the time-series fit. Consistent with the almost equal cross-sectional fit, the estimates in the risk-neutral drift are similar across the models. Specifically, the estimate of $k_{3,3}^Q$ is around -0.007 , indicating that $x_{3,t}$ is the most persistent factor affecting the long end of the yield curve. The result is in line with the previous work reporting that a realized volatility factor in the $A_1(3)$ model is related to long-term interest rates.

In Table 2, the estimates in $h_i(X_t)$ ($i = 1, 2, 3$) for the LSD model show an interesting pattern. First, it is not $x_{3,t}$ but μ_t that is the most influential for driving the volatility. In addition, the estimate of β_2^i is negative for all i , implying that the volatilities increase by decrease in μ_t . Second, the estimate of β_1^i ($i = 1, 2$) is also negative, implying an inverse relationship between the volatilities and the risk-free rate. These results are consistent with recent observations that hike in volatilities tends to be accompanied by fall in interest rates. Such an inverse relationship is difficult to capture using the $A_1(3)$ model because the volatilities are related positively to $x_{3,t}$, which is related positively to (long-term) interest rates. Third, the volatility of change in $x_{3,t}$ does not depend on r_t nor $x_{3,t}$. The result that a persistent factor has a relatively simple form of volatility is also confirmed in the QEV model shown in Table 3, where $l_3(X_t)$ is a dominant driver of the instantaneous variance of $x_{3,t}$.² Only the estimate of m_3^3 is significant, which is in contrast

²If $P = I$ (the identity matrix), $l_3(X_t)$ is exactly the instantaneous variance of $x_{3,t}$. Actually, the estimate of $\sin \varphi_i^P$ is close to zero for all i . Hence, an estimated P has the diagonal elements close to one and the off-diagonal elements close to zero.

to $l_i(X_t)$ with $i = 1, 2$, where the estimate of m_2^i is also significant. These results are consistent with Andersen and Benzoni (2010) showing that the volatilities of longer-term interest rates, which are more persistent than shorter-term interest rates, are more difficult to capture with the level of interest rates.

Figure 1 displays the time-series of realized factors. There is little discrepancy in the realized series across the models despite the difference in the covariance matrix. The behavior of $x_{3,t}$ does not look like that of typical volatility measures, which may be crucial for the $A_1(3)$ model but not for the LSD and QEV models. The similarity in realized factors is due to the rotation of factors proposed by CDGJ (2008), which clarifies ex-ante roles of factors. Specifically, r_t is the instantaneous risk-free rate, which is identified model-independently as the initial point of the yield curve. Since the measurement of r_t is accurate, its drift can also be measured accurately, nearly free of models. Though the ex-ante role of $x_{3,t}$ may be different between the original and proposed models, there is little room for different realizations of $x_{3,t}$. This is because the other two factors are almost identical and because the three interest rates used for extracting the factors are the same among the models.

4.2 Results for the $A_2(3)$ and $A_2(3)$ -LSD/QEV models

Table 4–6 present estimated parameters (standard errors) for the $A_2(3)$ and $A_2(3)$ -LSD/QEV models. In terms of the goodness-of-fit, a similar pattern to Tables 1–3 emerges. The value of LogL for the LSD and QEV models is larger than that for the $A_2(3)$ model but the estimate of ς is similar, around 6.15 bps, for all models. By the estimates of the risk-neutral drift, $x_{2,t}$ is the most persistent factor and r_t is the least for all models. The estimates of $k_{2,0}^Q$ and $k_{3,0}^Q$ for the LSD model, shown in Table 5, are 0.0044 and 0.096, respectively, both of which appear larger than those for the $A_2(3)$ model, 0.0034 and 0.068, shown in Table 4. The corresponding estimates for the QEV model, shown in Table 6, are similar to those for the $A_2(3)$ model. The difference in these estimates is related to the difference in realized factor values, which is addressed below.

In Table 5, the estimates in $h_i(X_t)$ ($i = 1, 2, 3$) for the $A_2(3)$ -LSD model show a similar pattern to that for the $A_1(3)$ -LSD model. First, a moderately persistent factor $x_{3,t}$ is the most influential for the volatility. Second, r_t is of no relevance to the volatility as β_1^i is set to zero for all i due to statistical insignificance. Consequently, the covariance matrix for the LSD model is driven by $x_{2,t}$ and $x_{3,t}$ alone, similarly to the $A_2(3)$ model. Third, the volatility of change in the most persistent factor $x_{2,t}$, again, has a simpler form. This can also be confirmed in the QEV model shown in

Table 6, where $l_2(X_t)$ is of the most simple form.

Figure 2 displays the time-series of realized factors. Between the $A_2(3)$ and QEV models, there is little difference in the realized series. However, it is noticed that the time-series of $x_{2,t}$ and $x_{3,t}$ for the LSD model are shifted upward with the shape unchanged. As suggested earlier, the upward shift is due to the larger estimates of $k_{2,0}^Q$ and $k_{3,0}^Q$ for the LSD model. Despite the shift, the difference between $x_{2,t}$ and $x_{3,t}$ is similar to that in the other models. This is not surprising because $x_{2,t} - x_{3,t}$ enters into the risk-neutral drift of change in r_t , which is measured model-independently. An important implication of the result is then that there is a room for identifying each volatility factor differently across models. But a certain combination of volatility factors has no such room when it is related to r_t . This implication is further confirmed by the results from the $A_3(3)$ and proposed models shown below.

4.3 Results for the $A_3(3)$ and $A_3(3)$ -LSD/QEV models

Tables 7–9 present estimates (standard errors) for the $A_3(3)$ and $A_3(3)$ -LSD/QEV models. A different pattern from the previous ones emerges. First of all, in Table 7, the overall goodness-of-fit of the $A_3(3)$ model is worse than that of the other affine models. The value of LogL is 21955, which is smaller by 321 than that for the $A_1(3)$ model. The smaller LogL is attributed to the loss in both the cross-sectional and time-series fit. The lower cross-sectional fit is evident in the estimate of ς , 6.22 bps. Though the increase of ς is less than 0.1 bps, it reduces the value of LogL because of a tight relationship in the cross-section of interest rates. The lower time-series fit appears as the failure in capturing the volatility, which is addressed in Section 5.3.

The primary reason for the worse fit of the $A_3(3)$ model is that all factors are restricted to be nonnegative. By the restriction, all elements of K_0^Q should be positive and the off-diagonal elements of K_1^Q should be nonnegative, which indeed is binding. In fact, only the estimate of $k_{2,3}^Q$ is positive, and the rest of the off-diagonal elements of K_1^Q are set to zero because they are negative when estimated as free parameters. If they are left unrestricted, which, however, is inadmissible from a theoretical point of view, the overall goodness-of-fit is equivalent to that of the $A_1(3)$ and $A_2(3)$ models. The result indicates that the admissibility conditions are sever for $m = 3$, but not for $m = 1, 2$, given $n = 3$. It will not be surprising, therefore, that the quadratic specification of volatility brings a significant improvement when introduced into the $A_3(3)$ model.

Tables 8 and 9 present the results for the $A_3(3)$ -LSD/QEV models. The value of LogL for the LSD and QEV models is 22289 and 22300, respectively, increased from that for the $A_3(3)$ model

by 334 and 345. Furthermore, the estimate of ς for the QEV model is 6.15 bps, which is of the same magnitude as in the previous cases. While the estimates of the risk-neutral drift are similar between the LSD and QEV models, they are different from those for the $A_3(3)$ model. Specifically, the value of $k_{3,0}^Q$ for the $A_3(3)$, LSD, and QEV models is 0.012, 0, and -0.143 , respectively. A negative estimate of $k_{3,0}^Q$ is allowed only in the proposed models. The difference in the value of $k_{3,0}^Q$ produces the difference in the realized series of $x_{3,t}$. Figure 3 shows that $x_{3,t}$ switches signs over time for the LSD model and remains negative for most of the time for the QEV model. In contrast, for the $A_3(3)$ model, it stays in a positive region without moving largely. Additionally, the realized time-series of $x_{2,t}$ also look different between the original and proposed models. Nevertheless, the time-series of the instantaneous risk-free rate, computed as $r_t = -1 + x_{1,t} - x_{2,t} + x_{3,t}$, are almost identical among the models. These results confirm the previous finding that though each factor can realize differently, a certain combination of factors such as related to r_t cannot.

In Table 8, the estimates in $h_i(X_t)$ ($i = 1, 2, 3$) for the LSD model show that $x_{2,t}$ alone, which is a moderately persistent factor implied by the estimate of $k_{2,2}^Q$, drives the covariance matrix. The estimates of ρ_{12} and ρ_{23} , the instantaneous correlations of $x_{2,t}$ with $x_{1,t}$ and $x_{3,t}$, respectively, are statistically different from zero, 0.264 and 0.948. Non-zero correlations are also implied in the QEV model. Specifically, in Table 9, the eigenvector matrix P is statistically different from the identity matrix as the estimates of $\sin \varphi_1^P$ and $\sin \varphi_3^P$ are both significant, -0.782 and -0.667 , respectively. Such non-zero instantaneous correlations are not allowed in the $A_3(3)$ model by the admissibility conditions, which is one of the reasons for reducing the value of LogL.

In summary, the quadratic specification of volatility improves the time-series fit for all cases. It also improves the cross-sectional fit for $m = 3$, where the admissibility conditions are binding. In the next two sections, we address where the improvement in the time-series fit appears, a higher predictive accuracy for the volatility and/or the level.

5 Results of volatility prediction

The volatility forecasting performance is evaluated by the RMSE and correlation criteria. As are the estimation results, the prediction results are reported for each m , the number of factors driving the volatility in the affine models, to highlight the effect of the quadratic specification that will be different in m .

5.1 Results for the $A_1(3)$ and $A_1(3)$ -LSD/QEV models

Table 10 presents the prediction results for the $A_1(3)$ and $A_1(3)$ -LSD/QEV models. Panel A presents the RMSEs in bps. As in CDGJ (2009), the Diebold and Mariano (DM) (1995) test is performed with the null hypothesis of equal predictive accuracy between the affine and each of the proposed models.³ The standard error is computed using the Newey and West method (with Bartlett weights). The truncation lag length is set at the nearest integer of \sqrt{T} ; $T = 640$ for in-sample and $T = 320$ for out-of-sample. This lag length is recommended by Coroneo and Iacone (2016) when a nonstandard (fixed-smoothing) asymptotic distribution of the test statistic is used, where the ratio of lag length to sample size is fixed at a constant. The use of this alternative distribution alleviates the problem of over-rejecting the null hypothesis when the asymptotic standard normal distribution is used. The critical values of the distribution are obtained from Kiefer and Vogelsang (2005, Table 1). Specifically, these values at the 5% significance level are 2.078 for in-sample ($T = 640$) and 2.127 for out-of-sample ($T = 640$), both of which are larger than 1.96 for the standard normal distribution. In Panel A of Table 10, the asterisk “*” indicates that the null hypothesis is rejected at the 5% significance level.

Looking at the left side of Panel A, we find that the in-sample forecasting performance of the proposed models is better than that of the $A_1(3)$ model. Specifically, the in-sample RMSEs for the LSD model are smaller than those for the $A_1(3)$ model for all h and τ by between 0.8 and 4.7 bps. The in-sample RMSEs for the QEV model are also smaller for all h except at $\tau = 10$. A higher predictive accuracy of the proposed models is more evident on the right side of Panel A showing the out-of-sample RMSEs. Generally, the relative accuracy tends to improve with τ , which holds for all h . For instance, at $h = 4$ and $\tau = 1$, the out-of-sample RMSEs for the $A_1(3)$ and LSD models are 37.1 and 35.6 bps, respectively, with the gap 1.5 bps. The gap increases to 10.4 bps at $\tau = 10$. For the QEV model, the out-of-sample RMSE at $h = 4$ and $\tau = 0.5$ is 46.2 bps, which is larger than that for the $A_1(3)$ model by 2.2 bps. At $\tau = 10$, however, the RMSEs for the $A_1(3)$ and QEV models are 49.6 and 31.2 bps, respectively, with the gap 18.4 bps. A large gap at $\tau = 10$ remains for longer h , 20.0 and 14.1 bps.

Panel B of Table 10 presents the correlations between the predicted and realized series of volatility. The better performance of the proposed models can also be found, which is robust to

³As noted by Diebold (2015), the DM test is designed to compare forecasts, but not models, as it treats forecast errors as primitive variables. Patton (2015) offers an interpretation of the DM test as the one comparing forecasts produced at estimated parameter values.

extending h and more evident in the out-of-sample period. On the left side of Panel B, a general pattern is that the in-sample correlations decrease with τ . But the rate of decrease is different among the models. It is faster for the $A_1(3)$ model than for the proposed models. Specifically, at $\tau = 0.5$, the in-sample correlations for the $A_1(3)$ model range between 0.54 ($h = 4$) and 0.60 ($h = 26$), which do not appear to differ much from those for the proposed models. But at $\tau = 10$, they decrease up to -0.35 ($h = 26$). The result of negative correlations is also reported by the previous work using recent samples; see, for example, Jacobs and Karoui (2009). In contrast, the in-sample correlations for the proposed models remain positive except the LSD model at $h = 26$ and $\tau = 10$. On the right side of Panel B showing the out-of-sample correlations, the performance of the QEV model is remarkable. At $h = 4$, the correlations range from 0.55 ($\tau = 0.5, 1$) to 0.71 ($\tau = 10$), exhibiting the reverse pattern in that the correlations tend to be higher with longer τ . Also, the correlations for the LSD model increase with τ from 0.20 ($\tau = 0.5$) to 0.62 ($\tau = 10$). These results are robust to extending h . In contrast, the out-of-sample correlations for the $A_1(3)$ model do not exceed 0.1 at $h = 4$. They are negative at $\tau = 0.5, 10$. The negative correlations disappear at $h = 26$, however, a large gap remains between the original and proposed models.

Figure 4 exhibits the time-series of four-week-ahead ($h = 4$) forecasts of the annualized standard deviation of a τ -year yield with $\tau = 0.5, 2, 10$ produced by the $A_1(3)$ model, together with the corresponding realized series. It is apparent that the forecast series do not fluctuate largely nor move in accordance with the realized series. In contrast, Figure 5 shows that the forecast series generated by the LSD model trace the realized series for all τ . More strikingly, in Figure 6, the QEV model not only traces the trend of the realized series but also predicts large variation in volatility observed in the out-of-sample period. It is worth noting that the parameter values are fixed at the in-sample estimates throughout the out-of-sample period.

As seen, the proposed models have a higher predictive accuracy both in-sample and out-of-sample. They overcome the difficulties of the $A_1(3)$ model, which is a preferred affine model suffering the least from the trade-offs between fitting the time-series and cross-section and between predicting the level and volatility. The $A_1(3)$ model fails to generate large variation in volatility because the volatility factor $x_{3,t}$ is realized as a persistent level factor that does not vary much by nature as a consequence of being used for matching the cross-section of interest rates. Small variation in $x_{3,t}$ also holds for the proposed models, as seen in Figure 1. However, the specification of volatility is different. In the $A_1(3)$ model, the conditional variances are linear in $x_{3,t}$ alone whereas in the proposed models they are quadratic in not only $x_{3,t}$ but the other factors, r_t and

μ_t . By the quadratic specification, it is possible to generate intense behavior of volatility simply because the factors affecting the volatility are squared.

The quadratic specification can also resolve the problem of negative correlations between the forecast and realized volatility in the $A_1(3)$ model. It is difficult for the $A_1(3)$ model to generate high volatility when the level of interest rates is low because the volatility factor $x_{3,t}$ is positively related to both the level and volatility. Then, the negative correlations may arise in a period when the actual volatility is high but the level is low: in this case, the volatility predicted by the $A_1(3)$ model is low. Indeed, the combination of high volatility and low level is a feature of our sample, including several hikes in volatility, as seen in Figures 4–6, with the level in a decreasing trend. The positive relationship implied by the $A_1(3)$ model between the level and volatility breaks down in the proposed models. Again, the reason is simple. By the quadratic specification, the volatility rises when the factors take extreme values in either a positive or negative direction: at least one direction is a state of low interest rates.

In this way, the proposed models can generate high volatility both when the level of interest rates is high and when it is low. The affine models work in the former period but not in the latter period. Looking at the U.S. data on interest rates, high volatility is observed in not only the period of high level (e.g., the late 70s and early 80s) but the period of low level (e.g., around 2008). It seems then that the proposed models are more robust to different samples than the affine models. In the next two subsections, we find the similar results in the $A_2(3)$ - and $A_3(3)$ -based models, which, therefore, are reported more briefly.

5.2 Results for the $A_2(3)$ and $A_2(3)$ -LSD/QEV models

Table 11 presents the prediction results for the $A_2(3)$ and $A_2(3)$ -LSD/QEV models. A higher predictive accuracy of the proposed models is also found both in-sample and out-of-sample. On the left side of Panel A, the in-sample RMSEs at $h = 4$ for the QEV model are smaller than those for the $A_2(3)$ model by between 1.4 ($\tau = 3$) and 2.1 bps ($\tau = 0.5$). The gap slightly widens by extending h . At $h = 26$, it ranges from 2.0 ($\tau = 10$) to 3.4 bps ($\tau = 1$). The predictive accuracy of the LSD model is higher at short maturities but lower at $\tau = 10$ than that of the $A_2(3)$ model, which holds for all h . The right side of Panel A shows that the out-of-sample RMSEs for the proposed models are smaller than those for the $A_2(3)$ model for all h and τ except the QEV model at $h = 26$ and $\tau = 0.5, 1$. The improvement by the proposed models is more evident at longer maturities. Specifically, at $h = 4$ and $\tau = 10$, the RMSE for the $A_2(3)$ model is 44.7 bps, and the

corresponding RMSEs for the LSD and QEV models are 34.1 and 35.1 bps, respectively, which are smaller by around 10 bps. Such a large gap at $\tau = 10$ remains for longer h .

In Panel B presenting the correlations between the predicted and realized series of volatility, the general pattern is similar to the previous one based on the $A_1(3)$ model. On the left side of Panel B, the in-sample correlations for the $A_2(3)$ model are similar to those for the proposed models at $\tau \leq 3$ for all h . However, they decrease rapidly by increasing τ . At $\tau = 10$, the correlations are negative and more so by extending h . The result indicates that even though the volatility is driven by a moderately persistent factor $x_{3,t}$ as well as the most persistent factor $x_{2,t}$, it is difficult to break down the positive relationship between the level and volatility of interest rates. Lower correlations at longer maturities are also implied by the proposed models. However, they remain equal or above 0.2 at $\tau = 10$ for all h . On the right side of Panel B, it is first noticed that the out-of-sample correlations are positive for all cases. At $h = 4$, the correlations for the QEV model range from 0.53 ($\tau = 1$) to 0.71 ($\tau = 10$). Though the correlations for the LSD model are lower than those for the QEV model at short-maturities, they are comparable at middle to long maturities. Relative to the QEV model, the correlations for the $A_2(3)$ model are low at $\tau = 0.5, 1$, comparable at $\tau = 2, 3$, and again low at $\tau = 5, 10$. This pattern is robust to extending h though the gap at the short maturities tends to shrink.

5.3 Results for the $A_3(3)$ and $A_3(3)$ -LSD/QEV models

Table 12 presents the prediction results for the $A_3(3)$ and $A_3(3)$ -LSD/QEV models. Panel A shows that the $A_3(3)$ model has much larger RMSEs at short to middle maturities than the proposed models for all h and both in-sample and out-of-sample. Such large RMSEs are due to the admissibility conditions for preventing the factors from being negative. As suggested in Section 4.3, when the $A_3(3)$ model is estimated without the admissibility conditions, it has a similar predictive accuracy to the other affine models, which in turn implies the severity of the admissibility conditions for $m = 3$. By introducing the quadratic specification of volatility, the predictive accuracy again improves. The RMSEs for the proposed models are smaller than those for the $A_3(3)$ model for most of the cases with a few exceptions (the in-sample RMSEs for the LSD model at $\tau = 10$).

Panel B of Table 12 shows that the $A_3(3)$ model when evaluated by the correlation criterion is not as bad as when evaluated by the RMSE criterion. The in-sample correlations for the $A_3(3)$ model are similar to those for the proposed models, implying that a positive relation between

the level and volatility can be weakened by increasing m . In the out-of-sample period, the proposed models have higher correlations especially at both ends of the maturity spectrum, a pattern consistent with the previous ones.

6 Results of level prediction

We have seen that the proposed models improve the predictive accuracy for the volatility of interest rates while maintaining the goodness-of-fit to the cross-section of interest rates. This section investigates whether they are at least as good at predicting the level of interest rates as the affine models. As noted earlier, the quadratic specification of volatility, aimed originally at capturing the volatility, has another advantage of removing the admissibility conditions. By this advantage, the proposed models can specify the physical drift of changes in factors as flexibly as the Gaussian term structure model. It is, therefore, natural to consider the Gaussian model, or $A_0(3)$, as a benchmark for the level prediction. The $A_1(3)$ model is also included in the horserace for both cases with and without the restriction that the free parameters in the physical drift are limited to the diagonal of K_1 . The comparison allows for uncovering how beneficial a more flexible physical drift is. The other affine models, $A_2(3)$ and $A_3(3)$, are not included as they are less flexible in the physical drift. Then, the predictive accuracy of the $A_1(3)$ -LSD/QEV models, both the restricted and unrestricted versions, is examined.⁴

Table 13 presents the RMSEs in bps. The structure of the table is the same as that of Panel A of Tables 10–12. In particular, the forecasting horizons are the same, $h = 4, 13, 26$, and both in-sample and out-of-sample analyses are conducted. Furthermore, the asterisk “*” indicates that the null hypothesis of equal predictive accuracy between the $A_0(3)$ model and the model indicated in each row is rejected at the 5% significance level. As in the volatility prediction, the critical values are obtained from Kiefer and Vogelsang (2005, Table 1). The unrestricted version of the models is distinguished by the upper subscript “F.”

Overall, the proposed models, both the restricted and unrestricted versions, do not outperform, nor are outperformed by, the affine models. At $h = 4$, the in-sample RMSEs are similar across the models. This also holds for the out-of-sample RMSEs except at $\tau = 0.5$, where the $A_0(3)$ model produces a smaller RMSE than the other models. Between the restricted and unrestricted versions,

⁴Since the parameters in the physical drift alone are different between the restricted and unrestricted versions, the results of volatility prediction as well as of parameter estimation of the covariance matrix and risk-neutral drift are similar between the two versions.

the latter have in general slightly smaller RMSEs both in-sample and out-of-sample. With such narrow gaps, it is difficult to find a significant merit of a more flexible physical drift. These results are robust to extending h . In sum, there is no decisive ranking of models with respect to predictive accuracy for the level of interest rates. Still, it can at least be concluded that the proposed models do not sacrifice level prediction for the sake of volatility prediction.

7 Concluding Remarks

No-arbitrage affine term structure models of interest rates face two trade-offs. The first is the trade-off between fitting the time-series and cross-section of interest rates and the second is the trade-off between predicting the level and volatility of interest rates. This study proposed term structure models that can mitigate the two trade-offs. The key feature of the models is that the covariance matrix of changes in factors is specified as quadratic functions of factors. The quadratic specification is shown to be able to capture the volatility even with spanned factors, mitigating the first trade-off. In particular, it is possible to predict high volatility when interest-rate level is low. The combination of low level and high volatility is a phenomenon observed in recent data but difficult to explain using affine term structure models. An additional advantage is that the proposed models allow for a flexible specification of the physical drift because the sign of factors does not need to be restricted. The proposed models are shown to have the predictive ability about the level of interest rates that is at least equal to that of the Gaussian term structure model, mitigating the second trade-off.

Appendix A. An approximation method of conditional moments and its application to the pricing of bonds

A1. Outline of the method

This section explains an approximation method for pricing no-arbitrage bond prices. The method generally allows for the computation of up to n -th conditional moments, if they exist, for a d -dimensional diffusion process. To ease the explanation, we use a specific case of $(n, d) = (2, 2)$, that is, conditional first and second moments of a two-dimensional diffusion process.

Let $X_t = (x_{t,1} \ x_{t,2})'$ be a two-dimensional diffusion process, which evolves according to the following stochastic differential equation (SDE):

$$dx_{t,i} = f_i(X_t)dt + \xi_i(X_t)'dW_t \quad (i = 1, 2), \quad (26)$$

where W_t is two-dimensional Brownian motion, and the drift and diffusion functions, f_i and ξ_i ($i = 1, 2$), satisfy certain technical conditions for the solution to equation (26) to exist for an arbitrary X_0 .

Let $\Psi_{s,t}$ be a vector consisting of the first and second moments of changes in X_t conditioned on time $s < t$:

$$\Psi'_{s,t} = E_s \left(x_{t,1} - x_{s,1} \quad x_{t,2} - x_{s,2} \quad (x_{t,1} - x_{s,1})^2 \quad (x_{t,2} - x_{s,2})^2 \quad (x_{t,1} - x_{s,1})(x_{t,2} - x_{s,2}) \right) .$$

The goal is to obtain an approximation of $\Psi_{s,t}$, which will turn out to be the solution to a system of ordinary differential equations.

By integrating equation (26) and taking the conditional expectation,

$$E_s[x_{t,i} - x_{s,i}] = E_s \left[\int_s^t f_i(X_u)du \right] . \quad (27)$$

By applying the Taylor expansion to $f_i(X_u)$ around X_s up to the second order

$$\begin{aligned} f_i(X_u) &= f_i(X_s) \\ &+ f_i^{(1,0)}(X_s)(x_{u,1} - x_{s,1}) + f_i^{(0,1)}(X_s)(x_{u,2} - x_{s,2}) + \frac{1}{2}f_i^{(2,0)}(X_s)(x_{u,1} - x_{s,1})^2 \\ &+ \frac{1}{2}f_i^{(0,2)}(X_s)(x_{u,2} - x_{s,2})^2 + f_i^{(1,1)}(X_s)(x_{u,1} - x_{s,1})(x_{u,2} - x_{s,2}) + e_i , \end{aligned} \quad (28)$$

where $f^{(k,l)} = \frac{\partial^{k+l} f}{\partial x_1^k \partial x_2^l}$, and e_i is a residual term. By substituting equation (28) into equation (27) and expressing the resulting equation in a vector form

$$\begin{aligned} E_s[x_{t,i} - x_{s,i}] &= f_i(t - s) \\ &+ \left(f_i^{(1,0)} \quad f_i^{(0,1)} \quad \frac{1}{2}f_i^{(2,0)} \quad \frac{1}{2}f_i^{(0,2)} \quad f_i^{(1,1)} \right) \int_s^t \Psi_{s,u} du + R_i , \end{aligned} \quad (29)$$

where X_s is omitted for notational convenience, and $R_i = E_s[e_i]$.

Next, by applying the Ito formula to $(x_{t,1} - x_{s,1})^2$ and taking the conditional expectation,

$$E_s[(x_{t,1} - x_{s,1})^2] = E_s \left[\int_s^t \{2f_1(X_u)(x_{u,1} - x_{s,1}) + g_{11}(X_u)\} du \right], \quad (30)$$

where $g_{11} = \xi_1' \xi_1$. By applying the Taylor expansion to $f_1(X_u)$ and $g_{11}(X_u)$ around X_s up to the first and second orders, respectively, the integrand of equation (30) becomes

$$\begin{aligned} & 2f_1(X_u)(x_{u,1} - x_{s,1}) + g_{11}(X_u) \\ &= g_{11}(X_s) + \{2f_1(X_s) + g_{11}^{(1,0)}(X_s)\}(x_{u,1} - x_{s,1}) + g_{11}^{(0,1)}(X_s)(x_{u,2} - x_{s,2}) \\ &+ \{2f_1^{(1,0)}(X_s) + \frac{1}{2}g_{11}^{(2,0)}(X_s)\}(x_{u,1} - x_{s,1})^2 + \frac{1}{2}g_{11}^{(0,2)}(X_s)(x_{u,2} - x_{s,2})^2 \\ &+ \{2f_1^{(0,1)}(X_s) + g_{11}^{(1,1)}(X_s)\}(x_{u,1} - x_{s,1})(x_{u,2} - x_{s,2}) + e_{11}, \end{aligned} \quad (31)$$

where $g^{(k,l)}$ is defined analogously with $f^{(k,l)}$, and e_{11} is a residual term. By substituting equation (31) into equation (30),

$$\begin{aligned} E_s[(x_{t,1} - x_{s,1})^2] &= g_{11}(t - s) \\ &+ \left(2f_1 + g_{11}^{(1,0)} \quad g_{11}^{(0,1)} \quad 2f_1^{(1,0)} + \frac{1}{2}g_{11}^{(2,0)} \quad \frac{1}{2}g_{11}^{(0,2)} \quad 2f_1^{(0,1)} + g_{11}^{(1,1)} \right) \\ &\times \int_s^t \Psi_{s,u} du + R_{11}, \end{aligned} \quad (32)$$

where $R_{11} = E_s[e_{11}]$. A similar manipulation is applied to $E_s[(x_{t,2} - x_{s,2})^2]$ and $E_s[(x_{t,1} - x_{s,1})(x_{t,2} - x_{s,2})]$. Expressing the resulting equations together in a vector form leads to

$$\Psi_{s,t} = A(X_s) \int_s^t \Psi_{s,u} du + b(X_s)(t - s) + R, \quad (33)$$

where

$$A = \begin{pmatrix} f_1^{(1,0)} & f_1^{(0,1)} & \frac{1}{2}f_1^{(2,0)} & \frac{1}{2}f_1^{(0,2)} & f_1^{(1,1)} \\ f_2^{(1,0)} & f_2^{(0,1)} & \frac{1}{2}f_2^{(2,0)} & \frac{1}{2}f_2^{(0,2)} & f_2^{(1,1)} \\ 2f_1 + g_{11}^{(1,0)} & g_{11}^{(0,1)} & 2f_1^{(1,0)} + \frac{1}{2}g_{11}^{(2,0)} & \frac{1}{2}g_{11}^{(0,2)} & 2f_1^{(0,1)} + g_{11}^{(1,1)} \\ g_{22}^{(1,0)} & 2f_2 + g_{22}^{(0,1)} & \frac{1}{2}g_{22}^{(2,0)} & 2f_2^{(0,1)} + \frac{1}{2}g_{22}^{(0,2)} & 2f_2^{(1,0)} + g_{22}^{(1,1)} \\ f_2 + g_{12}^{(1,0)} & f_1 + g_{12}^{(0,1)} & f_2^{(1,0)} + \frac{1}{2}g_{12}^{(2,0)} & f_1^{(0,1)} + \frac{1}{2}g_{12}^{(0,2)} & f_1^{(1,0)} + f_2^{(0,1)} + g_{12}^{(1,1)} \end{pmatrix},$$

$$b = (f_1 \quad f_2 \quad g_{11} \quad g_{22} \quad g_{12})',$$

$$R = (R_1 \quad R_2 \quad R_{11} \quad R_{22} \quad R_{12})'.$$

Equation (33) can be solved as

$$\Psi_{s,t} = \int_s^t e^{A(X_s)(t-u)} b(X_s) du + \hat{R}. \quad (34)$$

If, in addition, A is invertible,

$$\Psi_{s,t} = A^{-1}(X_s)\{e^{A(X_s)(t-s)} - I\}b(X_s) + \hat{R}. \quad (35)$$

Equations (33)–(35) hold for any (n, d) . In general, $\Psi_{s,t}$ consists of $\binom{n+d}{n} - 1 = (n+d)!/(n!d!) - 1$ elements. Correspondingly, up to n -th derivatives of f_i and g_{ij} ($i, j = 1, \dots, d$) are taken to compute the elements of $A(X_s)$. In the analysis, $(n, d) = (3, 3)$ is used for computing bond yields.

It is noted that R in equation (33) contains conditional expectations of derivatives of f_i higher than the first order and derivatives of g_{ij} higher than the second order. Then, if f_i and g_{ij} are linear and quadratic in X_s , respectively, there is no residual term. This is the case for the proposed models (both LSD and QEV), and conditional first and second moments, used for the quasi-maximum likelihood method, are exact.

A2. Application to bond prices

To apply the approximation method to the pricing of bonds, define

$$z_{s,t} = \exp \left\{ - \int_s^t r(X_u) du \right\}, \quad (36)$$

and the price of a zero-coupon bond at time t maturing at time T is equal to the conditional first moment of $z_{t,T}$ under the risk-neutral probability measure. This (actually $E_t^Q[z_{t,T} - z_{t,t}]$) is computed as one of the elements of the moment vector, $\Psi_{t,T}$. Specifically, a state vector is first extended as $\hat{X}_t = (X_t' z_{s,t})'$, where X_t is a d -dimensional diffusion process and $z_{s,t}$ is treated as the $(d+1)$ -th process. By the Ito formula,

$$dz_{s,t} = -r(X_t)z_{s,t}dt, \quad z_{s,s} = 1, \quad (37)$$

and

$$f_{d+1}(\hat{X}_t) = -r(X_t)z_{s,t}, \quad (38)$$

$$g_{i,d+1}(\hat{X}_t) = 0 \quad (i = 1, \dots, d+1). \quad (39)$$

Then, the elements of $A(\hat{X}_t)$ can be readily computed by taking appropriate derivatives of f_i ($i = 1, \dots, d+1$) (the risk-neutral drift functions here) and g_{ij} ($i, j = 1, \dots, d+1; i \leq j$). The accuracy of the approximation of $E_t^Q[z_{t,T}]$ is investigated in Appendix B.

Appendix B. Accuracy of the approximation of bond prices

The purpose of this section is to let the cost of using the approximation be known. By construction of the method, the accuracy becomes worse the longer the time interval, $t - s$. Here, the interval is up to ten years for pricing bonds, which may raise concerns with the application of this method. To check the accuracy of the approximation, two cases with and without closed-form solutions of no-arbitrage bond prices are considered.

B1. Comparison with closed-form solution for bond prices

The $A_1(3)$ model is treated as the true model having closed-form solution. The examination of the accuracy is conducted in three steps according to the degree of approximation involved. Let Θ_0 be the parameter vector of the $A_1(3)$ model, the elements of which are set at the estimates presented in Table 1. The extracted state vector can then be expressed as $X(Y_t^p; \Theta_0)$, where $Y_t^p = (y_{t,0.5} \ y_{t,2} \ y_{t,10})'$.

In the first step, we examine the accuracy of the approximation in terms of pricing bonds alone. Specifically, both Θ_0 and $X(Y_t^p; \Theta_0)$ are given as input for the approximation method. Then, compare:

$$Y(X(Y_t^p; \Theta_0), \tau; \Theta_0) \quad v.s. \quad \tilde{Y}(X(Y_t^p; \Theta_0), \tau; \Theta_0) \quad (\tau = \{0.5, 1, 2, 3, 5, 10\}) ,$$

where Y and \tilde{Y} stand for the closed-form and approximate functions, respectively (the tilde symbol is used for emphasizing that the approximation method is used).

In the second step, we examine the accuracy of the approximation in terms of extracting state variables as well as pricing bonds. Here, only Θ_0 is given. Using the approximation method, the state vector is first extracted, which is denoted as $\tilde{X}(Y_t^p; \Theta_0)$, and the rest of the yields are computed. Then, compare:

$$\begin{aligned} X(Y_t^p; \Theta_0) & \quad v.s. \quad \tilde{X}(Y_t^p; \Theta_0) , \\ Y(X(Y_t^p; \Theta_0), \tau; \Theta_0) & \quad v.s. \quad \tilde{Y}(\tilde{X}(Y_t^p; \Theta_0), \tau; \Theta_0) \quad (\tau = \{1, 3, 5\}) . \end{aligned}$$

Note that at $\tau = \{0.5, 2, 10\}$, both $Y(X_t, \tau; \Theta_0)$ and $\tilde{Y}(\tilde{X}_t, \tau; \Theta_0)$ are equal to the observed yields by construction of the inversion method.

In the last step, we examine the accuracy of the approximation in terms of estimating model parameters as well as pricing bonds and extracting state variables. Here, no prior information is given regarding the true value of the parameter or state vector. Instead, using the approximation

method, the parameter vector of the $A_1(3)$ model is first estimated; denote it as $\tilde{\Theta}_0$. Next, the state vector is extracted; denote it as $\tilde{X}(Y_t^p; \tilde{\Theta}_0)$. Finally, the rest of the yields are computed. Then, compare:

$$\begin{aligned} \Theta_0 & \quad v.s. \quad \tilde{\Theta}_0, \\ X(Y_t^p; \Theta_0) & \quad v.s. \quad \tilde{X}(Y_t^p; \tilde{\Theta}_0), \\ Y(X(Y_t^p; \Theta_0), \tau; \Theta_0) & \quad v.s. \quad \tilde{Y}(\tilde{X}(Y_t^p; \tilde{\Theta}_0), \tau; \tilde{\Theta}_0) \quad (\tau = \{1, 3, 5\}). \end{aligned}$$

It is noted that the accuracy in the third step, which is a more realistic setting, is not examined by Takamizawa and Shoji (2009).

Apart from the parameter vector, the key input for these comparisons is Y_t^p . The following nine observations of Y_t^p are selected. First, Y_t^p is transformed to the conventional level (lev_t), slope (slo_t), and curvature (cur_t) factors by

$$\begin{pmatrix} lev_t \\ slo_t \\ cur_t \end{pmatrix} = \begin{pmatrix} 0 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 2 & -1 \end{pmatrix} \begin{pmatrix} y_{t,0.5} \\ y_{t,2} \\ y_{t,10} \end{pmatrix}. \quad (40)$$

Then, from the whole sample period, three dates are selected in which lev_t takes the minimum, median, or maximum value. Likewise, the three dates are selected for each of the other proxies, leading to nine dates in total. In this way, the accuracy of the approximation is evaluated at both typical and atypical states.

Table B1 presents the differences in yields/factors between the approximate and closed-form solutions in basis point (bps). Panel A presents the results for the first step comparison, where the true values of the parameter and state vectors are given as input for the approximation method. For maturities of up to five years, the approximation errors are negligibly small at all states. Even for the ten-year maturity, the error exceeds 2 bps only at the maximum-level and slope states.

Panel B presents the results for the second-step comparison, where only the true value of the parameter vector is given. A systematic pattern is found in the approximation errors for the state variables. Specifically, the approximation undervalues both r and x_3 and overvalues μ . Again, the accuracy is the worst at the maximum-level and slope states. The approximation errors for the remaining yields, with maturities of up to five years, are small.

Panel C presents the results for the third-step comparison, where no prior information is given. Compared with Panel B, the magnitude of the approximation errors for the state variables is generally larger, reflecting also the difference in parameter estimates between the approximate and

closed-form solutions. This is the reality. Here, an error pattern is less clear though it is found that x_3 (μ) tends to be overvalued (undervalued). Also, the difficulty of the approximation method is not limited to the maximum-level and slope states. For example, the approximation error for x_3 exceeds 15 bps at the median-slope and maximum-curvature states as well as at the maximum-level state. The remaining yields are accurately computed as in the second-step comparison.

B2. Comparison with numerical solution for bond prices

The Monte Carlo (MC) method is employed to obtain the solution of no-arbitrage bond prices for the $A_1(3)$ -based LSD and QEV models. Let Θ be the true parameter vector, the elements of which are set at the estimates presented in Table 2 (LSD) and Table 3 (QEV). The state vector is extracted by the approximation method, but not the MC method, with which the extraction is computationally very demanding. The extracted state vector is denoted as $\tilde{X}(Y_t^p; \Theta)$. Then, compare:

$$Y(\tilde{X}(Y_t^p; \Theta), \tau; \Theta) \quad v.s. \quad \tilde{Y}(\tilde{X}(Y_t^p; \Theta), \tau; \Theta) \quad (\tau = \{0.5, 1, 2, 3, 5, 10\}) .$$

In the MC simulations, $\{X_s\}_t^{t+\tau}$ is generated from the risk-neutral distribution (1), where dt is replaced by $\Delta t = 1/1,000$, an interval corresponding roughly to four observations per day. The number of repetition is set at 10,000 with antithetic variates.

Table B2 presents the differences in yields between the approximate and MC solutions in bps. Generally, the error pattern is similar to that for the first-step comparison with the closed-form solution. For maturities of up to five years, the approximation errors are within 1 bp at all states for both models. For the ten-year yield, the approximation error exceeds 2 bps only at the maximum-level state for the LSD model and at the maximum-level and slope states for the QEV model.

It is noted, however, that this comparison scheme does not take into consideration the approximation errors in the parameter and state vectors. In reality, therefore, the approximation errors for the resulting yields would be larger, as is the case for the third-step comparison with the closed-form solution.

References

- Andersen, T. G., and L. Benzoni, 2010, Do bonds span volatility risk in the U.S. Treasury market? A specification test for affine term structure models, *Journal of Finance* 65, 603-653.
- Bikbov, R., and M. Chernov, 2009, Unspanned Stochastic Volatility in Affine Models: Evidence from Eurodollar Futures and Options, *Management Science* 55, 1292-1305.
- Bollerslev, T., 1990, Modelling the coherence in short-run nominal exchange rates: A multivariate generalized arch model, *Review of Economics and Statistics* 72, 498-505.
- Collin-Dufresne, P., and R. S. Goldstein, 2002, Do bonds span the fixed income markets? Theory and evidence for unspanned stochastic volatility, *Journal of Finance* 57, 1685-1730.
- Collin-Dufresne, P., R. S. Goldstein, and C. S. Jones, 2008, Identification of maximal affine term structure models, *Journal of Finance* 63, 743-795.
- Collin-Dufresne, P., R. S. Goldstein, and C. S. Jones, 2009, Can interest rate volatility be extracted from the cross section of bond yields? *Journal of Financial Economics* 94, 47-66.
- Coroneo, L., and I. Fabrizio, 2016, Comparing predictive accuracy in small samples using fixed-smoothing asymptotics, working paper, University of York.
- Dai, Q., and K. J. Singleton, 2000, Specification analysis of affine term structure models, *Journal of Finance* 55, 1943-1978.
- Diebold, F. X., 2015, Comparing predictive accuracy, twenty years later: A personal perspective on the use and abuse of Diebold-Mariano tests, *Journal of Business and Economic Statistics* 33, 1-9.
- Diebold, F. X., and R. S. Mariano, 1995, Comparing predictive accuracy, *Journal of Business and Economic Statistics* 13, 253-263.
- Duffee, G. R., 2002, Term premia and interest rate forecasts in affine models, *Journal of Finance* 57, 405-443.
- Engle, R. F., 2002, Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models, *Journal of Business and Economic Statistics* 20, 339-350.

- Fan, R., A. Gupta, and P. Ritchken, 2003, Hedging in the possible presence of unspanned stochastic volatility: Evidence from swaption markets, *Journal of Finance* 58, 2219-2248.
- Han, B., 2007, Stochastic volatilities and correlations of bond yields, *Journal of Finance* 62, 1491-1524.
- Jacobs, K., and L. Karoui, 2009, Conditional volatility in affine term-structure models: Evidence from Treasury and swap markets, *Journal of Financial Economics* 91, 288-318.
- Jarrow, R. A., H. Li, and F. Zhao, 2007, Interest rate caps smile too! But can the LIBOR market models capture the smile? *Journal of Finance* 62, 345-382.
- Kiefer, N. M., and T. J. Vogelsang, 2005, A new asymptotic theory for heteroskedasticity autocorrelation robust tests, *Econometric Theory* 21, 1130-1164.
- Litterman, R., and J. Scheinkman, 1991, Common factors affecting bond returns, *Journal of Fixed Income* 1, 54-61.
- Longstaff, F. A., P. Santa-Clara, and E. S. Schwartz, 2001, The relative valuation of caps and swaptions: Theory and empirical evidence, *Journal of Finance* 56, 2067-2109.
- Patton, A. J., 2015, Comment for Diebold (2015), *Journal of Business and Economic Statistics* 33, 22-24.
- Pérignon, C., and C. Villa, 2006, Sources of time variation in the covariance matrix of interest rates, *Journal of Business* 79, 1535-1549.
- Takamizawa, H., 2015, Predicting interest rate volatility using information on the yield curve, *International Review of Finance* 15, 347-386.
- Takamizawa, H., and I. Shoji, 2009, Modeling the term structure of interest rates with general diffusion processes: A moment approximation approach, *Journal of Economic Dynamics and Control* 33, 65-77.
- Thompson, S., 2008, Identifying term structure volatility from the LIBOR-swap curve, *Review of Financial Studies* 21, 819-854.

	Factor (Index)					
	r ($i = 1$)		μ ($i = 2$)		x_3 ($i = 3$)	
Risk-neutral drift						
$k_{\cdot,0}^Q$	0		0		0.0016	(0.0001)
$k_{\cdot,1}^Q$	0		-0.887	(0.025)	0	
$k_{\cdot,2}^Q$	1		-1.852	(0.051)	0	
$k_{\cdot,3}^Q$	0		1		-0.0064	(0.0019)
Physical drift						
$Diag(K_1)$	-0.238	(0.051)	-2.465	(0.430)	-0.057	(0.044)
Covariance matrix						
$s_{rr,0}$	0					
$s_{\cdot,\mu,0}$	0		0			
$s_{rr,3} \times 10^2$	0.112	(0.005)				
$s_{\cdot,\mu,3} \times 10^2$	-0.238	(0.021)	1.782	(0.100)		
$s_{3,3} \times 10^2$	0.012	(0.005)	0.156	(0.019)	0.155	(0.009)
$\varsigma \times 10^4$	6.150	(0.096)				
LogL	22276					

Table 1: Parameter estimates (standard errors) for $A_1(3)$

A state vector X_t consists of $X_t = (r_t \mu_{2,t} x_{3,t})'$, where r_t is the instantaneous risk-free rate, μ_t is the risk-neutral drift of change in r_t , and $x_{3,t}$ is a volatility factor. The parameters in the drift and covariance matrix are given in (3) and (4), respectively. Among the physical-drift parameters, only the diagonal elements of K_1 ($Diag(K_1)$) are estimated. To normalize $x_{3,t}$, $k_{\mu,3}^Q = 1$ is placed. Insignificant parameters are set to zero and parameters reaching the boundary (if any) are set to the boundary value before the final round of estimation. ς is the standard deviation of measurement errors and LogL is the log-likelihood value. In-sample data from January 4, 1991 to April 9, 2003 are used for the estimation (641 observations).

	Factor (Index)		
	r ($i = 1$)	μ ($i = 2$)	x_3 ($i = 3$)
Risk-neutral drift			
$k_{\cdot,0}^Q$	0	0	0.0016 (0.0001)
$k_{\cdot,1}^Q$	0	-0.883 (0.024)	0
$k_{\cdot,2}^Q$	1	-1.765 (0.047)	0
$k_{\cdot,3}^Q$	0	1	-0.0078 (0.0018)
Physical drift			
$Diag(K_1)$	-0.203 (0.046)	-2.364 (0.388)	-0.068 (0.042)
Volatility function $h_i(X_t)$			
β_0^i	-0.006 (0.001)	0	0.0096 (0.0004)
β_1^i	-0.049 (0.008)	-0.320 (0.048)	0
β_2^i	-0.084 (0.011)	-0.171 (0.038)	-0.043 (0.012)
β_3^i	0.286 (0.030)	0.790 (0.045)	0
Correlation matrix R			
ρ_{12}	-0.517 (0.037)		
$\rho_{.3}$	0.101 (0.043)	0.311 (0.033)	
$\varsigma \times 10^4$	6.132 (0.097)		
LogL	22346		

Table 2: Parameter estimates (standard errors) for $A_1(3)$ -LSD

A state vector X_t consists of $X_t = (r_t \ x_{2,t} \ x_{3,t})'$. The risk-neutral and physical drift vectors are the same as those in the $A_1(3)$ model. The covariance matrix Σ_t is decomposed as $\Sigma_t = H_t R H_t$, where R is a constant correlation matrix and H_t is a diagonal matrix with the i -th diagonal element specified as $h_i(X_t) = \beta_0^i + \beta^i X_t$ ($i = 1, 2, 3$). Insignificant parameters are set to zero before the final round of estimation. ς is the standard deviation of measurement errors and LogL is the log-likelihood value. In-sample data from January 4, 1991 to April 9, 2003 are used for the estimation (641 observations).

	Factor (Index)		
	r ($i = 1$)	μ ($i = 2$)	x_3 ($i = 3$)
Risk-neutral drift			
$k_{\cdot,0}^Q$	0	0	0.0017 (0.0001)
$k_{\cdot,1}^Q$	0	-0.915 (0.026)	0
$k_{\cdot,2}^Q$	1	-1.852 (0.051)	0
$k_{\cdot,3}^Q$	0	1	-0.0073 (0.0019)
Physical drift			
$Diag(K_1)$	-0.170 (0.032)	-2.455 (0.560)	-0.077 (0.044)
Eigenvalue function $l_i(X_t)$			
$c_0^i \times 10^3$	10^{-5}	10^{-5}	0.070 (0.007)
m_1^i	0	0	0
m_2^i	0.045 (0.008)	0.495 (0.181)	0
m_3^i	0.081 (0.013)	1.308 (0.368)	0.106 (0.027)
$\sin \varphi_1^i$	0.584 (0.025)	0.343 (0.118)	0
$\sin \varphi_2^i$	0.247 (0.055)	0.274 (0.101)	0.221 (0.087)
$\sin \varphi_3^i$	0.860 (0.042)	0.804 (0.103)	0.911 (0.032)
Eigenvector matrix P			
$\sin \varphi_i^P$	0.136 (0.008)	-0.244 (0.032)	0.065 (0.013)
$\varsigma \times 10^4$	6.139 (0.096)		
LogL	22388		

Table 3: Parameter estimates (standard errors) for $A_1(3)$ -QEV

A state vector X_t consists of $X_t = (r_t \ x_{2,t} \ x_{3,t})'$. The risk-neutral and physical drift vectors are the same as those in the $A_1(3)$ model. The covariance matrix Σ_t is decomposed as $\Sigma_t = PL_tP'$: P is an orthonormal eigenvector matrix given in (13), and L_t is a diagonal eigenvalue matrix with the i -th diagonal element specified as $l_i(X_t) = c_0^i + X_t' \Gamma^i X_t$ ($i = 1, 2, 3$), where $c_0^i > 0$ and Γ^i is a non-negative definite matrix given in (15)–(17). Insignificant parameters are set to zero and parameters reaching the boundary (if any) are set to the boundary value before the final round of estimation. ς is the standard deviation of measurement errors and LogL is the log-likelihood value. In-sample data from January 4, 1991 to April 9, 2003 are used for the estimation (641 observations).

	Factor (Index)					
	r ($i = 1$)		x_2 ($i = 2$)		x_3 ($i = 3$)	
Risk-neutral drift						
$k_{\cdot,0}^Q$	0		0.0034	(0.0004)	0.068	(0.009)
$k_{\cdot,1}^Q$	-1.075	(0.078)	0		0	
$k_{\cdot,2}^Q$	1		-0.013	(0.002)	0	
$k_{\cdot,3}^Q$	-1		0		-0.861	(0.057)
Physical drift						
$Diag(K_1)$	-1.310	(0.098)	-0.439	(0.223)	-0.673	(0.105)
Covariance matrix						
$s_{rr,0}$	0					
$s_{rr,2} \times 10^2$	0.025	(0.007)				
$s_{.2,2} \times 10^2$	0.005	(0.003)	0.094	(0.013)		
$s_{rr,3} \times 10^2$	0.035 (0.011)					
$s_{.3,3} \times 10^2$	0.099	(0.018)			0.854	(0.121)
$\varsigma \times 10^4$	6.151 (0.091)					
LogL	22278					

Table 4: Parameter estimates (standard errors) for $A_2(3)$

A state vector X_t consists of $X_t = (r_t \ x_{2,t} \ x_{3,t})'$, where r_t is the instantaneous risk-free rate and $x_{i,t}$ ($i = 2, 3$) are volatility factors. The parameters in the drift and covariance matrix are given in (3) and (4), respectively. Among the physical-drift parameters, only the diagonal elements of K_1 ($Diag(K_1)$) are estimated. To normalize $x_{2,t}$ and $x_{3,t}$, $k_{r,2}^Q = -k_{r,3}^Q = 1$ is placed. Insignificant parameters are set to zero and parameters reaching the boundary (if any) are set to the boundary value before the final round of estimation. ς is the standard deviation of measurement errors and LogL is the log-likelihood value. In-sample data from January 4, 1991 to April 9, 2003 are used for the estimation (641 observations).

	Factor (Index)		
	r ($i = 1$)	x_2 ($i = 2$)	x_3 ($i = 3$)
Risk-neutral drift			
$k_{.,0}^Q$	0	0.0044 (0.0005)	0.096 (0.013)
$k_{.,1}^Q$	-1.061 (0.058)	0	0
$k_{.,2}^Q$	1	-0.017 (0.002)	0
$k_{.,3}^Q$	-1	0	-0.867 (0.047)
Physical drift			
$Diag(K_1)$	-1.288 (0.083)	-0.414 (0.174)	-0.722 (0.083)
Volatility function $h_i(X_t)$			
β_0^i	0	0	0
β_1^i	0	0	0
β_2^i	0.034 (0.006)	0	0.064 (0.025)
β_3^i	0.017 (0.007)	0.093 (0.012)	0.131 (0.025)
Correlation matrix R			
ρ_{12}	0		
$\rho_{.3}$	0.404 (0.032)	0	
$\varsigma \times 10^4$	6.138 (0.090)		
LogL	22303		

Table 5: Parameter estimates (standard errors) for $A_2(3)$ -LSD

A state vector X_t consists of $X_t = (r_t \ x_{2,t} \ x_{3,t})'$. The risk-neutral and physical drift vectors are the same as those in the $A_2(3)$ model. The covariance matrix Σ_t is decomposed as $\Sigma_t = H_t R H_t$, where R is a constant correlation matrix and H_t is a diagonal matrix with the i -th diagonal element specified as $h_i(X_t) = \beta_0^i + \beta^{i'} X_t$ ($i = 1, 2, 3$). Insignificant parameters are set to zero before the final round of estimation. ς is the standard deviation of measurement errors and LogL is the log-likelihood value. In-sample data from January 4, 1991 to April 9, 2003 are used for the estimation (641 observations).

	Factor (Index)		
	r ($i = 1$)	x_2 ($i = 2$)	x_3 ($i = 3$)
Risk-neutral drift			
$k_{\cdot,0}^Q$	0	0.0032 (0.0006)	0.064 (0.020)
$k_{\cdot,1}^Q$	-1.084 (0.076)	0	0
$k_{\cdot,2}^Q$	1	-0.013 (0.002)	0
$k_{\cdot,3}^Q$	-1	0	-0.884 (0.058)
Physical drift			
$Diag(K_1)$	-1.301 (0.094)	-0.439 (0.229)	-0.676 (0.126)
Eigenvalue function $l_i(X_t)$			
$c_0^i \times 10^3$	10^{-5}	0.116 (0.019)	10^{-5}
m_1^i	0	0	0
m_2^i	0.0015 (0.0006)	0	0.032 (0.017)
m_3^i	0.224 (0.030)	0.029 (0.017)	0.115 (0.050)
$\sin \varphi_1^i$	0	0	0
$\sin \varphi_2^i$	-0.677 (0.041)	0	0.708 (0.121)
$\sin \varphi_3^i$	0.710 (0.019)	0.448 (0.080)	-1
Eigenvector matrix P			
$\sin \varphi_i^P$	0	-0.123 (0.013)	0
$\varsigma \times 10^4$	6.151 (0.092)		
LogL	22334		

Table 6: Parameter estimates (standard errors) for $A_2(3)$ -QEV

A state vector X_t consists of $X_t = (r_t \ x_{2,t} \ x_{3,t})'$. The risk-neutral and physical drift vectors are the same as those in the $A_2(3)$ model. The covariance matrix Σ_t is decomposed as $\Sigma_t = PL_tP'$: P is an orthonormal eigenvector matrix given in (13), and L_t is a diagonal eigenvalue matrix with the i -th diagonal element specified as $l_i(X_t) = c_0^i + X_t' \Gamma^i X_t$ ($i = 1, 2, 3$), where $c_0^i > 0$ and Γ^i is a non-negative definite matrix given in (15)–(17). Insignificant parameters are set to zero and parameters reaching the boundary (if any) are set to the boundary value before the final round of estimation. ς is the standard deviation of measurement errors and LogL is the log-likelihood value. In-sample data from January 4, 1991 to April 9, 2003 are used for the estimation (641 observations).

	Factor (Index)		
	x_1 ($i = 1$)	x_2 ($i = 2$)	x_3 ($i = 3$)
Risk-neutral drift			
$k_{\cdot,0}^Q$	0.0025 (0.0001)	0.001	0.012 (0.006)
$k_{\cdot,1}^Q$	-0.001	0	0
$k_{\cdot,2}^Q$	0	-0.989 (0.244)	0
$k_{\cdot,3}^Q$	0	7.780 (3.067)	-0.917 (0.220)
Physical drift			
$Diag(K_1)$	-0.005 (0.004)	-0.907 (0.248)	-0.784 (0.224)
Covariance matrix			
s_{ii}	0.0095 (0.0003)	0.035 (0.004)	0.030 (0.008)
$\varsigma \times 10^4$	6.220 (0.094)		
LogL	21955		

Table 7: Parameter estimates (standard errors) for $A_3(3)$

A state vector X_t consists of volatility factors alone: $X_t = (x_{1,t} \ x_{2,t} \ x_{3,t})'$. The instantaneous risk-free rate r_t is given by $r_t = -1 + x_{1,t} - x_{2,t} + x_{3,t}$. The parameters in the drift and covariance matrix are given in (3) and (4), respectively. Among the physical-drift parameters, only the diagonal elements of K_1 ($Diag(K_1)$) are estimated. Insignificant parameters are set to zero and parameters reaching the boundary (if any) are set to the boundary value before the final round of estimation. ς is the standard deviation of measurement errors and LogL is the log-likelihood value. In-sample data from January 4, 1991 to April 9, 2003 are used for the estimation (641 observations).

	x_1 ($i = 1$)		Factor (Index)		x_3 ($i = 3$)	
			x_2 ($i = 2$)			
Risk-neutral drift						
$k_{.,0}^Q$	0.031	(0.002)	0.103	(0.017)	0	
$k_{.,1}^Q$	-0.025	(0.002)	0		0	
$k_{.,2}^Q$	0		-0.676	(0.025)	0	
$k_{.,3}^Q$	0		0		-1.437	(0.072)
Physical drift						
$Diag(K_1)$	-0.028	(0.003)	-0.623	(0.039)	-1.479	(0.114)
Volatility function $h_i(X_t)$						
β_0^i	0		0.009	(0.004)	0.011	(0.004)
β_1^i	0		0		0	
β_2^i	0.066	(0.009)	0.179	(0.015)	0.180	(0.015)
β_3^i	0		0		0	
Correlation matrix R						
ρ_{12}	0.264	(0.030)				
$\rho_{.3}$	0		0.948	(0.012)		
$\varsigma \times 10^4$	6.178	(0.093)				
LogL	22289					

Table 8: Parameter estimates (standard errors) for $A_3(3)$ -LSD

A state vector X_t consists of $X_t = (x_{1,t} \ x_{2,t} \ x_{3,t})'$, and the instantaneous risk-free rate r_t is given by $r_t = -1 + x_{1,t} - x_{2,t} + x_{3,t}$. The risk-neutral and physical drift vectors are the same as those in the $A_3(3)$ model. The covariance matrix Σ_t is decomposed as $\Sigma_t = H_t R H_t$, where R is a constant correlation matrix and H_t is a diagonal matrix with the i -th diagonal element specified as $h_i(X_t) = \beta_0^i + \beta^i X_t$ ($i = 1, 2, 3$). Insignificant parameters are set to zero before the final round of estimation. ς is the standard deviation of measurement errors and LogL is the log-likelihood value. In-sample data from January 4, 1991 to April 9, 2003 are used for the estimation (641 observations).

	Factor (Index)					
	$x_1 (i = 1)$		$x_2 (i = 2)$		$x_3 (i = 3)$	
Risk-neutral drift						
$k_{.,0}^Q$	0.035	(0.002)	0.093	(0.014)	-0.143	(0.066)
$k_{.,1}^Q$	-0.026	(0.002)	0		0	
$k_{.,2}^Q$	0		-0.689	(0.024)	0	
$k_{.,3}^Q$	0		0		-1.500	(0.075)
Physical drift						
$Diag(K_1)$	-0.031	(0.003)	-0.609	(0.043)	-1.565	(0.087)
Eigenvalue function $l_i(X_t)$						
$c_0^i \times 10^3$	10^{-5}		0.221	(0.012)	1.822	(0.365)
m_1^i	0		0		0	
m_2^i	0		0		0	
m_3^i	0.0007	(0.0002)	0		0.972	(0.220)
$\sin \varphi_1^i$	0		0		0	
$\sin \varphi_2^i$	0		0		0	
$\sin \varphi_3^i$	-1		0		-0.580	(0.161)
Eigenvector matrix P						
$\sin \varphi_i^P$	-0.782	(0.009)	0		-0.667	(0.006)
$\varsigma \times 10^4$	6.148	(0.093)				
LogL	22300					

Table 9: Parameter estimates (standard errors) for $A_3(3)$ -QEV

A state vector X_t consists of $X_t = (x_{1,t} \ x_{2,t} \ x_{3,t})'$, and the instantaneous risk-free rate r_t is given by $r_t = -1 + x_{1,t} - x_{2,t} + x_{3,t}$. The risk-neutral and physical drift vectors are the same as those in the $A_3(3)$ model. The covariance matrix Σ_t is decomposed as $\Sigma_t = PL_tP'$: P is an orthonormal eigenvector matrix given in (13), and L_t is a diagonal eigenvalue matrix with the i -th diagonal element specified as $l_i(X_t) = c_0^i + X_t'\Gamma^i X_t$ ($i = 1, 2, 3$), where $c_0^i > 0$ and Γ^i is a non-negative definite matrix given in (15)–(17). Insignificant parameters are set to zero and parameters reaching the boundary (if any) are set to the boundary value before the final round of estimation. ς is the standard deviation of measurement errors and LogL is the log-likelihood value. In-sample data from January 4, 1991 to April 9, 2003 are used for the estimation (641 observations).

τ	In-sample					Out-of-sample						
	0.5	1	2	3	5	10	0.5	1	2	3	5	10
Panel A: Root mean squared errors												
$h = 4$												
A1	37.1	36.7	33.4	34.4	30.5	28.8	44.0	37.1	44.3	46.9	50.5	49.6
A1-LSD	32.4*	32.8	32.5	33.6	28.8	26.9	39.5	35.6	40.8*	40.5*	40.4*	39.2
A1-QEV	34.0	33.2	32.1	34.1	30.0	29.5	46.2	39.7	36.5	34.4	32.2	31.2
$h = 13$												
A1	32.9	31.1	26.0	26.0	23.5	22.4	40.9	33.2	41.8	45.3	49.0	46.8
A1-LSD	28.3*	27.1	24.7	24.7	21.7	19.7*	37.5	32.2	38.0*	38.5*	38.4*	35.7
A1-QEV	29.6	26.9	23.9	24.7	22.8	22.5	43.7	34.0	30.6	30.2*	29.2*	26.8
$h = 26$												
A1	32.8	28.5	22.1	21.9	20.2	18.7	40.5	31.6	40.6	44.4	47.0	41.7
A1-LSD	29.2	25.2	20.8	20.1	18.3	15.7*	37.9	31.4	37.3*	38.1*	37.5*	33.0
A1-QEV	30.3	24.7	20.3	20.4	19.5	18.7	44.5	33.1	31.3	32.2	31.7*	27.6*
Panel B: Correlations												
$h = 4$												
A1	0.54	0.44	0.19	0.08	-0.06	-0.16	-0.12	0.01	0.10	0.09	0.01	-0.09
A1-LSD	0.62	0.56	0.36	0.28	0.21	0.11	0.20	0.27	0.48	0.55	0.64	0.62
A1-QEV	0.52	0.54	0.45	0.35	0.28	0.13	0.55	0.55	0.62	0.66	0.70	0.71
$h = 13$												
A1	0.59	0.50	0.25	0.13	-0.09	-0.25	-0.05	0.05	0.13	0.11	0.04	-0.09
A1-LSD	0.68	0.63	0.45	0.37	0.22	0.10	0.14	0.27	0.51	0.58	0.67	0.64
A1-QEV	0.59	0.64	0.55	0.46	0.32	0.12	0.47	0.57	0.68	0.71	0.76	0.75
$h = 26$												
A1	0.60	0.54	0.29	0.18	-0.09	-0.35	0.00	0.10	0.18	0.18	0.18	0.13
A1-LSD	0.67	0.64	0.48	0.40	0.18	-0.03	0.06	0.22	0.47	0.54	0.64	0.57
A1-QEV	0.60	0.66	0.57	0.50	0.30	0.04	0.33	0.48	0.63	0.66	0.70	0.67

Table 10: Volatility prediction for $A_1(3)$ and $A_1(3)$ -LSD/QEV

Predicted is the h -week-ahead annualized standard deviation of a τ -year yield with $h = 4, 13, 26$ and $\tau = 0.5, 1, 2, 3, 5, 10$. Panel A presents the root mean squared error (RMSE) in bps and Panel B presents the correlation between predicted and realized series of volatility. The asterisk “*” indicates that the null hypothesis of equal predictive accuracy between the $A_1(3)$ and LSD/QEV models is rejected at the 5% significance level. The standard error is computed by the Newey and West method and the critical values of an asymptotic distribution of the test statistic are computed from Kiefer and Vogelsang (2005, Table 1) with the truncation lag parameter set at the nearest integer of square-root of the sample size (i.e., $\text{round}(\sqrt{T})$). The in-sample period is from January 9, 1991 to April 9, 2003 ($T = 640$), and the out-of-sample period is from April 16, 2003 to May 27, 2009 ($T = 320$).

τ	In-sample					Out-of-sample						
	0.5	1	2	3	5	10	0.5	1	2	3	5	10
Panel A: Root mean squared errors												
$h = 4$												
A2	37.8	37.1	32.9	33.5	29.1	27.5	46.5	38.8	38.7	39.5	42.5	44.7
A2-LSD	36.7*	36.3	32.9	33.5	28.8	28.2	45.4	38.4	36.2	35.1*	33.7*	34.1
A2-QEV	35.7	35.1	31.2	32.1	27.1	26.0	44.4	38.0	36.1	35.7*	35.1*	35.1
$h = 13$												
A2	36.4	35.1	27.4	25.8	22.4	20.8	48.1	40.3	35.5	35.9	38.8	40.7
A2-LSD	34.0*	33.0*	26.5	25.2	22.0	21.4	45.9*	38.4	32.6	31.2*	29.6*	29.7
A2-QEV	33.7	31.8*	24.6*	23.5	19.9	18.8	46.7	39.0	33.0	32.2*	31.3*	31.0
$h = 26$												
A2	41.5	37.8	25.9	22.1	19.1	16.7	54.5	45.3	35.6	34.8	36.2	36.0
A2-LSD	39.0*	35.6*	24.9	21.5	19.2	18.1	52.4	43.7	33.4	31.3	29.0*	27.8*
A2-QEV	39.0	34.4*	22.7*	19.4	16.5	14.7	54.6	45.4	34.5	32.6	30.9*	29.5
Panel B: Correlations												
$h = 4$												
A2	0.53	0.47	0.34	0.30	0.17	-0.10	0.23	0.37	0.54	0.56	0.46	0.16
A2-LSD	0.50	0.44	0.32	0.29	0.30	0.20	0.28	0.41	0.59	0.65	0.71	0.69
A2-QEV	0.46	0.49	0.41	0.34	0.34	0.22	0.62	0.53	0.58	0.62	0.70	0.71
$h = 13$												
A2	0.56	0.53	0.44	0.39	0.18	-0.18	0.26	0.41	0.58	0.59	0.48	0.18
A2-LSD	0.54	0.51	0.42	0.40	0.36	0.25	0.27	0.43	0.64	0.69	0.76	0.73
A2-QEV	0.52	0.55	0.50	0.45	0.41	0.28	0.49	0.48	0.60	0.64	0.73	0.74
$h = 26$												
A2	0.53	0.53	0.47	0.43	0.16	-0.28	0.21	0.37	0.55	0.55	0.48	0.27
A2-LSD	0.52	0.51	0.45	0.43	0.35	0.21	0.21	0.38	0.60	0.65	0.71	0.67
A2-QEV	0.50	0.55	0.52	0.50	0.42	0.27	0.25	0.33	0.52	0.57	0.65	0.64

Table 11: Volatility prediction for $A_2(3)$ and $A_2(3)$ -LSD/QEV

Predicted is the h -week-ahead annualized standard deviation of a τ -year yield with $h = 4, 13, 26$ and $\tau = 0.5, 1, 2, 3, 5, 10$. Panel A presents the root mean squared error (RMSE) in bps and Panel B presents the correlation between predicted and realized series of volatility. The asterisk “*” indicates that the null hypothesis of equal predictive accuracy between the $A_2(3)$ and LSD/QEV models is rejected at the 5% significance level. The standard error is computed by the Newey and West method and the critical values of an asymptotic distribution of the test statistic are computed from Kiefer and Vogelsang (2005, Table 1) with the truncation lag parameter set at the nearest integer of square-root of the sample size (i.e., $\text{round}(\sqrt{T})$). The in-sample period is from January 9, 1991 to April 9, 2003 ($T = 640$), and the out-of-sample period is from April 16, 2003 to May 27, 2009 ($T = 320$).

τ	In-sample					Out-of-sample						
	0.5	1	2	3	5	10	0.5	1	2	3	5	10
Panel A: Root mean squared errors												
$h = 4$												
A3	90.9	71.8	50.3	43.7	34.0	29.0	103.9	79.6	55.3	47.8	42.8	43.6
A3-LSD	37.4*	36.0*	32.8*	33.3*	29.0	29.2	48.6*	40.0*	35.5*	34.5*	34.2*	34.7*
A3-QEV	38.6*	39.7*	37.4*	36.7*	29.7*	26.7*	51.2*	47.8*	41.0*	36.9*	34.9*	39.6*
$h = 13$												
A3	86.9	68.6	45.1	36.3	27.3	21.9	100.0	77.5	51.9	44.0	39.0	39.3
A3-LSD	35.4*	32.5*	25.6*	24.6*	22.3	22.8	50.1*	40.3*	31.6*	30.7	30.6	30.9*
A3-QEV	39.5*	39.0*	31.7*	28.5*	22.9*	19.7	57.7*	52.5*	39.4	34.2	31.6*	35.6*
$h = 26$												
A3	86.6	68.2	42.6	32.2	23.7	17.7	100.0	78.4	51.3	42.8	37.1	35.6
A3-LSD	41.2*	34.8*	22.8*	20.3*	19.3	19.7	57.6*	45.5*	32.2*	30.7	30.0	28.7
A3-QEV	49.6*	44.1*	30.9*	25.1*	20.1	15.6	72.4*	62.7	43.1	36.3	31.8	32.6
Panel B: Correlations												
$h = 4$												
A3	0.40	0.45	0.34	0.31	0.30	0.10	0.03	0.21	0.41	0.44	0.39	0.21
A3-LSD	0.39	0.42	0.34	0.32	0.31	0.20	0.26	0.37	0.56	0.63	0.69	0.65
A3-QEV	0.38	0.36	0.28	0.24	0.25	0.15	0.29	0.37	0.53	0.60	0.67	0.60
$h = 13$												
A3	0.46	0.51	0.43	0.41	0.36	0.09	0.08	0.27	0.46	0.47	0.41	0.23
A3-LSD	0.46	0.49	0.44	0.42	0.38	0.24	0.23	0.38	0.61	0.66	0.73	0.68
A3-QEV	0.43	0.42	0.38	0.35	0.30	0.16	0.24	0.37	0.56	0.62	0.68	0.60
$h = 26$												
A3	0.46	0.49	0.45	0.44	0.34	0.00	0.10	0.27	0.44	0.44	0.38	0.24
A3-LSD	0.47	0.50	0.47	0.46	0.37	0.20	0.16	0.33	0.56	0.61	0.67	0.64
A3-QEV	0.41	0.43	0.39	0.37	0.28	0.11	0.12	0.27	0.48	0.54	0.59	0.53

Table 12: Volatility prediction for $A_3(3)$ and $A_3(3)$ -LSD/QEV

Predicted is the h -week-ahead annualized standard deviation of a τ -year yield with $h = 4, 13, 26$ and $\tau = 0.5, 1, 2, 3, 5, 10$. Panel A presents the root mean squared error (RMSE) in bps and Panel B presents the correlation between predicted and realized series of volatility. The asterisk “*” indicates that the null hypothesis of equal predictive accuracy between the $A_3(3)$ and LSD/QEV models is rejected at the 5% significance level. The standard error is computed by the Newey and West method and the critical values of an asymptotic distribution of the test statistic are computed from Kiefer and Vogelsang (2005, Table 1) with the truncation lag parameter set at the nearest integer of square-root of the sample size (i.e., $\text{round}(\sqrt{T})$). The in-sample period is from January 9, 1991 to April 9, 2003 ($T = 640$), and the out-of-sample period is from April 16, 2003 to May 27, 2009 ($T = 320$).

τ	In-sample					Out-of-sample						
	0.5	1	2	3	5	10	0.5	1	2	3	5	10
$h = 4$												
A0	22.0	28.7	30.7	30.3	28.4	26.2	23.7	28.0	27.7	29.1	30.9	31.0
A1	22.3	27.5	30.5	29.9	28.1	25.9	27.0*	29.7	29.8	29.6	30.8	31.3
A1 ^F	22.0	27.8	30.1	29.6	28.0	25.8	25.8*	28.8	28.8	28.9	30.2	31.2
A1-LSD	22.4	29.2	30.4	30.1	28.4	26.0	25.3	27.6	28.8	29.3	31.1	31.3
A1-LSD ^F	22.0	27.9	30.1	29.6	28.0	25.8	25.9*	28.8	28.9	28.8	30.3	31.3
A1-QEV	22.5	29.4	30.5	30.1	28.5	26.0	25.0	27.4	28.8	29.4	31.2	31.4
A1-QEV ^F	22.0	28.0	30.1	29.6	28.0	25.7	25.4*	28.8	28.6	28.8	30.4	31.4
$h = 13$												
A0	52.9	64.2	69.3	67.2	61.1	51.8	43.0	49.6	57.4	57.2	55.3	53.3
A1	53.5	61.0	65.7	62.5	56.1	47.8	57.8*	59.9*	64.0	60.2	56.3	54.0
A1 ^F	51.6	60.3	64.5	61.7	55.8	47.4	53.1*	55.4	60.1	57.0	54.4	54.3
A1-LSD	53.2	62.9	65.4	62.9	57.1	48.3	48.8	50.5	59.1	57.7	56.3	54.2
A1-LSD ^F	51.5	60.2	64.3	61.5	55.7	47.3*	53.1*	55.4	60.2	57.0	54.7	54.6
A1-QEV	53.7	63.6	65.6	63.1	57.1	48.2	47.8	49.9	59.3	58.0	56.6	54.4
A1-QEV ^F	51.7	60.7	64.4	61.6	55.7	47.3*	51.7*	54.6	59.4	56.7	54.7	55.0
$h = 26$												
A0	94.2	107.1	110.1	104.9	94.6	80.4	74.6	80.9	88.4	85.2	77.6	70.8
A1	94.1	99.8	100.0	92.3	81.1	69.2*	99.0*	98.0	96.6	86.7	74.7	66.4
A1 ^F	89.4	96.8	96.9	90.1	79.5*	67.1*	90.4	89.2	89.1	80.6	72.3	70.4
A1-LSD	91.1	99.8	98.9	93.1	83.4	70.7	77.4	76.7	85.9	81.4	74.9	67.6
A1-LSD ^F	88.9	96.4	96.3	89.4	79.0*	66.8*	89.9	88.7	89.0	80.6	72.8	71.1
A1-QEV	92.2	101.1	99.4	93.5	83.5	70.6*	76.1	76.0	86.3	81.9	75.1	67.7
A1-QEV ^F	89.4	97.2	96.8	89.8*	79.3*	66.9*	88.5	87.5	87.8	80.1	72.8	71.8

Table 13: Level prediction for $A_1(3)$ and $A_1(3)$ -LSD/QEV

Predicted is the h -week-ahead level of a τ -year yield with $h = 4, 13, 26$ and $\tau = 0.5, 1, 2, 3, 5, 10$. The root mean squared error (RMSE) is presented in bps. An upper subscript “F” corresponds to a model with no constraint on the physical drift. The asterisk “*” indicates that the null hypothesis of equal predictive accuracy between the $A_0(3)$ and each model in the first row is rejected at the 5% significance level. The standard error is computed by the Newey and West method and the critical values of an asymptotic distribution of the test statistic are computed from Kiefer and Vogelsang (2005, Table 1) with the truncation lag parameter set at the nearest integer of square-root of the sample size (i.e., $\text{round}(\sqrt{T})$). The in-sample period is from January 9, 1991 to April 9, 2003 ($T = 640$), and the out-of-sample period is from April 16, 2003 to May 27, 2009 ($T = 320$).

Panel A: First-step comparison		$y_{0.5}$	y_1	y_2	y_3	y_5	y_{10}
Minimum	Level	0.00	0.00	0.00	0.00	0.00	0.09
	Slope	0.00	0.00	0.00	0.00	0.00	0.20
	Curvature	0.01	0.01	0.00	0.00	0.00	0.42
Median	Level	0.00	0.00	0.00	0.00	0.06	1.46
	Slope	0.00	0.00	0.00	0.01	0.07	1.41
	Curvature	0.00	0.00	0.00	0.00	0.03	0.71
Maximum	Level	0.00	0.00	0.00	0.01	0.10	2.57
	Slope	0.00	0.00	0.00	0.01	0.14	2.62
	Curvature	0.00	-0.01	0.00	0.01	0.06	1.15

Panel B: Second-step comparison		r	μ	x_3	y_1	y_3	y_5
Minimum	Level	-0.03	0.19	-0.13	0.01	-0.02	-0.06
	Slope	-0.07	0.41	-0.28	0.02	-0.05	-0.13
	Curvature	-0.15	0.85	-0.57	0.05	-0.09	-0.25
Median	Level	-0.48	2.87	-1.98	0.15	-0.32	-0.81
	Slope	-0.46	2.78	-1.92	0.14	-0.30	-0.78
	Curvature	-0.23	1.41	-0.97	0.07	-0.16	-0.40
Maximum	Level	-0.84	5.02	-3.46	0.26	-0.55	-1.43
	Slope	-0.85	5.12	-3.52	0.26	-0.56	-1.41
	Curvature	-0.38	2.27	-1.56	0.11	-0.24	-0.63

Panel C: Third-step comparison		r	μ	x_3	y_1	y_3	y_5
Minimum	Level	-1.02	6.44	-2.66	0.27	-0.45	-0.91
	Slope	-0.40	2.76	6.12	0.08	-0.07	-0.12
	Curvature	0.11	-0.11	5.42	-0.10	0.25	0.38
Median	Level	0.49	-3.08	8.82	-0.12	0.11	-0.06
	Slope	0.52	-3.53	15.21	-0.11	0.09	0.01
	Curvature	0.02	-0.18	5.61	0.01	-0.08	-0.30
Maximum	Level	1.00	-6.15	17.64	-0.28	0.38	0.38
	Slope	0.68	-4.56	8.08	-0.14	0.02	-0.44
	Curvature	0.38	-2.75	15.09	-0.07	0.01	-0.08

Table B1: Accuracy of approximation using $A_1(3)$

Approximation errors, defined as the difference in yields/factors between the approximate and closed-form solutions, are presented in bps. The errors are evaluated at nine states taken from the actual data from January 4, 1991 to May 27, 2009, where the level, slope, and curvature factors take the minimum, median, or maximum value. Panel A presents the first-step comparison, where the true values of both parameter and state vectors are given as input for the approximation method. Panel B presents the second-step comparison, where the true value of only the parameter vector is given. Panel C presents the third-step comparison, where no prior information is given.

Maturity		0.5	1	2	3	5	10
Panel A: LSD model							
Minimum	Level	0.12	0.07	0.02	-0.01	-0.01	0.04
	Slope	0.17	0.14	0.09	0.06	0.07	0.34
	Curvature	0.34	0.29	0.18	0.06	-0.02	0.44
Median	Level	-0.04	-0.07	-0.07	-0.11	-0.09	1.37
	Slope	-0.16	-0.13	-0.10	-0.02	0.09	1.09
	Curvature	-0.04	-0.05	-0.04	-0.03	0.04	0.94
Maximum	Level	0.05	0.07	0.18	0.25	0.42	2.13
	Slope	-0.18	-0.20	-0.26	-0.24	-0.13	1.91
	Curvature	-0.22	-0.19	-0.16	-0.13	-0.12	0.54
Panel B: QEV model							
Minimum	Level	0.12	0.14	0.14	0.10	0.04	0.21
	Slope	0.20	0.02	-0.26	-0.36	-0.41	-0.23
	Curvature	0.33	0.22	-0.13	-0.23	-0.14	1.53
Median	Level	-0.01	0.00	0.05	0.04	0.04	1.36
	Slope	-0.19	-0.18	-0.14	-0.13	-0.02	1.14
	Curvature	-0.02	-0.01	-0.03	-0.01	-0.02	0.58
Maximum	Level	0.02	-0.04	-0.19	-0.26	0.00	2.80
	Slope	-0.16	-0.12	-0.16	-0.13	0.10	2.88
	Curvature	-0.26	-0.24	-0.14	-0.02	0.24	1.14

Table B2: Accuracy of approximation using $A_1(3)$ -LSD/QEV

Approximation errors, defined as the difference in yields between the approximate and Monte Carlo solutions, are presented in bps. The errors are evaluated at nine states taken from the actual data from January 4, 1991 to May 27, 2009, where the level, slope, and curvature factors take the minimum, median, or maximum value. The parameter values of the LSD and QEV models are given in Tables 2 and 3, respectively.

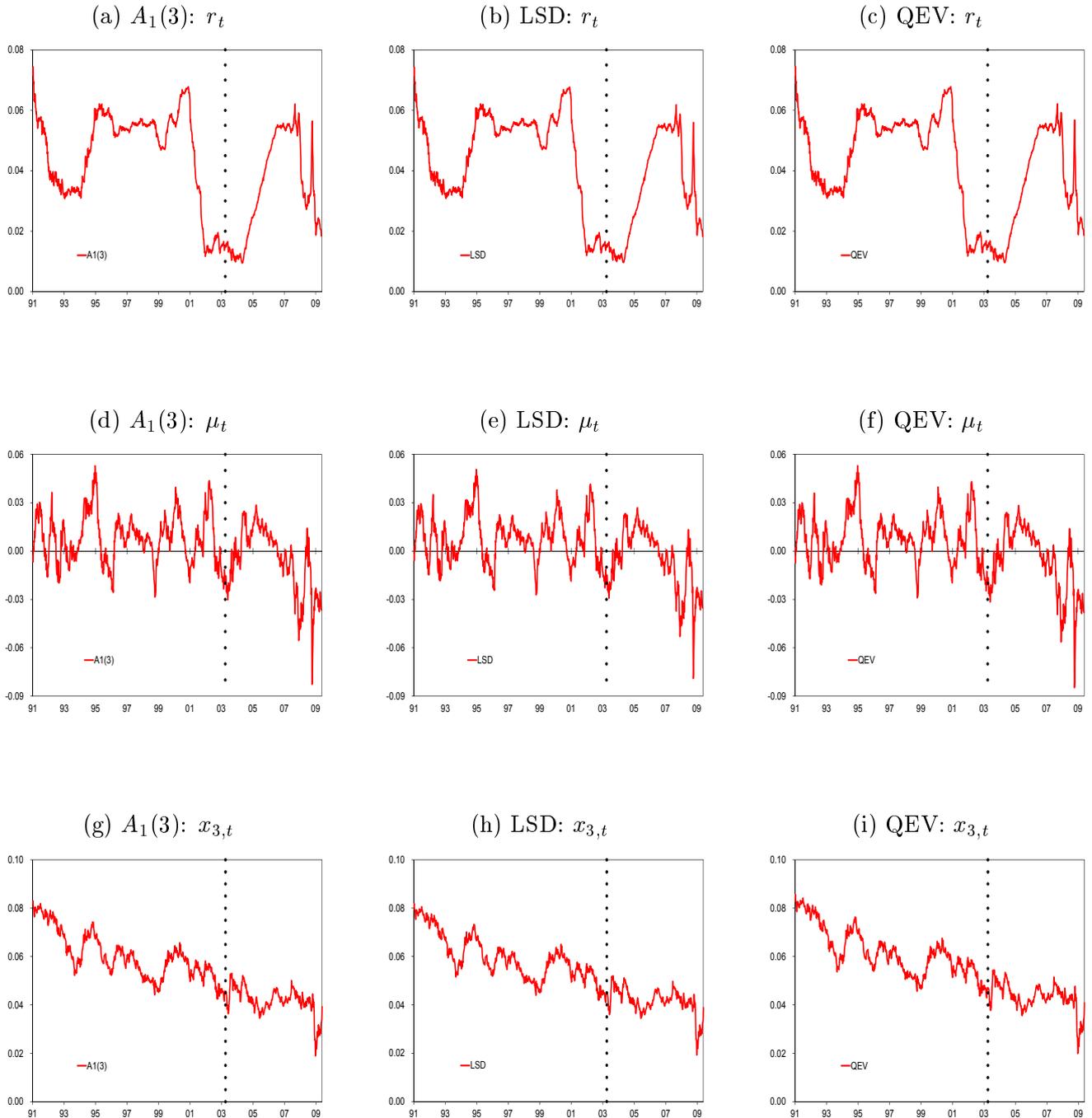


Figure 1: Time series of factors for $A_1(3)$ and $A_1(3)$ -LSD/QEV

Time-series of factors extracted through the $A_1(3)$ (left column), LSD (middle column), and QEV (right column) models are presented. The factors are r_t (the risk-free rate; first row), μ_t (the risk-neutral drift of change in r_t ; second row), and $x_{3,t}$ (volatility factor; third row). The vertical dotted line separates the in-sample and out-of-sample periods.

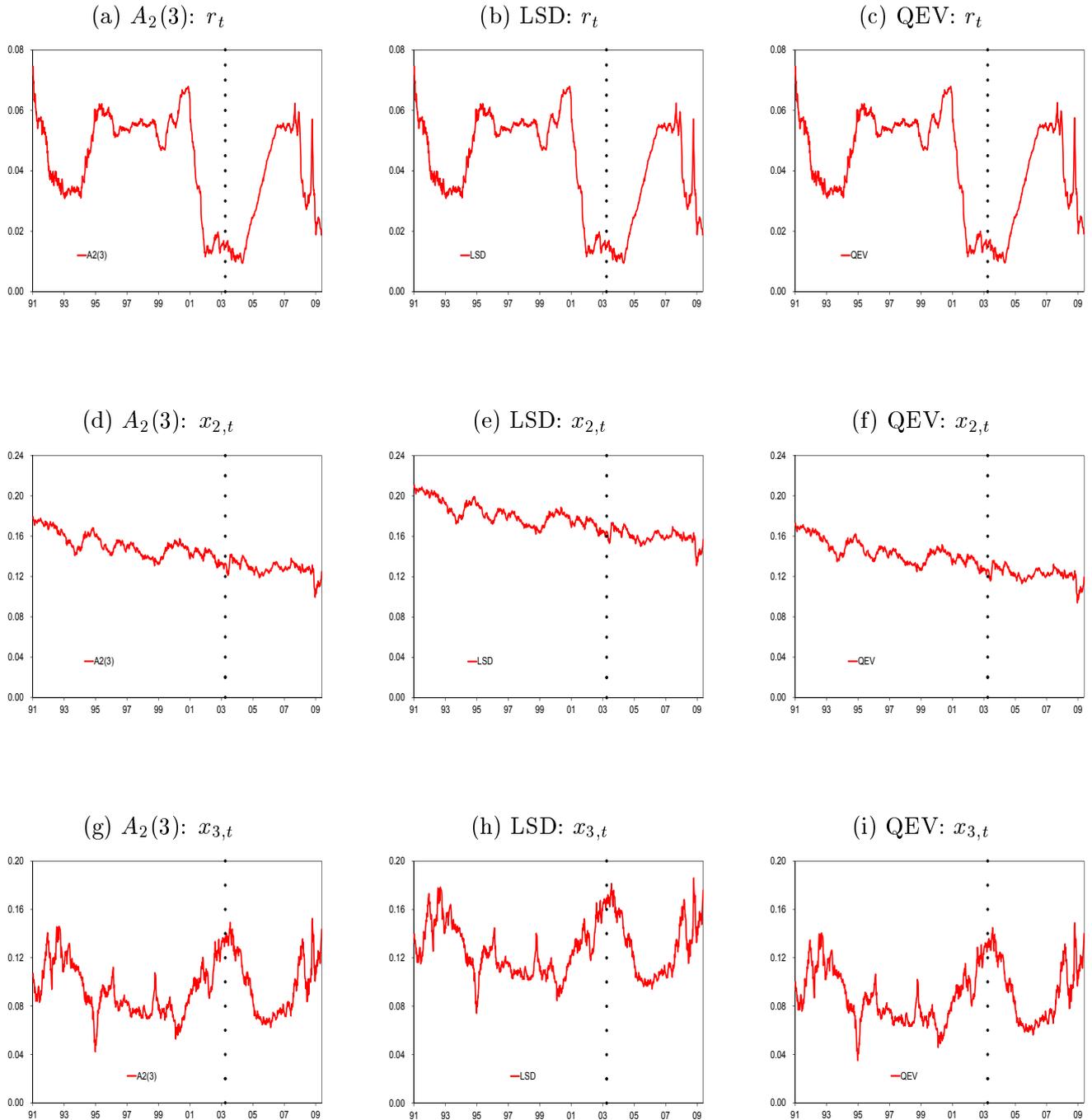


Figure 2: Time series of factors for $A_2(3)$ and $A_2(3)$ -LSD/QEV

Time-series of factors extracted through the $A_2(3)$ (left column), LSD (middle column), and QEV (right column) models are presented. The factors are r_t (the risk-free rate; first row) and $x_{i,t}$ ($i = 2, 3$) (volatility factors; second and third rows). The vertical dotted line separates the in-sample and out-of-sample periods.

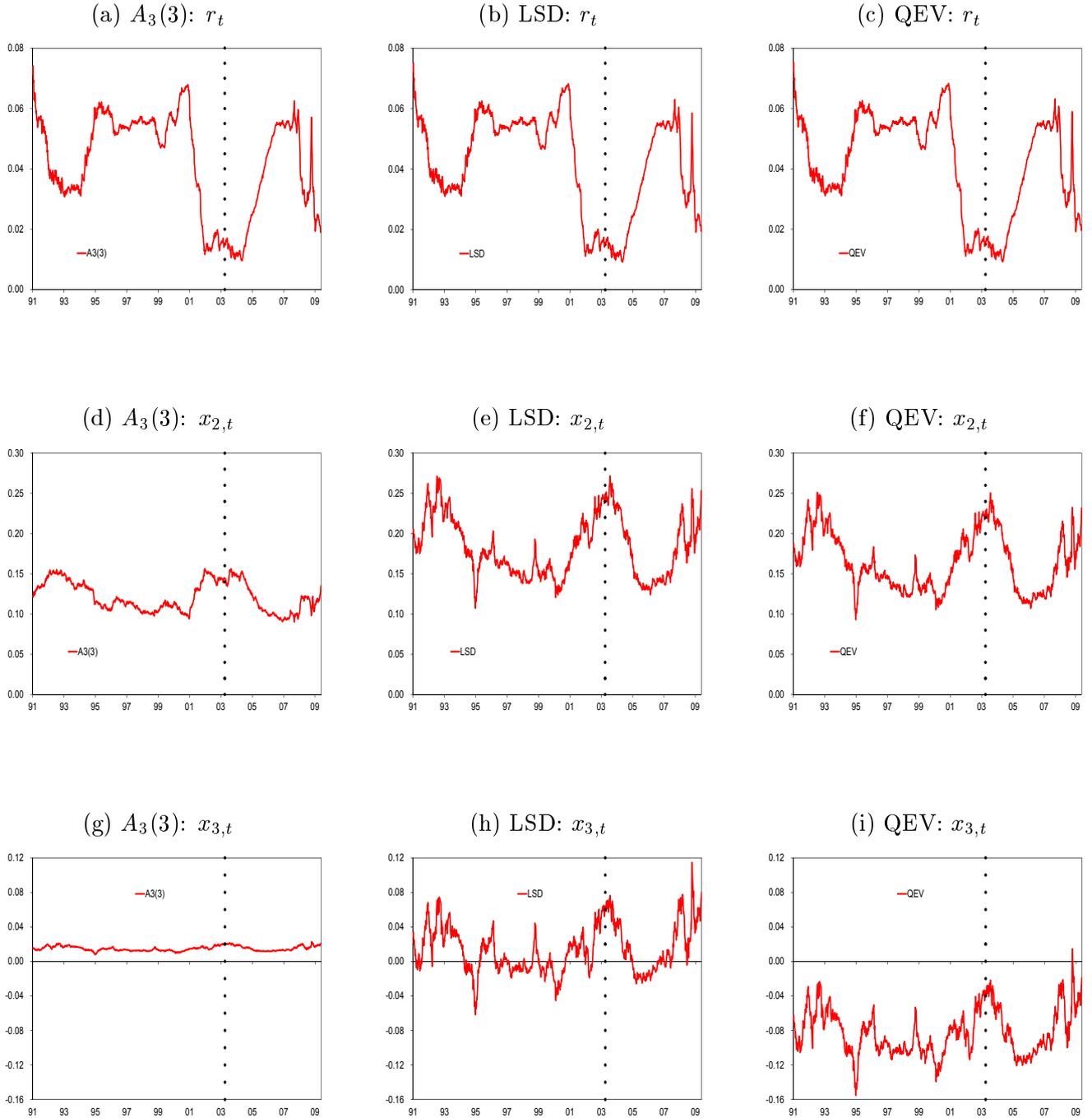


Figure 3: Time series of factors for $A_3(3)$ and $A_3(3)$ -LSD/QEV

Time-series of factors extracted through the $A_3(3)$ (left column), LSD (middle column), and QEV (right column) models are presented. The factors are r_t (the risk-free rate; first row), computed by $r_t = -1 + x_{1,t} - x_{2,t} + x_{3,t}$, and $x_{i,t}$ ($i = 2, 3$) (volatility factors; second and third rows). The vertical dotted line separates the in-sample and out-of-sample periods.

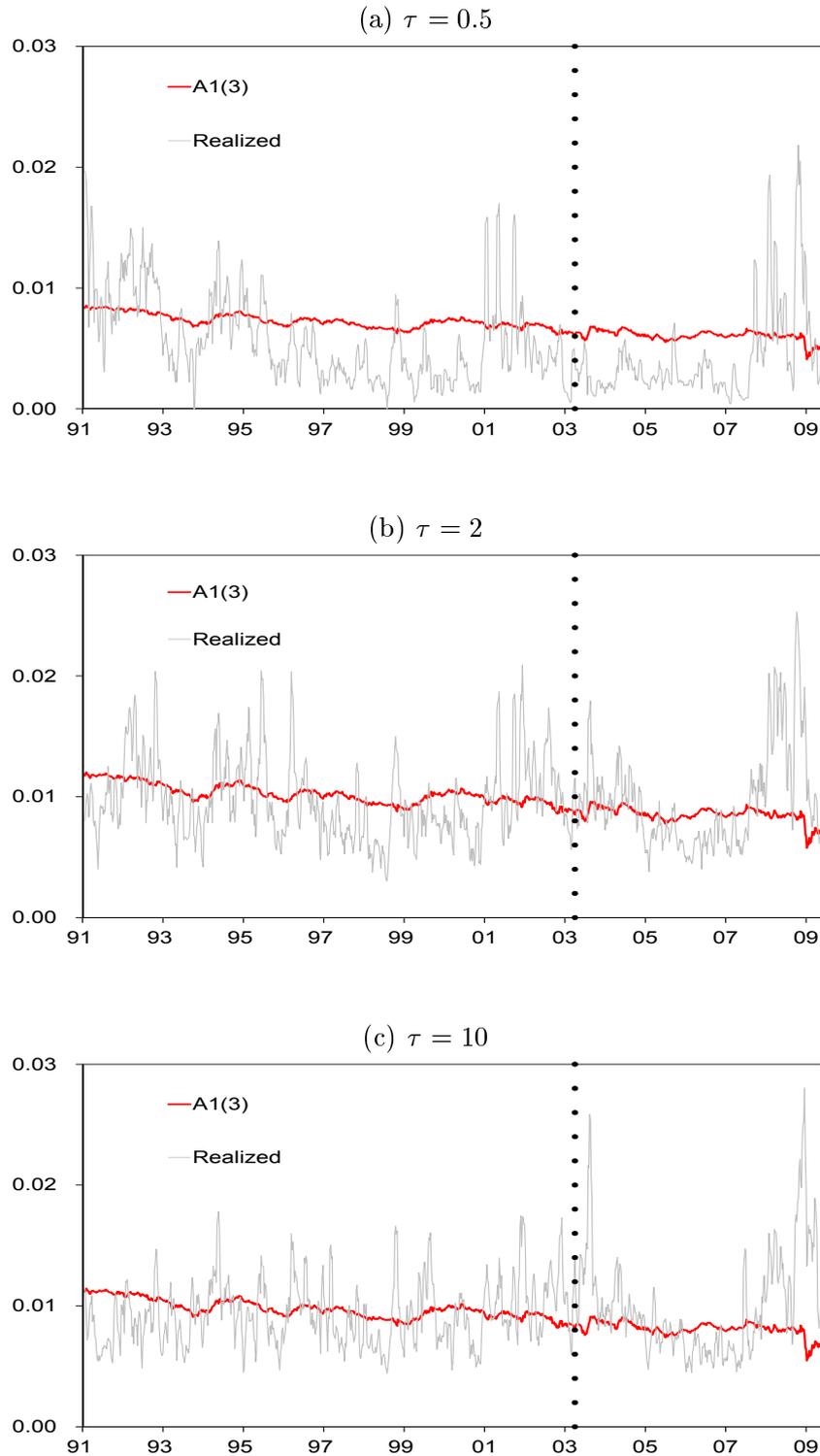


Figure 4: Time series of four-week-ahead volatility forecasts by $A_1(3)$

Predicted by the $A_1(3)$ model is the four-week-ahead annualized standard deviation of a τ -year yield with $\tau = 0.5, 2, 10$. The forecast is in solid line and the realized is in thin line. The vertical dotted line separates the in-sample and out-of-sample periods.

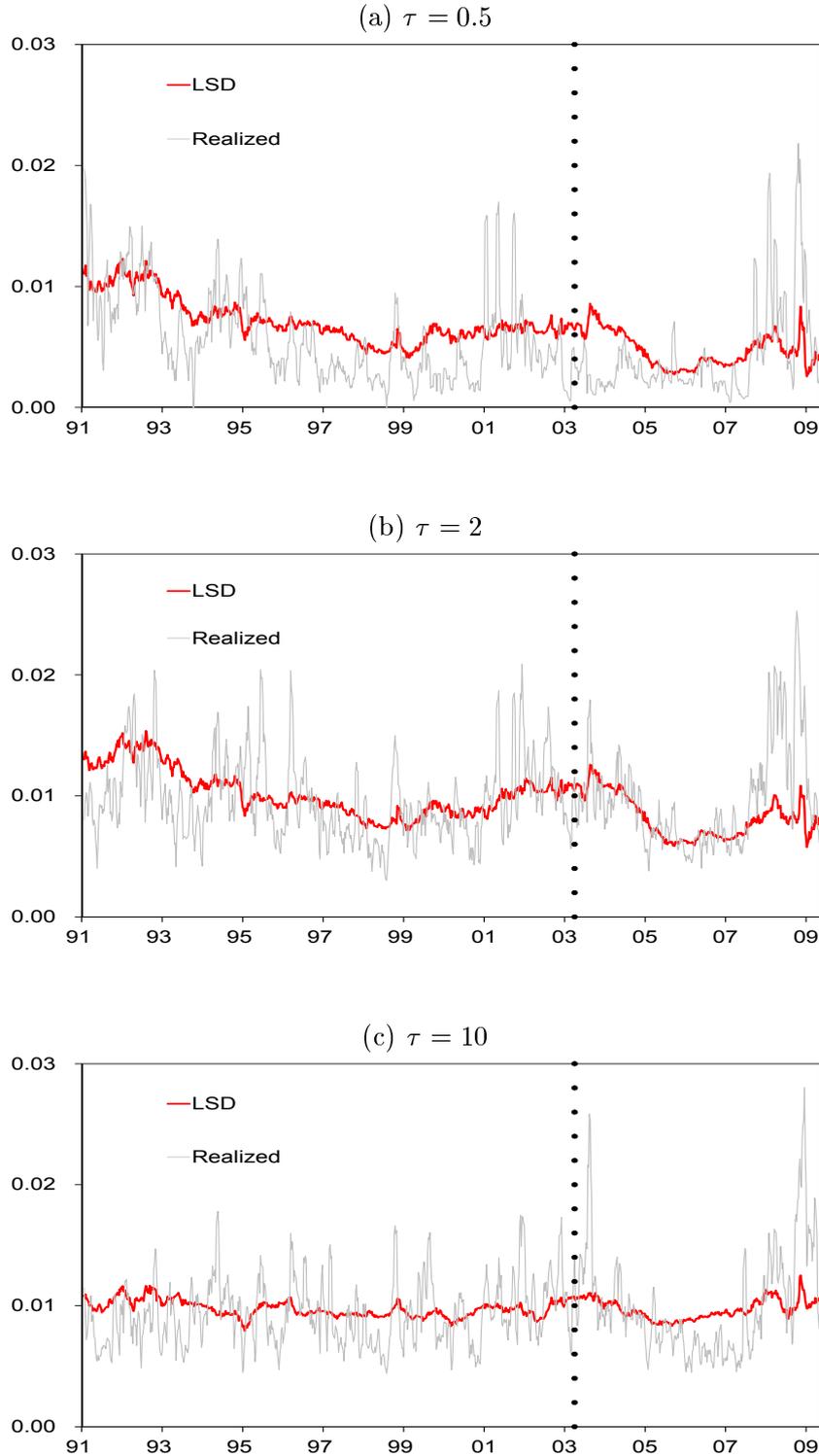


Figure 5: Time series of four-week-ahead volatility forecasts by $A_1(3)$ -LSD

Predicted by the $A_1(3)$ -LSD model is the four-week-ahead annualized standard deviation of a τ -year yield with $\tau = 0.5, 2, 10$. The forecast is in solid line and the realized is in thin line. The vertical dotted line separates the in-sample and out-of-sample periods.

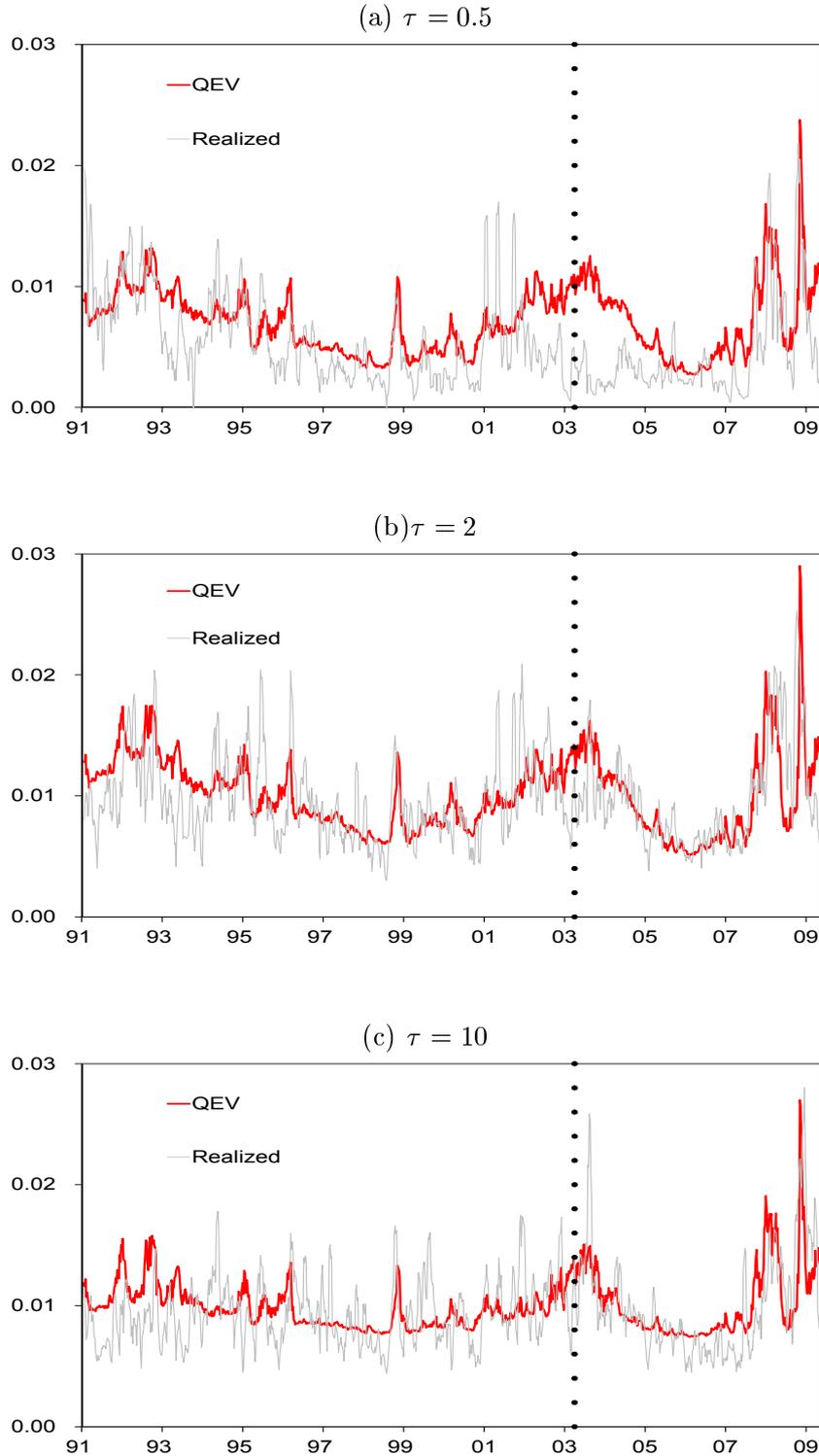


Figure 6: Time series of four-week-ahead volatility forecasts by $A_1(3)$ -QEV

Predicted by the $A_1(3)$ -QEV model is the four-week-ahead annualized standard deviation of a τ -year yield with $\tau = 0.5, 2, 10$. The forecast is in solid line and the realized is in thin line. The vertical dotted line separates the in-sample and out-of-sample periods.